STATISTICAL SEGMENTATION AND REGISTRATION OF MEDICAL ULTRASOUND DATA

NILS MATTIAS HANSSON

LUND INSTITUTE OF TECHNOLOGY
Lund University
Centre for Mathematical Sciences
Mathematics
Major funding for this thesis was provided by Malmö University, Sweden.
Preface

Papers which constitute thesis:

Active Contour models for cardiac ultrasound segmentation


Bayesian MCMC model for cardiac ultrasound segmentation


Ultrasound Similarity measures for feature description and registration of ultrasound data


Acknowledgments

First of all I would like to acknowledge my supervisors Yuanji Cheng and Sami S. Brandt, whose support and encouragement has helped me finish this work. Yuanji Cheng has been especially supportive in the finalization of the thesis, where he has shown a keen eye for detail. My most long-term collaborator Sami S. Brandt has provided guidance throughout most of my time as a graduate student. This guidance has been especially important when I hit a bump in the road, which has been the case on many occasion. My other main collaborator Tassilo Klein of TU München has been very important for my research; working with him continues to be effortless even over long distances. I would also like to thank Nassir Navab for inviting me to the CAMP group at TU München, and for providing me with assistance during my research stays in Munich. My fellow graduate students Alma Mašić and Ketut Fundana have been of tremendous support during this, at times, very challenging time of my life. I have also collaborated in research with Ketut, where he has proved to be insightful and creative. Johan Lindström provided crucial advice in theoretical matters, taking the time to break down difficult problems, which I highly appreciate. My medical partners Petri Gudmundsson, Amra Jujić and Andreas Malmgren have provided crucial clinical advice, in the nicest way possible. I would also like to acknowledge the help and support of my PhD advisors during the first two years of graduate studies: Anders Heyden and Niels C. Overgaard. Funding for this PhD Thesis has been provided by Malmö University and Lund University. The financial support of these institutions has been crucial and is highly appreciated.

Finally I would like to thank my mother Aino and my brother Mikael for their continuing support.
# Contents

0.1 Organisation of the Thesis ........................................... 1  
0.2 Author contributions ................................................. 2  

1 Medical Ultrasound ......................................................... 5  
1.1 Introduction ............................................................... 5  
1.2 Physical Background ..................................................... 5  
1.3 Transmission to Display of Ultrasound Echo ......................... 10  
1.4 3D Freehand Ultrasound ................................................. 13  
1.5 Reconstruction of 3D Data ............................................. 14  
1.6 Applications in Cardiology ............................................ 18  
1.7 Applications in Neurology ............................................ 22  

2 Statistical Modeling of the Ultrasound Signal ........................ 27  
2.1 Statistical Modeling of Ultrasound Data .............................. 27  
2.2 Mixture Models ........................................................... 33  
2.3 Bayesian Statistical Formulations .................................... 34  
2.4 Maximum Likelihood Estimators ..................................... 35  
2.5 EM Algorithm ............................................................... 36  
2.6 EM-Gamma ................................................................. 40  
2.7 EM-MAP ................................................................. 44  
2.8 Similarity Measures ....................................................... 45  

3 Segmentation of the Endocardium - Variational Approach .......... 47  
3.1 Background ................................................................. 47  
3.2 Dimensionality Reduction .............................................. 52  
3.3 Level set representation ................................................ 54  
3.4 Functions of Bounded Variation ...................................... 55  
3.5 First Necessary Condition for Relative Minimum ..................... 56  
3.6 Gâteaux variation of functional ....................................... 57  
3.7 $L^2$ Shape Gradients ................................................. 61  
3.8 Construction of Locally Optimal Shape Prior Model ................. 61
# List of Figures

1.1 Specular scattering condition .................................................. 8  
1.2 Diffusive scattering condition .................................................. 9  
1.3 Transmission to Reception of Echo .............................................. 11  
1.4 Reception to Display .............................................................. 12  
1.5 System setup for a 3D freehand system ....................................... 14  
1.6 Parasternal Long Axis View ...................................................... 20  
1.7 Apical 2-chamber View of Heart ............................................... 20  
1.8 Apical 4-chamber view of heart ................................................ 21  
1.9 Parasternal Short Axis view of heart ......................................... 21  
1.10 Main diagnostic views of heart ................................................. 22  
1.11 MRI slices of the brain .......................................................... 23  
1.12 Schematic illustration of midbrain and structures within ............... 24  
1.13 Ultrasound slices obtained from transcranial reconstructions ......... 26  

2.1 Large first bin due to quantization ............................................ 33  
2.2 Fitted Nakagami mixture distribution to US neck data .................. 44  

3.1 Example PCA ............................................................................ 53  
3.2 Evolving level set front .......................................................... 54  
3.3 Level set representation .......................................................... 55  
3.4 Transformation of domain ....................................................... 58  
3.5 Domain parametrized by polar coordinates ................................ 59  
3.6 Update procedure ................................................................. 82  
3.7 Synthetic Example 1 .............................................................. 84  
3.8 US Example 1 ........................................................................ 86  
3.9 US Example 2 ........................................................................ 87  
3.10 Rayleigh model in endocardium segmentation ........................... 87  
3.11 Segmentation Example 1 .......................................................... 89  
3.12 Histogram/pdf fits Example 1 .................................................. 90
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.13 Segmentation Example 2</td>
<td>91</td>
</tr>
<tr>
<td>3.14 Histogram/pdf fits Example 2</td>
<td>92</td>
</tr>
<tr>
<td>4.1 Latent variables</td>
<td>107</td>
</tr>
<tr>
<td>4.2 Regions used in likelihood formulation</td>
<td>109</td>
</tr>
<tr>
<td>4.3 Illustration censored mixture model</td>
<td>109</td>
</tr>
<tr>
<td>4.4 Discretization of endocardium and atrial region</td>
<td>110</td>
</tr>
<tr>
<td>4.5 Apex Region</td>
<td>110</td>
</tr>
<tr>
<td>4.6 Summary of proposed algorithm</td>
<td>113</td>
</tr>
<tr>
<td>4.7 Example Bayesian probability map</td>
<td>117</td>
</tr>
<tr>
<td>4.8 Evaluation Landmarks</td>
<td>121</td>
</tr>
<tr>
<td>4.9 Unsigned distances from sample to expert points</td>
<td>121</td>
</tr>
<tr>
<td>4.10 Graphical results</td>
<td>122</td>
</tr>
<tr>
<td>4.11 Challenges US segmentation</td>
<td>123</td>
</tr>
<tr>
<td>5.1 Pair-wise and triple-site cliques</td>
<td>131</td>
</tr>
<tr>
<td>5.2 Feature histograms for high/low threshold FLBP</td>
<td>135</td>
</tr>
<tr>
<td>5.3 Circular Square LBP neighbourhood</td>
<td>136</td>
</tr>
<tr>
<td>5.4 Illustration LBP and FLBP for US image</td>
<td>139</td>
</tr>
<tr>
<td>5.5 Distance map of the feature descriptor</td>
<td>141</td>
</tr>
<tr>
<td>5.6 Scheme of US MRF model</td>
<td>143</td>
</tr>
<tr>
<td>5.7 US human neck</td>
<td>143</td>
</tr>
<tr>
<td>5.8 Texture data for training DistBoost</td>
<td>147</td>
</tr>
<tr>
<td>5.9 Distribution of the interaction parameters</td>
<td>147</td>
</tr>
<tr>
<td>5.10 The threshold parameter defining a probability interval</td>
<td>150</td>
</tr>
<tr>
<td>5.11 Confidence map and corresp. RF image</td>
<td>154</td>
</tr>
<tr>
<td>5.12 Block matching and derivation of the similarity maps</td>
<td>155</td>
</tr>
<tr>
<td>5.13 Global registration performance</td>
<td>155</td>
</tr>
<tr>
<td>5.14 Similarity maps</td>
<td>156</td>
</tr>
<tr>
<td>5.15 Distributional change for cones within single view of midbrain</td>
<td>158</td>
</tr>
<tr>
<td>5.16 TCUS reconstructions 1</td>
<td>162</td>
</tr>
<tr>
<td>5.17 TCUS reconstructions 2</td>
<td>163</td>
</tr>
<tr>
<td>5.18 TCUS reconstructions 3</td>
<td>164</td>
</tr>
<tr>
<td>5.19 TCUS reconstructions 4</td>
<td>165</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Statistical models for varying density and organization of scatterers . . 28
3.1 Results previous works . . . . . . . . . . . . . . . . . . . . . . . . . . 95
3.2 Average distance results - Convex Active Contour Model . . . . . . 96
4.1 Evaluated distribution models . . . . . . . . . . . . . . . . . . . . . . . . 118
4.2 Mean AMDL score for distribution models . . . . . . . . . . . . . . . 118
4.3 Manual Inter- and intraobserver errors cardiac US segmentation . . . 127
4.4 Results of works on 2D+T cardiac segmentation . . . . . . . . . . . . 128
5.1 Median errors (pixels) of random reg. study and similarity metrics . . . 156
5.2 Median errors (mm) and std. dev. of registration study . . . . . . . . 161
Abstract

The interpretation of ultrasonic imagery is typically not straightforward and of quite subjective nature and therefore strongly dependent on the expertise of its users. Thus the development of algorithms which aid in the interpretation of ultrasonic data is a highly relevant topic. This thesis examines aspects of segmentation and registration of ultrasonic data, utilizing the fact that the ultrasound signal can be modeled statistically. The object of segmentation is the endocardium in the left-ventricular long-axis view of the human heart in clinical B-mode ultrasound (US) image sequences, while similarity measures for feature descriptors and registration are applied to the envelope-detected radio frequency US data of the human neck and brain. Locally and globally optimal variational active contour methods and a Bayesian Markov Chain Monte Carlo sampling method are applied to the segmentation problem, utilizing prior formulations for shape and regularization. A feature descriptor is proposed which combines global data statistics, by a maximum-likelihood-estimated distribution, with local pattern characteristics, employing Markov Random Field interaction parameters. For registration we propose two approaches. Firstly, a hybrid procedure incorporating global statistics, by Hellinger distance between distribution in images, and local textural features by a statistics-based extension of Fuzzy Local Binary Patterns. Secondly, we explore the registration of 3D freehand US data, where view dependency of ultrasound is addressed by modeling speckle statistics, using a finite mixture model. The proposed methods for segmentation, feature description and registration are evaluated through experiments and/or comparative experiments to state-of-the-art models.
Sammanfattning


I denna avhandling presenterar vi metoder för segmentering och registrering av ultraljudsdata. Segmentering är en potentiellt tidskrävande process där olika anatomiska strukturer märks ut i ultraljudsdata. Vi presenterar tre olika segmenteringalgoritmer för endokardiet, dvs hjärtsäcken. Segmentering av endokardiet är relevant för att kunna approximera hjärtvolym, nödvändigt som hjälpmedel vid diagnos av hjärtsjukdom. I konstruktionen av dessa algoritmer har vi utnyttjat att ultraljudssignalen kan modelleras statistiskt och inkluderat forminformation av hjärtskamare, som inhämtats från manuella segmenteringar. Registrering, dvs sammanfogan av bilder eller sekvenser, är avgörande för att bedöma utseendet av strukturer i kroppen. En viktig applikation är spårning av sjukdomsförlopp; det är uppenbart att detta kräver korrekt rekonstruerade bilddata för att kunna genomföras. För att en korrekt registrering ska kunna ske krävs att det är möjligt att kunna mäta likhet mellan datapunkter, och därför har vi lanserat flera matematiska mått för att mäta likhet i ultraljudsdata.

Vi har i experiment visat användbarheten av våra segmenteringsalgoritmer genom jämförelse med motsvarande manuella segmentering av biomedicinisk analytiker och resultat från tidigare studier. Registreringsalgoritmerna har applicerats på ultraljudsbilder av mänsklig nacke, och för rekonstruktion av hjärna från en mängd ultraljudsbilder som inhämtats med en ultraljudsgivare placerad vid tinningarna på patienten. Vi har observerat signifikant förbättrade resultat i jämförelse med standardalgoritmer för registrering.

De utvecklade metoderna har potential att underlätta bedömningen av olika sjukdomstillstånd. Segmenteringsalgoritmerna underlättar bedömningen av hjärtvolym, vilket är avgörande för diagnos av ischemisk hjärtsjukdom, medan registreringsalgoritmer kan
underlätta diagnos av degenerativa sjukdomar. En av de utvecklade metoderna appliceras på rekonstruktion av mellanhjärnan, specifikt substantia nigra vars utveckling över tid är en viktig del av diagnos av Parkinsons sjukdom.
Nomenclature

Mathematical

$H$  
Heaviside function

$$H(x) = \begin{cases} 
1, & x \leq 0 \\
0, & \text{otherwise.}
\end{cases}$$

$\hat{H}_\epsilon$  
Regularised Heaviside function

$$\hat{H}_\epsilon(x) = \begin{cases} 
1, & \text{if } x > \epsilon \\
0, & \text{if } x < -\epsilon \\
\frac{1}{2} \left[1 + \frac{x}{\epsilon^2} + \frac{1}{\pi} \sin \left(\frac{\pi x}{\epsilon}\right)\right], & \text{if } |x| \leq \epsilon.
\end{cases}$$

Alternatively

$$\hat{H}_\epsilon(x) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan \left(\frac{x}{\epsilon}\right)\right).$$

where $\hat{H}_\epsilon \in C^2$ and $\hat{H}_\epsilon \in C^\infty$.

$D$  
regular open bounded set corresponding to the planar image domain, s.t. $D \subset \mathbb{R}^2$.

$\delta$  
The Kronecker delta

$$\delta(x) = \begin{cases} 
1, & x = 0 \\
0, & \text{otherwise.}
\end{cases}$$
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega$</td>
<td>subset of image region $D$</td>
</tr>
<tr>
<td>$</td>
<td>\Omega</td>
</tr>
<tr>
<td>$</td>
<td>\Gamma</td>
</tr>
<tr>
<td>g.c.d</td>
<td>greatest common divisor</td>
</tr>
<tr>
<td>$</td>
<td></td>
</tr>
<tr>
<td>$</td>
<td></td>
</tr>
<tr>
<td>$\text{det}(A)$</td>
<td>determinant of matrix $A$</td>
</tr>
<tr>
<td>$I$</td>
<td>identity matrix</td>
</tr>
<tr>
<td>$</td>
<td></td>
</tr>
<tr>
<td>$f \propto g$</td>
<td>the functions $f$ and $g$ are proportional</td>
</tr>
<tr>
<td>$H$</td>
<td>Hilbert space</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Dirac delta</td>
</tr>
<tr>
<td>a.e.</td>
<td>almost everywhere</td>
</tr>
</tbody>
</table>

**Probability**

- **MLE**: Maximum Likelihood Estimate
- **a.s**: an event occurs a.s (almost surely) if it has probability 1 of occurring
- **iid**: independent and identically distributed
- $f(x|\theta)$: density of $X$ conditional on the parameter $\theta$
- $X \sim f(x|\theta)$: $X$ is distributed according to density $f(x|\theta)$
- $\chi$: state space
- **pdf**: probability density function
- **cdf**: cumulative density function
- **MCMC**: Markov Chain Monte Carlo

**Abbreviations**

- **PCA**: Principal Component Analysis
- **EM**: Expectation Maximization
- **US**: Ultrasound
- **RF**: Radio frequency
- **TGC**: Time Gain Compensation
- **B-mode**: Brightness-mode
- **LV**: Left ventricle
Overview of Thesis

Medical ultrasound is, relative to other modalities like magnetic resonance imaging (MRI) and computed tomography (CT), inexpensive and readily available in most medical institutions around the world. Using ultrasound it is possible to image most tissue in the body in a noninvasive manner. Ultrasound equipment has gone through radical improvements in the last 30 years, moving from initial two-dimensional imagery to three-dimensional color flow sequences. It must be pointed out that the great majority of clinical users of ultrasound only have access to basic ultrasound equipment, which produces sequences of 2-dimensional images which are subject to proprietary post-processing not under the user’s control. Three-dimensional ultrasound is still quite rare and often unavailable in developing nations. Ultrasound data is not however uncomplicated to acquire and interpret. This is related to the physics of ultrasound and takes the form of speckle contamination (the granular pattern which is characteristic), view-dependency, bone-reflections etc. Being highly dependent on specific conditions under which the data is acquired, it follows that US is highly user-dependent, further complicating matters. The degree of processing that the data has undergone will also affect its statistical properties. Our goal in this thesis is to describe approaches for dealing with the problems that are inherent to ultrasound data, in both segmentation and registration problems. We present models that address the problems that occur when dealing highly processed clinical US data, as well as very lightly processed envelope-detected radio frequency data. Obtained results indicate the potential clinical usefulness of the proposed approaches.

0.1 Organisation of the Thesis

In Chapter 1 the physical foundations of ultrasound are described, and a general overview is given of two- and three-dimensional ultrasound equipment used in our experiments. The details of how 3D data is reconstructed from 2D data is presented. Furthermore we describe the specific applications of ultrasound in cardiology and neurology.

Chapter 2 presents a description of the statistical modeling of the ultrasound signal. Mathematical tools for estimating mixture models to model the US signal are described.
Chapter 3 describes locally optimal and convex active contour models for segmentation of the endocardium in B-mode US cardiac sequences, incorporating a prior for shape model of the endocardium.

In Chapter 4 an overview of Markov Chain Monte Carlo theory is given, and a statistical model for segmenting the endocardium in B-mode US cardiac sequences is described.

Chapter 5 presents basic Markov Random Field theory and an associated model which provides a feature descriptor for envelope detected ultrasound data. Two registration models are described, one using statistical Local Binary Patterns in the case of 2D images, and for the case of 3D freehand US data, a model which takes into account view-dependency.

0.2 Author contributions

In compliance with the regulations of Lund University, we here state the contributions of the author of this thesis. The following abbreviations are used for contributing authors: Mattias Hansson (MH), Sami S. Brandt (SB), Ketut Fundana (KF), Tassilo Klein (TK), Finn Lindgren (FL), Johan Lindström (JL), Yuanji Cheng (YC), Anders Heyden (AH), Niels Christian Overgaard (NCO) and Nassir Navab (NN).

Chapter 3

In Sec. 3.10, based on the paper Rayleigh Segmentation of Ultrasound Images [85], a locally optimal method, depending on the simultaneous evolution of a pair of interacting active contours, is proposed for purposes of segmenting the endocardium in left-ventricle in B-mode US images. MH, NCO and AH developed the theory, MH implemented the model and performed experiments. Manuscript was written and reviewed by MH, NCO and AH.

Sec. 3.11 is based on the paper Convex Spatio-Temporal Segmentation of the Endocardium in Ultrasound Data using Distribution and Shape Priors [84], where a convex variational active contour model with shape priors, for spatio-temporal segmentation of ultrasound B-mode image sequences of the endocardium in the left-ventricle of the human heart, is solved by Continuous Cuts. To our knowledge this is the first time convex shape prior segmentation is applied to ultrasound. MH and KF developed the main theory with minor contribution of SB. MH implemented the model and performed experiments, with advice from KF. Manuscript was written and reviewed by MH and KF.
Chapter 4

Sec. 4.2 is based on the paper *Evaluation of Cardiac Ultrasound Data by Bayesian Probability Maps* [83], which introduces a Bayesian Markov Chain Monte Carlo sampling model for determining the position distribution of the endocardium in a sequence of B-mode ultrasound images. A three region latent variable model is proposed, which represents the endocardium, atrial region and background. A censored Gamma mixture model is proposed to address quantization artifacts. Misclassification of the apex of the endocardium, due to foreshortening related to a highly curved apex of the endocardium, is handled by down-weighting of the likelihood in this region. MH and SB developed the main theory, with key contributions of JL. MH implemented the model and performed experiments, with advice of SB. Manuscript was written and reviewed by MH, SB and YC. A previous version of the model presented in Sec 4.2, was presented in: (1) *Evaluation of Cardiac Ultrasound Data by Bayesian Probability Maps* [82] and (2) *Bayesian Probability Maps for Evaluation of Ultrasound Data* [81]. In (1) and (2) MH and SB developed the main theory, with some advice provided by FL in (2). In (1) and (2) MH implemented the model and performed experiments, with advice from SB. Manuscript in (1) and (2) was written and reviewed by MH and SB.

Chapter 5

Sec. 5.6 is based on the paper *Spatial Statistics Based Feature Descriptor For RF Ultrasound Data* [115], which introduces a feature descriptor envelope detected ultrasound data, using global data statistics in terms of a Maximum-Likelihood-Estimated distribution. Local pattern characteristics are incorporated in the model through Markov Random Field interaction parameters.

Sec. 5.7, based on the paper *Registration of RF Ultrasound Data using Hybrid Local Binary Patterns* [114], introduces a hybrid procedure, incorporating global statistics and local textural features, for the registration of envelope detected radio frequency ultrasound. On the global scale this is achieved by Hellinger distance between distribution in images, and on the local scale by a statistics-based extension of Fuzzy Local Binary Patterns.

Sec. 5.8, based on the paper *Modeling of Multi-View 3D Freehand Radio Frequency Ultrasound* [116], introduces a statistical parametric approach in conjunction with a pseudo-distance, Pseudo-J-Divergence, applied to the registration of 3D envelope detected radio frequency ultrasound data. The proposed approach models view dependency in ultrasound by speckle statistics using a finite mixture model. Reconstruction of 3D data on Gamma manifold reduces artefacts compared to standard methods. Registration is finally achieved by the ultrasound specific distribution based pseudo-distance.

In the three forementioned papers, which were written in collaboration with the Computer Aided Medical Procedures (CAMP) group at TU München, MH and TK developed the theory. TK implemented the main components of the model and per-
formed experiments, with minor tasks performed by MH. Manuscript was written and reviewed by MH, TK and NN.
Chapter 1

Medical Ultrasound

1.1 Introduction

Ultrasonography is more accessible, mobile and inexpensive compared to other imaging techniques. Therefore it has become a widely used diagnostic method in various medical disciplines in recent years. However, the interpretation of ultrasound imagery is typically not straightforward, as well as of quite subjective nature. This largely stems from, among other things, the inherent process of ultrasound imaging that is view dependent and subject to noise. Ultrasound (US) noise patterns, referred to as speckle noise, is characteristic for various types of tissue. The noise is the dependent on the many factors such as spatial arrangement and size of speckle generating particles.

Speckle can be treated either as noise, which is to be reduced or removed, or as a feature in data, to be used to indicate different tissue types. In this thesis we adopt the approach of treating speckle noise as a feature, which can be modeled in a statistical framework. Algorithms for segmentation and registration have to address the issue of speckle in order to be successful.

In the sequel we will briefly review statistical models for the backscattered echo from tissues in ultrasound, beginning with a physical background in Sec. 1.2 briefly detailing the formation of ultrasound data. Statistical models for the ultrasound signal envelope are reviewed in Sec. 2.1.

1.2 Physical Background

1.2.1 Frequency, Pressure and Intensity of an Ultrasound Wave

Ultrasound is a high-frequency sound wave, which as it moves through the medium, causes particles to oscillate back and forth in a sinusoidal motion, creating areas of rar-
CHAPTER 1. Medical Ultrasound

efaction and compression. The (displacement) amplitude of a wave can be expressed as peak excess pressure (Pa), that is the maximum difference between the rest pressure in the medium and the pressure induced by the wave.

In the case of an idealised incompressible medium the pressure $p$ of the ultrasound waves are modeled, in the one-dimensional case, by the wave equation

$$\frac{\partial^2 p}{\partial t^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial y^2} \quad (1.1)$$

where $c$ is the speed of sound in the medium, $y$ spatial index and $t$ time index. The speed $c$ at which ultrasound travels through the medium depends on medium density $\rho$ and stiffness $k$, through the relation $c = \sqrt{k/\rho}$. The average speed of sound in tissue is 1540 m$s^{-1}$. The wavelength $\lambda$ of the soundwave is influenced by the speed at which sound travels in the medium, by $\lambda = c/f$, where frequency of the wave is denoted by $f$. Typical wavelength used in ultrasound are 1 – 10 MHz. Related to frequency is the wavenumber, expressed as $K = \frac{\omega}{c}$, where $\omega = 2\pi f$ is angular frequency.

The phase of the wave describes the pattern of displacement of particles as the sound wave passes through the medium. The phase difference between waves needs to be corrected, in order to sum individual waves into the RF signal.

The intensity $I$, stated in unit ($W m^{-2}$), of the beam is defined as the power flowing through a unit area of the beam cross section, orthogonal to the direction of propagation. The intensity is proportional to the square of the pressure of the wave, i.e. $I \propto p^2$.

The response of a medium as the wave is sent through it, is described by the acoustic impedance of the medium, defined as $z = p/v$, where $p$ and $v$ is local pressure of the wave and local particle velocity, respectively. The acoustic impedance varies for different tissue and organ types, for example blood: $1.67 \times 10^6 kg m^{-2} s^{-1}$ and fat: $1.33 \times 10^6 kg m^{-2} s^{-1}$.

The loss of intensity as the ultrasound beam travels through tissue is called attenuation (dB). The decay of intensity is exponential, attenuation coefficient (expressed dB cm$^{-1}$) describes the rate of intensity decay with distance. The attenuation is due to absorption (conversion of intensity into heat), scattering and divergence of the beam over distance. The attenuation coefficient increases linearly with frequency, and so attenuation in tissue is commonly expressed in units dB cm$^{-1}$ MHz$^{-1}$. The coefficient varies for different types of tissue, e.g. blood $\alpha = 0.2$ while cardiac tissue $\alpha = 0.52$ and bone $\alpha = 9.94$. The attenuation is (attenuation coefficient) $\times$(distance traveled in tissue)$\times$(frequency), and thus if a 2 MHz ultrasound beam has traveled 5 cm in blood is attenuated $0.2\alpha \times 2 MHz \times 5 cm = 2$ dB. Attenuation is corrected in modern medical ultrasonic machines by Time Gain compensation, by increasing the amplitude with distance from the transducer.
1.2 Physical Background

1.2.2 Reflection and Scattering

After the pulse is transmitted into the tissue, it is reflected from structures within. The reflected pulse is known also referred to as an echo. As the ultrasound beam strikes the tissue, the manner in which it will scatter, depends on the size of the structure relative to beam width.

To describe the relation of beam width to structure, it is instructive to consider three simple cases of scattering, which can occur in homogeneous media. Clearly tissue is heterogeneous, and so these cases cannot be directly transferred to tissue scattering, but they are nevertheless instructive.

When the ultrasound beam hits a particle, the beam is scattered in all directions. The total scattering cross section for a particle is given by

$$\sigma_s = \frac{W_s}{I_i}$$

(1.2)

where $W_s$ is the total powered scattered and $I_i$ is the incident intensity. $W_s$ is a function of $a/\lambda$, where $a$ is the radius of the particle. It follows then that amount of scattered amplitude is proportional to $\sqrt{\sigma_s}$.

There are three different cases of scattering: specular, diffuse and diffractive. Specular scattering refers to the case when the object is relatively smaller than the wavelength of the beam. In this case the reflected echo will be an approximate image of the object encountered by the beam. In Fig. 1.1 an example is given of specular scattering. A plane wave of radius $b$ encounters particle of radius $a$. The scattering cross section of the wave as it hits the particle is $2\pi a^2$. The reflection factor is

$$R = \frac{Z_s - Z_m}{Z_s + Z_m}$$

(1.3)

which $Z_s$ and $Z_m$ is the impedance the sphere and medium, respectively. The relation here between $I_i$, the incident intensity, and overall backscattered intensity $I_r$ is, as described in [92, 202, 165],

$$\frac{I_r}{I_i} = \frac{\pi b}{4\pi r} |R|^2 = |R|^2 \frac{b^2}{4r^2}$$

(1.4)

where $r$ denotes the radius of the sphere of backscattered intensity. As the backscattered sphere expands ($r$ grows) the fraction $\frac{I_r}{I_i} \rightarrow 0$. Scattering is here frequency independent, as is seen in Eq. 1.4.

Diffusive scattering refers to the case of the wavelength of the beam being relatively larger than the object. This case often occurs in tissue, as for example, blood cells are much smaller than the size of the beam. An illustration of this is given in Fig. 1.2 and the corresponding relation to Eq. 1.4 is given by

$$\frac{I_r}{I_i} = \frac{k^4 a^6}{9r^2} \left[ \frac{3(1 - \rho_p/\rho_m) \cos(\theta)}{1 + 2\rho_p/\rho_m} \left( \frac{1 - \kappa_m}{\kappa_p} \right) \right]^2$$

(1.5)
see [92, 202, 165]. Here \( \rho \) denotes density, while \( \kappa_p \) and \( \kappa_m \) denotes particle and medium compressibility correspondingly. The angle \( \theta = 0 \) along the axis of forward propagation.

For rigid sphere \( \rho_p/\rho_m \to \infty \) and \( \kappa_p/\kappa_m \to \infty \)

\[
\frac{I_s}{I_i} = \frac{k^4 a^6}{9 r^2} \left[ 1 - \frac{3 \cos \theta}{2} \right]^2
\]  

(1.6)

where \( I_s \) is the intensity scattered from the sphere.

For the sphere, backscattering occurs for \( \theta = \pi \) and thus the ratio of backscattered to incident intensities is given by

\[
\frac{I_s}{I_i} = \frac{25 k^4 a^6}{36 r^2}
\]  

(1.7)

The scattering cross section is

\[
\sigma = \frac{7 \pi k^4 a^6}{9},
\]  

(1.8)

see [92, 202, 165].

Scattered intensity/magnitude is dependent on the power of \( 4/2 \) on frequency, since \( I \propto \rho^2 \). Diffusive scatterers are often referred to as Rayleigh scatterers, characterized by this type of dependence on frequency and magnitude.
1.2. Physical Background

The constructive and destructive interference between a multitude of diffusive scatterers in the resolution cell, results in the condition known as speckle. A small point mass will return an echo with amplitude \( \sigma_s \) upon being hit by an ultrasound beam. When several adjacent point masses are hit by the beam, their respective echoes will interfere and summation of the echoes will create a complex signal. The interference creates a granular pattern, commonly referred to as speckle, when we observe the signal in an image.

Finally, diffractive scattering refers to the case in between specular and diffuse. Many different scattering regimes are contained in this category.

Total pressure is the sum of incident \( (p_i) \) and scattered fields \( (p_s) \).

\[
p(r, t) = p_i(r, t) + p_s(r, t) \tag{1.9}
\]

Small fluctuations \( (f) \) from a homogeneous region in e.g. tissue can be modeled using Born approximation [92].

Scattered pressure for a sphere of radius \( a \) is given by

\[
p_s(r) = \frac{\exp(ikr)}{r} \frac{k^2}{k_s^2} (\gamma_k + \gamma_p \cos \theta) (\sin k_s a - k_s a \cos k_s a), \tag{1.10}
\]
where \( k_s = 2k \sin(\theta/2) \), see [92, 202, 165].

Scattering cross section for sphere under diffractive scattering is

\[
\sigma_s(\theta) = \left[ \frac{k^2}{k_s^2} (\gamma_\kappa + \gamma_\rho \cos \theta)(\sin k_s a - k_s a \cos k_s a) \right]^2
\]  

(1.11)

The ultrasound echo is highly view-dependent, as described in [230] and [87]. The reflected intensity \( R(x) \), that is the intensity of the echo, is here modeled by

\[
R(x) = \rho(x) \cdot I_i(x) \cdot (\cos \phi(x))^m,
\]  

(1.12)

\( \phi(x) \) is the incident angle at position \( x \) and \( I_i \) the associated incident intensity. Furthermore, the reflection coefficient \( \rho \) is defined

\[
\rho = \left( \frac{z(x) - z(x - \delta d)}{z(x) + z(x - \delta d)} \right)^2,
\]  

(1.13)

with \( z \) denoting the acoustic impedance and \( \delta d \) the distance between points on a scan-line, see [92, 202, 165]. In this formula tissue heterogeneity, from specular to diffusive scattering, is handled by the exponent \( m \).

Another common problem of ultrasound is phase aberration. Phase aberration is a phenomena that occurs as the acoustic wave passes through inhomogeneous tissue, as the speed of the wave is retarded differently over different parts of the tissue. The misalignment of the echoes that occurs results in non-coherent summation of echoes and in decreased echo amplitude.

### 1.3 Transmission to Display of Ultrasound Echo

Medical ultrasound is used to image organ structures, by transmission an ultrasound into the body and then imaging the received echo. The process from transmission of an ultrasound echo into tissue, to the final display image is complex, and a detailed description is beyond the scope of this thesis. Thus we will here limit ourselves to a brief overview. For a more in depth treatment the reader is referred to works such as [92, 202, 165].

The transmission of the ultrasound signal into tissue (here represented by a stylized heart), and the subsequent reception of scattered echoes is shown in Fig. 1.3, and Fig. 1.4 shows the steps between reception of signal to the final display image are shown. We now give a more detailed description of the forementioned steps. A transducer with an array of piezoelectric elements, mounted on a probe, transmits a acoustic impulse into tissue and receives echoes from scattering structures. Dynamic range is adjusted before A/D conversion, where the analog signal is sampled. In A/D conversion the signal amplitude is measured according to a linear or non-linear function (e.g. a logarithmic transform).
and represented by a discrete number of histogram bins. This is called process is called quantization. A signal quantized into 4 bits, is represented by a histogram of 16 bins. Quantization can introduce problems, especially when the rate of quantization is too low. Quantization noise is a well-known problem; this degradation of the data can be described as additive noise. A specific problem observed in this thesis is due the effects of quantization on an US signal with low echo return.

The amount of gain varies with depth, referred to as Time Gain Compensation (TGC). Each sampled signal is delayed in order to focus in a certain direction and depth; this functionality is denoted as phased array, referring to the phase of the signal.

The delayed signal is weighted for purposes of getting appropriate apodization, that is enhancing signal at the center of the transducer to enhance signal quality, and beam profile. The phase of all weighted delayed signals is adjusted, so that all of them are in phase. The signals are then summed into one single signal: the so called radio frequency (RF) signal. The RF-signal is then bandpass filtered so that the frequency range of interest is obtained. In the majority of ultrasound equipment, the RF signal is not available to the user. RF signal is in ultrasound setting used to denote a signal with no loss of frequency information. Following bandpass filtering the RF signal is sampled at sampling rate $\Delta t$.

In order to be able to recover the original signal the waveform must be sampled minimally at the Nyqvist rate, $i.e.$ $\Delta t = 1/2f_{\text{max}}$ where $f_{\text{max}}$ is the maximum frequency of the waveform.
The sampled RF signal is decomposed into a real and imaginary component, through IQ-demodulation. This may be done by an inverse Discrete Fourier Transform, which is defined as follows. Given a periodic sequence of number \( \{f_k\}_{k=0}^{N-1} \) of period \( N \), the Discrete Fourier Transform of the sequence is defined as

\[
X[f_k] = F_n = \sum_{k=0}^{N-1} f_k \exp(-2\pi ink/N)
\]  

The inverse Discrete Fourier Transform is given by

\[
X^{-1}[F_n] = f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n \exp(2\pi ink/N)
\]  

The complex signal \( h(n) \), whose real part corresponds to the RF signal (albeit of course filtered), is obtained by applying \( X^{-1} \) on the complex spectrum \( H(m) \) of the sampled RF signal,

\[
h(n) = \left[ \frac{2}{\Delta t} \right] X^{-1}[H(m)]
\]  

\( \Delta t \) is the inter-sample spacing for integer values.
1.4. 3D Freehand Ultrasound

The same result may be achieved by applying the Hilbert transform to the sampled RF-signal, which is the general practice in application. The demodulation step also includes low-pass filtering, which drops the negative frequency components. In the case of the Hilbert transform, this is a side-effect of its application.

The In-phase $I = \text{REAL}[h(n)]$ and Quadrature $Q = \text{IMAG}[h(n)]$ components are then combined to form the envelope of the RF-signal,

$$\text{ENV}[h(n)] = \sqrt{I^2 + Q^2}$$

The envelope is then log-compressed in order to reduce its dynamic range to within the range of human visual perception. The envelope is remapped to a 2D spatial grid, using interpolation. Depending on machine architecture and user setting, several (possibly non-linear) post-processing steps follow, which may include e.g. further compression of the envelope, colorization of high and low amplitude areas etc. Finally the digital signal is converted to analog, through D/A conversion, for on-screen display.

1.3.1 Higher Level Signal Processing Functions

Examples of high level signal processing functions are Time Gain compensation (previously mentioned), frequency and spatial compounding. Frequency compounding is a speckle reduction method, which combines signals at different frequencies. This has the potential to achieve greater contrast since speckle may be more destructive at one frequency than at another. Spatial compounding reduces effects of speckle and adds new backscatter information by combining a various angles of incidence.

1.3.2 Types of Ultrasound Data Considered

The term B-mode is short for Brightness-mode, due to the fact that the strength of an echo is reflected in the brightness of area in the display. We will refer to the displayed B-mode ultrasound image on standard medical clinical ultrasound equipment, as Clinical B-mode ultrasound. In this thesis we deal with two types of data: clinical B-mode ultrasound and envelope detected RF data. Most works consider B-mode data since this is (generally) more compressed and thus leading to decreased computational load. In addition it is much more widely available for study [152], since the vast majority of available ultrasound equipment in medical institutions only delivers B-mode data.

1.4 3D Freehand Ultrasound

The 3D US RF freehand data used in this thesis is obtained using the optical tracking system NDI Spectra in conjunction with an Ultrasonix MDP US machine. The three-dimensional RF datasets were acquired using a phased-array probe with a frequency of
1.5 Reconstruction of 3D data from 2D data with spatial information

In this section we describe the way data is obtained for the experiments described in Sec. 5.8, essentially following the review of methods in Klein [113].
1.5.1 Background

Reconstruction methods for 3D volumes from 2D data can be divided into basically three categories: Voxel-based, pixel-based and function-based.

**Voxel-based** (also known as backward-warpping) reconstruction uses a 3D voxel grid space covering all spatially organized tracked 2D slices. McCann et al. [134], Sherebrin et al. [193] and Rohling et al. [177] uses a nearest neighbor scheme, where a voxel in the volume grid is assigned pixel value according to which image pixel is the closest in distance to the voxel. Some voxels may be left blank if the volume is not sampled densely enough. An extension of this is proposed in Trobaugh et al. [207] where, for a given voxel, the two nearest surrounding slices are determined by orthogonal projection. The intensity value of the voxel is then given by distance weighted interpolation. Coupe et al. [46] incorporates information about probe trajectory in the interpolation process, instead of using orthogonal projection. The limitation of the method described in Coupe et al. [46] is the assumption of constant trajectory speed, claimed to be reasonable by the authors when using a frame rate above 10 MHz. An efficient reconstruction scheme is proposed in Wein et al. [221], using four different weighting functions: Inverse Distance Weighting, Gaussian Weighting, Nearest Neighbour and Weighted Median.

Interpolating from scan-converted 2D images causes error accumulation due to multiple interpolation steps. To alleviate this, nearest neighbor 1D scanline interpolations can be used, as proposed in [19, 132]. 3D interpolation is achieved in Thune et al. [205] by first 1D scanline and then trilinear interpolation.

In **Pixel-Based** (or forward-warpping) reconstruction each pixel from the US slices is associated with a corresponding voxel in the volume, by using the transformation obtained by position information from the optical tracker. There exists several proposed methods for the assignment of intensity information to the voxel. In Hottier and Billon [93] the nearest neighbour approach is adopted, thus the voxel is assigned the value of the transformed pixel closest to the center of the voxel. This strategy can be quite unrobust w.r.t. to noise, so in [148, 74] a number of pixels neighbouring the voxel center are used to compute an average pixel value which is then assigned to the voxel. Pixel-based methods have a general weakness in that they may produce holes in the reconstructed volumes, since one generally sets a limit on what is to be regarded as a close neighbour. Averaging [148] may avoid this if one averages over a large enough neighbourhood. Other strategies to avoid holes include max [153] or median in a local neighborhood [180]. Another approach is to use a 3D kernel [181] on each voxel in order to assign an intensity value. Thus the intensity value is calculated by distance-weighting intensities in a local neighbourhood of the voxel center. A great benefit of pixel-based method is the speed at which it can be implemented, since the entirety of the data is not required to perform reconstruction, and so real-time methods have been proposed for CPU in [74] and GPU in [110]. Voxel-based methods do not have such a benefit.

**Function-based** reconstruction uses interpolation functions to reconstruct voxel intensities, from intensities sampled in an irregular spaced grid. The simplest form of
CHAPTER 1. Medical Ultrasound

interpolation function would be a polynomial, and then the coefficient of the polynomial would be computed in order for it pass through the pixel values. However, this simple strategy comes at the cost of high complexity. Thus it would be desirable to formulate an interpolation function which is not overly time-complex, while still providing a dense and smooth 3D reconstruction. The use of Radial Basis Functions (RBF) was proposed by [177]. In Sanches and Marquez [182] the Rayleigh distribution is used to model speckle noise in data formation. An optimal Bayes criterion is then used in filtering and interpolated the data, in reconstruction. This model was further improved in [183], especially w.r.t. speed, by proposing a multi-scale extension. Spatial compounding of ultrasound data, that is the composition of volumes using images from various transducer angles, to reduce SNR and thereby increase data quality is used in [190, 222, 14].

In the field of neurosurgery and neurology, US methods face challenges due to the skull bone, which often defocuses the beam and causes loss in spatial resolution and increased blur, largely due to phase aberration, see Sec. 1.2. This is addressed in [103, 104, 50], where real time phase aberration correction is used for a 2D array probe, and in Smith et al. [195] for use with a helmet of multiple transducers. Unsgaard et al. [208] provides an overview of the range of applications of 3D US in neurosurgery. One recent application of Doppler ultrasound in neurosurgery, specifically brain shift analysis, can be found in Reinertsen et al. [170] where a vessel tree is reconstructed from Doppler ultrasound images and then registered to the corresponding tree obtained from MRI.

1.5.2 Backward-Warping Reconstruction

We here give a short description of the fundamentals of voxel-based 3D freehand ultrasound reconstruction. In three-dimensional ultrasound reconstruction, the goal is to create a 3D voxel volume on a Cartesian grid, from a series of not-necessarily aligned 2D ultrasound images by use of position information obtained by a tracker. See Fig. 1.5 for an example of a freehand system for bilateral reconstructions, where three-dimensional data is obtained by an ultrasound transducer used in conjunction with a tracking system. A tracking device attached to the transducer and spatial information is obtained by a optical tracking system or an electro-magnetic sensor.

In the sequel, we describe the 3D reconstruction process. Denote the transformation from coordinate frame A to coordinate frame B as $B^T_A$. The following abbreviations are used: (T)racker, (P)robe (U)S Plane, (C)alibration phantom.

The transformations are Euclidean with 6 degrees of freedom (DOF) consisting of 3 rotations ($\alpha, \beta, \gamma$) and 3 translations ($x, y, z$), such that $B^T_A$ decomposes to

16
1.5. Reconstruction of 3D Data

\[ BTA(\alpha, \beta, \gamma, x, y, z) = \begin{pmatrix}
\cos \alpha \cos \beta & \cos \alpha \sin \beta \sin \gamma - \sin \alpha \cos \gamma & \cos \alpha \sin \beta \cos \gamma + \sin \alpha \sin \gamma \\
\sin \alpha \cos \beta & \sin \alpha \sin \beta \sin \gamma + \cos \alpha \cos \gamma & \sin \alpha \sin \beta \cos \gamma - \cos \alpha \sin \gamma \\
- \sin \beta & \cos \beta \sin \gamma & \cos \beta \cos \gamma \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
x \\\
y \\\
z
\end{pmatrix} \]  

(1.18)

A tracking object is a rigidly attached to the probe, allowing the spatial position to be determined w.r.t. the tracking device coordinate system, \( TTP \). A calibration of the system is needed, since it is necessary to correlate the pixel positions in each 2D US image with position in the 3D space. This calibration gives the coordinate system \( TTU \). A review of the main different calibration methods is given in [167, 95, 135]. Single wall calibration has frequently been preferred due to its simplicity and relatively high accuracy.

The calibration setup is as follows: A nylon membrane, representing a planar object, is immersed in a water, and is imaged at a variety of angles and positions, and appears as a (distorted) line in the ultrasound image. The origin of the phantom coordinate system is placed within the plane \( z = 0 \), which yields the equation system

\[ \begin{pmatrix}
x_k \\
y_k \\
z_k
\end{pmatrix} = CTT \cdot TTP \cdot PTU \cdot \begin{pmatrix}
s_x u_k \\
s_y v_k \\
0 \\
1
\end{pmatrix}, \]  

(1.19)

where \( x_k, y_k, z_k \) denote the 3D positions on the \( k \):th plane in the world coordinates, under constraint of \( z = 0 \). The B-mode image then provides the ultrasound plane, whose coordinates are \( u_k, v_k \). Scale factors \( s_x, s_y \) are used to convert pixel scale to metric scale (mm). From Eq. 1.19 we see that an imaged line generates two equations. For calibration 11 parameters are estimated: \( CTP \) (2 rotations, 1 translation), \( TTU \) (3 rotations, 3 translations) and the scale factors \( s_x \) and \( s_y \). A series of ultrasound images are acquired and each put into Eq. 1.19, yielding an overdetermined set of equations. This set of equations is then solved using an optimizer e.g. Levenberg-Marquardt. The calibrated probe and tracking data then allows us to spatially locate each pixel in the ultrasound plane w.r.t. a 3D reference coordinate system. Specifically, by using the estimated matrices \( TTP \) and \( PTU \), we can use the relation \( TTP \cdot PTU = TTU \), to find the spatial location of each pixel in the coordinate system of the tracker. By acquiring enough to ultrasound images to cover the three-dimensional object we wish to image, we then have a large number of data points and their corresponding spatial location. We use so called backward warping to compute voxel intensities. Backward warping amounts to accumulating, by some distance-weighted transform, intensity information by traversing a grid which has been placed on top of the measurements. Forward warping on the other hand maps the pixel
measurements directly into the volume, which can result in gaps. Some forward warping uses averaging over voxels, but this is still problematic should the voxel be empty, since this would produce a hole in the reconstruction. Commonly considered distance-weighted transforms are described for use in backward warping to reconstruct intensities: Nearest-Neighbour (NN) and Gaussian-kernel smoothing, inverse-distance weighting and weighted median. We describe the first two, since the first one is the most simple and commonly used, and since we apply the second one in our work. A description of the remaining methods can be found in [221].

To define these transforms \( f \) we equip each measurement \( y_i \) with a scalar \( d_i \), defined as the distance from the spatial location of the measurement to the discrete position \( x_i \) in the 3D volume consisting of all the ultrasound slices obtained during acquisition. The set of distance-intensity tuples \( (y_i, d_i) \) is denoted \( A = \{(y_i, d_i)\} \).

**Nearest-Neighbour (NN)**

The simplest and most computationally effective intensity reconstruction method. NN assigns the measurement \( y_i \) which is closest to the position \( x_i \):

\[
 f(A) = y_i \mid \min_i d_i.
\]  

A deficiency of NN is that it often results in blocky and non-smooth reconstructions.

**Gaussian-Kernel Smoothing**

Gaussian kernel smoothing is defined as

\[
 f(A) = \frac{\sum_{i=1}^{n} y_i e^{-d_i^2/\sigma^2}}{\sum_{i=1}^{n} e^{-d_i^2/\sigma^2}},
\]  

where \( \sigma \) determines the size of the Gaussian kernel.

**1.6 Applications in Cardiology**

**1.6.1 Anatomy and Function of the Human Heart**

The human heart is a serially connected muscle pump responsible for maintaining circulation of blood to the organs of the body. The atrium and chamber of the left side of the heart provides blood to the entire body via the arteries. The blood then returns to the right side of the heart via the veins. Subsequently the right side of the heart pumps the deoxygenated blood to be oxygenated in the lungs, after which it is returned to the left side of the heart. Finally the newly oxygenated blood is pumped out into the body. The valves of the heart prevent the blood from flowing in the wrong direction in the cycle described above.
1.6. Applications in Cardiology

The heart is dependent on an electrical pulse in order to be able to contract. In the so-called systolic state, both chambers of the heart contract and pump blood in each of its circulatory tracts, that is lungs and body. In the diastolic state the chambers expand, and atria are filled with blood and contract, at which point the chambers are once again filled. The electrical impulses of the heart start at a group of specialized cells, the sinoatrial node, which initiates contraction and then travels to the atria which in turn pushes blood into the chamber. The AV node, which is located between atrium and chamber, then delays the impulse so that the chambers have enough time to fill with blood. After the chambers have filled the AV-node forwards the electrical signal to the chambers which then contract and pump the blood further into the body. The blood supply is maintained by the regular contraction of the heart at a frequency of approx. 60-70 beats per minute. The phase at which the heart contracts, that is when the left chamber is the smallest, is called end-systole. After each contraction, the muscle relaxes and cavities of the heart are filled with blood. This is known as the end-diastolic phase.

1.6.2 Heart Disease and Ultrasound

The heart is subject to many potential ailments, including arrhythmia, heart muscle diseases, valve diseases and circulatory diseases, which are often a consequence of each other. Heart diseases may be congenital, a consequence of volume- and pressure stress or poor circulation in the heart muscle (ischemia). Using ultrasound it is possible to access regional muscle contractability. A measure of the systolic function of the heart is ejection fraction (EF), which describes the estimated percentage amount of blood that the heart expels during a heart beat. EF estimation is central in the assessment of the heart’s systolic function, but is highly subjective and therefore need to be interpreted with care.

Ejection fraction is defined as

\[ EF = \frac{(\text{end-diastolic volume}) - (\text{end-systolic volume})}{(\text{end-diastolic volume})} \times 100. \quad (1.22) \]

The four main diagnostic views of the heart are parasternal long and short axis, and apical two- and four chamber, see Fig. 1.10. The three main views that have been studied in the context of ultrasound segmentation algorithm are two and four chamber apical views and parasternal short axis views (SAX). The apical views of the heart are four-chamber (4C), two-chamber (2C) and long-axis view (APLAX). We here give examples of end-systolic and end-diastolic frames of each of the forementioned views, see Fig. 1.6, 1.7, 1.8 and 1.9. Typically SAX has been preferred in US segmentation, due to the less pronounced drop-out of heart chamber walls. However obtaining SAX images is of less use to the clinician, since the most common way to obtain EF is through LAX images, due the high difficulty of obtaining the correct plane in SAX images.

The most commonly used 2D measurement for volume measurements is the biplane method of disks (modified Simpson’s rule) and is the currently recommended method of
CHAPTER 1. Medical Ultrasound

Figure 1.6: Parasternal Long Axis end-systole (left), end-diastole (right)

Figure 1.7: Apical 2-chamber end-systole (left), end-diastole (right)

An overview of other methods for measuring volume is also given in Lang et al. [120]. The underlying principle of this method is that the total left ventricular (LV) volume is calculated from the summation of a stack of elliptical disks. The height of each disk is calculated as a fraction (usually 1/20) of the LV long axis based on the longer of the two lengths from the 2- and 4-chamber apical views. The cross-sectional area of the disk is based on the two diameters obtained from the 2- and 4-chamber views. When two adequate orthogonal views are not available, a single plane can be used and the area of the disk is then assumed to be circular. Extensive wall-motion abnormalities severely limit the usefulness of a single plane.

There are many approaches to solving the problem of segmentation of anatomical structures by use of computer vision algorithms, e.g. the endocardium in the left-ventricle of the heart. These include Active Contours [141], Variational Level Set formulations [227, 127, 42], Bayesian [28, 198, 223, 67, 138], and variants of Active Appearance Motion models [26, 142]. An excellent overview of a large part of existing methods is given in Noble et al. [151]. Many works concern the segmentation of the endocardium.
in the parasternal short axis view of the heart cf. [117, 23, 141, 138, 227, 127]. The more challenging apical long-axis view, more subject to signal dropout and out-of-plane motion (resulting in foreshortening, which affects the appearance of heart walls) has been considered in [26, 233, 28, 42]. This view commonly suffers from partially missing heart walls, which have been dealt by use of shape prior formulations e.g. in [26, 233, 42] and spatio-temporal continuity [26, 233, 28, 42].

Several works have focused on modeling the speckle noise as an integral part of the segmentation model. cf. [152, 151] for an exhaustive survey. In Zhu et al. [233], mixtures of Nakagami densities are used to model the blood and background distribution, and a mixture of shifted Rayleigh distribution is used in Zhu et al. [232]. In Paragios et al. [158] the Exponential distribution is used to model the blood pool, while the Gaussian distribution is used for the tissue/walls. Many previous works use single component distributions, such as the Rayleigh [57, 68], Shifted Rayleigh [139], Gaussian [28, 6, 171, 80], Gamma [203], and Beta [133] distribution.
1.6.3 2D B-Mode Cardiac US Data

The 2D B-mode cardiac US data in this thesis was obtained, at the setting 1.4, 2.4 or 3.2 MHz, using the echocardiogram machines Philips Sonos 7500, Philips iE33 or GE Vivid 7, from consecutive adult patients referred to the echocardiography laboratory at Malmö University hospital, Sweden which has a primary catchment area of 250,000 inhabitants. Expert outlines of the endocardium in the sequences have been provided by the same hospital.

1.7 Applications in Neurology

In the sequel we will give a short overview of the brain anatomy, and then how ultrasound may be used as a diagnostic tool in neurology.
1.7. Applications in Neurology

1.7.1 Anatomy and Function of the Human Brain

Brain is the central control organ of the body, consisting of the forebrain, the brainstem, and the hindbrain, see Fig. 1.11. The forebrain consists of the telencephalon, cerebral cortex, the thalamus and hypothalamus. The telecephalon in turn consists of cerebral cortex and hypothalamus. The cerebral cortex and thalamus processes the information which is received by sensory organs. The hypothalamus regulates hunger, thirst and arousal. The brainstem is composed of the midbrain and hindbrain, and relays information back and forth between the forebrain and sensory organs. The hindbrain consists of the medulla oblongata, the pons and cerebellum. The medulla governs involuntary functions such as respiration and cardiac processes, while the cerebellum is responsible for coordination of movement. A easily accessible description of this topic can be found in Haycock et al. [86].

Having given a general overview of the major regions of the brain, we will now focus on a specific part of the midbrain, namely the substantia nigra. Substantia nigra is latin for "black substance". The black color is due to the melanin in the dopaminergic neurons. Parkinson's disease is associated to changes in substantia nigra, specifically to death in dopaminergic neurons. The midbrain is illustrated in Fig. 1.12.

1.7.2 Ultrasound in Neurology

We here give an overview of works of ultrasound in neurology, essentially following the review in Klein [113]. Transcranial ultrasound has emerged as a potentially useful technique in providing early diagnosis of Parkinson's disease. By directing the ultrasound beam...
CHAPTER 1. Medical Ultrasound

through the narrow bone window at the temporal lobe, image data is obtained of the mid-brain area, see Fig. 1.11. Since the mid 90’s this analysis has been done using 2D images [13]. Among the diseases that can be diagnosed using this ultrasound-based technique are Parkinson’s [12], atypical Parkinson syndromes [15, 219, 218], depression [13, 11], Dystonia [147] as well as multiple sclerosis [91, 220]. Development of Parkinson’s disease is in 90% of the cases correlated with hyperechogenic substantia nigra [17, 217], specifically with the agglomeration of ferrite deposits [18], which are responsible for the hyperechogenicity. The difficulties in using 2D imagery when performing Parkinson’s disease detection, are chiefly due to specific machine architecture and skill of operator. In order to diagnose by ultrasound, one needs to obtain a specific plane cutting through the midbrain, at which the circumference of the hyperechogenic substantia nigra is measured. Different machine and settings will require different thresholds when making a diagnosis, and crucially the selection of the diagnostic plane by the operator will play a huge role in whether a useful diagnosis can be rendered using 2D imagery. This results in increased inter- and intra-observer variability [194, 212]. These factors of uncertainty combined lead to a concurrence of 48-100% of ultrasound diagnosis compared to the corresponding clinical diagnosis. A complicating factor is the quality of the bone-window through which the probe is directed; poor quality leads to inconclusive diagnosis [213]. This motivates the use of 3D ultrasound, as it provides a wider field-of-visions circumventing some of the drawbacks of the 2D approach. In Fig. 1.5 the setup for acquisition of bilateral 3D freehand scans is detailed. A reference target is attached to the patient and the US data volume is reconstructed w.r.t. to the target. Thus multiple ultrasound sweeps can be merged into a single volume, e.g. into a bilateral volumes which are clinically useful, due to the clear visibility of the skull in these volumes. The skull being visible aids in obtaining orientation, as well as making registration tasks easier. See Fig. 1.13 for a 3D bilateral transcranial scan reconstruction and midbrain structures therein.

There exist several potential benefits with using 3D data generated from 2D images by means of a tracked transducer: (1) reduction in subjectivity of diagnosis, since no
specific ultrasound plane has to be selected for classification and (2) reduction of drop out rate related to poor imaging of intracranial structures. In order to track the change in hyperechogenic substantia nigra regions in patients over time, one needs to recover virtually the same plane every time, complicating the 2D approach.

A first pilot study of 3D US segmentation of the midbrain and the substantia nigra was conducted by Plate et al. [162]. This study consisted of 23 subjects, where 11 patients had the diagnosis Parkinson's disease and another 11 were healthy. The final subject was inconclusive due to no or partially visualizable intracranial structures. Blinded segmentations were performed on the remaining subjects and used for evaluation.

A support vector machine (SVM) algorithm using Radial Basis Functions (RBF) is used to classify the data. The parameters of the SVM are variance $\sigma$, complexity $C$ controlling the smoothness of the classifying hyper-plane. Automatic grid search gives the parameter ranges $\sigma \in [10^{-8}, 10^6]$ and $C \in [10^{-2}, 10^9]$.

Out of both planar and volumetric measures the best classification results of 90.9% sensitivity and 72.7% specificity, are obtained by a multi-dimensional feature vector consisting of the 3D freehand reconstructed volumes of substantia nigra echogenicities in the left and right hemispheres, and the midbrain volume. Sensitivity describes the test’s ability to identify positive results and specificity relates the ability of the test to identify negative results. Using 3D volumes for classification radically reduces the drop-out rate, compared to the 2D method, from around 10-20% [217] to around 4%. Since large (100+ subjects) independent studies of substantia nigra area classification on 2D B-modes images have yielded sensitivity and specificity of around 90% [16], using volumetric measures may further improve these results, since volumetric measures outperform planar in our study. Also since drop-out rate is improved, using 3D reconstructed volumes for classification promises less subjectivity and dependence on bone window quality in diagnosis.
Figure 1.13: Ultrasound slices obtained from transcranial reconstructions. (a) bilateral 3D free-hand ultrasound transcranial brain scan, midbrain (red). (b) magnified midbrain area with hyper-echogenic substantia nigra (red). Images reprinted with permission of Tassilo Klein, TU München, Germany.
Chapter 2

Statistical Modeling of the Ultrasound Signal

2.1 Statistical Modeling of Ultrasound Data

The statistical properties of the received echo signal at the transducer has been shown to depend on the density (number of scatterer in the resolution cell of the system) and spatial distribution of scatterers. Different types of biological tissue have different characteristics w.r.t. density and scatterer distribution, and hence this can be exploited in statistical modeling of tissue types.

A key characteristic of an ultrasound image is its granular appearance, also known as speckle. Speckle is an interference pattern, where the imagined data contains a multitude of diffusive scatterers. The speckle pattern is completely deterministic, given that the location of all scatterers in the medium and transducer settings are known. Since in real application the location of all subresolvable scatterers is not possible to determine, their location is considered random. Due to this, one often refers to speckle as a random deterministic interference phenomena [151]. Speckle can either be removed, through speckle reduction methods, or treated as a property of the ultrasound data. In this thesis we adopt the latter view and model the data statistically, letting the speckle remain in our data.

When modeling the echo envelope statistically two factors primarily need to be considered: density and organization of the scatterers. In Table 2.1 an overview is given of appropriate distributions for the four different conditions of density and organization for envelope detected Radio-Frequency (RF) data. When the ultrasound data has been subject to post-processing (e.g. quantization, log-compression) the distribution of the data is affected. This case will be addressed in Sec. 2.1.1. Naturally Table 2.1 presents a simplification, since density and level of coherence is a sliding scale.
Table 2.1: Statistical models for varying degrees of density and organization of scatterers in medium.

<table>
<thead>
<tr>
<th>Level of organization</th>
<th>density</th>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>K</td>
<td>Nakagami</td>
<td>General</td>
</tr>
<tr>
<td>high</td>
<td>General</td>
<td>Rice</td>
<td>General Nakagami</td>
</tr>
</tbody>
</table>

The general models, which can cover all four conditions, are the homodyned K-distribution (see e.g. [106] and [61]), the generalized K-distribution, the generalized Nakagami distribution, the Nakagami-generalized inverse Gaussian distribution and Rician Inverse of Gaussian distribution [64]. An advantage of these models is that parameters can be given physical meaning [55]. However, the analytical complexity associated with these families is significant.

Two cases of scattering admit relatively simple analysis, namely high density scattering with high and low coherence. It is instructive to consider these cases to give an insight into the statistical modeling of the echo envelope. Following this we relate results which show that the Nakagami distribution, our distribution of choice for envelope detected data, models exactly or approximates three of the forementioned configurations of density and coherence.

In the following derivations concerning the Rayleigh and Rice distribution we essentially follow the exposition given in Dutt et al. [60].

For the special randomly located scatterers of high density, the statistics of the envelope signal shows a Rayleigh distribution. In this case the speckle is called fully developed. The Rayleigh distribution is arrived at as follows. Consider the backscattered echo is the sum of echoes from the scatterer in the resolution cell. Each scattered echo is a phasor of amplitude $x_j$ and angular frequency $\omega_0$, given that the input signal has angular frequency $\omega_0$. The backscattered echo is then

$$X = \sum_{k=0}^{N-1} x_k e^{i\theta_k} e^{i2\pi\omega_0t}$$

(2.1)

The component time-varying component $e^{i2\pi\omega_0t}$ is dropped as it does affect
2.1. Statistical Modeling of Ultrasound Data

the statistics of the backscattered echo. Normalizing, we express $X$ as

$$X = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \xi_k \exp\{i\theta_k\} \quad (2.2)$$

where $\xi_k = \frac{x_k}{\sqrt{N}}$. Assuming that $\xi_j$ and $\theta_j$ are independent and identically distributed (iid), then the expectations of the real ($X_r$) and imaginary ($X_i$) parts of $X$ are

$$E\{X_r\} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} E\{\xi_k\} E\{\cos \theta_k\} \quad (2.3)$$

and

$$E\{X_i\} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} E\{\xi_k\} E\{\sin \theta_k\} \quad (2.4)$$

If the angles $\theta_j$ are uniformly distributed on $[-\pi, \pi]$, then $E\{\cos \theta_j\} = E\{\sin \theta_j\} = 0$ which in turn implies that $E\{X_r\} = E\{X_i\} = 0$. The second order moments of $X_i$ and $X_r$ are

$$E\{X_r^2\} = \frac{1}{N} \sum_{j,k=0}^{N-1} E\{\xi_j\xi_k\} E\{\cos \theta_j \cos \theta_k\} \quad (2.5)$$

and

$$E\{X_i^2\} = \frac{1}{N} \sum_{j,k=0}^{N-1} E\{\xi_j\xi_k\} E\{\sin \theta_j \sin \theta_k\} \quad (2.6)$$

Since $\xi_j$ are assumed independent,

$$E\{X_r^2\} = E\{X_i^2\} = \frac{E\{X^2\}}{2} = \sigma^2 \quad (2.7)$$

It is easily seen that $E\{X_r X_i\} = 0$, that is $X_r$ and $X_i$ are uncorrelated, and thus the central limit theorem gives that the joint distribution of $X_r$ and $X_i$ is

$$p(X_r, X_i) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{X_r^2 + X_i^2}{2\sigma^2}\right\} \quad (2.8)$$

The echo envelope is the amplitude of the complex phasor,

$$X = \sqrt{X_r^2 + X_i^2} \quad (2.9)$$

Changing to polar coordinates, Eq. 2.8 becomes

$$p(X, \phi) = \frac{X}{2\pi\sigma^2} \exp\left\{-\frac{X^2}{2\sigma^2}\right\} \quad (2.10)$$
CHAPTER 2. Statistical modeling of the ultrasound signal

where $X = \sqrt{X_r^2 + X_i^2}$ and $\phi = \tan^{-1}(\frac{X_i}{X_r})$. Marginalizing over $\phi$,

$$f_{\text{rayl}}(X \mid \sigma^2) = \int_{-\pi}^{\pi} p(X, \phi) d\phi = \frac{X}{\sigma^2} \exp\{-\frac{X^2}{2\sigma^2}\}$$

(2.11)

we arrive at the Rayleigh density. Total backscattered energy under fully formed speckle conditions is $\sigma$.

The Signal-to-Noise-Ratio (SNR) of the Rayleigh distribution is given by

$$\text{SNR}_{\text{rayl}} = \frac{E\{X\}}{\sqrt{E\{X^2\} - E^2\{X\}}} = \frac{\sqrt{\frac{\pi}{2} \sigma}}{\sqrt{2 - \frac{\pi^2}{2} \sigma}} \approx 1.91$$

(2.12)

and thus can be used as an indicator of fully formed speckle.

The Rice distribution has been suggested [216, 101] for the condition where there exists some coherence or structure in subresolvable scatterers, or if there is strong specular scattering. This model however does not account for the case where scatterer density is low. To derive the Rice distribution, the same phasor model as in fully formed speckle case is used, with the modification that a deterministic coherent component $s$ is added,

$$X = s + \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \xi_k \exp\{i\theta_k\}.$$  

(2.13)

The joint distribution function of $X_r$ and $X_i$ is

$$p(X_r, X_i) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{(X_r + s)^2 + X_i^2}{2\sigma^2}\}$$

(2.14)

This can be reformulated using the modified Bessel of the first kind of order zero,

$$I_0(z) = \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m + 1)} \left(\frac{z}{2}\right)^{2m}.$$  

(2.15)

s.t.

$$f_{\text{rice}}(X \mid s, \sigma) = \frac{X}{\sigma^2} \exp\left(-\frac{(X^2 + s^2)}{2\sigma^2}\right) I_0\left(\frac{sX}{\sigma^2}\right),$$

(2.16)

which is the Rice Distribution with parameters $s$ and $\sigma$. Now define the ratio of coherent to diffuse scattering as $k = \frac{s}{\sigma}$. It is clear that the Rayleigh and Rice distribution are related, if $k = 0$, the Rice distribution becomes Rayleigh. When $k \to \infty$ the Rice distribution becomes Gaussian.

For the Rice distribution the SNR is given by

$$\text{SNR}_{\text{rice}} = \frac{\sqrt{\pi/2} \exp\{-\frac{k^2}{4}\}\{(1 + \frac{k^2}{2})I_0(\frac{k^2}{4}) + \frac{k^2}{2} I_1(\frac{k^2}{4})\}}{\sqrt{2 + k^2 - \frac{\pi^2}{2} \exp(-\frac{k^2}{4})\{(1 + \frac{k^2}{2})I_0(\frac{k^2}{4}) + \frac{k^2}{2} I_1(\frac{k^2}{4})\}^2}}$$

(2.17)
2.1. Statistical Modeling of Ultrasound Data

\( SNR_{\text{rice}} \) will increase linearly with the coherent component \( s \), for \( SNR_{\text{rice}} > 1.91 \). For large \( k (k >> 1) \) \( SNR_{\text{rice}} \) will be approximately \( k \).

In the case of low scatter density (partially developed speckle), the K-distribution \([106, 188, 189, 106]\) has been shown to be appropriate. The density of the \( K \) distribution is

\[
f_K(A \mid \sigma^2, \alpha) = \frac{4A^{\alpha-1+n/2}}{(2\sigma^2)^{(\alpha+n/2)/2}\Gamma(\alpha)\Gamma(n/2)} K_{\alpha-n/2}(\sqrt{\frac{a}{\sigma^2}}A) \tag{2.18}
\]

where \( \alpha > 0, \sigma^2 > 0 \) and \( K_p \) is the modified Bessel function of the second kind of order \( p \). Here \( \alpha \) is the effective density of random scatterers. However the K-distribution does not account for the presence of structure (coherence) in the speckle.

Often one is able to model an echo by a simpler model than K-distribution, like Nakagami or Gamma, but in these cases the parameters have no physical meaning. Since we do not apply our distributional modeling to a characterization problem, in which the meanings of distributional parameters may be employed, we do not need employ, the often harder estimate, models like the K-distribution.

The Nakagami distribution \([188, 146]\) is given by

\[
f_{\text{nak}}(x \mid \mu, \omega) = \frac{2\mu^\mu}{\Gamma(\mu)\omega^\mu} x^{2\mu-1} \exp \left( -\frac{x^2}{\omega} \right), \tag{2.19}
\]

where \( \mu \) and \( \omega \) is the shape and scale parameters, respectively. The Nakagami distribution covers three of the above scattering cases, namely high and low density coherent scattering, and low coherence random scattering. In order to see this one can show that the Nakagami distribution can either model exactly or approximate the Rayleigh, Rice and K-distribution. Putting \( \mu = 1 \) one obtains

\[
f_{\text{nak}}(x \mid 1, \omega) = \frac{2x}{\omega} \exp \left( -\frac{x^2}{\omega} \right), \tag{2.20}
\]

which is the Rayleigh distribution.

The following theorem states the relationship between Nakagami and Rice distribution in terms of the Kullback-Leibler Distance (defined in Sec. 2.8),

**Theorem 2.1.1.** \((2.5)\) Let \( \mu = \frac{x^2+2\sigma^2}{4\sigma^2(x^2+\sigma^2)} \) and \( \Omega = e^2 + 2\sigma^2 \) Then

\[
KL(f_{\text{Ri}}(\epsilon, \sigma^2), f_{\text{nak}}(\mu, \Omega)) \leq 0.02 \tag{2.21}
\]

Lastly,

**Theorem 2.1.2.** \((2.5)\) Let \( \mu = \frac{a}{\sigma+1} \) and \( \Omega = 2\sigma^2e \) Then

\[
KL(f_K(\epsilon, \sigma^2), f_{\text{nak}}(\mu, \Omega)) \leq 0.0035 \tag{2.22}
\]
Thus Nakagami distribution can thus cover a large part of possible scattering conditions, excepting low-density coherent scattering. For our uses we have found Nakagami to be sufficiently flexible. The Gamma distribution may be used to model Nakagami distributed data through the connection

\[
\text{If } Y \sim \text{Gam}(\mu_{\text{gam}}, \omega_{\text{gam}}) \text{ and } X \sim \text{Nak}(\mu_{\text{nak}}, \omega_{\text{nak}}) \text{ then } X^2 = Y(\mu_{\text{nak}}, \omega_{\text{nak}}/\mu_{\text{nak}}) \quad (2.23)
\]

2.1.1 B-mode Ultrasound Data

In most clinical ultrasound machines, the RF signal is not available to the user. Instead the user is presented with data that has undergone various forms of manipulation. The most basic of these is log-compression, which is essential for viewing the data, since the range of the envelope detected data is too great to admit viewing it on a screen. Log-compression is however a non-linear transform, making the process of recovering the original uncompressed signal more complicated. This has been addressed in Shankar et al. [191], where the Nakagami parameters where estimated from log-compressed envelope detected RF data. Note that no quantization or censoring was taken into account in this model.

Log-compression is however not the only transform applied in the formation of the data available to the user standard medical ultrasound equipment. An insidious factor is the left-censoring of data, which means that a significant part of the data has been discarded and set to a constant value. Visually, this affects the data only marginally, but it obviously affects the statistical modeling of the data. The data may also be further compressed to allow more efficient storage. Modeling B-mode data requires an experimental approach, e.g. when selecting distribution type, due to the unknown factors behind its formation.

2.1.2 On the modeling of clinical B-mode data in this thesis

We have used different approaches to modeling clinical B-mode data in our work. In Sec. 3.10 we model the data by a single distribution model, in this case a Rayleigh distribution to model the inside and outside of the endocardium. In Sec. 3.11 we employ a mixture of Rayleigh densities to model to do the same. Finally in Sec. 4.2 we also use a mixture model, but we take into account the fact that the data has been censored at a, to the user, unknown cut-off point due e.g. quantization (resulting in an unusually large first bin in the histogram, see Fig. 2.1). This way of modeling has evolved over the time we have been working with this data, and we believe the latest model, that is the one described in Sec. 4.2 to be the one that most faithfully models our B-mode data, as it takes into account censoring.

In a mixture model setting we have, in this thesis, dealt with this in two ways: (1) By modeling the first bin as a point distribution (Dirac-delta), and thus assuming that the intensities in the first bin are separate form the intensities in the remainder of the
2.2 Mixture Models

Assume that the random variables $X_i$ are distributed according to the density $f_j$ with probability $\pi_j$. The probability $\pi_j$ is commonly referred to as a mixture weight or mixing proportion, and the density $f_j$ is referred to as a component density.
The densities \( f(x_i) \) of the random variables \( X_i \) can be written in the form

\[
f(x_i) = \sum_{j=1}^{d} \pi_j f_j(x_i),
\]

where \( f_j(x_i) \) are densities and \( \pi_j \) are non-negative relative frequencies, that together sum to one, that is \( 0 \leq \pi_j \leq 1 \) s.t. \( \sum_k \pi_k = 1 \).

A sample of independent random variables \( (X_1, \ldots, X_n) \) then has the likelihood

\[
\prod_{i=1}^{n} \{ \pi_1 f_1(x_i) + \cdots + \pi_d f_d(x_i) \}
\]

Mixture models are flexible and have found application in various scenarios, discriminant analysis, image analysis and survival analysis. They occupy the border area between parametric and non-parametric models. Mixture models are parametric in that component densities are parametric, but non-parametric if we allow the number of components to vary freely.

Maybe the most intuitive application of mixture models is to the categorization of distinct groups in a population, s.t. each component corresponds to a distinct group in the data. Another case is where a group in the data may correspond to a mixture of component densities, this commonly occurs in medical imaging, e.g. an organ which is composed of several type of tissue. The advent of the EM algorithm (1977) has played a huge role in popularizing the use of mixture models, since EM makes it possible to easily estimate MLE’s for mixture models.

### 2.3 Bayesian Statistical Formulations

First assume that \( \theta \) is a parameter of a model and \( x \) is observed data. Bayesian formulations differ from standard frequentist formulation, in that The Bayesian statistician considers the probability the distribution of the parameter \( \theta \) conditioned on the observed data \( x \), i.e \( p(\theta \mid x) \), while the frequentist draws inference from how much an estimate, say \( \hat{x} \), of the parameter varies over repeated samples of data \( x \).

Bayesian formulations at the core concern probability statements about parameters given data. Consider the joint distribution \( p(\theta, x) \) and the apply Bayes rule

\[
p(\theta, x) = p(\theta)p(x \mid \theta),
\]

where \( p(\theta) \) is prior distribution and \( p(x \mid \theta) \) is referred to as the sampling distribution.

The posterior density is now defined as

\[
p(\theta \mid x) = \frac{p(\theta, x)}{p(x)} = \frac{p(\theta)p(x \mid \theta)}{p(x)}
\]

34
2.4 Maximum Likelihood Estimators

where

\[ p(x) = \int p(\theta)p(x \mid \theta) \, d\theta \]  

(2.28)

The unnormalized posterior density is

\[ p(\theta \mid x) \propto p(\theta)p(x \mid \theta) \]  

(2.29)

and so

\[ p(\theta, x) \propto p(\theta \mid x) \]  

(2.30)

Thus inferences about the joint probability \( p(x, \theta) \) can be made through the posterior \( p(\theta \mid x) \) since the two quantities are equal up to a constant.

2.4 Maximum Likelihood Estimators

Assume that \( x = (x_1, \ldots, x_n) \) is an iid sample from a population with density \( f(x \mid \theta_1, \ldots, \theta_k) \). In this case the likelihood function is

\[ L(\theta \mid x) = L(\theta_1, \ldots, \theta_k \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta_1, \ldots, \theta_n) . \]  

(2.31)

When \( X_i \) is not iid, the likelihood is

\[ L(\theta \mid x) = f(x_1, \ldots, x_n \mid \theta) \]  

(2.32)

taken as a function of \( \theta \). Thus the likelihood function is the joint density function. The value \( \hat{\theta} \) which is the parameter at which \( L(\theta \mid x) \) attains its maximum, is the maximum likelihood estimator (MLE) \([169]\).

Example 2.4.1. Gamma MLE.
Recall the Gamma distribution

\[ G(x \mid k, \theta) = \frac{x^{k-1}}{\Gamma(k)} e^{-\frac{x}{\theta}} . \]  

(2.33)

where \( k, \theta > 0 \).

The log-likelihood is now

\[ \log L(k, \theta) = \log \prod_{i=1}^{n} G(x_i \mid k, \theta) = \]

\[ (k - 1) \sum_{i=1}^{n} \log x_i - \log \Gamma(k) - k \log \theta - \frac{1}{\theta} \sum_{i=1}^{n} x_i \]  

(2.34)
We find the stationary point of the log-likelihood w.r.t. the scale parameter $\theta$

$$\frac{\partial \log L(k, \theta)}{\partial \theta} = 0 \iff \hat{\theta} = \frac{\bar{x}}{k}$$

(2.35)

By substituting $\hat{\theta}$ into Eq. 2.34 we get

$$\log L(k, \hat{\theta}) = (k - 1) \sum_{i=1}^{n} \log x_i - \log \Gamma(k) - k \log \bar{x} - k \log k - k \frac{1}{n} \sum_{i=1}^{n} x_i$$

(2.36)

This is a non-linear objective function of one variable, that is the shape parameter $k$, which can be solved e.g. by a direct search method ad the Nelder-Mead method. This way of dimension reduction is sometimes called concentration of the likelihood [143].

In the case of mixtures densities, MLE's can be obtained using the EM algorithm.

### 2.5 EM Algorithm

The Expectation Maximization (EM) [53] algorithm is a procedure for computing MLE where data is absent, specifically knowledge of which class data belongs. Data is thus grouped into classes following a parametric density.

If each data point is sampled from a randomly select class, then data is distributed according to

$$p(x \mid \Psi) = \sum_{i=1}^{d} \pi_i f(x \mid \theta_i)$$

(2.37)

where $f$ is a parametric density, and $\Psi = \{\theta_k, \pi_k\}_{k=1}^{d}$ is a set of distributions parameters $\theta_k$ and mixture weights $\pi_k$.

The density of the data distribution takes on a much simpler form, if true class $y_j$ of $x_i$ is known. Let $\Psi_{y_j}$ denote the parameters and mixture weights, or relative frequencies, associated with class $y_j$. The density is then expressed as

$$p(x, y \mid k, \Psi) = \prod_{j=1}^{d} \pi_{y_j} f(x \mid \theta_{y_j})$$

(2.38)

By augmenting the data with (unknown) class indicators $W = \{w_1, \ldots, w_N\}$ we can compute a maximum likelihood estimate, since it is possible to then formulate the likelihood in the simpler manner as is the case if the true class was known. Thus for every data element $x_i$ that a corresponding $w_i$ exists, indicating which class data element $x_i$ belongs to.

Let us now express this formulation with unknown class indicator, commonly known as a latent variable. Let $X = \{X_1, \ldots, X_N\}$ be a random vector corresponding to observed
It is assumed that the data can be divided into \( K \) classes, and that each of these classes follow an a priori known parametric distribution \( f(·|\Theta_k) \). The distribution parameter(s) \( \theta_k \) of each of distribution and the relative frequency \( \alpha_k \) of each class are assumed unknown, and to be estimated. Let \( \Theta = \{\theta_1, \ldots, \theta_K\} \) be a set of distribution parameters, and the vector \( \alpha = (\alpha_1, \ldots, \alpha_K) \) denote the relative frequencies of the \( K \) classes, thus the frequencies fulfill the constraint \( \sum_k \alpha_k = 1 \). The set \( \Psi = \{\alpha, \Theta\} \) completely defines the mixture model. Let \( W = (W_1, \ldots, W_N) \) be a random vector of values interpreted as missing data, corresponding to \( X \). The missing data can be e.g. be interpreted as missing information about which category each data point belongs to, say \( w_i \in \{1, \ldots, K\} \). Let \( x = (x_1, \ldots, x_N) \) and \( w = (w_1, \ldots, w_N) \) be the realizations of \( X \) and \( W \), respectively.

The complete likelihood of the data is now
\[
p(x, w|\Psi) = \prod_{i=1}^N \prod_{k=1}^K [\alpha_k f(x_i|\theta_k)]^\delta(w_i - k) .
\]

The EM-algorithm consists of two steps: first the likelihood is computed according to the current parameter estimates (E-step), then the next estimate is the stationary point of the computed likelihood (M-step).

The process is the following. On the \( m \):th E-step (Expectation) the expected complete data log-likelihood is built, conditioned on the measured data and the previous parameter estimates, or
\[
\chi(\Psi, \hat{\Psi}^{(m-1)}) = E_{W|X, \hat{\Psi}^{(m-1)}} \{\log p(x, w|\Psi)\}
\]
\[
= \sum_{i=1}^N \sum_{k=1}^K P(W_i = k|x_i, \hat{\Psi}^{(m-1)}) \log \left( \alpha_k f(x_i|\theta_k) \right) .
\]

where the posterior probabilities of class \( k \in K \) is defined as
\[
P(W_i = k|x_i, \hat{\Psi}^{(m-1)}) = \frac{\alpha_k^{(m-1)} f(x_i|\theta_k^{(m-1)})}{\sum_k \alpha_k^{(m-1)} f(x_i|\theta_k^{(m-1)})} .
\]

On the M-step (Maximization), the expected complete data loglikelihood is maximized to obtain an update for the parameters
\[
\hat{\Psi}^{(m)} = \arg \max_\theta \chi(\Psi, \hat{\Psi}^{(m-1)}) ,
\]
and the steps are iterated until convergence.
2.5.1 Initialization EM Algorithm

The EM algorithm can be shown [53] to converge to a local maximum of the marginal a posteriori probability function \( p(\theta \mid x) = p(x \mid \theta)p(\theta) \). Biernacki et al. [22] propose a strategy for initialization, which generates \( p \) initial points using random selection, stochastic EM [36] and classification EM [37]. EM is then started from these \( p \) initial points accepting the result with the highest likelihood.

A common [24] strategy for initialization is the K-means algorithm [130], by which \( K \) clusters are generated on which an MLE for parameters is computed. In detail, for a \( K \) component mixture model, \( K \) clusters are generated on which MLE estimates of parameters are generated. The mixture weights are the given by the relative size of the clusters. These estimates are then used as initial point for the EM. In our work with EM this is the strategy we have adopted. We also adopt the strategy of generating \( p \) starting points, and accepting the result with the highest likelihood.

Example 2.5.1. (Gaussian N-dimensional distribution)
Assume that the set \( x = (x_1, ..., x_p) \) is composed of \( l \) vectors of dimension \( N \), which follow a Gaussian distribution with mean and covariance of class \( k \), \( \mu_k \) and \( \Sigma_k \), \( k = 1, ..., K \). Let \( \Psi_k = \{\Theta_k, \alpha_k\} \), where \( \Theta_k = \{\mu_k, \Sigma_k\} \) and relative frequencies \( \alpha_k \) of class \( k \) s.t. \( \sum_k \alpha_k = 1 \).

The distribution of data point \( x_i \) of class \( k \) is distributed according to
\[
 f(x_i|\Theta_k) = (2\pi)^{-N/2}|\Sigma_k|^{-1/2} \exp \left\{ -\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right\}. \tag{2.43}
\]

The E-step of the EM algorithm is in this case
\[
 \chi(\Psi, \Psi^{(m)}) = E_{W|x, \Psi^{(m-1)}} \left\{ \log \left( \prod_k \prod_{i=1}^l (\alpha_k f(x_i|\Theta_k)) \delta(w_i - k) \right) \right\}
 = E_{W|x, \Psi^{(m-1)}} \left\{ \sum_k \sum_{i=1}^l \delta(w_i - k) \log (\alpha_k f(x_i|\Theta_k)) \right\}
 = \sum_k \sum_{i=1}^l \left( \log \alpha_k + \log f(x_i|\Theta_k) \right) P(W_i = k|\Psi^{(m-1)}, x) \tag{2.44}
 = \sum_k \sum_{i=1}^l \left( \log \alpha_k - \frac{N}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma_k| \right.
 - \frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \big) \times P(W_i = k|\Psi^{(m-1)}, x).
\]

To obtain updates for parameters we maximize \( \chi(\psi, \psi^{(m)}) \) w.r.t. to parameter \( \mu_k \) and \( \Sigma^{-1} \). Thus we solve
\[
 \frac{\partial \chi(\Psi, \Psi^{(m-1)})}{\partial \mu_k} = 0 \tag{2.45}
\]
yielding update equation

\[ \mu_k^{(m)} = \frac{1}{N\alpha_k^{(m)}} \sum_{i=1}^{l} P(W_i = k|\Psi^{(m-1)}, x)x_i. \] (2.46)

To obtain updates for \( \Sigma_k \) for all parameters we instead solve

\[ \frac{\partial\chi(\Psi, \Psi^{(m)})}{\partial \Sigma_k^{-1}} = 0 \] (2.47)

Thus

\[
\frac{\partial\chi(\Psi, \Psi^{(m-1)})}{\partial \Sigma_k^{-1}} = \sum_{i=1}^{l} \left[ \frac{1}{2} \Sigma_k^T + \frac{1}{2}(x - \mu_k)(x - \mu_k)^T \right] P(W_i = k|\Psi^{(m-1)}, x) = 0
\]

\[ \Longleftrightarrow \Sigma_k = \frac{1}{N\alpha_k} \sum_{i=1}^{l} (x_i - \mu_k)(x_i - \mu_k)^T P(W_i = k|\Psi^{(m-1)}, x) \] (2.48)

In (2.47) we have used

\[ \frac{d(\det(A))}{dA} = \det(A)(A^{-1})^T \] (2.49)

and

\[ \frac{da^T Ab}{dA} = ab^T, \] (2.50)

where \( a, b \) are row vectors and \( A \) a square matrix. This yields the update equation

\[ \Sigma_k^{(m)} = \frac{1}{N\alpha_k^{(m)}} \sum_{i=1}^{N} (x_i - \mu_k^{(m)})(x_i - \mu_k^{(m)})^T P(W_i = k|\Psi^{(m-1)}, x). \] (2.51)

Finally, the updates for \( \alpha_k \) are found by solving

\[ \frac{\partial\chi(\Psi, \Psi^{(m-1)})}{\partial \alpha_k} = 0, \] (2.52)

under the constraints \( \alpha_k > 0, \forall k \) and \( \sum_j \alpha_j = 1 \):
\[ \frac{\partial}{\partial \alpha_k} \left( \chi(\Psi, \Psi^{(m-1)}) + \lambda(1 - \sum_k \alpha_k) \right) = \left( \frac{1}{\alpha_k} \sum_{i=1}^{n} P(W_i = k|\Psi^{(m-1)}, x) \right) - \lambda = 0 \]

\[ \iff \frac{1}{\alpha_k} \left( \sum_{i=1}^{l} P(W_i = k|\Psi^{(m-1)}, x) \right) - \lambda = 0 \]

\[ \iff \sum_k \left( \sum_{i=1}^{l} P(W_i = k|\Psi^{(m-1)}, x) \right) = \lambda \sum_k \alpha_k \]

\[ \iff l = \lambda \quad (2.53) \]

Thus

\[ \alpha_k = \frac{1}{l} \sum_{i=1}^{l} P(W_i = k|\Psi^{(m-1)}, x) \quad (2.54) \]

and the update equation of \( \alpha_k \) is thus

\[ \alpha_k^{(m)} = \frac{1}{l} \sum_{i=1}^{l} P(W_i = k|\Psi^{(m-1)}, x) \quad (2.55) \]

### 2.6 EM-Gamma

We here follow the derivation provided in Destrempes et al. [56]. First recall the Gamma distribution

\[ f_{\text{gam}}(x | k, \theta) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)} \quad k, \theta > 0 \text{ and } \mathbb{R}_+ \quad (2.56) \]

In the following let \( \Delta = \{ \pi, \Psi \} \), and index the \( N \) data points by \( j \) and class cate-
gories by $i$. We compute the expected log-likelihood

$$
\chi(\Delta, \Delta^{(n)}) = EW \left\{ \log \left( \prod_i \prod_j \left( \pi_i f_{\text{gam}}(x_j \mid \Psi^{(n)}) \right) \right) \right\}
$$

$$
= EW \left\{ \sum_i \sum_j \delta(W_j - i) \left( \log(\pi_i) + \log f_{\text{gam}}(x_j \mid \Psi^{(n)}) \right) \right\}
$$

$$
= \sum_i \sum_j EW \left\{ \delta(W_j - i) \left( \log(\pi_i) + \log f_{\text{gam}}(x_j \mid \Psi^{(n)}) \right) \right\}
$$

$$
= \sum_i \sum_j \{ \log(\pi_i) + \log f_{\text{gam}}(x_j \mid \Psi^{(n)}) \} P(W_j = i \mid X, \Psi^{(n)})
$$

$$
= \sum_i \sum_j \{ \log \pi_i + (k_i - 1) \log x_j - \frac{x_j}{\theta_i} - k_i \log \theta_i - \log \Gamma(k_i) \} \times \frac{\pi_i^{(n)} f_{\text{gam}}(x_j \mid \Psi_i^{(n)})}{\sum_l \pi_l^{(n)} f_{\text{gam}}(x_j \mid \Psi_l^{(n)})}
$$

(2.57)

where

$$
P(W_j = i \mid X, \Psi^{(n)}) = \frac{\pi_i^{(n)} f_{\text{gam}}(x_j \mid \Psi_i^{(n)})}{\sum_l \pi_l^{(n)} f_{\text{gam}}(x_j \mid \Psi_l^{(n)})}.
$$

(2.58)

The mixture weights fulfill the condition $\sum_i \pi_i = 1$. To incorporate this into (2.57) we use the theory of Lagrange multipliers, adding a term $\lambda(\sum_l \pi_l - 1)$ to (2.57), where the $\lambda$ is a non-zero constant.

Thus

$$
\chi(\Delta, \Delta^{(n)}) = \sum_i \sum_j \{ \log \pi_i + (k_i - 1) \log x_j - \frac{x_j}{\theta_i} - k_i \log \theta_i - \log \Gamma(k_i) \} \times \frac{\pi_i^{(n)} f_{\text{gam}}(x_j \mid \Psi_i^{(n)})}{\sum_l \pi_l^{(n)} f_{\text{gam}}(x_j \mid \Psi_l^{(n)})} + \lambda(\sum_l \pi_l - 1)
$$

(2.59)

Next we define the digamma function $\psi$ as the derivative of the logarithm of the Gamma function

$$
\psi(x) = \frac{d}{dx} \log \Gamma(x).
$$

(2.60)
Furthermore for ease of notation we define

\[ A_i = \sum_j \log x_j P(W_j = i \mid X, \Psi_i^{(n)}) , \]

\[ B_i = \sum_j P(W_j = i \mid X, \Psi_i^{(n)}) \quad \text{and} \]

\[ C_i = \sum_j x_j P(W_j = i \mid X, \Psi_i^{(n)}) . \]

Next we determine the stationary points of \( \chi \) w.r.t. \( k_i, \theta_i \) and \( \pi_i \),

\[ \frac{\partial \chi}{\partial k_i} = \sum_j \{ \log x_j - \log \theta_i - \psi(k_i) \} P(W_j = i \mid X, \Psi_i^{(n)}) \]

\[ = \sum_j \log x_j P(W_j = i \mid X, \Psi_i^{(n)}) - \log \theta_i \sum_j P(W_j = i \mid X, \Psi_i^{(n)}) \quad (2.61) \]

\[ - \sum_j P(W_j = i \mid X, \Psi_i^{(n)}) \psi(k_i) = A_i + B_i \log \theta_i + B_i \psi(k_i) = 0 \]

\[ \frac{\partial \chi}{\partial \theta_i} = \sum_j \left\{ \frac{x_j}{\theta_i^2} - \frac{k_i}{\theta_i} \right\} P(W_j = i \mid X, \Psi_i^{(n)}) \]

\[ = \frac{1}{\theta_i^2} \sum_{j=1}^N x_j P(W_j = i \mid X, \Psi_i^{(n)}) - \frac{k_i}{\theta_i} \sum_j P(W_j = i \mid X, \Psi_i^{(n)}) \quad (2.62) \]

\[ = \frac{1}{\theta_i^2} C_i - \frac{k_i}{\theta_i} B_i = 0 \Leftrightarrow C_i \theta_i - k_i B_i = 0 \Leftrightarrow \theta_i = \frac{C_i}{k_i B_i} \]

Insert \( \theta_i = \frac{C_i}{k_i B_i} \) into Eq. 2.61:

\[ A_i - B_i (\log \left( \frac{C_i}{k_i B_i} \right) + \psi(k_i)) = 0 \]

\[ A_i - B_i (\log C_i + \log k_i - \log B_i + \psi(k_i)) = \quad (2.63) \]

\[ A_i - B_i (\log C_i + B_i \log k_i + B_i \log B_i - B_i \psi(k_i)) = \]

\[ A_i - B_i \log C_i + B_i (\log k_i - \psi(k_i)) + B_i \log B_i = 0 \]

\[ \Leftrightarrow \]

\[ B_i (\log k_i - \psi(k_i)) = B_i (\log C_i - \log B_i) - A_i \Leftrightarrow \]

\[ \log k_i - \psi(k_i) = \log \left( \frac{C_i}{B_i} \right) - \frac{A_i}{B_i} \geq \frac{A_i}{B_i} - \frac{A_i}{B_i} = 0 \] (2.64)
since \( \log \) is a concave function we can apply Jensen's inequality.

The asymptotic series for the digamma function \( \psi \) is

\[
\psi(x) = \log x - \frac{1}{2x} - \sum_{i=1}^{\infty} \frac{B_{2n}}{2n \cdot x^{2n}} \tag{2.65}
\]

where \( B_{2n} \) are Bernoulli numbers.

Since all the terms in the asymptotic series

\[
\log x - \psi(x) = \frac{1}{2x} + \sum_{i=1}^{\infty} \frac{B_{2n}}{2n \cdot x^{2n}} \tag{2.66}
\]

are positive, and thus

\[
\lim_{x \to \infty} (\log x - \psi(x)) = 0 \tag{2.67}
\]

and

\[
\lim_{x \to 0} (\log x - \psi(x)) = \infty , \tag{2.68}
\]

and so it follows that \( \log x - \psi(x) \) is monotonically decreasing. Thus we can employ binary search in order to solve (2.63) for \( k_i \).

Thus in the \( n \)th iteration we solve

\[
\log k_i - \psi(k_i) = \log \left( \frac{C_i}{B_i} \right) - \frac{A_i}{B_i} \tag{2.70}
\]

for \( k_i \) using binary search, and set \( k_i^{(n+1)} = k_i^* \) where \( k_i^* \) is the stationary point. Then \( \theta_i^{(n+1)} \) follows from \( \theta_i = \frac{C_i}{k_i B_i} \). Eq. 2.69 provides \( \pi_i^{(n+1)} \).
Examples of fitted Nakagami mixture distributions  The above algorithm, described in Sec. 2.6, may be used to fit mixture distributions of Nakagami, through the connection in Eq. 2.23. In Fig. 2.2 two examples of fitted Nakagami mixtures to ultrasound images of the human neck.

2.7 EM-MAP

In the standard formulation of the Expectation Maximization algorithm the log-likelihood $p(x|\theta)$ is maximized,

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta \in \Omega} \log p(x \mid \theta)$$  \hspace{1cm} (2.71)
2.8. Similarity Measures

But since there may be many spurious maxima, it is sometimes desirable to impose a prior \( p(\theta) \) on the parameters, and instead maximize the posterior density \( p(\theta \mid x) \)

\[
\hat{\theta}_{\text{MAP}} = \arg \max_{\theta \in \Omega} \log p(\theta \mid x) = \arg \max_{\theta \in \Omega} \log p(x \mid \theta)p(\theta)
\]

\[
= \arg \max_{\theta \in \Omega} \log p(x \mid \theta) + \log p(\theta)
\]

\[\text{(2.72)}\]

2.8 Similarity Measures

A metric on a class of probability measures \( X \) is a function \( \rho : X \times X \rightarrow [0, \infty) \) s.t. three properties are fulfilled by \( \rho \): (1) \( \rho(P, Q) = 0 \) iff \( P = Q \) for \( P, Q \in X \) (Uniqueness), (2) \( \rho(P, Q) = \rho(Q, P) \), \( \forall P, Q \in X \) (Symmetry) and (3) \( \rho(P, R) \leq \rho(P, Q) + \rho(Q, R) \) for all \( P, Q, R \in X \) (Triangle inequality). If properties (1) and (2) are fulfilled, but not (3), then \( \rho \) is referred to as a distance. Thus all metrics are defined a distance, but not all distance function are metrics.

We detail the similarity measures used in this thesis: Assume that \( P, Q \) are absolutely continuous probability measure on \( \mathbb{R}^p, p \geq 1 \) with corresponding distribution functions \( F \) and \( G \), and densities \( f \) and \( g \).

**Definition 2.8.1. (Hellinger Distance [47])** The Hellinger metric is defined as

\[
H(P, Q) = \left[ \int \left( \sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx \right]^{1/2}
\]

Despite its name, it is a metric.

**Definition 2.8.2. (Kullback-Leibler Distance [119])** The Kullback-Leibler Distance is defined as

\[
KL(P, Q) = KL(f, g) = \int f(x) \log \frac{f(x)}{g(x)} dx
\]

**Definition 2.8.3. (J-divergence or Symmetric Kullback-Leibler Metric [107])** The Symmetric Kullback-Leibler Metric is defined as

\[
J(P, Q) = J(f, g) = \int f(x) \log \frac{f(x)}{g(x)} dx + \int g(x) \log \frac{g(x)}{f(x)} dx
\]

There exists a number of other metrics not employed in this thesis: e.g. Kolmogorov metric, Total Variation Metric and Bhattacharia distance. For an overview of these and other metrics and distances, see [72].
CHAPTER 2. Statistical modeling of the ultrasound signal
Chapter 3

Segmentation of the Endocardium in B-mode Ultrasound - Variational approach

3.1 Background

In this chapter we present locally optimal and global variational shape prior models for the segmentation of the endocardium in B-mode ultrasound images. In this section we will give a background on locally and globally optimal segmentation models.

One powerful way of obtaining segmentations of structures in an image, or sequence of images, or volumes, is to apply variational segmentation models. These formulate the segmentation problem as an energy integral, where different terms correspond to different desired properties of the curve, such as smoothness, data dependence etc. One then proceeds to minimized this integral using Calculus of Variations. This problem may also, as we shall see in the sequel, be formulated using Shape derivation.

In very general terms, variational segmentation models work by formulating a functional

$$ F(\Gamma) = \int_{\Omega_{\text{in}}} f(x, y) dx dy. $$

(3.1)

Here the region $\Omega_{\text{in}}$ is defined to be the region inside $\Gamma$, and the integrand may be either independent or dependent on the domain $\Omega_{\text{in}}$, i.e. $f(x, y) = f(x, y, \Omega_{\text{in}})$. To obtain
the desired segmentation one the solves the problem
\[
\arg\min_{\Gamma} F(\Gamma).
\] (3.2)

From the functional \( F \) one derives a velocity \( v \), e.g. through use of the Euler-Lagrange equation, and the solves the PDE
\[
\frac{\partial \Gamma(s,t)}{\partial t} = v \quad \text{with} \quad \Gamma(t = 0) = \Gamma_0 ,
\] (3.3)

where \( \Gamma_0 \) is an initial curve. Convergence is measured by some error criterion, e.g. in \( L^2 \) norm between successively computed curves.

In the following we assume that \( I : \mathbb{R}^2 \to \mathbb{R} \) is a gray scale image. Furthermore we assume that the curve \( \Gamma \) is parametrized in arc length and time, s.t. \( \Gamma(s,t) \) where \( 0 \leq s \leq 1 \) and \( t \in \mathbb{R}_+ \). The curve is assumed closed, i.e. \( \Gamma(0,t) = \Gamma(1,t) \). In the two-dimensional case, \( \Gamma = \{(x(s,), y(s,)) : s \in [0,1] \to \mathbb{R}^2 \} \).

The Active contour/Snake model was first proposed by Kass et al. [111]. The curve energy is expressed by
\[
E_{\text{snake}}(\Gamma) = \int_{0}^{1} \alpha ||\Gamma'(s,t)||^2 + \beta ||\Gamma''(s,t)||^2 + \gamma ||\nabla(G_\sigma \ast I(\Gamma(s,t)))||^2 ds ,
\] (3.4)

where the first two terms relate to deformation of the curve, regularization and preventing of discontinuities and overlap in evolution. The third term relates to the image, containing the gradient of the image smoothed by a Gaussian kernel \( G_\sigma \), attracting the curve to edges in the image.

The snake model cannot however handle topological changes, without adding special procedures for curve merging and splitting. In [34, 131, 35] topological changes in curve evolution are handled by use of level sets, using the method described by Osher and Sethian [155]. These works employ curvature-based flows in order to impose a smooth evolution, avoiding the second-order smoothness term in (3.4).

The geodesic active contour model (GAC) [35], solved by use of level sets, consists of the functional
\[
E_{\text{geodesic}}(\Gamma) = \int_{0}^{l(\Gamma)} g(||\nabla I(\Gamma(s)))||) ds ,
\] (3.5)

where the curve \( \Gamma \) is sought, which is minimal in length w.r.t. to weighted arc-length, by \( g(\cdot) \), where \( \lim_{t \to \infty} g(t) = 0 \). The data term \( g(\cdot) \) is edge-based, as in the snake-model.

A very general variational segmentation model is the continuous version of Potts model [166] which is defined as
\[
\min_{\{\Omega_i\}_{i=1}^{n}} \sum_{i=1}^{n} \int_{\Omega_i} f_i(x) dx + \lambda \sum_{i=1}^{n} |\partial \Omega_i| \quad \text{s.t.} \quad \cup_{i=1}^{n} \Omega_i = D, \quad \Omega_k \cap \Omega_l = \emptyset, \forall k \neq l ,
\] (3.6)

48
where $|\partial \Omega_i|$ is the length of boundary of subdomain $\Omega_i$, with the functions $f_i$ are defined on $\Omega$. In 3.6 the region $D$ is subdivided in $n$ non-overlapping regions, the smoothness of which are controlled by the parameter $\lambda$.

A special case of the continuous Potts model was proposed by Mumford et al. [144],

$$E_{\text{mumford}}(\Gamma) = \sum_i \int_{\Omega_i} (I - g_i)^2 dx + \nu |\Gamma| ,$$

(3.8)

where $g_i : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function and $I : \mathbb{R}^2 \to \mathbb{R}$, $|\cdot|$ denotes length of curve $(\cdot)$ and $\nu$ is a parameter. The functions $f_i$ of the Potts model (3.6) are here $f_i = (I - g_i)^2$. In segmentation one often speak of phases or segments into which the image is decomposed into, for example a two-phase segmentation would then be the decomposition of the image into and background and foreground, which would in (3.8) correspond to $m = 2$.

Another important variational method is the ROF (Rudin-Osher-Fatemi) [179] model, which was introduced for denoising images using total variation

$$\inf_{u(x) \in BV(\mathbb{R}^n)} \int_D |\nabla u| + \lambda_1 (u - u_0) + \lambda_2 (u - u_0)^2 dx ,$$

(3.9)

where $D$ is the image domain, $u_0$ is the image to be approximate by $u$, and first term enforces smoothness. The function space of bounded variation is defined in 3.4. Note that it is reminiscent of (3.8) excepting the fact that the smoothness is enforced differently in the two models.

Chan and Vese [39] used level sets to solve the two-phase Mumford-Shah model (3.8),

$$CV(c_1, c_2, \Gamma) = \mu |\Gamma| + \lambda_1 \int_{\Omega} (I(x) - c_1)^2 dx + \lambda_2 \int_{\Omega} (I(x) - c_2)^2 dx ,$$

(3.10)

where $\Omega$ is a region enclosed by the curve $\Gamma$. The original formulation of Chan and Vese also had a term penalizing the area in the region $\Omega$, which we have chosen to exclude here.

A very desirable property among functionals is convexity, as it implies that the solution one obtains is globally optimal. As an aid to the reader we provide definition of a convex set and convex functional:

**Definition 3.1.1. (Convex set)** A set $A$ in a vector space is convex if the line $(1 - t)x + ty$ belongs to $A$, for all $x, y \in A$ and $0 \leq t \leq 1$.

**Definition 3.1.2. (Convex functional)** A functional $f$ which does not assume the value $-\infty$ on convex set $A$ is convex on $A$ if

$$f((1 - \alpha)x + \alpha y) \leq (1 - \alpha)f(x) + \alpha f(y)$$

(3.11)

for $x, y \in A$ and $0 \leq \alpha \leq 1$. 

49
CHAPTER 3. Segmentation of the Endocardium - Variational approach

The Mumford-Shah functional does not fulfill the condition of convexity, since

\[ E_{\text{mumford}}(\Sigma_1 + \Sigma_2, c_1, c_2) \ngeq E_{\text{mumford}}(\Sigma_1, c_1, c_2) + E_{\text{mumford}}(\Sigma_2, c_1, c_2) \]  

(3.12)

The functionals (3.4), (3.5) and (3.10) can be shown to be non-convex in the same manner.

The ROF functional (3.9) is convex, but when used for binary image denoising is a non-convex minimization problem,

\[ \min_{\Sigma \subset \mathbb{R}^n} \int_{\mathbb{R}^n} |\nabla \chi_{\Sigma}| + \lambda (\chi_{\Sigma}(x) - \chi_{\Omega}(x))^2 \, dx. \]  

(3.13)

Note that in Eq. 3.13 the first order term has been dropped from Eq. 3.9. Finding optimal region \( \Sigma^* \) which minimized the energy is thus performed over a non-convex set of functions, that is the set of characteristic function.

Prior information is also an important aspect when constructing a segmentation functional, indeed the vast majority of methods developed for ultrasound segmentation incorporate prior information. In Qian et al. [168] a complex segmentation model is proposed involving a prior model for signal dropouts, using the log-normal distribution to model image intensities. In addition a novel search technique called “tunneling descent” was used to avoid trapping in local minima. Chen et al. [43] combines an edge-based active contour with shape priors to obtain segmentations of ultrasound and MRI images, making no use of statistical image models. Furthermore an intensity prior is incorporated into the model.

The work most closely related to the locally optimal shape prior segmentation model proposed in Sec. 3.10 is Dydenko et al. [62]. They propose a two-step method is proposed which combines an initial affine “alignment” of the previous segmentation result to the current frame, followed by a Rayleigh segmentation where the aligned contour was used as initialization. Only local rigid transformations were used in contrast to our proposed model which allows global affine “alignment” of the prior, to guide the active contour over greater distances towards the desired object. For more examples of incorporating shape priors in segmentation see [48, 174, 178, 123, 43] and references therein.

In the last decade there has been much work done in the field of convex segmentation models. Examples of this are [32, 102, 40, 52, 163, 94, 79, 125, 8]. Of these [102, 52, 163] are discrete MRF models, while [32, 40, 79, 125, 8, 94] are continuous. Chan et al. [40] used convex relaxation to obtain global solutions in case of binary image denoising and the two-phase piece-wise constant Mumford-Shah model. Using a convex prior function Ishikawa [102] formulated a globally optimal MRF model. In Pock et al. [163], a spatially continuous version of the model of Ishikawa [102] is formulated. The authors use functional lifting, i.e. formulating an equivalent higher dimensional functional, and convex relaxation is then use in minimizing this functional. Brown et al. [32] proposed a globally optimal solution to the multi-phase Mumford-Shah problem, by employing the piecewise constant level set method (PCLSM) proposed by Lie et
3.1. Background

al. [126] and convexification approach introduced by Pock et al. [163]. Lie et al. [125] presents a level set method where curves are represented by discontinuities of piecewise constant level set functions, instead of the traditional definition of level set which defines the curve as the zero-level set. The minimizing functional of the model is smooth and convex. Bae et al. [8] proposes a dual method for solving the convex relaxed continuous Potts model, i.e. continuous multi-labeling problem. A smooth convex version of the dual problem is introduced, and shown to produce solutions which are, under certain conditions, also solutions to the non-convex Potts model. In Gu et al. [79] an extension of algorithm in [8] is presented. The proposed method is a primal-dual approach to global minimization of Potts model, and applied to Mumford-Shah model for multi-phase segmentation. Does not use convex relaxation, i.e. allowing labeling functions to take values ∈ [0, 1]. A development of augmented Langrangian-based methods (ALM), providing significantly less parameters than in original ALM.

A further challenge is to incorporate shape prior knowledge into the convex segmentation process. Convex graph cuts methods with shape priors are proposed by [211, 214, 199]: in Veksler et al. [211] a single star-convex prior, in Vu et al. [214] multiple shape priors and a 3D shape prior in Song et al. [199]. A probabilistic shape prior was described by Cremers et al. [49] for variational shape prior segmentation, under the assumption that probabilistic shape functions are Lipschitz continuous. The probabilistic shape prior concept is in Andrews et al. [4] extended to multi-region prior segmentation. In Fundana et al. [70], the prior is "frozen" during minimization and updated between rounds of minimization. The minimization problem is relaxed according to the strategy of Chan et al. [40].

A few works in convex segmentation have been applied to the segmentation of ultrasound data. Huang et al. [96] formulates a MLE based model with assumption of Fisher-Tippett distribution of the intensities of ultrasound images. Like Chan etal [40] the authors use relaxation to convexify functional.

To our knowledge we are the first to apply a convex shape prior model to ultrasound data. The globally optimal shape prior segmentation model is described in Sec. 3.11. As in Houhou et al. [94], the segmentation model consists of histogram/pdf competition. A fast texture segmentation model using Kullback-Leibler measure is used to find the maximum distance between background and textures of interest. This works very well on images that have distinct distributions; an assumption which mostly does not hold for ultrasound images.

We therefore propose a segmentation model that consists of minimizing the distance between estimated and prior pdf’s, estimated from annotated data. In Besson et al. [108] a prior histogram of the region of interest is used to perform segmentation, which is reminiscent of our method except we employ both estimated and training data in the construction of the prior pdf’s. Our model is formulated using convex relaxation, as described in [40], and solved using the Split-Bregman method [75].
Principal Component Analysis (PCA) is a well-known approach for dimensionality reduction. In PCA the dimension of the data is reduced while the maximum amount of variance is preserved in the space of reduced dimension.

Our use of PCA is the dimension reduction of covariance matrices, as in Sec. 4.2, and the determination of eigenspaces spanned by principal eigenvectors of binary data sets of shapes as in Sec. 3.11.

In the sequel we follow the derivation of PCA given in Hyvärinen et al. [98]. Let $x = [x_1, ..., x_n]^T$ be a vector of mean-subtracted data points. The goal is now to determine the projection $w^T x$ which preserves the maximum amount of variance. The variance of $w^T x$ is

$$E[(w^T x)^2] = E[w^T (xx^T) w] = w^T E(xx^T)w = (N - 1)w^T Cw,$$  \hspace{1cm} (3.14)

where $C = \frac{1}{N-1}xx^T$ is empirical covariance matrix. To find the non-trivially maximizing vector $w$, the constraint $||w||_2 = 1$, and the PCA problem thus becomes

$$\arg \max_{w: ||w||_2 = 1} w^T Cw.$$  \hspace{1cm} (3.15)

The covariance matrix $C$ is symmetric and so it can be decomposed as $C = UDU^T$, where $U$ is an orthogonal matrix and $D$ a diagonal matrix. The columns of $U$ are eigenvectors of $C$ and $D = (\lambda_1, ..., \lambda_n)$, where $\lambda_1, ..., \lambda_n$ are the eigenvalues corresponding to the eigenvectors. Now let $v = U^T w$. Then

$$w^T Cw = w^T UDU^T w = w^T Dv = \sum_i v_i^2 \lambda_i.$$  

Since $U$ is orthogonal the vectors $w$ and $v$ are equal in norm: $|v|_2 = |U^T w|_2 = |w|_2$. By setting $n_i = v_i^2$ the problem (3.15) is expressed equivalently as

$$\arg \max_{n_i \geq 0, \sum_i n_i = 1} \sum_i n_i \lambda_i.$$  \hspace{1cm} (3.16)

It is easy to see that the solution of problem (3.16) is achieved by setting $n_{i_{\text{max}}} = 1$ for the largest eigenvalue (indexed by $i_{\text{max}}$). Thus the sought vector $w$ in problem (3.15) is the eigenvector of $C$ with the largest eigenvalue.

Since all the eigenvectors of $C$ are orthogonal, finding the component of second largest variance is achieved by solving (3.16) with $m_{\text{max}} = 0$. Thus the second principal component is the eigenvector with the second largest eigenvalue. Repeating this process $m$ times, where $m < n$, the $m$ principal components are obtained.

Using the $m$ principal components it is the possible to project the $n$-dimensional vector $x$ onto the $m$-dimensional space $M$ spanned by the principal components, that is

$$a_i = w_i^T x \in M \hspace{0.5cm} i = 1, ..., m$$  \hspace{1cm} (3.17)
3.2. Dimensionality Reduction

Figure 3.1: Data point cloud and plane spanned by the two principal eigenvectors of the points.

where \( w_i \) is the \( i \)th principal component.

**Truncation of Covariance Matrix**

We use the principal eigenvectors to truncate the covariance matrix \( C \), for example in high-dimensional problems and/or to reduce noise by retaining only the most significant principal components. By limiting the number of eigenvectors used to construct \( W \), we can define

\[
\tilde{C} = \tilde{W}^T \tilde{\Lambda} \tilde{W}
\]

(3.18)

where \( \tilde{W} = [w_1, \ldots, w_p] \). Here \( w_p \) correspond to the \( p \) largest eigenvalues. The matrix \( \tilde{\Lambda} \) is similarly an truncated version of \( \Lambda \), s.t. \( \tilde{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p) \).
CHAPTER 3. Segmentation of the Endocardium - Variational approach

3.3 Level set representation

A popular way to represent a curve $\Gamma$ is to describe it implicitly by the zero level set of a Lipschitz function $\phi$, s.t. $\Gamma = \{x(s,t), t; \phi(x(s,t), t) = 0\}$ where $x$ are spatial coordinates and $t$ is (algorithmic) time. Using this implicit representation allows for splitting and merging of contours without the need for reparametrization. Furthermore, we assume that $\phi$ is negative inside its zero level set and positive on the outside. Let $n$ and $v$ be the normal and velocity vector field resp. of the curve $\Gamma$. The deformation of the curve $\Gamma$ in normal direction, to each point on the curve, is described by $\frac{\partial \Gamma}{\partial t} = V n$, where $V = \langle v, n \rangle$.

Differentiating the zero level set $\phi(x(s,t), t) = 0$, in the two-dimensional case, w.r.t. time yields:

$$\frac{d\phi}{dt} = \frac{\partial \phi}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial \phi}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial \phi}{\partial t} = \langle \nabla \phi, \frac{\partial \Gamma}{\partial t} \rangle + \frac{\partial \phi}{\partial t} = 0. \quad (3.19)$$

Since $\frac{\partial \Gamma}{\partial t} = V n$, (3.19) becomes

$$\frac{\partial \phi}{\partial t} = -V \langle \nabla \phi, n \rangle. \quad (3.20)$$

To finalize the level set evolution equation we want to express the normal vector field using the level set function $\phi$. First we observe that taking the derivative of the zero level set w.r.t. the spatial coordinate $s$ yields

$$\frac{\partial \phi}{\partial s} = \frac{\partial \phi}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial \phi}{\partial y} \frac{\partial y}{\partial s} = \langle \nabla \phi, \frac{\partial \Gamma}{\partial s} \rangle = 0. \quad (3.21)$$

Figure 3.2: Region deformation along normal direction $n$ with speed provided by velocity field $V$. 
This implies that $\nabla \phi$ is normal to $\Gamma$, since $\frac{\partial \Gamma}{\partial s}$ is tangent to $\Gamma$. Assuming $n$ is oriented outward, then the normal can be written as $n = \frac{\nabla \phi}{|\nabla \phi|}$. Substituting this into (3.20) we arrive at the evolution equation expressed in terms of the level set function $\phi$:

$$ \frac{\partial \phi}{\partial t} = -V|\nabla \phi|. \quad (3.22) $$

### 3.4 Functions of Bounded Variation

**Definition 3.4.1.** ([73]) Let $\Omega \subset \mathbb{R}^n$ be an open set and $u \in L^1(\Omega)$. The Total Variation of $u$ is defined by

$$ TV(u) = \int_{\Omega} |\nabla u| dx = \sup \left\{ \int_{\Omega} u \text{div} \phi \, dx : \phi \in C_0^1(\Omega, \mathbb{R}^n) \land |\phi(x)| \leq 1, x \in \Omega \right\}. \quad (3.23) $$

**Definition 3.4.2.** ([73]) If the total variation of a function $u \in L^1(\Omega)$ is bounded, i.e. $TV(u) < \infty$, then $u \in BV(\Omega)$.

**Theorem 3.4.1.** ([73]) Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with Lipschitz boundary, and let $(u_n)_{n \geq 1}$ be a bounded sequence in $BV(\Omega)$. Then there exists a subsequence $(u_{n_k})_{k \geq 1}$ s.t. $\lim_{k \to \infty} u_{n_k} = u$ in $L^1(\Omega)$, for an element $u \in BV(\Omega)$. 

55
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Definition 3.4.3. Let \( \Omega \subset \mathbb{R}^n \) and \( f : \Omega \to \mathbb{C} \) a Lebesgue integrable measurable function. The function \( f \) is locally integrable if
\[
\int_K |f| < \infty
\]
(3.24)
on all compact subsets \( K \) of \( \Omega \). The set of all locally integrable functions is
\[
L_{\text{loc}}^1(\Omega) = \{ f : \Omega \to \mathbb{C} | f \in L_1(K), \forall K \subset \Omega, K \text{ compact} \}.
\]

Theorem 3.4.2. ([73]) Let \( \Omega \subset \mathbb{R}^n \) be an open set and \( (u_n)_{n \geq 1} \) a sequence of functions in \( BV(\Omega) \) s.t. \( u_k \overset{L_{\text{loc}}^1}{\to} u \), where \( u \) is a function in \( L^1(\Omega) \). Then \( u \in BV(\Omega) \), and moreover
\[
TV(u) = \int_\Omega |\nabla u| dx \leq \liminf_{j \to \infty} \int_\Omega |\nabla u_j| dx,
\]
(3.26)in other words \( TV \) is lower semi-continuous.

Definition 3.4.4. ([73]) A set \( \Omega \subset \mathbb{R}^n \) has finite perimeter if and only if the characteristic function \( \chi_\Omega \) of \( \Omega \) belongs to \( BV(D) \). We have
\[
\text{Per}(\Omega) = TV(\chi_\Omega) = \int_D |\nabla \chi_\Omega| dx < \infty,
\]
(3.27)

Theorem 3.4.3. (Co-area formula) [73] Let \( u \in BV(\Omega) \) and define
\[
U_t = \{ x \in \Omega : u(x) > t \}.
\]
(3.28)
Then
\[
TV(u) = \int_\Omega |\nabla u| dx = \int_{-\infty}^\infty dt \int_\Omega |\nabla \chi_{U_t}| dx.
\]
(3.29)

3.5 First Necessary Condition for Relative Minimum

Theorem 3.5.1. First necessary condition for a relative minimum
Assume that \( y_0 \in C^1[a,b] \) is a minimizer to the integral
\[
F(y) = \int_a^b f(x, y(x), y'(x)) dx,
\]
where \( y \in C^1(a,b) \) and \( f \in C^1(R) \). Here \( R \) is the region of all triples \( (x, y(x_0), y'(x_0)) \) for \( x_0 \in [a,b] \). Then \( y_0(x) \) is a solution to
\[
f_{y'}(x, y(x), y'(x)) = \int_a^x f_y(t, y(t), y'(t)) dt + C
\]
(3.30)
for \( x \in [a,b] \).
3.6. Gâteaux variation of functional

**Corollary 3.5.1.** Under the assumptions made in Theorem 3.5.1 for $F(y)$ it follows that a solution $y = y_0(x)$ to (3.30) satisfies the equation

$$f_y(x, y, y') - \frac{d}{dx} f_{y'}(x, y, y') = 0$$

(3.31)

for $x \in [a, b]$.

Equation (3.31) is referred to as the Euler-Lagrange equation.

For $n$ independent variables $x_1, ..., x_n$ with $m$ unknown functions $u_1, ..., u_m$, the Euler-Lagrange equation takes the form

$$f_{u_{\mu}} - \sum_{\nu=1}^{n} \frac{\partial}{\partial x_{\nu}} f_{u_{\mu} x_{\nu}} = 0, \quad \mu = 1, 2, 3, ..., m.$$  

(3.32)

### 3.6 Gâteaux variation of a functional defined over regular open bounded domains

The set of image regions is not a vector space; therefore deformations of a regular open bounded region cannot be handled simply by vector addition/subtraction. In this section we follow the description of Gâteaux variations provided in Sokolowski and Zolésio [196].

Let $E$ be a domain in $\mathbb{R}^n$ with boundary $\partial E$ which is piecewise $C^k$ for a given $k \geq 0$. Now we define a one-to-one mapping $T_t : \overline{E} \to \bar{E}$ where $T_t, T_t^{-1} \in C^k(\bar{E}, \mathbb{R}^n)$. It is also possible to assume that $T^t$ and $T^{-t}$ are in the Sobolev space $W^{1,\infty}(\bar{E}, \mathbb{R}^n)$. See Adams and Fournier [2] for a formal definition of Sobolev spaces.

Furthermore the mappings $t \to T_t(x)$ and $T_t^{-1}(x)$ are continuous on the interval $[0, \epsilon]$, i.e. $t \to T_t(x), T_t^{-1}(x) \in C([0, \epsilon]), \forall x \in \bar{E}$. Summarizing, $(t, x) \to T_t(x) \in C([0, \epsilon); C^k(\bar{E}, \mathbb{R}^n))$, or $(t, x) \to T_t(x) \in C([0, \epsilon); W^{1,\infty}(\bar{E}, \mathbb{R}^n))$, with $\epsilon > 0$.

Now let $T_t : D \to D$ with $t \in [0, \epsilon]$ define a family of transformations, describing the change of the domain $\Omega \subset D$:

$$X \in \Omega \to x = T_t(X) \equiv x(t, X).$$

(3.33)

We only consider small perturbations of the domain $\Omega$, s.t. for $X \in \Omega$

$$T_t(X) = x(t) = X + t\Theta(X)$$

(3.34)

where $t \geq 0$ and $\Theta \in \mathbb{R}^n$ is a smooth vector field. The transformed geometry is given by:

$$\Omega_t = T_t(\Omega).$$

(3.35)
The velocity vector field at point \( x(t) \) is given by
\[
v(t, x) = \frac{\partial}{\partial t} T_t(t, T_t^{-1}(x)).
\]
(3.36)

Because of the assumptions on \( T_t \) and \( T_t^{-1} \) the vector field
\[
v(t)(x) = v(t, x) \in C([0, \epsilon), W^{1,\infty}(\bar{E}, \mathbb{R}^N)).
\]
(3.37)

The new transformed domain indexed by \( t \) is determined by \( x = x(t, X) \), which is the solution to the system of ordinary differential equations
\[
\begin{align*}
\frac{d}{dt} x(t, X) &= v(t, x(t, X)), \\
x(0, X) &= X.
\end{align*}
\]
(3.38)

Considering (3.38), it is thus possible to interpret \( X \) as a Langrangian coordinate, while \( x \) is the Eulerian coordinate when describing a flow. If the vector field \( v(t, x(t, X)) \) \( \in W^{1,\infty}(\mathbb{R}^n, \mathbb{R}^n) \) then the problem (3.38) is well-posed.

**Example 3.6.1.** We give an example of a transformation of a star-shaped domain. Let \( f \) be a periodic function \( f \in C^k[0, 2\pi] \), and define a domain \( \Omega(f) = \{(\rho, \theta) \in \mathbb{R}^2|0 \leq \theta \leq 2\pi, 0 \leq \rho < f(\theta)\} \). For any admissible \( g \in C^k[0, 2\pi] \) with \( g(0) = g(2\pi) \) there exists \( \epsilon > 0 \), s.t.
\[
\min\{f + tg(\theta)|\theta \in [0, 2\pi]\} > 0
\]
(3.39)
for $|t|<\epsilon$. Denote the transformed domain $\Omega_t = \Omega(f + tg)$, corresponding to transformation

$$T_t(\rho, \theta) = (\rho \frac{f + tg(\theta)}{f(\theta)}, \theta) \quad (3.40)$$

The inverse transformation is then

$$T_t^{-1}(\rho, \theta) = (\rho \frac{f(\theta)}{f + tg(\theta)}, \theta) \quad (3.41)$$

Using $T_t$ and $T_t^{-1}$ we obtain the velocity

$$v(t, x) = \frac{\partial}{\partial t} T_t(T_t^{-1}(\rho, \theta)) = (\rho \frac{g(\theta)}{f + tg(\theta)}, 0). \quad (3.42)$$

Let $F(\Omega) = \int_D f(x, \Omega) dx$ be a domain functional, where $f$ is a Lebesgue integrable function. An example of this is $f(x, \Omega) = \chi_{\Omega}(x)$.

**Definition 3.6.1.** ([196]) The Gâteaux derivative (or Eulerian derivative) of $F(\Omega) = \int_\Omega f(x, \Omega) dx$ in the direction of $v$, denoted $dJ(\Omega; v)$, is equal to

$$dJ(\Omega; v) = \lim_{t \to 0} \frac{F(\Omega_t) - F(\Omega)}{t}. \quad (3.43)$$

**Definition 3.6.2.** ([196]) Let $E$ be a domain in $\mathbb{R}^n$, with boundary $\partial E$ is piecewise $C^k$, $k \geq 1$. Outward normal fields exists a.e. on $\partial E$.

$$v^k(E) = \{v \in C^k(\mathbb{R}^n; \mathbb{R}^n); \langle v, n \rangle_{\mathbb{R}^n} = 0 \text{ a.e. on } \partial E\}.$$

**Definition 3.6.3.** ([196]) The material derivative $\dot{f}$ of $f(x, \Omega) \in W^{1, \infty}(\Omega)$ in the direction $v \in C([0, \epsilon); v^k(E))$ is

$$\dot{f}(x, \Omega; v) = \lim_{t \to 0} \frac{1}{t} (f(T_t(v)(x), \Omega) - f(x, \Omega)). \quad (3.43)$$
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Definition 3.6.4. ([196]) The point-wise shape derivative of \( f(x, \Omega) \), denoted \( f'(x, \Omega, \mathbf{v}) \), is equal to

\[
\begin{align*}
\lim_{t \to 0} \frac{f(x, \Omega_t) - f(x, \Omega)}{t}.
\end{align*}
\]

The shape derivative is furthermore defined in terms of the material derivative,

Definition 3.6.5. ([196]) The shape derivative \( f'(x, \Omega; \mathbf{v}) \) of \( f(x, \Omega) \) in the direction \( \mathbf{v} \) is defined to be

\[
\begin{align*}
f'(x, \Omega; \mathbf{v}) = \dot{f}(x, \Omega; \mathbf{v}) - \nabla f(x, \Omega) \cdot \mathbf{v}(0, x)
\end{align*}
\]

(3.44)

Theorem 3.6.1. ([196]) \( \text{(Differentiation with respect to parameter)} \)

Let \( I(\mu) = \int_{\Omega_t} g(x, \mu) dx \), then

\[
\frac{dI}{d\mu}(\mu) = \int_{\Omega_t} \frac{\partial g}{\partial \mu}(x, \mu) dx - \int_{\partial \Omega_t} g(x, \mu) \mathbf{v}(t, x) \cdot \mathbf{n}(x) da(x)
\]

(3.45)

where \( \mathbf{n} \) is the unit inward normal to the boundary \( \partial \Omega \), and \( da(x) \) is the area element.

The first necessary condition for a relative minimum of \( F(\Omega) \) is thus

\[
\int_{\Omega} f'(x, \Omega, \mathbf{v}) dx - \int_{\partial \Omega} f(x, \Omega) \mathbf{v}(0, x) \cdot \mathbf{n}(x) da(x) = 0 \quad \forall \mathbf{v}.
\]

(3.46)

The corresponding Euler-Lagrange equation \( EL = 0 \) is obtained by transferring (3.46) to the form

\[
\int_{\partial \Omega} EL \cdot (\mathbf{v}(0, x) \cdot \mathbf{n}(x)) da(x) .
\]

(3.47)

The exact form of the Euler-Lagrange equation will vary on form of the functional \( F(\Omega) \), e.g. on the region dependent and independent features of the functional.

Theorem 3.6.2. ([196]) The Gâteaux derivative of the functional \( F(\Omega) = \int_{\Omega} f(x, \Omega) dx \), in the direction of \( \mathbf{v} \), is

\[
\begin{align*}
dF(\Omega; \mathbf{v}) &= \int_{\Omega} f'(x, \Omega, \mathbf{v}) dx - \int_{\partial \Omega} f(x, \Omega) \mathbf{v}(0, x) \cdot \mathbf{n}(x) da(x).
\end{align*}
\]

(3.48)

Proof: By definition \( F(\Omega_t) = \int_{\Omega_t} f(x, \Omega_t) dx \), then Theorem 3.6.1 implies

\[
\begin{align*}
dF(\Omega; \mathbf{v}) &= \frac{dJ(\Omega_t)}{dt}(0) \\
&= \int_{\Omega_t} \frac{\partial f}{\partial t}(x, \Omega_t) dx|_{t=0} - \int_{\partial \Omega_t} f(x, \Omega_t) \mathbf{v}(t, x) \cdot \mathbf{n}(x) da(x)|_{t=0} \\
&= \int_{\Omega} f'(x, \Omega, \mathbf{v}) dx - \int_{\partial \Omega} f(x, \Omega) \mathbf{v}(0, x) \cdot \mathbf{n}(x) da(x),
\end{align*}
\]

(3.49)

since \( \Omega_0 = \Omega \) and \( f'(x, \Omega, \mathbf{v}) = \frac{\partial f}{\partial t}(x, \Omega_t)|_{t=0} \) by definition of material derivative. \( \blacksquare \)
3.7 \( L^2 \) Shape Gradients

Using Riesz representation theorem, see e.g. [172], one can express the Gâteaux derivative as the inner product \( \langle \cdot | \cdot \rangle_{\mathcal{H}} \)

\[
d F(\Omega; v) = \langle u | v \rangle_{\mathcal{H}}
\]

(3.52)

where \( u, v \in \mathcal{H} \) are vector fields in the Hilbert space \( \mathcal{H} \). If \( d F(\Omega; v) \) is continuous then \( u \in \mathcal{H} \) is unique. The vector field \( u \) is denoted by \( u = \nabla \mathcal{H} F \), and is also called the (shape) gradient of \( \mathcal{H} \). In our work we make the simplifying assumption that \( \mathcal{H} = L^2 \). When considering (3.47) we note that this is expression \( L^2 \) inner product of the deformation fields \( EL \) and \( v(0, x) \cdot n(x) \). In Charpiat et al. [41] the use of other inner products is explored.

In the case of region integrals, which is largely our concern, one can also apply the same framework using \( L^2 \)-gradients. Let \( \Omega \) denote the region contained inside the curve \( \Gamma \). If \( g(x) : \mathbb{R}^{m+1} \to \mathbb{R} \) and

\[
F(\Omega) = \int_{\Omega} g(x) \, dx
\]

(3.53)

The Gâteaux derivative of \( F(\Omega) \) is computed in direction \( v \):

\[
d F(\Omega; v) = \langle g(x), v(0, x) \cdot n(x) \rangle_{\partial \Omega}
\]

(3.54)

so \( \nabla_{L^2} F = g(x) \).

3.8 Construction of Locally Optimal Shape Prior Model

3.8.1 Energy Terms in Segmentation Functional

There are many ways to define an energy functional when setting up a variational segmentation problem; in general the functional is broken up into separate parts which define different desired properties of the final segmenting curve/surface, and of the area/volume within resp. without the curve/surface. The properties are e.g. distribution of data within/without the curve, geometrical shape of curve and curve length. In the three-dimensional case, substitute curve for surface and area for volume in the preceding.

We define a variational segmentation functional \( E_{\text{Total}}(\Omega) \) over the image domain \( D \), which is made up of three different energy terms: a data energy term \( E_{\text{Data}} \), a length energy term \( E_{\text{Length}} \) and a shape prior energy \( E_{\text{Shape}} \):

\[
E_{\text{Total}}(\Omega) = E_{\text{Data}}(\Omega) + E_{\text{Shape}}(\Omega) + E_{\text{Length}}(\Omega).
\]

(3.55)
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Data Term

The data term, or fidelity term, is related to properties of the data, varying from simple, like mean, or more sophisticated like distributional, like Hellinger distance between a prior density and an online estimated density of a region. Statistical data terms are especially appropriate for data sets which are best described in terms of statistical distributions; ultrasound data is a prime example of this. In Chp. 2 we have detailed the various distributions used to model the backscattered echo of ultrasound, under different scattering conditions. It therefore stands to reason that a data term should employ these statistics.

Example 3.8.1. Let the total likelihood inside of the curve $C$ be $P_{\text{in}}(C) = \prod_{i \in \Omega} p(I_i)$, and outside $P_{\text{out}}(C) = \prod_{i \in \Omega^c} p(I_i)$, where the region inside and outside $C$ is denoted by $\Omega$ and $\Omega^c$, respectively. The maximum likelihood curve $C$ is thus given by the curve that maximizes the likelihood $P(I|C) = P_{\text{in}}(C)P_{\text{out}}(C)$, or equivalently minimizing the log-likelihood

$$-\log P(I|C) = -\log P_{\text{in}}(C) - \log P_{\text{out}}(C).$$ (3.56)

Assuming that the image can be modeled by a Rayleigh distribution, we have $p(I|\sigma) = \frac{I}{\sigma^2} \exp\left(-\frac{I^2}{2\sigma^2}\right)$. The parameter $\sigma^2$ may be estimated using the technique proposed in Chesnaud et al. [44], where $\hat{\sigma}^2 = \frac{1}{\Omega} \int \Omega I^2(x) dx$. Here $\Omega$ is the region over which the likelihood is computed. Thus Eq. 3.56 becomes

$$-\log P(I|C) = -\int_{\Omega^c} \log p(I(x)|\sigma_{\text{in}}) dx - \int_{\Omega} \log p(I(x)|\sigma_{\text{out}}) dx$$

$$= -\int_{\Omega} \log I(x) dx + |\Omega| \log \left( \frac{1}{|\Omega|} \int_{\Omega} I^2(x) dx \right)$$

$$+ |\Omega^c| \log \left( \frac{1}{|\Omega^c|} \int_{\Omega^c} I^2(x) dx \right) - \frac{|\Omega|}{2}$$ (3.57)

The segmentation functional $E_{\text{data}}$ is formulated from the terms dependent on the position of the curve $C$, i.e. terms containing $\Omega^c$ and $\Omega$, hence

$$E_{\text{data}}(\Omega) = |\Omega| \log \left( \frac{1}{|\Omega|} \int_{\Omega} I^2(x) dx \right) + |\Omega^c| \log \left( \frac{1}{|\Omega^c|} \int_{\Omega^c} I^2(x) dx \right)$$ (3.58)

Length Regularization Term

Assuming as in previous sections, that $\Gamma = \Gamma(s,t), 0 \leq s \leq 1, t \in \mathbb{R}^+$ is a family of close curves, which bound the region $\Omega_t$. A length regularisation term can then be formulated as

$$E_{\text{length}}(\Omega_t) = \int_0^1 \left\| \frac{\partial \Gamma}{\partial s} \right\| db.$$ (3.59)
3.8. Construction of Locally Optimal Shape Prior Model

Let \( C = C(s, t) \) be a smooth family of closed curves where \( t \) parametrizes the family and \( p \) the given curve, say \( 0 \leq s \leq 1 \), and assuming \( C(0, t) = C(1, t) \). Using a penalty on length of curve is commonly used in variational segmentation e.g. in [35, 144, 39].

Another way to impose regularization is to represent the domain of interest by a characteristic function, and measure the total variation of this of the domain. Specifically, the perimeter of a set \( \Sigma \in \mathbb{R}^n \) is defined

\[
\text{Per}(\Sigma) = \int_{\mathbb{R}^n} |\nabla \chi_\Sigma(x)|. \tag{3.60}
\]

In the plane the two formulations, (3.60) and (3.59), for measuring curve length are equivalent. Representing the length of the contour delineating the domain, using (3.60), is convenient when formulating convex segmentation functionals, since (3.60) is convex.

**Remark 3.8.1.** \( C \) is a 1-dim subset of \( \mathbb{R}^n \), and \( \Sigma \) is a \((n-1)\)-dim subset.

**Shape Prior Term**

Ideally image data is noise free, and problem of identifying the sought object(s) is reduced to e.g. detecting edges. However, noise free image data is a rare occurrence. A widely employed strategy is to use a shape prior term which reduces space of admissible shapes that the curve can take. A shape prior constraint may taken many forms, an intuitive form is given by

\[
E_{\text{Shape}}(\Omega, \Omega_0) = \int_D \left( \chi_\Omega(x) - \chi_{\Omega_0}(x) \right)^2 dx \tag{3.61}
\]

where

\[
\chi_\Omega(x) = \begin{cases} 
1, & \text{if } x \in \Omega, \\
0, & \text{otherwise}. 
\end{cases} \tag{3.62}
\]

In this formulation the prior shape \( \Omega_0 \) can be fixed over the entirety of the process curve evolution, or be updated in each iteration to take into account the current state of the curve, e.g. the current shape of the area \( \Omega \) may be projected onto the eigenspace of training shapes, in order to obtain the closest possible shape within a training database. This may be handled in many ways, for example by PCA.

In order to explain the concept of pose-invariant shape priors we now introduce the Affine groups.
CHAPTER 3. Segmentation of the Endocardium - Variational approach

The Affine Group

Let \( T(\rho)(x) = A(x-a) \) with \( \rho = (\mu, \theta, s_x, s_y, a_x, a_y) \in \mathbb{R}^+ \times [-\pi, \pi] \times \mathbb{R}^4 \) and \( a = (a_x, a_y) \). The matrix \( A \) is defined
\[
A = \mu \begin{pmatrix}
\cos \theta & \sin \theta \\
-sin \theta & \cos \theta
\end{pmatrix}
\begin{pmatrix}
1 & s_x \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
s_y & 1
\end{pmatrix}
= \mu J(\theta)S_xS_y .
\] (3.63)

We use \( y = T(\rho)x \) as a shorthand for an orientation preserving planar affine map; the set of these type of maps form a group under composition.

We hare now ready to introduce a pose-invariant (pi) shape prior
\[
E_{\text{Shape}}(\Omega, \Omega_0) = \int_D \left( \chi_{\Omega}(x) - \chi_{\Omega_0}(T(\rho)(x)) \right)^2 dx .
\] (3.64)

While pose-invariance can be very helpful in guiding the contour to its desired optima, it does however have some drawbacks, chiefly consisting of that it is not clear in what order the transformation parameters should be optimized, and the fact that different optimization schemes of parameters will lead to different outcomes in the segmentation process.

3.8.2 First Variation of Segmentation Functional

We wish to solve the minimization problem
\[
\arg \min_{\Omega} E_{\text{Total}}(\Omega) = \arg \min_{\Omega} \left( E_{\text{Data}}(\Omega) + E_{\text{Length}}(\Omega) + E_{\text{Shape}}(\Omega) \right)
\] (3.65)

In order to accomplish this we proceed to compute the Gâteaux derivative of each term in the direction \( v \).

3.8.3 Data Term

In the sequel we define \( \Omega^c = D \setminus \Omega \), where \( D \) is the image domain. For the case of \( E_{\text{data}}(\Omega) \) we first reformulate the term, for ease of computation:
\[
E_{\text{Data}}(\Omega) = |\Omega| \log \left( \frac{1}{|\Omega|} \int_{\Omega} I^2(x) \, dx \right) + |\Omega^c| \log \left( \frac{1}{|\Omega^c|} \int_{\Omega^c} I^2(x) \, dx \right)
\] (3.66)

where
\[
L_1(G_1(\Omega), G_2(\Omega)) = G_2(\Omega) \log \left( \frac{G_1(\Omega)}{G_2(\Omega)} \right)
\] (3.67)
3.8. Construction of Locally Optimal Shape Prior Model

\[ G_1(\Omega) = \int_{\Omega} I^2(x) \, dx \] and \[ G_2(\Omega) = \int_{\Omega} dx. \] (3.68)

Correspondingly, we define

\[ L_2(G_3(\Omega), G_4(\Omega)) = G_4(\Omega) \log \left( \frac{G_3(\Omega)}{G_4(\Omega)} \right), \] (3.69)

and

\[ G_3(\Omega) = \int_{\Omega^c} I^2(x) \, dx = \int_{D} I^2(x) \, dx - \int_{\Omega} I^2(x) \, dx, \] (3.70)

and

\[ G_4(\Omega) = \int_{\Omega^c} dx = \int_{D} dx - \int_{\Omega} dx. \] (3.71)

Now we are ready to calculate the Gâteaux derivative of \( E_{\text{Data}}(\Omega) \) in the direction \( v \)
using shape derivation:

\[ dE_{\text{Data}}(\Omega; v) = dL_1(\Omega; v) + dL_2(\Omega; v). \] (3.72)

The Gâteaux derivative of \( L_1 \) is

\[ dL_1(\Omega; v) = dG_2(\Omega; v) \left[ \log \left( \frac{G_1(\Omega)}{G_2(\Omega)} \right) - 1 \right] + dG_1(\Omega; v) \frac{G_2(\Omega; v)}{G_1(\Omega; v)}. \] (3.73)

Similarly,

\[ dL_2(\Omega; v) = dG_4(\Omega; v) \left[ \log \left( \frac{G_3(\Omega)}{G_4(\Omega)} \right) - 1 \right] + dG_3(\Omega; v) \frac{G_4(\Omega; v)}{G_3(\Omega; v)}. \] (3.74)

Thus

\[ dE_{\text{Data}}(\Omega; v) = \int_{\partial \Omega} \left[ \log \left( \frac{\sigma_{\text{ext}}}{\sigma_{\text{int}}} \right) + I^2(x) \left( \frac{1}{\sigma_{\text{int}}} - \frac{1}{\sigma_{\text{ext}}} \right) \right] v(0, x) \cdot n(x) \, da(x) \] (3.75)

where

\[ \sigma_{\text{int}} = \frac{1}{|\Omega|} \int_{\Omega} I^2(x) \, dx, \] (3.76)

and

\[ \sigma_{\text{ext}} = \frac{1}{|\Omega^c|} \int_{\Omega^c} I^2(x) \, dx. \] (3.77)
3.8.4 Length Term

We now proceed to compute the Gâteaux derivative of $E_{\text{length}}(\Omega) = \int_0^1 \left\| \frac{\partial \Gamma}{\partial s} \right\| \, ds$ in direction $v$:

$$dE_{\text{length}}(\Omega; v) = \frac{d}{dt} \int_0^1 \left\| \frac{\partial \Gamma}{\partial s} + t \cdot \frac{\partial \nu}{\partial s} \right\| \, ds \bigg|_{t=0} = \int_0^1 \frac{\partial \Gamma}{\partial s} \cdot \frac{\partial \nu}{\partial s} \, ds \text{ part.int.} = \int_0^1 \kappa \cdot v \, ds,$$

(3.78)

where $\kappa = \frac{\partial}{\partial s} \left[ \frac{\partial \Gamma}{\partial s} \right]$ is the standard curvature and $ds$ is the arc length element.

3.8.5 Shape Prior Term

We use explicit minimization here due to its simplicity, being aware that more sophisticated formulations exist, proposed by Cremers [48] et al. and Overgaard [157] et al.

Two major problems exist: (1) the time step of gradient descent require tuning (addressed in Overgaard et al. [157]), (2) it is unclear how to alternate between updates (addressed in Cremers et al. [48]).

Let $E_{\text{shape}}(\Omega, \Omega_0) = \int_D (\chi_\Omega - \chi_{\Omega_0})^2 \, dx$, then we may rewrite it as

$$E_{\text{shape}}(\Omega, \Omega_0) = \int_\Omega (1 - 2\chi_{\Omega_0}) \, dx + \int_D \chi_{\Omega_0} \, dx$$

(3.80)

In the frame work of level set method, we have

$$E_{\text{shape}}(\phi, \phi_0) = \int_D (1 - 2H(\phi_0(x)))H(\phi(x)) \, dx + \int_D H(\phi_0(x)) \, dx$$

(3.81)

$$= \int_D (1 - 2H(\phi(x)))H(\phi_0(x)) \, dx + \int_D H(\phi(x)) \, dx.$$ (3.82)

Now we consider the affine transformation $T : x \mapsto A^{-1}x + a$, where

$$A = \mu \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & s_x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ s_y & 1 \end{pmatrix} = \mu JS_x S_y,$$

we look for the expression of $E_{\text{shape}}(\Omega, T\Omega_0)$ in the frame work of the Level set method. If $\tilde{\phi}$ is the associated level set function for $T\Omega_0$ then we have

$$E_{\text{shape}}(\Omega, T\Omega_0) = \int_D (1 - 2H(\phi(x)))H(\tilde{\phi}(x)) \, dx + \int_D H(\phi(x)) \, dx.$$ (3.83)
3.8. Construction of Locally Optimal Shape Prior Model

On the other hand, since \( \phi_0(T^{-1}x) = 0 \) for any \( x \in \partial T \Omega_0 \), and \( \phi_0(T^{-1}x) > 0 \) for all \( x \in T \Omega_0 \), so we may choose \( \hat{\phi}(x) = \phi_0(T^{-1}x) \).

Because \( T^{-1}x = A(x - a) \), we derive

\[
E_{\text{shape}}(\Omega, T \Omega_0) = \int_D (1 - 2H(\phi(x)))H(\phi_0(A(x - a))) \, dx + \int_D H(\phi(x)) \, dx.
\] (3.84)

The Gâteaux derivative of (3.80) is

\[
dE_{\text{shape}}(\Omega, \Omega_0; \nu) = -\int_{\partial \Omega} (1 - 2\chi_{\Omega_0}) \nu(0, x) \cdot n(x) \, da(x).
\] (3.85)

In order to calculate the derivatives of \( E_{\text{shape}} \) with respect to the affine transformation parameter \( a, \mu, \theta, s_x, s_y \) we observe that \( \frac{\partial A}{\partial \mu} = JS_xS_y, \frac{\partial A}{\partial \theta} = J_0 A, \frac{\partial A}{\partial s_x} = \mu JS_y \) and \( \frac{\partial A}{\partial s_y} = \mu JS_x \) where

\[
J_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad S_y = \begin{pmatrix} s_y & 1 \\ 0 & 0 \end{pmatrix}, \text{ and } S_x = \begin{pmatrix} s_x & 0 \\ 1 & 0 \end{pmatrix}.
\]

Consequently

\[
\frac{\partial E_{\text{shape}}}{\partial a} = -\int_D (1 - 2H(\phi(x))) \delta(\phi_0(A(x - a))) \nabla \phi_0 A \, dx; \quad (3.86)
\]

\[
\frac{\partial E_{\text{shape}}}{\partial \mu} = \int_D (1 - 2H(\phi(x))) \delta(\phi_0(A(x - a))) \nabla \phi_0 JS_xS_y(x - a) \, dx \quad (3.87)
\]

\[
\frac{\partial E_{\text{shape}}}{\partial \theta} = \int_D (1 - 2H(\phi(x))) \delta(\phi_0(A(x - a))) \nabla \phi_0 J_0 A(x - a) \, dx \quad (3.88)
\]

\[
\frac{\partial E_{\text{shape}}}{\partial s_x} = \mu \int_D (1 - 2H(\phi(x))) \delta(\phi_0(A(x - a))) \nabla \phi_0 JS_xS_y^-(x - a) \, dx \quad (3.89)
\]

\[
\frac{\partial E_{\text{shape}}}{\partial s_y} = \mu \int_D (1 - 2H(\phi(x))) \delta(\phi_0(A(x - a))) \nabla \phi_0 JS_x^-(x - a) \, dx. \quad (3.90)
\]

where \( \delta \) is the dirac function at the origin.

3.8.6 Evolution Equation

Next we would minimize the functional (3.55). The standard approach to obtain a locally optimal solution \( \Gamma^* \), using gradient descent to evolve the curve \( \Gamma \). Starting with an initial curve \( \Gamma(0) = \Gamma_0 \) and evolving the contour according to the evolution equation

\[
\frac{\partial \Gamma}{\partial t} = -\nabla_H E_{\text{Total}}.
\] (3.91)
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Here $\mathcal{H}$ denotes the Hilbert space in which the curve $\Gamma$ is embedded. In this work we assume that $\mathcal{H} = L^2$, the space of square integrable functions.

Compare Eq. 3.22 and Eq. 3.91. Eq. 3.22 is governed by the velocity function $V$, while the evolution described by Eq. 3.91 is governed by several energies such data, shape and length of curve.

To be able to apply this approach we thus need to obtain an expression for the gradient. To achieve this we compute the Gâteaux derivative of the functional $E_{\text{Total}}$, and through this obtain the Euler-Lagrange equation which provides an expression of the gradient.

### 3.9 Construction of Convex Segmentation Shape Prior Model

#### 3.9.1 Convex Segmentation Model for Ultrasound Data

We assume a image $I$ is a function in $L^\infty(D)$, where $D$ is a regular open bounded set corresponding to the image domain. The set $U$ denotes all image regions in $D$, i.e. the set of regular open bounded sets of $D$. Let $\chi_\Omega$ be the characteristic function of the set $\Omega$

$$
\chi_\Omega(x) = \begin{cases} 1, & x \in \Omega \in U, \\ 0, & \text{otherwise} \end{cases}
$$

(3.92)

Segmentation is achieved by minimizing the energy

$$
E(\Omega) = E(\chi_\Omega) = \lambda(HL_{in}(\chi_\Omega) + HL_{out}(\chi_\Omega)) + \gamma E_{\text{Shape}}(\chi_\Omega, \chi_\Omega_0) + \mu L(\chi_\Omega)
$$

(3.93)

where $\Omega_0$ is a fixed shape prior for the sought region $\Omega$, and parameters $\lambda, \gamma, \mu > 0$. The segmentation functional is composed of the Hellinger distance ($HL$) between the prior inside and outside distribution $\hat{p}_{in}$ and $\hat{p}_{out}$ and the corresponding Parzen window [159] estimated distributions $p_{in}$ and $p_{out}$.

The Parzen estimated distribution $p_{in}$ and $p_{out}$ are defined by

$$
p_{\text{in}}(x, \chi_\Omega) = \frac{\int_D G_\sigma(x - I(y))\chi_\Omega(y)dy}{\int_D \chi_\Omega(y)dy}
$$

(3.94)

and

$$
p_{\text{out}}(x, \chi_\Omega) = \frac{\int_D G_\sigma(x - I(y))(1 - \chi_\Omega(y))dy}{\int_D (1 - \chi_\Omega(y))dy},
$$

(3.95)

where $G_\sigma(x) = e^{-x^2/2\sigma^2}$ is a Gaussian kernel. One can also consider an asymmetric kernel e.g. Gamma. The parameter $\sigma$ of the kernel is selected ad-hoc.
3.9. Construction of Convex Segmentation Shape Prior Model

The Hellinger distance between a prior distribution \( \hat{p}_{\text{in}} \) and \( p_{\text{in}} \) is provided by

\[
HL_{\text{in}}(\chi_{\Omega}) = \int_{\mathbb{R}^+} \left( \sqrt{\hat{p}_{\text{in}}(x)} - \sqrt{p_{\text{in}}(\chi_{\Omega}(x))} \right)^2 \, dx ,
\]  

with \( HL_{\text{out}}(\chi_{\Omega}) \) defined in an analogous manner.

The same shape prior term (3.80) as in the locally optimal model is used, i.e.

\[
E_{\text{Shape}}(\chi_{\Omega}, \chi_{\Omega_0}) = \int_D (1 - 2\chi_{\Omega_0}(x))\chi_{\Omega}(x) \, dx + \int_D \chi_{\Omega_0}(x) \, dx .
\]  

Finally the regularity of the region \( \Omega \) is measured by Total variation, i.e

\[
E_{\text{TV}}(\chi_{\Omega}) = \int_D |\nabla \chi_{\Omega}(x)| \, dx
\]  

In order to formulate a convex optimization problem from the energy (3.93) we need to do the following: 1) Determine that there exists at least one minimizer of \( E(\Omega) \), 2) Propose an equivalent convex formulation of (3.93).

3.9.2 Gâteaux Derivatives of Energy Terms of Convex Segmentation Model

The Gâteaux derivative of \( HL_{\text{in}} \) in direction \( v \) is

\[
dHL_{\text{in}}(\Omega; v) = \int_{\partial \Omega} \left\{ \int_{\mathbb{R}^+} \frac{1}{[\Omega]} \left( \sqrt{\frac{p_{\text{in}}(y)}{\hat{p}_{\text{in}}(y, \chi_{\Omega})}} - 1 \right) G_{\sigma}(y - I(x)) \, dy \right\} \times
\]

\[
v(0, x) \cdot n(x) da(x)
\]  

where \( V_{\text{in}}(x, \Omega) \) is the velocity of \( HL_{\text{in}} \) at point \( x \in \partial \Omega \). Eq. 3.99 follows from Thm. 3.6.2. For \( HL_{\text{out}} \), \( dHL_{\text{out}}(\Omega; v) \) is analogously determined.

The Gâteaux derivative of \( E_{\text{shape}} \) in direction \( v \) is provided in (3.85), while the Gâteaux derivative of \( E_{\text{TV}} \) is

\[
dE_{\text{TV}}(\Omega; v) = - \int_{\partial \Omega} \text{div} \left( \frac{\nabla \chi_{\Omega}(x)}{|\nabla \chi_{\Omega}(x)|} \right) v(0, x) \cdot n(x) da(x) .
\]  

69
3.9.3 Existence of Minimizing Solution of Proposed Segmentation Functional

**Theorem 3.9.1.** The minimization problem

$$\min_{\chi_\Omega \in BV(D)} \left\{ \lambda \left( HL_{in}(\chi_\Omega) + HL_{out}(\chi_\Omega) \right) + \gamma E_{Shape}(\chi_\Omega, \chi_{\Omega_0}) + \lambda L(\chi_\Omega) \right\}, \lambda, \gamma, \mu > 0 \tag{3.101}$$

has a solution in $BV(D)$.

**Proof:**

A) To show that

$$\lim_{j \to \infty} HL(\hat{p}(\cdot, p(\cdot, I, \chi_{\Omega_{kj}}))) = HL(\hat{p}(\cdot, p(\cdot, I, \chi_{\Omega_*}))), \tag{3.102}$$

we analyse the case of $HL_{in}$, as $HL_{out}$ is worked out analogously.

Since $I \in L^\infty(D)$, we see clearly that there are constants $c, C_0 > 0$ such that

$$C_0^{-1} G_\sigma(x + c) \leq G_\sigma(x - I(y)) \leq C_0 G_\sigma(x - c), \quad \forall (x, y) \in \mathbb{R}^+ \times D, \tag{3.103}$$

thus for all domain $\Omega$ and all $x \in \mathbb{R}^+$

$$C_0^{-1} G_\sigma(x + c) \leq p_{in}(\chi_\Omega)(x) \leq \int_D C_0 G_\sigma(x - c) \chi_\Omega(y) \frac{dy}{\int_D \chi_\Omega(y) dy} = C_0 G_\sigma(x - c). \tag{3.104}$$

By definition

$$|p_{in}(\chi_{\Omega_{kj}}) - p_{in}(\chi_{\Omega_*})| \leq \int_D \frac{G_\sigma(x - I(y))}{|\Omega_*||\Omega_{kj}|} |\chi_{\Omega_{kj}}(y) - \chi_{\Omega_*}(y)|\, dy$$

$$+ \chi_{\Omega_*}(y) \int_D |\chi_{\Omega_{kj}}(t) - \chi_{\Omega_*}(t)| \, dt \, dy \tag{3.105}$$

$$\leq \left( \frac{C_0}{|\Omega_*|} + \frac{C_0}{|\Omega_{kj}|} \right) G_\sigma(x - c) |\chi_{\Omega_{kj}} - \chi_{\Omega_*}|. \tag{3.105}$$

Once again by the definition, we may rewrite

$$HL_{in}(\chi_{\Omega_{kj}}) - HL_{in}(\chi_{\Omega_*}) = \int_{\mathbb{R}^+} Df_j(x) \, dx$$

$$= \int_0^M Df_j(x) \, dx + \int_M^\infty Df_j(x) \, dx, \tag{3.106}$$
where

\[ Df_j(x) = \left[ 2 \sqrt{p_n} + \sqrt{p_n(\chi_{\Omega_j})} + \sqrt{p_n(\chi_{\Omega_j})}(\sqrt{p_n(\chi_{\Omega_j})} - \sqrt{p_n(\chi_{\Omega}))}. \right. \]

(3.107)

Let \( g(x) = \sqrt{C_0 G_\sigma(x - c)}, g_0(x) = \sqrt{C_0^{-1} G_\sigma(x + c)}, \) then it follows from (3.104) that \(|Df_j(x)| \leq 4[\sqrt{p_n(x)} + g(x)]|g(x)|, \) and from (3.105) that

\[ |Df_j(x)| \leq \left[ 1 + g_0^{-1}(x)\sqrt{p_n(x)}\right]g_0(x)|\Omega_{kj}|^{-1} + |\Omega_*|^{-1}|\chi_{\Omega_j} - \chi_{\Omega_*}|, \]

(3.108)

For any \( \varepsilon > 0, \) it follows from the above estimate that there is \( M > 0 \) such that \( \int_M^{\infty} |Df_j(x)|\, dx \leq \frac{\varepsilon}{2}. \) Furthermore, there is \( N > 0 \) such that

\[ (|\Omega_{kj}|^{-1} + |\Omega_*|^{-1})|\chi_{\Omega_j} - \chi_{\Omega_*}| \leq \frac{\varepsilon}{2M_0}, \quad \forall j \geq N, \]

(3.109)

where \( M_0 = \int_0^M (1 + g_0^{-1}(x)\sqrt{p_n(x)})g^2(x)\, dx, \) and thereafter for all \( j \geq N \)

\[ \int_0^M |Df_j(x)|\, dx \leq M_0 \cdot \frac{\varepsilon}{2M_0} = \frac{\varepsilon}{2} \quad \Rightarrow \quad \int_0^\infty |Df_j(x)|\, dx < \varepsilon, \]

(3.110)

which shows that \( H\{\chi_{\Omega_j}\} \to H\{\chi_{\Omega_*}\} \) as \( j \to \infty. \)

B) Let \( \{\chi_{\Omega_k}\} \) be a minimizing sequence of (3.93), \( i.e. \)

\[ \lim_{k \to \infty} E(\chi_{\Omega_k}) = \min_{\chi \in BV(D)} E(\chi). \]

(3.111)

It follows from Definition 3.4.2 that there is a constant \( M > 0 \) such that

\[ |\nabla \chi_{\Omega_k}|_{L^1(D)} \leq M, \forall k \geq 1. \]

(3.112)

Therefore \( \{\chi_{\Omega_k}\} \) is a bounded sequence in \( BV(D). \) Following the Theorem 3.4.2, there is a subsequence \( \chi_{\Omega_k} \) and a function \( u_0 \in BV(D) \) such that \( \chi_{\Omega_k} \to u_0 \) in \( L^1(D). \)

Furthermore, since \( \chi_{\Omega_k} \in \{0, 1\}, \) and \( \chi_{\Omega_k} \to u_0 \) a.e. in \( D, \) thus \( u_0 \in \{0, 1\} \) a.e. in \( D \) and hence there is a domain \( \Omega_* \subset D \) such that \( u_0 = \chi_{\Omega_*}. \)
CHAPTER 3. Segmentation of the Endocardium - Variational approach

C) Since $E_{\text{shape}}(\chi_{\Omega} \mid \chi_{\Omega_0})$ is linear in $\chi_{\Omega_0}$, it follows from the $L^1$-convergence that

$$\lim_{j \to \infty} E_{\text{shape}}(\chi_{\Omega_j} \mid \chi_{\Omega_0}) = E_{\text{shape}}(\chi_{\Omega} \mid \chi_{\Omega_0}).$$

(3.113)

Combining A) - C), we derive $E(\chi_{\Omega}) = \min_{\chi_{\Omega} \in BV(D)} E(\chi_{\Omega})$, i.e., $\chi_{\Omega}$ is a minimizer of problem 3.93.

3.9.4 Convexification of Energy Functional

Consider the functional

$$E(\Omega) = \text{Per}(\Omega; D) + \int_{\Omega} I_1(x) dx + \int_{\Omega^c} I_2(x) dx$$

(3.114)

with $I_1, I_2 \in L^\infty(D)$ and defined the non-convex minimization problem

$$\min_{\Omega \subseteq D} E(\Omega).$$

(3.115)

Define a corresponding minimization problem

$$\min_{u \in K} \tilde{E}(\Omega)$$

(3.116)

s.t.

$$\tilde{E}(u) = \int_D |\nabla u(x)| + (I_1 - I_2)(x)u(x) dx$$

(3.117)

and

$$K = \{u(x); 0 \leq u \leq 1, x \in D, u \in BV(D)\}.$$  

(3.118)

Remark 3.9.1. $E(\Omega)$ can be also expressed in terms of characteristic functions.

$$E(\Omega) = \int_D |\nabla \chi_{\Omega}(x)| + \int_D I_1(x)\chi_{\Omega}(x) + I_2(x)(1 - \chi_{\Omega}(x))$$

$$= \int_D |\nabla \chi_{\Omega}(x)| + \int_D (I_1(x) - I_2(x))\chi_{\Omega}(x) + \int_D I_2(x) \text{ constant.}$$

(3.119)

(3.120)

Since $\{\chi_{\Omega}(x) : \Omega \subseteq D\}$ is not a convex set, so the minimization of (3.115) is not a convex minimization problem.

In Theorems 3.9.2 - 3.129 we show that that the problem (3.116) is a convex minimization problem, and its solutions are also minimizing solutions to the problem (3.115).

Theorem 3.9.2. $\tilde{E}(u)$ is convex and (3.116) has (unique) minimizer $u_*$.  

72
3.9. Construction of Convex Segmentation Shape Prior Model

Proof: Let \( u_1, u_2 \in \mathcal{K} \) and \( \alpha, \beta \geq 0, \alpha + \beta = 1 \) then \( \alpha u_1 + \beta u_2 \in \mathcal{K} \) and

\[
\tilde{E}(\alpha u_1 + \beta u_2) = \int_D |\nabla(\alpha u_1 + \beta u_2)| + (I_1 - I_2)(\alpha u_1 + \beta u_2)dx
\]

\[
= \int_D |\alpha \nabla u_1 + \beta \nabla u_2| + \alpha(I_1 - I_2)u_1 + \beta(I_1 - I_2)u_2dx
\]

\[
\leq \int_D \alpha |\nabla u_1| + \beta |\nabla u_2| + \alpha(I_1 - I_2)u_1 + \beta(I_1 - I_2)u_2dx
\]

\[
= \alpha \tilde{E}(u_1) + \beta \tilde{E}(u_2). \tag{3.121}
\]

Clearly \( \tilde{E}(u) \) is bounded from below i.e. \( \tilde{E}(u) \geq -\int_D (|I_1| + |I_2|)dx \forall u \in \mathcal{K} \). To show the existence of minimizer \( u_* \). Let \( \{u_k\} \in \mathcal{K} \) be a minimizing sequence

\[
\tilde{E}(u_k) \to \min_{u \in \mathcal{K}} \tilde{E}(u), \tag{3.122}
\]

then \( \{u_k\} \in BV(D) \) is a bounded sequence and by Theorem 3.4.1 there is a subsequence \( \{u_{k_j}\} \) and a \( u_* \in BV(D) \) s.t. \( u_{k_j} \to u^* \) in \( L^1(D) \) and

\[
\int_D |\nabla u_*|dx \leq \lim_{j \to \infty} \int_D |\nabla u_{k_j}|dx, \tag{3.123}
\]

and thus \( u_{k_j} \to u_* \) a.e. in \( D \). Since \( 0 \leq u_{k_j} \leq 1 \), it is clear \( 0 \leq u_* \leq 1 \) a.e. in \( D \) i.e. \( u_* \in \mathcal{K} \),

\[
\tilde{E}(u_*) \leq \lim_{j \to \infty} E(u_{k_j}) = \min_{u \in \mathcal{K}} \tilde{E}(u) \tag{3.124}
\]

i.e. \( u_* \) is a minimizer. ■

**Theorem 3.9.3.** \( \min_{\Omega \subseteq D} E(\Omega) \) has a global solution.

*Proof:* Same argument as in proof of Theorem 3.9.2 applies. ■

**Theorem 3.9.4.**

\[
\min_{\Omega \subseteq D} E(\Omega) = \min_{u \in \mathcal{K}} \tilde{E}(u) + C_0 \tag{3.125}
\]

where \( C_0 = \int_D I_2(x)dx \). Moreover i) if \( u \) is a minimizer of (3.116), then for a.e. \( \mu \in [0,1] \), \( \Omega_\mu = \{x \in D; u(x) > \mu\} \) is a minimizer of (3.115). ii) If \( \Omega_* \) is a minimizer of (3.115), then \( \chi_{\Omega_*} \) is a minimizer of (3.116).

*Proof:* First we show (3.125). By the co-area formula we may rewrite

\[
\int_D |\nabla u|dx = \int_0^1 \text{Per}(\Omega_\mu; D)d\mu \tag{3.126}
\]
CHAPTER 3. Segmentation of the Endocardium - Variational approach

where

\[ \Omega_\mu = \{ x \in D; u(x) > \mu \} \]  \hspace{1cm} (3.127)

and

\[ \int_D (I_1 - I_2) u(x) dx = \int_D (I_1 - I_2) \int_0^{u(x)} d\mu \]

\[ = \int_0^1 d\mu \int_{D \cap \{ x : u(x) > \mu \}} (I_1 - I_2) dx d\mu \Rightarrow \]

\[ \tilde{E}(u) = \int_0^1 \left[ \text{Per}(\Omega_\mu; D) + \int_{\Omega_\mu} I_1(x) dx + \int_{\Gamma^e} I_2(x) dx - \int_D I_2(x) dx \right] d\mu \]

\[ = \int_0^1 E(\Omega_\mu) d\mu - C_0 , \]  \hspace{1cm} (3.128)

where \( C_0 = \int_D I_2(x) dx \). Thus we have

\[ \min_{u \in K} \tilde{E}(u) \geq \min_{\Omega \subseteq D} E(\Omega) - C_0 . \]  \hspace{1cm} (3.129)

On the other hand

\[ \min_{\Omega \subseteq D} E(\Omega) = \min_{u \in [0,1]} \tilde{E}(u) + C_0 \geq \min_{u \in [0,1] u \in BV(D)} \tilde{E}(u) + C_0 . \]  \hspace{1cm} (3.130)

Since \{ u \in BV(D); u(x) \in \{0,1\} \} \subset K, (3.129) and (3.130) imply that the equality (3.125) holds.

To show i) we let \( u_* \) be a minimizer of (3.116), then by Theorem 3.125

\[ \min_{\Omega \subseteq D} E(\Omega) = \tilde{E}(u_*) + C_0 = \int_0^1 E(\Omega_\mu) d\mu \]  \hspace{1cm} (3.131)

\[ \Rightarrow \int_0^1 [E(\Omega_\mu) - \min_{\Omega \subseteq D} E(\Omega)] d\mu = 0 . \]  \hspace{1cm} (3.132)

But \( E(\Omega_\mu) - \min_{\Omega \subseteq D} E(\Omega) \geq 0 \) therefore \( E(\Omega_\mu) = \min_{\Omega \subseteq D} E(\Omega) \) a.e. in \([0,1]\]. To show ii) let \( \Omega_* \) be a minimizer of (3.115) then \( E(\Omega_*) = \tilde{E}(\chi_{\Omega_*}) + C_0 \), by Theorem 3.125

\[ E(\Omega_*) = \min_{\Omega \subseteq D} E(\Omega) = \min_{u \in K} \tilde{E}(u) + C_0 \Rightarrow \]

\[ C_0 + \min_{u \in K} \tilde{E}(u) = C_0 + \tilde{E}(\chi_{\Omega_*}) + C_0 \]  \hspace{1cm} (3.133)

\[ \Rightarrow \min_{u \in K} \tilde{E}(u) = \tilde{E}(\chi_{\Omega_*}) . \]  \hspace{1cm} (3.135)
3.9. Construction of Convex Segmentation Shape Prior Model

3.9.5 Convexification of Level Set Formulation

In order to define a convex level set formulation, we as in e.g. Chan et al. [40], consider the evolution equation of (3.93) in terms of level sets. Then we propose a convex energy which has the same evolution equation, and thus the solution to the convex energy is a solution to the original energy.

First we formulation $HL_{in}$ in terms of level sets

$$HL_{in}(\phi(x)) = \int_{\mathbb{R}^+} \left( \sqrt{p_{in}(x)} - \sqrt{p_{in}(\phi(x))} \right)^2 dx.$$  \hspace{1cm} (3.136)

with

$$p_{in}(\phi(x)) = \frac{\int_D G_{\sigma}(x - I(y))H_\epsilon(\phi(y))dy}{\int_D H_\epsilon(\phi(y))dy},$$  \hspace{1cm} (3.137)

where $H_\epsilon$ is a compactly supported approximation of Heaviside function.

$$E_1(\phi) = \int_D \lambda [HL_{in}(\phi(x))H_\epsilon(\phi(x)) + HL_{out}(\phi(x))(1 - H_\epsilon(\phi(x)))]$$

$$+ \int_D \gamma (1 - 2H_\epsilon(\phi_0(x)))H_\epsilon(\phi(x)) + \int_D |\nabla H_\epsilon(\phi(x))| dx$$

$$= \int_D \lambda HL_{out}(\phi(x)) + \gamma H_\epsilon(\phi_0(x)) + |\nabla H_\epsilon(\phi(x))| dx$$

$$+ H_\epsilon(\phi(x)) [HL_{in}(\phi(x)) - HL_{out}(\phi(x)) - 2H_\epsilon(\phi_0(x))] dx.$$  \hspace{1cm} (3.138)

The velocity of the term $HL_{in}(\phi(x))$ is denoted by $V_{HL_{in}}(\phi(x))$, with $V_{HL_{out}}(\phi(x))$ defined correspondingly, and is the level set formulation of $V_{HL}(x, \Omega)$ defined in Sec. 3.9.2.

Assuming $V_{HL}$ and $\phi_0$ are fixed, and updated iteratively between each evolution step, the evolution equation corresponding to (3.138) becomes

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \{ \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda \left[ V_{HL_{in}} + V_{HL_{out}} \right] + 2\gamma H_\epsilon(\phi_0(x)) \} ,$$  \hspace{1cm} (3.139)

which has the same steady state solution as

$$\frac{\partial \phi}{\partial t} = \{ \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda \left[ V_{HL_{in}} + V_{HL_{out}} \right] + 2\gamma H_\epsilon(\phi_0(x)) \} ,$$  \hspace{1cm} (3.140)

since $\delta_\epsilon(\phi) \geq 0$.

Now we propose a energy which has the same first variation, and hence same evolution equation as (3.138)

$$E_2(\phi) = \int_D |\nabla \phi| + \left[ \lambda (HL_{in} - HL_{out}) + 2\gamma H_\epsilon(\phi_0(x)) \right] \phi(x)dx$$  \hspace{1cm} (3.141)
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Assume that $\psi \in C_0^1(\Omega)$. Then first variation of $E_2(\phi)$ is

$$d(E_2[\phi]; \psi) = dE[\phi + s\psi]|_{s=0} = \int_D \frac{d}{ds}(|\nabla (\phi + s\psi)|)|_{s=0}$$

$$+ \left[ \lambda (H_{\text{Lin}} - H_{\text{Out}}) + 2\gamma H_e(\phi_0(x)) \right] \phi + s\psi]\big|_{s=0}$$

$$= \int_D \left( - \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \left[ \lambda (V_{\text{Lin}}^{\text{in}} - V_{\text{HL}}^{\text{out}}) + 2\gamma H_e(\phi_0(x)) \right] \right) \psi dx$$

(3.142)

since integration by parts provides that

$$\int_D \nabla \phi \cdot \nabla \psi = -\int_D \psi \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) .$$

(3.143)

The energy (3.141) is homogeneous of degree one, and thus it may not have a global minimizer. For a standard level set formulation this is non-problematic, since in level set terms we are only interested in the zero-level set of $\phi$. By constraining the $\phi$ s.t. $0 \leq \phi(x) \leq 1$, the energy (3.141) is converted into a convex form (i.e. convexified).

The evolution equation of the segmentation functional (3.93), again assuming fixed velocities of Hellinger functionals and fixed shape prior updated iteratively, expressed in terms of level sets becomes

$$\frac{\partial \phi}{\partial \tau} = \left( \lambda (V_{\text{HL}}^{\text{in}} + V_{\text{HL}}^{\text{out}}) + \gamma (1 - 2\phi_0) + \mu \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right) |\nabla \phi| ,$$

(3.144)

The gradient of the shape prior term $E_{\text{Shape}}$ is

$$\nabla E_{\text{Shape}}(\phi, \phi_0) = 1 - 2\phi_0 .$$

(3.145)

The steady state of (3.146) is the same as:

$$\frac{\partial \phi}{\partial \tau} = \lambda (V_{\text{HL}}^{\text{in}} + V_{\text{HL}}^{\text{out}}) + \gamma (1 - 2\phi_0) + \mu \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) .$$

(3.146)

The variational model that provides (3.148) is also given by:

$$\min_{\phi \in [0,1]} \int_D -\lambda (V_{\text{HL}}^{\text{in}} + V_{\text{HL}}^{\text{out}}) \phi - \gamma (1 - 2\phi_0) \phi + |\nabla \phi|$$

(3.147)

$$= \min_{\phi \in [0,1]} \int_D -\lambda \left[ (V_{\text{Lin}}^{\text{in}} + V_{\text{HL}}^{\text{out}}) + \frac{\gamma}{\lambda} (1 - 2\phi_0) \right] \phi + |\nabla \phi|$$

(3.148)

This is a convex minimization problem which has at least one solution.
3.10 Local Endocardial Segmentation Model

This section is based on the paper *Rayleigh Segmentation of Ultrasound Images* [85]. The theoretical prerequisites are provided in:

- Sec. 1.2 (Physical Background)
- Sec. 1.3 (Transmission to Display of Ultrasound Echo)
- Sec. 1.6 (Application in Cardiology)
- Sec. 2.1 (Statistical Modeling of Ultrasound Data)
- Sec. 2.4 (Maximum Likelihood Estimators)
- Sec. 3.3 (Level Set Representation)
- Sec. 3.4 (Functions of Bounded Variation)
- Sec. 3.5 (First necessary condition for relative minimum)
- Sec. 3.6 (Gâteaux Variation of Functional)
- Sec. 3.7 ($L^2$ Shape Gradients)
- Sec. 3.8 (Construction of Locally Optimal Shape Prior Model)

### 3.10.1 Model

We define our local model, which we denote as the *Couple Active Contour* model, as follows:

$$
\Gamma^* = \arg \min_{\Gamma, T} \left\{ E_{CAC}(\Gamma, TT_0) := \alpha E_{Ray}(\Gamma) + \beta E_{Ray}(TT_0) + \gamma E_I(\Gamma, TT_0) \right\}, \tag{3.151}
$$

where $\Gamma$ denotes the active contour, $T_0$ a prior contour and $\alpha, \beta, \gamma > 0$ are weight parameters. $T$ ranges over a group of transformations, so that the interaction becomes pose invariant. Here $TT_0$ denotes the transformed contour, and $\gamma$ is a coupling constant which determines the strength of the contour interaction $E_I$. $E_{Ray}$ denotes a Rayleigh functional, see Sec. 3.10.1. In this way we obtain a model in which the active- and prior contours are treated on an (almost) equal footing, hence we speak of *Coupled Active Contours*. We refer to $TT_0$ as the *sensitive prior*, because it interacts with the image. The latter may be a shape prior, or a so-called *deformable shape prior*, as in our case, cf. Sec. 3.10.1.
CHAPTER 3. Segmentation of the Endocardium - Variational approach

The idea is to let the sensitive prior guide the active contour towards the desired object, resulting in a more robust and rapidly converging segmentation. The CAC model is compared to a standard prior-based model (SPS),

\[
\Gamma^* = \arg \min_{\Gamma, T} \left\{ E_{Ray}(\Gamma) + \gamma E_{\Delta I}(\Gamma, T_{\Gamma_0}) \right\},
\]

(3.152)

with interaction as in Cremers et al. [48], cf. Eq. 3.160 in Sec. 3.10.1, and allowing affine transformation of the prior.

Rayleigh Segmentation

Our main impetus for using the Rayleigh distribution in modeling the ultrasonic B-scan images comes from Sarti et al. [184]. In the continuous setting, a gray scale image is considered to be a real valued function \( I: D \rightarrow [0, \infty) \) defined on the image domain \( D \subset \mathbb{R}^2 \). A point \( x \in D \) is referred to as a pixel, \( I(x) \) as the corresponding gray scale value. We use \( | \cdot | \) to denote length or area, depending on whether \( \cdot \) is a region or curve. Thus \( |\text{int}(\Gamma)| \) denotes the area of the interior of the curve \( \Gamma \), and \(|\text{ext}(\Gamma)|\) is the area of the exterior of the curve \( \Gamma \). In the case of a curve \( \Gamma \), then \( |\Gamma| \) signifies the length of the curve.

Rayleigh segmentation is an active contour model where the idea is to find a contour \( \Gamma \), such that the image \( I \) is optimally approximated by the following statistical image model:

\[
I(x) \sim \begin{cases} 
\text{Ray}(\sigma_{\text{int}}^2), & \text{if } x \in \text{int}(\Gamma) \\
\text{Ray}(\sigma_{\text{ext}}^2), & \text{if } x \in \text{ext}(\Gamma) 
\end{cases}
\]

(3.153)

Here \( X \sim \text{Ray}(\sigma^2) \) means that the random variable \( X \) is Rayleigh distributed with parameter \( \sigma^2 \), i.e. \( X \) has the probability density function

\[
f_X(x) = \begin{cases} 
(x/\sigma^2) \exp(-x^2/2\sigma^2), & x \geq 0 \\
0, & \text{otherwise}
\end{cases}
\]

(3.154)

The Rayleigh functional is the given by,

\[
E_{Ray}(\Gamma) = \nu |\Gamma| + |\text{int}(\Gamma)| \log(\sigma_{\text{int}}^2(\Gamma)) + |\text{ext}(\Gamma)| \log(\sigma_{\text{ext}}^2(\Gamma)),
\]

(3.155)

where \( \nu > 0 \) is a regularization parameter,

\[
\sigma_{\text{int}}^2 = \frac{1}{|\text{int}(\Gamma)|} \int_{\text{int}(\Gamma)} I(x)^2 \, dx
\]

and \( \sigma_{\text{ext}}^2 \) is computed similarly. Let the contour \( \Gamma \) be represented as the zero-level set of the time dependent function \( \phi = \phi(t, x) \) as \( \Gamma = \{ x \in \mathbb{R}^2 : \phi(t, x) = 0 \} \) with
3.10. Local Endocardial Segmentation Model

\[ \text{int}(\Gamma) = \{ x : \phi(t, x) < 0 \} \], \cite{[156]}. Then the contour is evolved in time towards a (local) minimum by solving the level set PDE,

\[
\frac{\partial \phi}{\partial t} = \left[ \nu \kappa + \log \left( \frac{\sigma^2_{\text{int}}}{\sigma^2_{\text{ext}}} \right) + I(x)^2 \left( \frac{1}{\sigma^2_{\text{int}}} - \frac{1}{\sigma^2_{\text{ext}}} \right) \right] |\nabla \phi|, \quad (3.157)
\]

where \( \kappa \) is the curvature of \( \Gamma \).

**Interaction: The Deformable Shape Prior**

The affine pose-invariant interaction \( E_I \) between two contours \( \Gamma \) and \( \Gamma_0 \) is defined by the integral

\[
E_I(\Gamma, T\Gamma_0) = |\det T^{-1}| \int_{\text{int}(\Gamma)} \left( \frac{\phi_0(T^{-1}x)}{W} \right)^3 dx. \quad (3.158)
\]

where \( \phi_0 = \phi_0(y) \) denotes the signed distance function for \( \Gamma_0 \), and the parameter \( W > 0 \) defines the reach of the interaction. The associated \( L^2 \)-gradient is:

\[
\nabla_{\Gamma} E_I(\Gamma, T\Gamma_0) = |\det T^{-1}| \left( \frac{\phi_0(T^{-1}x)}{W} \right)^3. \quad (3.159)
\]

For points close to the sensitive prior, the \( L^2 \)-gradient is small. Therefore the interaction between nearby contours will be weak. This is desirable because it allows the active contour to adapt to the image information in a neighbourhood of width \( \approx W \) around the prior.

**Interaction: Area of the Set Symmetric Difference**

We compare the proposed method and the standard prior-based method (3.152) with the interaction given by the symmetric set difference as in \( e.g. \) \cite{[48]}:

\[
E_{1\triangle}(\Gamma, \Gamma_0) = \text{area}(\text{int}(\Gamma) \triangle \text{int}(\Gamma_0)) = \\
= \int |H(-\phi(x)) - H(-\phi_0(x))|^2 dx, \quad (3.160)
\]

where \( H \) is the usual Heaviside function. The \( L^2 \)-gradient is

\[
\nabla_{\Gamma} E_{1\triangle}(\Gamma) = 1 - 2H(-\phi_0(x)). \quad (3.161)
\]

We refer to segmentation using the pose-invariant functional in (3.152) with \( E_1 = E_{1\triangle} \) as **Standard Prior segmentation (SPS)**.
CHAPTER 3. Segmentation of the Endocardium - Variational approach

The CAC Gradient Descent Flow

The minimization problem (3.151) becomes
\[(\Gamma^*, \rho^*) = \arg \min_{\Gamma, \rho} E_{CAC}(\Gamma, T(\rho)\Gamma_0),\]
which is solved using a standard gradient descent method. The transformation \(T(\rho)\) is defined in Sec. 3.8.1. That is, we find the solution \(t \mapsto (\Gamma(t), \rho(t))\) of the following system of differential equations
\[
\frac{\partial \phi}{\partial t} = \left[ \alpha \nabla \Gamma E_{Ray}(\Gamma) + \gamma \nabla \Gamma E_I(\Gamma, T(\rho)\Gamma_0) \right] |\nabla \phi|, \tag{3.162}
\]
\[
\dot{\rho} = -\beta \nabla \rho E_{Ray}(T(\rho)\Gamma_0) - \gamma \nabla \rho E_I(\Gamma, T(\rho)\Gamma_0). \tag{3.163}
\]
The \(L^2\)-shape gradients with respect to \(\Gamma\) of the Rayleigh functional and the interaction term were computed in (3.157) and (3.159), respectively. The symbol \(\nabla\) denotes the usual (finite dimensional) gradient w.r.t. the group parameters \(\rho\). The \(\rho\)-derivatives are straightforward to compute and look similar to (but are not the same as) those computed in Yezzi et al. [228]. Eqns (3.86)-(3.90) in Sec. 3.8.5 in Chp. 3 describe how to differentiate a functional, in this case \(E_{shape}(\Omega, T\Omega_0)\), w.r.t. transformations parameters.

3.10.2 Level Set Implementation

We here describe the level set implementation [187] of the variational shape prior model described in Sec. 3.10, that is Eq. 3.162. After the level set has been updated, Eq. 3.163 is computed. The obtained equations are analogous to Eqns. 3.86-3.90 in Sec. 3.8.5.

The level set scheme (3.164) corresponding to Eq. 3.162 is
\[
\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^n}{\Delta t} = |\nabla \phi_{i,j}^n| \left[ \alpha \nabla \cdot \left( \nabla \phi_{i,j}^n \right) \right] + \left[ \log \left( \frac{\sigma_{int}^2(\phi_{i,j}^n)}{\sigma_{ext}^2(\phi_{i,j}^n)} \right) + I(x)^2 \left( \frac{1}{\sigma_{int}^2(\phi_{i,j}^n)} - \frac{1}{\sigma_{ext}^2(\phi_{i,j}^n)} \right) \right] + \gamma |\det T^{-1}| \left( \frac{\phi_0(T^{-1}\mathbf{x})}{W} \right)^3 \) \tag{3.164}
\]
First we define the finite difference operators
\[
\Delta^x \phi_{i,j} = \phi_{i+1,j} - \phi_{i,j}, \quad \Delta^y \phi_{i,j} = \phi_{i,j+1} - \phi_{i,j} \tag{3.165}
\]
\[
\Delta^x \phi_{i,j} = \phi_{i+1,j} - \phi_{i,j-1}, \quad \Delta^y \phi_{i,j} = \phi_{i,j+1} - \phi_{i,j} \tag{3.166}
\]
\[
\Delta^{2x} \phi_{i,j} = \phi_{i+1,j} - \phi_{i-1,j}, \quad \Delta^{2y} \phi_{i,j} = \phi_{i,j+1} - \phi_{i,j-1} \tag{3.167}
\]
and shift operators
\[
E_i^x \phi_{i,j} = \phi_{i+1,j}, \quad E_i^x \phi_{i,j} = \phi_{i-1,j} \tag{3.168}
\]
\[
E_i^y \phi_{i,j} = \phi_{i+1,j}, \quad E_i^y \phi_{i,j} = \phi_{i,j-1} \tag{3.169}
\]
3.10. Local Endocardial Segmentation Model

The leading term $|\nabla (\phi^n)|$ is used instead of $\delta(\phi^n)$, since the replacement does not affect the steady state solution, but removes stiffness near the zero level set [231].

Expanding the divergence term, one arrives at the expression

$$\nabla \cdot \left( \frac{\nabla \phi}{|\nabla \phi|} \right)_{j,k} = \left[ \left( \frac{\phi_x}{|\nabla \phi|} \right)_{j+1/2,k} - \left( \frac{\phi_x}{|\nabla \phi|} \right)_{j-1/2,k} \right]/\nabla x + \left[ \left( \frac{\phi_y}{|\nabla \phi|} \right)_{j+1/2,k} - \left( \frac{\phi_y}{|\nabla \phi|} \right)_{j-1/2,k} \right]/\nabla y \quad (3.170)$$

and furthermore

$$\begin{align*}
\left( \frac{\nabla \phi_x}{|\nabla \phi|} \right)_{j+1/2,k} &= \frac{\Delta^x \phi_{i,j}/\Delta x}{\sqrt{\left( \frac{\Delta^x \phi_{i,j}}{\Delta x} \right)^2 + \left( \frac{1}{2} \left( \frac{\Delta^x \phi_{i,j}}{2\Delta y} + \frac{E_x \Delta^x \phi_{i,j}}{2\Delta y} \right) \right)^2}} \\
\left( \frac{\nabla \phi_x}{|\nabla \phi|} \right)_{j-1/2,k} &= \frac{\Delta^x \phi_{i,j}/\Delta x}{\sqrt{\left( \frac{\Delta^x \phi_{i,j}}{\Delta x} \right)^2 + \left( \frac{1}{2} \left( \frac{E_x \Delta^x \phi_{i,j}}{2\Delta y} + \frac{\Delta^x \phi_{i,j}}{2\Delta y} \right) \right)^2}} \\
\left( \frac{\nabla \phi_y}{|\nabla \phi|} \right)_{j,k+1/2} &= \frac{\Delta^y \phi_{i,j}/\Delta y}{\sqrt{\left( \frac{1}{2} \left( \frac{\Delta^y \phi_{i,j}}{\Delta x} + \frac{E_x \Delta^y \phi_{i,j}}{2\Delta y} \right) \right)^2 + \left( \frac{\Delta^y \phi_{i,j}}{\Delta y} \right)^2}} \\
\left( \frac{\nabla \phi_y}{|\nabla \phi|} \right)_{j,k-1/2} &= \frac{\Delta^y \phi_{i,j}/\Delta y}{\sqrt{\left( \frac{1}{2} \left( \frac{E_x \Delta^y \phi_{i,j}}{2\Delta x} + \frac{\Delta^y \phi_{i,j}}{2\Delta y} \right) \right)^2 + \left( \frac{\Delta^y \phi_{i,j}}{\Delta y} \right)^2}}. \quad (3.171)
\end{align*}$$

The level set function $\phi$ is a signed distance function, which needs recomputation (reinitialization) in order to remain a distance function, during the solution process. The velocity field $F$ also needs updating in order to remain smooth. In order to achieve this, a reinitialization step is performed.

A temporary signed distance function $\phi^{\text{temp}}$ is constructed and extension velocity $F^{\text{ext}}$ s.t.

$$\nabla \phi^{\text{temp}} \cdot \nabla F^{\text{ext}} = 0. \quad (3.175)$$

$\phi^{\text{temp}}$ and $F^{\text{ext}}$ coincide with $\phi$ and $F$ on the zero level set and interface respectively. This is achieved by solving the Eikonal equation

$$|\nabla T| = 1 \quad (3.176)$$

on either side of the interface, and setting $T = \phi^{\text{temp}}$.

If $\phi, F \in C^\infty$ the condition (3.175) guarantees that $\phi$ remains a signed distance function.

In the evolution process, it is only necessary to keep track of the values of $\phi$ in the immediate vicinity of the interfaces (the zero level set of $\phi$). Thus the values of $\phi$ are
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Figure 3.6: Update procedure.

only updated in narrow band around the zero level set. The idea is to fix a narrow region around the zero level set and to recompute the values of \( \phi \) only when zero level set touches the boundary of the region. There are several options for implementing narrow band methods, see [1].

3.10.3 Fast Marching Method

Fast Marching is a Finite difference technique for solving the Eikonal equation

\[
|\nabla T| F = 1, \quad T = 0 \quad \text{on} \quad \Gamma .
\]

(3.177)

Fast Marching works by constructing a solution using only upwind values. A two-dimensional Fast Marching scheme has the following appearance,

\[
\left[ \max(D_{ij}^{+x}T, -D_{ij}^{-x}T, 0)^2 + \max(D_{ij}^{+y}T, -D_{ij}^{-y}T, 0)^2 \right]^{1/2} = \frac{1}{F_{ij}},
\]

(3.178)

where

\[
D_{ij}^{+x}u = \frac{u(x + h, y, t) - u(x, y, t)}{h},
\]

(3.179)

\[
D_{ij}^{-x}u = \frac{u(x, y, t) - u(x - h, y, t)}{h}.
\]

(3.180)

\(D_{ij}^{+y}u\) and \(D_{ij}^{-y}u\) are expressed analogously. See [1, 186] for other variants of this type of scheme.

The Fast Marching method builds the solution outward from the smallest \( T \) value, stepping away from the boundary condition in downwind direction. Define three sets: Known, Trial and Far. Known contains grid points on the interface, Trial contains the points located one point away from the interface, and Far contains the remaining points. For each side of the interface, \( \phi^{\text{temp}} \) is updated using Fast Marching, and at the same time \( F^{\text{ext}} \) is obtained by solving (3.175).
3.10. Local Endocardial Segmentation Model

3.10.4 Algorithm

Using the scheme 3.164 we construct an algorithm for obtaining a solution:

(i) Compute initial signed distance function $\phi_0$.
(ii) Compute $\sigma^2_{\text{int}}(\phi^n)$ and $\sigma^2_{\text{ext}}(\phi^n)$.
(iii) Solve for $\phi^{n+1}$ using (3.164) using $\phi^n$.
(iv) Reinitialize signed distance function and update velocity field.

3.10.5 CFL Condition

To ensure stability the front must not cross more than one grid per iteration. This can be stated as the CFL (Courant-Friedrichs-Lewy) condition,

$$\max_D F \Delta t \leq \Delta x \quad (3.181)$$

where $D$ is the domain of computation, $\Delta t$ is time-step and $\Delta x$ spatial step.

3.10.6 Experiments

We demonstrate our method in three experiments: one synthetic and two using ultrasound images. The B-mode ultrasound data was obtained from the database described in Sec. 1.6.3.

The parameter $\gamma$ controls the interaction between the two contours. The active contour can move almost freely in a band of approximately $W = \frac{1}{\sqrt{\gamma}}$ pixels. Too large $\gamma$ will result in the two contours being very closely aligned and fixed in the image and too small $\gamma$ will result in the contours evolving independently of each other. The evolution is stopped when the active contour has remained stable for 20 iterations. We give only approximate ($\approx$) number of iterations. The notation CAC = 'Coupled Active Contour' and SPS = 'Standard Prior Segmentation' is used in this section.

**Synthetic Experiment**

A synthetic image of intensity 100 was subjected to Rayleigh noise with $\sigma^2_{\text{int}} = 1$ inside of a object (rectangle-like structure minus an oval disc) and $\sigma^2_{\text{ext}} = 50$ in the background, cf. Fig. 3.7. We wish to segment the structure without the oval inside. The prior $\Gamma_0$ initially has the same oval shape as the segmenting contour. This is sometimes relevant in medical applications, like endocardium segmentation, where we wish to segment only the endocardium and not structures within or close to the heart wall. Rayleigh segmentation without the use of a prior is ineffective. CAC could move the prior through the oval because it is sensitive to image data. The SPS approach was not equally effective; either the circle is not excluded from segmentation ($\gamma \leq 0.24$) or the active contour becomes trapped by the oval ($\gamma > 0.24$). Varying the curvature parameter $\nu$ did not improve...
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Figure 3.7: Synthetic experiment. Goal was to segment the structure without the oval inside. Prior (white) and active contour (black). (a) Initial position of prior and active contour, (b) Rayleigh segmentation without prior, \( \approx 300 \) its, (c) CAC, \( \approx 250 \) its, (d) SPS, \( \approx 2200 \) its, (e) SPS, \( \approx 2000 \) its, (f) SPS, \( \approx 500 \) its.

the performance of SPS; it only moves the cut-off point (see (d) and (e)). The CAC parameters are \( \beta = 100 \) and \( \gamma = (\frac{1}{25})^3 \) i.e. \( W = 25 \). The parameters for both methods are \( \alpha = 1 \) and \( \nu = 2 \).

**Segmentation of B-mode Ultrasound Images**

Before segmentation, graphics displaying the time of recording etc. was removed from the ultrasound image by applying a mask. The prior \( \Gamma_0 \) has a generic shape, roughly approximating the heart chamber. It was constructed by manually smoothing the mean shape of nine endocardium outlines from different patients at a specific moment in the cardiac cycle. In both experiments the transformation of the prior was affine, i.e. has 6 degrees of freedom. The curvature parameter \( \nu \) was set high to ensure regular segmentations. Varying \( \nu \) did not alter the results for SPS significantly, except the smoothness of the
3.11. Convex Endocardial Segmentation Model

segmented contour.

US Experiment 1

For both SPS and CAC the active contour and prior were initialized in the center of the endocardium. In CAC segmentation the prior guides the active contour past the irrelevant dots in the chamber and ends up close to the outline given by clinician. Furthermore, the active contour was allowed to deviate from the prior, whereby segmentation of surrounding tissue is avoided. As is clearly seen in the top row of Fig. 3.8 SPS cannot achieve a useful segmentation, since it cannot get past the irrelevant dots. If \( \gamma \) is set low \((\approx 0.1)\) then the prior does not help the segmentation at all, and if set higher it still cannot help the active contour past the dots. This is due to the fact that the prior (in the standard method) does not interact with image. The CAC parameters are \( \beta = 20 \) and \( \gamma = \left(\frac{1}{3}\right)^3 \) \( i.e. W = 3 \). Parameters for both methods are \( \alpha = 1 \) and \( \nu = 7 \).

US Experiment 2

We demonstrate the performance of CAC vs. SPS with respect to poor initialization. In Figure 3.9 we see that CAC converged nicely, while SPS fails. In SPS the prior is unable to guide the active contour to the chamber. For \( \gamma < 1 \) SPS fails, since the active contour disappears. For \( \gamma > 1 \) we obtained similar results as for \( \gamma = 1 \). The ultrasound image used here was taken from a different part of the heart sequence than in Experiment 1. CAC parameters: \( \beta = 30 \), \( \gamma = \left(\frac{1}{3}\right)^3 \) \( i.e. W = 3 \). Parameters for both methods: \( \alpha = 1 \), \( \nu = 15 \).

Rayleigh Model in Endocardium Segmentation

We have observed that the Rayleigh statistics, used in both CAC and SPS model, did not sufficiently describe the endocardium. This is illustrated in Fig. 3.10, where we have used CAC to segment the endocardium. It is clear that the upper part of the chamber is not modeled by the Rayleigh. This is a shortcoming of the statistical model, which influences the performance of CAC, since both active contour and sensitive prior depend on the Rayleigh model.

3.11 Convex Endocardial Segmentation Model

This section is based on the paper Convex Spatio-Temporal Segmentation of the Endocardium in Ultrasound Data using Distribution and Shape Priors [84].

– Sec. 1.2 (Physical Background)

– Sec. 1.3 (Transmission to Display of Ultrasound Echo)
Figure 3.8: US Experiment 1: CAC vs. SPS with favorable initialization of active contour and prior. Clinician outline of heart chamber (dashed white), prior (red) and active contour (white). (a) Initialization, (b) Rayleigh segmentation without prior (c) CAC, $\approx 300$ its, (d) SPS, $\approx 2000$ its, (e) SPS, $\approx 1000$ its, (f) SPS, $\approx 1500$ its, (g) SPS, $\approx 2000$ its.

- Sec. 1.6 (Application in Cardiology)
- Sec. 2.1 (Statistical Modeling of Ultrasound Data)
- Sec. 2.2 (Mixture Models)
- Sec. 2.4 (Maximum Likelihood Estimators)
3.11. Convex Endocardial Segmentation Model

Figure 3.9: US Experiment 2: CAC vs. SPS w.r.t. poor initialization. Clinician outline of heart chamber (dashed white), prior (red) and active contour (white). (a) Initialization, (b) CAC, \( \approx 500 \) its, (c) SPS, \( \approx 600 \) its.

Figure 3.10: Rayleigh model in endocardium segmentation. Clinician outline of heart chamber (dashed white) and active contour (white). CAC segmentation of three different echocardiographic images. Note that upper part of endocardium is not segmented when using the Rayleigh model.

- Sec. 2.5 (EM Algorithm)
- Sec. 2.8 (Similarity measures)
- Sec. 3.2 (Dimensionality Reduction)
- Sec. 3.3 (Level Set Representation)
- Sec. 3.4 (Functions of Bounded Variation)
- Sec 3.5 (First Necessary Condition for Relative Minimum)
- Sec. 3.6 (Gâteaux Variation of Functional)
- Sec. 3.7 (\( L^2 \) Shape Gradients)

87
3.11.1 Mixture Speckle Model

Here we have chosen to employ the Rayleigh distribution to model the speckle in our ultrasound images. The Rayleigh probability density function is given by

$$p_{\text{rayl}}(I|\sigma) = \frac{I}{\sigma} \exp\left(-\frac{I^2}{2\sigma}\right),$$

where $I$ denotes intensity values and $\sigma > 0$ is a parameter. The inside and outside of the endocardium is modeled as a mixture model of Rayleigh densities (to account for speckle), including a Dirac measure at zero $\delta_0$ (to account for signal dropout). Thus our proposed mixture model is defined as

$$p(x|\alpha, \sigma) \propto \alpha_1 \delta_0 + \sum_{i=2}^{n} \alpha_i p_{\text{rayl}}(x|\sigma_i) 1_{(x>0)},$$

where $\alpha_i > 0$, s.t. $\sum \alpha_i = 1$, are the mixture weights, $\delta_0$ denotes the Dirac delta at zero, and $1$ is the indicator function.

Empirically we have observed that the pixel intensities of ultrasound images can be divided into 4 categories: (1) signal dropout, (2) echocardiographic artifacts, (3) blood, and (4) tissue. Signal dropout implies no echo returned to the transducer, resulting in zero intensity. This may also be due to post-processing where very low intensities are quantized to zero. By artifacts we refer to areas with tissue-like intensity caused by chordae, papillary muscles, ribs or local increase in signal strength. The inside of the endocardium is assumed to consist of a mixture of categories (1), (2) and (3), while the outside is taken to be a mixture of the 4 categories.

Given an input sequence of $N$ ultrasound images we wish to determine the mixing weights of each category on the inside and outside of the endocardium, respectively, denoted by $\alpha_{in}^j$ and $\alpha_{out}^j$, $j = 1,...,4$. Let

$$\alpha_{inp}^j = \alpha_{out}^j + \alpha_{in}^j,$$

where $\alpha_{inp}^j$ is the weight, determined by Expectation Maximization (EM), for category $j$ in the input sequence. We use annotated data to calculate $\alpha_{in}^j$, $j = 1,2,3$. By assumption $\alpha_{in}^4 = 0$. Let $\alpha_{in}^j$ be the fraction, estimated by EM, of category $j$ on the inside of the endocardium of annotated sequence $i$. Furthermore, let $A_{seq}^i$ denote the number of pixels in annotated sequence $i$, and $A_{ine}^i$ the number of pixels inside the endocardium in the same sequence. Then we get

$$\alpha_{in}^j = \frac{1}{N} \sum_{i=1}^{N} \alpha_{in}^j \cdot \frac{A_{ine}^i}{A_{seq}^i}, \quad j = 1,...,3.$$

Finally $\alpha_{out}^j$ is determined by (3.184).
3.11. Convex Endocardial Segmentation Model

**Figure 3.11:** Sequence A. Calculated contour (*solid white*) and ground truth (*dashed red*). End-Systole/Diastole - left/right

### 3.11.2 Convex Segmentation Model with Shape Prior

Segmentation is achieved by minimizing the distance between the prior inside/outside distribution $\hat{p}_{in}/\hat{p}_{out}$ and the corresponding estimated ones $p_{in}/p_{out}$, which are stated in equations (3.94) and (3.95).

We use the Hellinger distance, see Sec. 2.8, to measure closeness between estimated
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Figure 3.12: Sequence A. Histogram of ground truth distributions inside (top)/outside (bottom) of endocardium, and estimated pdf (solid red). Right - Corresponding shape priors.

and prior pdf’s, which has also been proposed in Besson et al. [108]. The symmetric Kullback-Leibler distance may also be used. The Hellinger distance between the prior inside/outside distribution \( \hat{p}_{in}/\hat{p}_{out} \) and the corresponding Parzen estimated [159] ones \( p_{in}/p_{out} \) is given in Eq. 3.96.

We formulate a two-step algorithm in which we first solve (3.149) keeping the prior \( \phi_0 \) fixed. Next \( \phi_0 \) is updated, and (3.149) is solved again. Thus, finding the final segmentation \( \phi \) consists of solving sequential minimization problems, with the prior \( \phi_0 \) being updated in between each of them. Since step 3 in (1) is convex any initial segmentation \( \phi_{init} \) can be chosen. The shape prior \( \phi_0 \) is initialized with the mean of binary training shapes \( \bar{x} \) (see below).

The update procedure projEig of the shape prior is defined as follows. We have chosen to use principal eigenvectors of binary training shapes to build our prior \( \phi_0 \), as proposed in Dambreville et al. [51]. It is argued in this paper that the construction of a space of shapes built from applying PCA on binary training shapes makes it possible to restrict contour evolution in a way that remains closer to the shape space, when compared to representation of shapes by signed distance functions. Let \( X = \{x_1, ..., x_N\} \subset S \) be a set of transitionally registered binary training shapes, where \( S \) is the space of shapes constructed from binary maps and \( m \) is the length of training shape vectors. The shapes are transitionally registered to a user-selected point. Furthermore,
3.11. Convex Endocardial Segmentation Model

Figure 3.13: Sequence B. Calculated contour (solid white) and ground truth (dashed red). End-Systole/Diastole - left/right
CHAPTER 3. Segmentation of the Endocardium - Variational approach

Figure 3.14: Sequence B. Histogram of ground truth distributions inside (top)/outside (bottom) of endocardium, and estimated pdf (solid red). Right - Corresponding shape priors.

Algorithm 1 Convex Segmentation

1: Input: $\phi(0) = \phi_{\text{init}}$, $\phi_0^{(0)} = \bar{x}$, $k = 0$
2: while not converged do
3: $\phi^{(k+1)} \leftarrow$ Solve (3.149) with $VL_{\text{in}} = VL_{\text{in}}(\phi^{(k)})$, $VL_{\text{out}} = VL_{\text{out}}(\phi^{(k)})$.
4: $\phi_0^{(k+1)} \leftarrow \text{projEig}(\phi^{(k+1)})$
5: $k \leftarrow k + 1$
6: end while

let $P = [P_1, ..., P_n]$ be the $n$ principal eigenvectors of the covariance matrix of $X$. A segmentation vector $u$ is projected onto the eigenspace $F$ by $\phi(u) = P^T(u - \bar{x}) = y$, where $\bar{x}$ is the mean training shape. The shape prior $\phi_0$ is then obtained by projecting $y$ onto $S$, i.e. $\phi_0 = Py + \bar{x}$, and is the shape in the training set which is closest, under our assumptions, to our current segmentation $\phi$. This process is summarized in Algorithm 2. See right column in Fig. 3.12 and 3.14 for examples of $\phi_0$ corresponding to $\phi$.  

92
3.11. Convex Endocardial Segmentation Model

Algorithm 2 Update Shape prior \( \phi_0 = \text{proj}Eig(\phi) \)

1. Project onto \( \mathcal{F} \): \( y = P^T(\phi - \bar{x}) \)
2. Project \( y \) onto \( \mathcal{S} \): \( \phi_0 = Py + \bar{x} \)
3. Return \( \phi_0 \)

3.11.3 Fast Minimization Using Split-Bregman

In this section we shall give a brief description of the Split-Bregman method, and refer the reader to Goldstein et al. [75] for complete details. The Split-Bregman method is a technique for minimization of \( L^1 \) regularized functionals and has been previously applied for segmentation and surface reconstruction and shown to out-perform well-known methods, such as graph cuts and methods based on duality.

The problem we wish to solve, previously stated in Eq. 3.149, is

\[
\min_{0 \leq \phi \leq 1} \int_{\Omega} -\lambda R(\lambda, \gamma)\phi + |\nabla \phi| \tag{3.186}
\]

In the following, let \(| \cdot |_1 \) and \(| \cdot |_2 \) denote the \( L^1 \) and \( L^2 \) norm, respectively. First an auxiliary variable \( d = \nabla \phi \) is introduced. This is done to split the problem in a convex and \( L^1 \)-part, hence the name Split-Bregman. The auxiliary variable gives an equality constraint and thus (3.186) now becomes

\[
(\phi^*, d^*) = \arg \min_{0 \leq \phi \leq 1, d} |d|_1 + \lambda R(\lambda, \gamma)\phi + \beta \frac{1}{2} |d - \nabla \phi|^2_2 \tag{3.187}
\]

The second part of the name of Split-Bregman is due to the Bregman iteration technique [76], which is applied to the unconstrained problem (3.187) by adding the vector \( b \) inside the quadratic penalty, giving an alternating minimization scheme:

\[
(\phi^{k+1}, d^{k+1}) = \arg \min_{0 \leq \phi \leq 1, d} |d|_1 + \lambda R(\lambda, \gamma)\phi + \frac{\beta}{2} |d - \nabla \phi - b^k|^2_2 \tag{3.188}
\]

\[
b^{k+1} = b^k + \nabla \phi^{k+1} - d^{k+1}, \tag{3.189}
\]

starting from \( b^0 = d^0 = \phi^0 = 0 \).

The Euler-Lagrange equation of (3.188) is

\[
\Delta \phi = \frac{\lambda}{\beta} R(\lambda, \gamma) + \text{div}(d^k - b^k). \tag{3.190}
\]

Disregarding the constraints on \( \phi \), Eq. 3.188 is quadratic in \( \phi_{i,j} \) (all other elements \( \phi_{k,l} \) s.t. \((k, l) \neq (i, j) \) are kept fixed). This is the convex part of the problem referred to above. Thus we can obtain a solution to Eq. 3.188 by solving the Euler-Lagrange equations.
equation (3.190) for $\phi_{i,j}$. If the obtain solution is within the constraints, $0 \leq \phi_{i,j} \leq 1$, then it is accepted. Otherwise the true solution lies at the closest endpoint of the interval $[0, 1]$ to the obtained solution, since the equation is quadratic and hence monotonic.

The solution to the Euler-Lagrange equation (3.190) is computed by using a Gauss-Seidel iterative method:

$$
\alpha_{i,j} = d^x_{i-1,j} - d^x_{i,j} + d^y_{i,j-1} - d^y_{i,j} - (b^x_{i-1,j} - b^x_{i,j} + b^x_{i-1,j} - b^x_{i,j})$, \\
\beta_{i,j} = \frac{1}{4} \left( \phi_{i-1,j} + \phi_{i,j+1} + \phi_{i,j-1} + \phi_{i,j+1} - \frac{\lambda}{\beta} R(\gamma_1, \lambda) + \alpha_{i,j} \right), \\
\phi_{i,j} = \max \left\{ \min \{ \beta_{i,j}, 1 \}, 0 \right\}. 
$$

The $L^1$ part of the problem, that is the solution of (3.188) w.r.t. $d$, can be obtained in various ways, in Houhou et al. [94] it is given by soft-wavelet thresholding

$$
d^{k+1} = \text{sign}(\nabla \phi^{k+1} + b^k) \max(|\nabla \phi^{k+1} + b^k| - \lambda^{-1}, 0). 
$$

and in Goldstein et al. [75] by

$$
d^{k+1} = \max \{ \nabla \phi^{k+1} + b^k, 0 \} \frac{\nabla \phi^{k+1} + b^k}{|\nabla \phi^{k+1} + b^k|^2}.
$$

The equations in (3.192) and (3.193) are known as shrink operators and we use the term $\text{shrink}(\cdot, \lambda)$ to denote them.

This estimate $d^{k+1}$ is then plugged

$$
b^{k+1} = b^k + \nabla \phi^{k+1} - a^{k+1}
$$

Thus the convex minimization problem (3.149) is solved by the Split-Bregman optimization scheme described in Algorithm 3.

where $\text{projEig}(\phi_0^k)$ refers to the projection of $\phi_0^k$ onto an eigenspace of binary training shapes, as described in Sec. 3.11, and $\text{GS}(R^k, d^k, b^k)$ denotes updating all elements according to the Gauss-Seidel formula (3.191).

### 3.11.4 Experiments

Our algorithm for segmenting the endocardium in ultrasound sequences can be divided into three main parts. Firstly, mixture distribution parameters are estimated by the EM algorithm from the ultrasound sequence. Secondly, the estimated parameters are used to construct prior distributions for the inside and outside of the endocardium. In the final step, segmentation is achieved by minimizing the distance between the prior and online estimated pdf’s by a Continuous cut integrated with a statistical shape prior.
Algorithm 3 Split-Bregman Optimization Scheme

1: while $|\phi^{k+1} - \phi^k|_2 > \epsilon$ do
2: Define $R^k = (V_{\text{in},k}^{\text{HL}} + V_{\text{out},k}^{\text{HL}}) + \frac{2}{\lambda}(1 - 2\phi^k_0)$
3: $\phi^{k+1} = \text{GS}(R^k, d^k, b^k)$
4: $d^{k+1} = \text{shrink}(\nabla \phi^{k+1} + b^k, \lambda)$
5: $b^{k+1} = b^k + \nabla \phi^{k+1} - d^{k+1}$
6: Find $\Omega^{k+1} = \{x : \phi^{k+1}(x) > \mu\}$
7: Update $V_{\text{in},k+1}^{\text{HL}} = V_{\text{in}}^{\text{HL}}(x, \Omega^{k+1})$
8: $V_{\text{out},k+1}^{\text{HL}} = V_{\text{out}}^{\text{HL}}(x, \Omega^{k+1})$
9: $\phi^0_{k+1} = \text{projEig}(\phi^0_k)$
10: end while

<table>
<thead>
<tr>
<th>Author</th>
<th>Modality</th>
<th>Average Distance (AD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen et al. [42]</td>
<td>LAX 2C</td>
<td>1.97 pxl</td>
</tr>
<tr>
<td>Milic et al. [140]</td>
<td>SAX</td>
<td>1.15 ± 0.2 mm</td>
</tr>
<tr>
<td>Boukerroui et al. [28]</td>
<td>LAX 4C</td>
<td>5.96 ± 2.20 pxl</td>
</tr>
<tr>
<td>Lin et al. [127]</td>
<td>SAX</td>
<td>1.643 ± 0.503 mm</td>
</tr>
<tr>
<td>Chalana et al. [38]</td>
<td>SAX</td>
<td>3.61 ± 1.61 mm</td>
</tr>
</tbody>
</table>

The data was obtained from the database described in Sec. 1.6.3. We divided our data, cardiac cycles from 25 distinct patients, into two sets: training set and validation set. The training set consisted of 20 cardiac cycles. The training set was further divided into subsets, corresponding to parts of the cardiac cycle. The validation set consisted of 5 cardiac cycles.

Training data from annotated ultrasound sequences was used both in constructing the prior distributions, and the statistical shape prior. The shape prior is transitionally registered to a user-selected point from an end-diastolic frame, which approximates the center of the endocardium, due to low chamber deformation at end-diastole.

The mixture parameters $\theta = \{\alpha_i, \sigma_i; i = 2, 3, 4\}$, corresponding to the categories (2) echocardiographic artifacts, (3) blood, and (4) tissue, was estimated by Expectation Maximization (EM) [53]. The mixture weight for signal dropout $\alpha_1$ is readily computed without EM, this is simply the proportion of zero intensity values in the sequence. In practice, we substituted the signal dropout intensities with random samples from the blood category, since regions of blood intensity return the least echo and are most subject to dropout.

We applied the fast Split-Bregman method to our convex minimization problem (3.149), for a description of the algorithm see Sec. 3.11.3.
Given estimated mixture parameters, we compared our proposed convex segmentation method to an equivalent non-convex formulation implemented using Fast Marching Level Sets \cite{187}. Average running time per frame for Convex/Non-Convex formulation was 10/120 sec on a Intel 2.4 GHz Duo processor. Both were implemented using C++. One of the main advantages of using the convex formulation is its robustness with respect to initialization. This is not the case for the non-convex formulation which depends highly on the initialization, in this case proximity to the endocardial wall.

Quantitative results for end-systole/diastole are shown in Table 3.2, and by images for 2 sequences in Fig. 3.11 and 3.13. We use Average Distance (AD), employed by comparable works \cite{151}, to measure the distance between points on the endocardial wall in the expert contour and the one produced by our method. Points on the valve are excluded, since these are not clinically relevant. Chen et al. \cite{42} and Boukerroui et al. \cite{28} work with Long Axis sequences, four (4C) and two chamber (2C) respectively, while Mikic et al. \cite{140}, Chalana et al. \cite{38} and Lin et al. \cite{127} work on Short Axis (SAX) sequences and thus these results are somewhat less useful for comparison. See Noble et al. \cite{151} for a summary of these works.

Average distance is defined as follows, \( AD = \frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^{n} d(a_i, B) + \frac{1}{m} \sum_{i=1}^{m} d(b_i, A) \right) \), where \( A = \{a_1, a_2, ..., a_n\} \) and \( B = \{b_1, b_2, ..., b_m\} \) are sets of contour points, and \( d(a_i, B) = \min_j ||b_j - a_i|| \).

Our convex minimization scheme produced results (Table 3.2) comparable to those in Table 3.1.

### 3.12 Conclusion

In our active contour formulations we have focused on shape prior and distribution formulations.

In our local formulation a prior template is affinely deformed, whilst interacting with image data as well as the current state of the contour. The local prior model is shown to be more robust to image artifacts than a prior model employing the area set symmetric difference measure for the interaction between prior and active contour. Furthermore we have observed the proposed model is more tolerant to poor initialization and exhibits higher convergence rate. The local model is demonstrated by experiments where affine
3.13 Future work

The shape prior formulations used in the locally optimal and convex active contour models are extendible to more complex shape priors, e.g., using manifold learning of shapes [66], Multilinear Principal Component Analysis (MPCA) [209] and Multilinear Independent Component Analysis (MICA) [210], which may improve the efficacy of the algorithms. The censored Gamma mixture model used in our work on MCMC segmentation could be extended to our work on active contours. The initial positioning of the shape prior in the convex model should be expanded from manual to semi-automatic or fully-automatic. An option would be to sample the translation distribution of the mean shape, using the method developed for the MCMC Segmentation method.
Chapter 4

Segmentation of the Endocardium in B-mode Ultrasound - Bayesian MCMC approach

4.1 Markov Chain Monte Carlo (MCMC)

A Markov Chain Monte Carlo method is a stochastic process with Markov property. This means that the current state of the process $x^{(t)}$ depends only on the previous state $x^{(t-1)}$.

Formally the Markov property is stated as

$$P(X^{(m+1)}|X^{(1)}, \ldots, X^{(m)}) = P(X^{(m+1)}|X^{(m)})$$

for random variables $X^{(m+1)}, \ldots, X^{(1)}$.

MCMC methods are typically used to sample the density $P(x) = P^*(x)/Z$, where $P^*(x)$ is known. The obvious question is then: Why not simple determine the normalizing constant $Z = \int P^*(x) \, dx$ directly? The problem is that the sample space may be very high-dimensional, so that this is not easily achieved. Even if it is possible to determine $Z$, the high-dimensionality can make sampling $P(x)$ very difficult. All or most states have to be enumerated and in a very high-dimensional sample space the cost of performing this is prohibitive [129]. MCMC methods provide a way to sample these hard-to-sample densities.

There are several well-known methods which fall in the MCMC category, e.g. importance sampling, rejection sampling, slice sampling, Gibbs sampling, and the Metropolis-Hastings algorithm. We will here detail the last two, as these are employed in this thesis.
Markov Chain Basics

A Markov Chain (MC) is a stochastic process in which future states are independent of other states than the present. Formally, MC is a discrete time stochastic process with an arbitrary state space, which has the Markov property, and called time-homogeneous if the chain has stationary transition probabilities.

If the conditional distribution of $X_n$ given $X_{n-1}$ is independent on the index $n$, then the MC has stationary transition probabilities, thus the conditional distribution of $X_n$ given $X_{n-1}$ together with the initial distribution completely defines the chain.

Before we proceed, we review some basic facts of measure theory. A general measurable space is a pair $(\mathcal{X}, B(\mathcal{X}))$, where $\mathcal{X}$ is an abstract set of points and $B(\mathcal{X})$ is a $\sigma$-field of subsets of $\mathcal{X}$, that is

(i) $\mathcal{X} \in B(\mathcal{X})$
(ii) if $A \in B(\mathcal{X})$ then $A^c \in B(\mathcal{X})$
(iii) If $A_k \in B(\mathcal{X})$, $k = 1, 2, 3, ...$ then $\bigcup_{k=1}^{\infty} A_k \in B(\mathcal{X})$.

In the discrete case the transitions are defined using a transition matrix $K$ with entries $P_{xy} = P(X_n = y \mid X_{n-1} = x)$, $x, y \in \chi$, (4.2)

where $\chi$ is the state space of the chain.

The $n$-step transition kernel is in the discrete case defined by the matrix

$$ K^{(n)} = (k^{(n)}_{ij})_{ij} $$

(4.3)

with $k^{(n)}_{ij} = P(X_{t+n} = j \mid X_t = i)$.

For general state spaces the conditional probability is defined in terms of a transition kernel $K : \chi \times \chi \to \mathbb{R}$ for all measurable sets $A \subset \chi$ and all points $x \in \chi$, defined by

$$ P(X_t \in A \mid X_{t-1} = x) = \int_A K(x, y) dy . $$

(4.4)

For a measurable $A \subset \chi$ the $n$-step probability of the Markov chain is provided by

$$ P(X_{t+n} \in A \mid X_t = x_t) = \int_{\chi^{n-1} \times A} \prod_{k=t+1}^{t+n} K(x_{k-1}, x_k) dx_{t+1} \ldots dx_{t+n} . $$

(4.5)

To simplify this expression one can use the $n$-step transition kernel. This is defined as

$$ K^{(n)}(x_t, x_{t+n}) = \int_{\chi^{n-1}} \prod_{k=t+1}^{t+n} K(x_{k-1}, x_k) dx_{t+1} \ldots dx_{t+n-1} , $$

(4.6)

and (4.5) then may be written in the simpler form

$$ P(X_{t+n} \in A \mid X_t = x_t) = \int_A K^{(n)}(x_t, x_{t+n}) dx_{t+n} . $$

(4.7)
As the chain progresses it may reach an equilibrium state, a so called invariant distribution. This is defined in the discrete and general case as follows

**Definition 4.1.1.** [175] (Invariance Discrete case) A probability distribution \( \pi \) is invariant for a transition probability matrix \( K \) i.f.f. \( \pi = \pi K \).

**Definition 4.1.2.** [175] (Invariance General case) A \( \sigma \)-finite measure \( \pi \) is invariant on \( B(\chi) \) for the transition kernel \( K(.,.) \) (and for associated chain) if

\[
\pi(B) = \int_X K(x, B)d\pi(x), \forall B \in B(\chi)
\]  

(4.8)

If \( \pi \) is a probability measure then the chain is stationary in distribution, since \( X_0 \sim \pi \) implies that \( X_n \sim \pi \) for every \( n \).

Assume that the Markov chain has an associated invariant distribution \( \pi \), s.t. when the chain remains unchanged after having achieved this distribution. In order for the chain to converge to the invariant distribution, it needs to be irreducible and aperiodic. Irreducibility means that it must be possible to reach any state of the chain, given any initial state. The chain is aperiodic if there is no regularity at which a state is visited. Together then irreducibility and aperiodicity mean that there exists a positive probability to reach any state in a finite number of steps.

Now let \( \nu_A \) be the number of times the set \( A \) is visited, i.e. \( \nu_A = \sum_{l=1}^{\infty} 1_A(X_k) \), and \( \chi_x \) and \( P_x \) is the expectation and distribution of the Markov Chain, given the initial distribution provided by the point mass \( \delta_x \) at the point \( x \).

Irreducibility and aperiodicity is defined as

**Definition 4.1.3.** (Irreducibility - Discrete case [175, 137]) The chain is irreducible if all states communicate,

\[
P_x(\tau_y < \infty) > 0 \quad \forall x, y \in \chi,
\]  

(4.9)

where \( \tau_y \) is the first time \( y \) is visited.

**Definition 4.1.4.** [175, 137] (Irreducibility - General case) Given a distribution \( \phi \), a MC is said to be \( \phi \)-irreducible if for all points \( x \in \chi \) and all measurable sets \( A \) s.t. \( \phi(A) > 0 \) there exists some \( t \) s.t.

\[
\int_A K^{(t)}(x, y)dy > 0.
\]  

(4.10)

Here we will let \( \phi = \pi \), the stationary distribution of the chain.

**Example 4.1.1.** (Irreducibility) If \( \{X_n\}_{n>0} \) are iid with distribution \( \nu \), then it is \( \nu \)-irreducible.

Next we define the concept of periodicity

**Definition 4.1.5.** [175, 137] (Periodicity Discrete case) The period \( d(\omega) \) of state \( \omega \in \chi \) is defined as

\[
d(\omega) = \text{g.c.d.}\{m \geq 1; K^{(m)}(\omega, \omega) > 0\}.
\]  

(4.11)
A chain is periodic if all states have common period $d$ greater than 1, if the common period is 1 then the chain is aperiodic.

In the general case we talk of cycles. To be able to define cycles we need to define atoms of a kernel and small sets of a chain.

**Definition 4.1.6.** [175, 137] (Atoms) A Markov chain with transition kernel is said to have an atom, $\alpha \subset \chi$, if there is some probability measure $\nu$, s.t.

$$\forall x \in \chi : \int_A K(x, y)dy = \int_A \nu(y)dy .$$

**Definition 4.1.7.** [175, 137] (Small sets) A set $C$ is termed small for a given Markov chain if there exists some positive integer $s$, and a nonzero measure $\nu$, s.t. for all $x \in C$,

$$\int_A K(x, y)dy \geq \epsilon \int_A \nu(y)dy \quad (4.12)$$

**Definition 4.1.8.** [175, 137] (Cycles) A $\phi$-irreducible Markov chain has a cycle of length $d$ if there exists a small set $C$, an associated integer $M$ and some probability distribution $\nu_s$, which has positive mass on $C$, i.e., $\int_C \nu_s(x)dx > 0$, s.t. d = g.c.d. $\{s \geq 1 : C$ is small for some $\nu_s \geq d_s \nu_M$ with $d_s > 0\}$.

**Definition 4.1.9.** ([175, 137]) (Recurrence) Set $A \subset \chi$, for $\phi$-irreducible general space Markov chains. A set is recurrent if $\forall x \in A$ : $\chi_x[\nu_A] = \infty$. A general state space MC is recurrent if

(1) The chain is $\phi$-irreducible for some distribution $\phi$.

(2) For every measurable set $A \subset \chi$ s.t. $\int_A \nu(y)dy > 0$ for a measure $\nu$, and $\chi_x[\nu_A] = \infty$ for every $x \in A$.

**Definition 4.1.10.** [175, 137] (Harris Recurrence) A set is Harris recurrent if $P_x(\nu_A = \infty) = 1$ for every $x \in A$. A MC is Harris recurrent if there exists some distribution $\phi$ w.r.t. which it is irreducible and every set $A$ s.t. $\int_A \mu(x)dx > 0$ is Harris recurrent.

**Definition 4.1.11.** [175, 137] (Positive chain) If the chain $\phi$ is $\pi$-irreducible and has an invariant probability measure $\pi$, then it is called a positive chain.

**A Convergence Result for Markov Chains**

By examining the limiting behavior of the Markov chain, we may under certain conditions extract useful statistics such as the expected value of the chain. In other words there exists a Law of Large Numbers for the case of Markov Chains. That such a result holds is however non-trivial. First, we need to make sure that the chain is Harris recurrent and positive (sometimes referred to as positive Harris). Second, there needs to exist a unique invariant distribution for the chain (otherwise there would be nothing to converge to). The above can be stated in the following theorem,
Theorem 4.1.1. ([137]) (Law of Large Numbers for Harris Positive Markov Chains) Assume that the chain \( \Phi = \{X_0, X_1, \ldots\} \) is Harris recurrent and positive, with invariant distribution \( \pi \). Then for each integrable function \( f \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \pi(f) = E_\pi[f(X_0)], \text{ a.s}
\]

(4.13)

Metropolis-Hastings Algorithm

In the Metropolis-Hastings (MH) algorithm [136], a proposal density \( q \) is sampled from in lieu of a more complex target density \( f \). Thus the proposal density is typically much simpler than the target density, and chosen for feasibility of sampling. The great benefit of Metropolis Hastings is that the proposal density need not be remotely similar to the target density. The sampling scheme for MH is the following: i) Generate a new proposal state \( x' \) by sampling the proposal density \( q(x' | x^{(t)}) \). Accept state \( x' \) if

\[
\alpha(x', x^{(t)}) = \frac{f(x')}{f(x^{(t)})} \cdot \frac{q(x^{(t)} | x')}{q(x' | x^{(t)})} \geq 1
\]

(4.14)

ii) Accept state \( x' \) with probability \( \alpha \) if \( \alpha < 1 \), otherwise take \( x^{(t)} \) with probability \( 1 - \alpha \).

Gibbs Sampling

In its basic version, as described here, Gibbs sampling is a special case of Metropolis-Hastings with acceptance probability set to 1. Suppose we want to sample from the distribution \( p(x_1, \ldots, x_N) \). In each step, a given variable \( x_i \) is updated by drawing from the distribution of \( x_i \) conditioned on the variables \( x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N \), that is from \( p(x_i | x_{\backslash i}) \), as seen in Algorithm 4.

Algorithm 4 Gibbs sampling

1: Initial state \( \{x_i^{(1)} : i = 1, \ldots, N\} \).
2: for \( m = 1 \) to \( M \) do
3: Sample \( x_1^{(m+1)} \sim p(x_1 | x_2^{(m)}, \ldots, x_N^{(m)}) \)
4: Sample \( x_2^{(m+1)} \sim p(x_2 | x_1^{(m+1)}, x_3^{(m)}, \ldots, x_N^{(m)}) \)
5: \ldots
6: Sample \( x_j^{(m+1)} \sim p(x_j | x_1^{(m+1)}, \ldots, x_{j-1}^{(m+1)}, x_{j+1}^{(m)}, \ldots, x_N^{(m)}) \)
7: \ldots
8: Sample \( x_N^{(m+1)} \sim p(x_N | x_1^{(m+1)}, \ldots, x_{N-1}^{(m+1)}) \)
9: end for
Convergence Metropolis-Hastings

By first examining the transition kernel of the Metropolis-Hastings algorithm, we establish that it satisfies the detailed balance condition, since this guarantees the existence of an invariant distribution for the chain.

**Lemma 4.1.1.** Transition kernel of Metropolis Hastings algorithm is

$$K(x, y) = \alpha(x, y)q(y | x) + (1 - \alpha(x))\delta_x(y),$$  \hspace{1cm} (4.15)

where

$$\alpha(x) = \int \alpha(x, y)q(y | x)dy \quad \text{and} \quad \alpha(x, y) = \min \left\{ \frac{f(y)}{f(x)} \frac{q(x | y)}{q(y | x)}, 1 \right\}.$$  \hspace{1cm} (4.16)

with proposal density $q$ and function $\alpha(x, y)$ as defined in (4.14).

**Definition 4.1.12.** A Markov chain with a transition kernel $K$ satisfies the detailed balanced condition if there exists a function $f$ satisfying

$$K(y, x)f(y) = K(x, y)f(x)$$  \hspace{1cm} (4.17)

for every $(x, y)$.

**Theorem 4.1.2.** ([175]) Let $(X(t))$ be the Markov chain generated by the Metropolis Hastings algorithm sampling from the target density $f$. For every proposal distribution $q$ s.t. $\text{supp}(f) \subset \text{supp}(q)$,

(i) the kernel of the chain satisfies detailed balance condition with $f$ ;

(ii) $f$ is a stationary distribution of the chain.

**Theorem 4.1.3.** ([175]) Suppose that Markov chain with transition kernel $K$ satisfies the detailed balanced condition with a probability density function $\pi$. Then the density $\pi$ is the invariant density of the chain.

After this we establish Harris recurrence, through irreducibility. We use that a sufficient condition for $\pi$-irreducibility of the Metropolis Hastings chain, is that the proposal density is strictly positive,

$$q(y|x) > 0 \quad \text{for every } (x, y) \in \eta \times \eta,$$  \hspace{1cm} (4.18)

where $\eta$ is the support of $\pi$ (see [175]).

**Theorem 4.1.4.** ([206]) If the Metropolis-Hastings chain $(X(t))$ is $\pi$-irreducible, then it is Harris recurrent.

Since the Metropolis-Hastings algorithm is $\pi$-irreducible with invariant distribution $\pi$, the chain is per definition positive. Thus it is possible to apply Theorem 4.1.1, to the Markov Chain $(X_0, X_1, ...)$ generated by Metropolis-Hastings algorithm.
4.1. Markov Chain Monte Carlo (MCMC)

Convergence of Gibbs Sampling Algorithm

Assume we wish to sample the joint distribution \( f(x_1, \ldots, x_p) \). Then the following propositions and theorems guarantee that by using Gibbs sampling we are indeed sampling from the desired distribution.

As for the case of the Metropolis-Hastings algorithm, we first establish that an invariant distribution exists for the chain.

**Corollary 4.1.1.** \((175)\) The joint distribution \( f(x_1, \ldots, x_p) \) is the invariant distribution of the Markov Chain \((X^{(0)}, X^{(1)}, \ldots)\) generated by the Gibbs sampler.

The irreducibility of the chain is established by using the positivity condition.

**Definition 4.1.13.** (Positivity condition) A distribution with density \( f(x_1, \ldots, x_p) \) and marginal densities \( f_{X_i}(x_i) \) is said to satisfy the positivity condition if \( f_{X_i}(x_i) > 0 \) for all \( x_1, \ldots, x_p \) implies that \( f(x_1, \ldots, x_p) > 0 \).

**Theorem 4.1.5.** \((206)\) If the joint distribution \( f(x_1, \ldots, x_p) \) satisfies the positivity condition, the Gibbs sampler yields an irreducible Markov chain.

Harris recurrence is now established through showing irreducibility of the chain.

**Theorem 4.1.6.** \((176)\) Every \( \phi \)-irreducible deterministic or random-scan Gibbs sampler is Harris recurrent.

Using Corollary 4.1.1, Theorem 4.1.6 and finally Theorem 4.1.1 the convergence of the Markov Chain \((X_0, X_1, \ldots)\), which is generated by Gibbs Sampling, follows.

Convergence Alternating Gibbs Metropolis-Hastings Sampler

We have seen that under certain conditions both Gibbs sampling and Metropolis-Hastings algorithm, limit theorems may be applied. Is it then the case that we have the same convergence when these are combined? Here we consider the case of alternating between Gibbs sampling and Metropolis-Hastings, which we have employed in Hansson et al. [83], also described in Sec. 4.2. In Nobile et al. [150] this strategy is described for a Bayesian multinomial probit model, using the terminology hybrid Markov chain. We talk about alternating Gibbs Metropolis-Hastings, to avoid confusion with the Hamiltonian Monte Carlo method, which is also known as hybrid Monte Carlo. With minor modifications to the proof presented in Nobile et al. [150], Appendix p. 241-242, for the convergence of the Markov chain associated with the alternating Gibbs-Metropolis algorithm, we are able to state

**Theorem 4.1.7.** The Law of Large Numbers is valid for the Markov Chains resulting from the algorithm for sampling the posterior distribution in Hansson et al. [83].
CHAPTER 4. Segmentation of the Endocardium - Bayesian MCMC

Proof: \( \chi \) is the state space of the sampling chain, and \( \chi \) is a Borel \( \sigma \)-field on \( \chi \). \( P_{\chi} \) is the transition probability of moving from \( x \in \chi \) to \( A \in \chi \). Assume that \( P_G \) and \( P_{MH} \) are the transition kernels for the Gibbs sampler and Metropolis-Hastings algorithm, respectively. Furthermore let \( \phi \) be the posterior that is to be sampled from. The idea of the proof in Nobile et al. [150] consists of proving that (i) \( \phi \) is invariant to \( (P_G P_{MH}) \), (ii) For every \( x \in \chi \), for every \( A \in \chi \) with \( \phi(A) > 0 \), \( (P_G P_{MH}(x, A) > 0 \), (iii) \( P_G P_{MH} \) is \( \phi \)-irreducible, (iv) \( P_G P_{MH} \) is aperiodic and (v) \( P_G P_{MH} \) is absolutely continuous w.r.t. to \( \phi \). Using (i) - (v) and Corollary 1 and Theorem 1 in Tierney [206], it follows that the hybrid sampler generates a positive and Harris recurrent Markov chain that converges to a unique stationary distribution.

The argument is identical for the proposed sampler in Hansson et al. [83], excepting (ii). Recall that \( T^t \) denotes the translation operator. Then \( B = \{ y : P_{MH}(y, A) > 0 \} = \{ v : v = T^t(u); u \in A, t \in C \} \), where \( C \) is the group of all spatial translations, and \( A \) the set of all admissible latent variable images. By admissible latent variable images we refer to all images which contain all three categories of latent variables: 1, 2 and 3. It is clear that \( A \subset B \), and \( \phi(A) > 0 \) implies \( \phi(B) > 0 \). Since \( \phi \) consists of strictly positive densities, \( P_G(x, A) > 0 \). The transition kernel of the hybrid chain is

\[
(P_G P_{MH})(x, A) = \int P_G(x, dy) P_{MH}(y, A) > 0 .
\]

\[
(4.19)
\]

4.2 Evaluation of Ultrasound data by Bayesian Probability Maps

This section is based on the paper Evaluation of Cardiac Ultrasound Data by Bayesian Probability Maps [83]. Previous version of the model was presented in: “Evaluation of Cardiac Ultrasound Data by Bayesian Probability Maps” [82] and “Bayesian Probability Maps for Evaluation of Ultrasound Data” [81].

The theoretical prerequisites are provided in:

– Sec. 1.2 (Physical Background)
– Sec. 1.3 (Transmission to Display of Ultrasound Echo)
– Sec. 1.6 (Application in Cardiology)
– Sec. 2.1 (Statistical Modeling of Ultrasound Data)
– Sec. 2.2 (Mixture Models)
– Sec. 2.4 (Maximum Likelihood Estimators)
4.2. Bayesian Probability Maps

Figure 4.1: The symbol $u$ represents the $m$ latent variable images $I_u$ stacked into a single vector. Each $I_u$ (top row) corresponds to an image in the US sequence (bottom row) of length $m$, where $z$ denotes the spatial coordinates. Three regions are defined: one signifies the endocardium, two atrial region and three background. The notation $I_u$ is used in the latent variable prior, defined in Sec. 4.2.

- Sec. 2.3 (Bayesian Statistical Formulations)
- Sec. 2.5 (EM Algorithm)
- Sec. 2.7 (EM-MAP)
- Sec. 4.1 (Markov Chain Monte Carlo).

Model

Our goal here is to determine the position of the endocardium in a clinical B-mode US sequence. In order to make this process more robust we incorporate prior knowledge of the position of the atrium into our model. Due to signal dropout the true anatomical form of the atrium is rarely visible and instead we model the low-intensity region connected to atrium, in the sequel referred to as the atrial region.

Posterior

Assume that there are $N$ pixels in a sequence of US images. For each pixel $i = 1, \ldots, N$, the latent variables $u_i$ and $w_i$ are defined. The latent variables $u_i$ indicates to which region it belongs, $u_i = 1$ indicates endocardium, $u_i = 2$ atrial region and $u_i = 3$ the background, see Fig. 4.1. The latent variables are stacked into the vector $u$. The vector

107
\( \mathbf{u}_{\text{endo}} \) of equal size to \( \mathbf{u} \) is defined by \( \mathbf{u}_{\text{endo}} = (\delta(u_1 - 1), \delta(u_2 - 1), ..., \delta(u_N - 1)) \), where \( \delta \) is the Kronecker delta. The elements of \( \mathbf{u}_{\text{endo}} \) are set to zero, if the associated latent variables do not belong to the endocardium. In the same manner, \( \mathbf{u}_{\text{attr}} \) and \( \mathbf{u}_{\text{bkgr}} \), are defined by \( \mathbf{u}_{\text{attr}} = (\delta(u_1 - 2), \delta(u_2 - 2), ..., \delta(u_N - 2)) \) and \( \mathbf{u}_{\text{bkgr}} = (\delta(u_1 - 3), \delta(u_2 - 3), ..., \delta(u_N - 3)) \), respectively. Let \( \mathbf{U} \) and \( \mathbf{U}_{\text{endo}} \) be random vectors, corresponding to \( \mathbf{u} \) and \( \mathbf{u}_{\text{endo}} \), respectively. The latent variables \( w_i \) are defined in the same way as \( u_i \).

Our approach is two step. First the class probabilities \( \gamma_i \) of \( w_i \) are computed, assuming data independence. In the second step the posterior distribution of \( u_i \) is estimated, given priors on \( u_i \) and where the class probabilities \( \gamma_i \) of \( w_i \) are taken as fixed parameters.

The problem of determining the position of the endocardium can thus be formulated as the determination of the marginal posterior distribution of the latent variables, where the nuisance parameters \( \mathbf{u}_{\text{attr}} \) and \( \mathbf{u}_{\text{bkgr}} \) have been integrated out, or

\[
P(\mathbf{U}_{\text{endo}} = \mathbf{u}_{\text{endo}} \mid \mathbf{x}, \mathbf{w}, \Delta) \propto \sum_{\mathbf{u}_{\text{attr}} \cdot \mathbf{u}_{\text{bkgr}}} L(\mathbf{x} \mid \mathbf{U} = \mathbf{u}, \mathbf{W} = \mathbf{w}, \Theta, \pi, \sigma_{\text{apex}}) P(\mathbf{U} = \mathbf{u} \mid \Lambda),
\]

(4.20)

where \( \mathbf{x} = \{x_1, ..., x_N\} \) represent gray level intensities, corresponding to the pixels \( i = 1, ..., N \), stacked into a single vector, and \( \Delta = \{\Theta, \Lambda, \pi, \sigma_{\text{apex}}\} \) are parameters.

**Likelihood**

The likelihood of the posterior (4.20) is defined as

\[
L(\mathbf{x} \mid \mathbf{u}, \mathbf{w}, \Theta, \pi, \sigma_{\text{apex}}) = \prod_i [\gamma_i^{1(u_i \in \mathbb{R}_{12})} \times (1 - \gamma_i)^{1(u_i \in \mathbb{R}_3)}]^A(x_i \mid \sigma_{\text{apex}}),
\]

(4.21)

where \( u_i \) are the realizations of latent variables \( U_i \) corresponding to \( x_i \), \( \gamma_i \) are the posterior class probabilities of the latent variables \( w_i \), and \( 1 \) is an indicator variable. The endocardium and atrial region, and background is denoted by \( \mathbb{R}_{12} \) and \( \mathbb{R}_3 \), respectively (see Fig. 4.2). The posterior class probabilities \( \gamma_i \) are defined as

\[
\gamma_i = P(W_i \in \mathbb{R}_{12} \mid x_i, \Theta, \pi) = \frac{\pi_{12} p(x_i \mid \Theta_{12})}{\pi_{12} p(x_i \mid \Theta_{12}) + \pi_3 p(x_i \mid \Theta_3)},
\]

(4.22)

where the parameters \( \Theta = \{\Theta_{12}, \Theta_3\} \) of the likelihood describe the mixture distributions \( p(x \mid \Theta_{12}) \) and \( p(x \mid \Theta_3) \), and \( \pi = \{\pi_{12}, \pi_3\} \), s.t. \( \pi_{12} + \pi_3 = 1 \) are prior weights on the relative size of regions, obtained from training data, cf. Sec. 4.2. It follows that

\[
1 - \gamma_i = 1 - P(W_i \in \mathbb{R}_{12} \mid x_i, \Theta, \pi) = P(W_i \in \mathbb{R}_3 \mid x_i, \Theta_3, \pi).
\]

(4.23)
4.2. Bayesian Probability Maps

**Figure 4.2:** Regions used in likelihood formulation: $R_{12}$ signifies endocardium and atrial region, while $R_3$ denotes the background.

**Figure 4.3:** Histogram of B-mode gray level intensities and estimated mixture (red) distribution $p(x \mid \Theta)$. Component distributions of mixture (blue, yellow, purple, cyan). The green area, given by $p(b_1 \mid \Theta)$, indicates the probability of the censored gray level intensities in the first bin $b_1$. The censored gray level intensities are also indicated by the green area in the B-mode US image (top left) corresponding to the histogram.

The apex sampling weight $A(z_i \mid \sigma_{\text{apex}})$ is introduced to deal with the fact that the endocardium is problematic to model in the apical long axis view. When the apex of the heart is highly curved, this may appear as thickening of the cardiac wall due to foreshortening.
The apex sampling weight $A(z_i | \sigma_{\text{apex}})$ is given by

$$ A(z_i | \sigma_{\text{apex}}) = P(W_i \in R_{12} | x, \Theta, \pi) \cdot (1 - G_{\sigma_{\text{apex}}}(z_i - z_{\text{apex}})) , \quad (4.24) $$

where $G_{\sigma_{\text{apex}}}$ is a 2-dimensional Gaussian density, $z_{\text{apex}}$ the apex coordinate of the endocardium, and $z_i$ the spatial coordinate of pixel $i$. The influence of the down-weighting term decays according to Gaussian distribution with parameter $\sigma_{\text{apex}}$. In the formulation of the likelihood, we need to take into account the left-censoring of the data, seen in Fig. 4.3, which is the result of pooling all the data equal and below an (unknown) positive intensity level $a > 0$ into the first bin of the histogram of the data. This form of quanti-
zation affects the visual appearance only slightly, since the affected intensities are often in the low range, but this type of censoring must be taken into account when formulating a statistical model for US data. Censored data models with unknown censoring limits commonly occur in survival analysis, see e.g. [112]. The censored mixture model $p(x \mid \Theta)$ is defined as follows. Assume that the log-compressed intensities $\mathbf{x} = \{x_1, \ldots, x_N\}$ are independent censored at an unknown point $a$, and subsequently quantized (i.e. binned).

Let $b_1$ denote the value of the first bin.

Dependent on how many bits are used in quantization and on the degree of nonlinearity, the number $b_1$ will change. For a very high bit count $b_1 = 0$, since the quantization error is then $< 0.5$. As the bit count decreases, the number $b_1$ will increase; this of course disregards any nonlinear binning which may affect the result. The number $a + b_1$ is thus the range over which the signal is censored.

The distribution of the censored intensities $x$ is expressed as a censored mixture model of $q$ Gamma densities,

$$p(x \mid \Theta) = P_1(\Theta)\delta_0(x - b_1) + g(x \mid \Theta),$$

where

$$P_1(\Theta) = \sum_{j=1}^q \alpha_j \int_0^{a+b_1} f_{\text{gam}}(s \mid k_j, \theta_j)ds$$

and

$$g(x \mid \Theta) = \begin{cases} 0, & \text{if } x \leq b_1 \\ \sum_{j=1}^q \alpha_j f_{\text{gam}}(x + a \mid k_j, \theta_j), & \text{if } x > b_1 \end{cases}$$

In (4.25) $\delta_0$ denotes the Dirac delta function, not to be confused with Kronecker delta used in Sec. 4.2.

The shape parameter is restricted s.t. $k_j > 1$. This is to avoid components which solely model the first bin (akin to a Dirac "spike"), as we regard distributions with such components as non-informative.

Here the Gamma density function is denoted by $f_{\text{gam}}$, and parameters $\Theta = \{\Theta_{12}, \Theta_3\}$. The elements of $\Theta$ correspond to the endocardium and atrial region, and the background. Thus $\Theta_{12} = \{\alpha_j, k_j, \theta_j; j = 1, \ldots, q_{12}\}$ are the parameters of a $q_{12}$ component censored mixture distribution for the endocardium and atrial region, and $\Theta_3 = \{\alpha_j, k_j, \theta_j; j = q_{12} + 1, \ldots, q\}$ are the parameters for the $q_3 = q - q_{12}$ components mixture for the background. The mixture parameters $\alpha_j > 0$ fulfill $\sum_j \alpha_j = 1$.

The mode of component $j$ is denoted by $m_j$. We assume that the sum of modes of components associated with region $R_{12}$ is less than those associated with region $R_3$, i.e. $\sum_{j=1}^{q_{12}} m_j < \sum_{j=q_{12}+1}^q m_j$. 

111
CHAPTER 4. Segmentation of the Endocardium - Bayesian MCMC

Distribution Parameter Priors

The distribution parameter prior is given by

\[ P(\Theta) = P(k_1, \theta_1, \ldots, k_q, \theta_q) . \] (4.28)

To achieve the best possible separation between components we impose priors on the shape parameters and scale parameters corresponding to the two regions \( R_{12} \) and \( R_3 \). For the shape and scale parameters of the Gamma densities a uniform prior is defined,

\[ P(k_j, \theta_j) \propto \begin{cases} 1, & \text{if } k_j, \theta_j \in K_{12}; j = 1, \ldots, q_{12} \land k_j, \theta_j \in K_3; j = q_{12} + 1, \ldots, q \end{cases}, \] (4.29)

where \( K_{12} \) and \( K_3 \) are each a simplex, corresponding to region \( R_{12} \) and \( R_3 \) constructed from shape and scale parameters estimates obtained from training data.

Latent Variable Priors

The priors on the latent variables are defined as

\[ P(\mathbf{u} | \Lambda) = P_{sd} (\mathbf{u} | \lambda_{sd}) P_{\text{shape|sd}} (\mathbf{u} | \lambda_{\text{shape}}, \lambda_{\text{mean}}) , \] (4.30)

where \( \Lambda = \{ \lambda_{\text{shape}}, \lambda_{\text{mean}}, \lambda_{sd} \} \) are weighting parameters controlling the influence of the priors.

The two components of the prior characterize different kinds of properties preferred. The spatial derivative \( P_{sd} \) prior enforces spatial and temporal smoothness for latent variable images. Possible shape variations around the mean shape are described by \( P_{\text{shape|sd}} \) constructed from eigenshapes of manually segmented images through \( P_{\text{shape|sd}} \). The sequence of images is divided into subsequences, to take the temporal variations of the endocardium into account, and so for each part of the sequence a corresponding set of eigenshapes and mean is used.

Let \( I_{1(\mathbf{u} \in R_{12})}(\mathbf{z}; t) \) be the 3-dimensional array of latent variables corresponding to an image in the US sequence, where endocardial and atrial region latent variables are set to one. In Fig. 4.1, the array \( I_{1}(\mathbf{z}, t) \) is illustrated. Here \( \mathbf{z} \) and \( t \) are spatial and temporal coordinates, respectively (see Fig. 4.1). The spatial derivative prior is then given by

\[ P_{sd} (\mathbf{u} | \lambda_{sd}) \propto \exp \{ -\lambda_{sd} || I_{1(\mathbf{u} \in R_{12})} \ast f ||_{L_1} \} , \] (4.31)

where \( f \) is a three dimensional Laplacian kernel and \( \ast \) denotes convolution. By increasing \( \lambda_{sd} \) our sampling is regularised towards smoothness.

Inspired by the Bicycle Chain model in Sommer et al. [197], we build our shape prior as follows. For each image in the US sequence, there is a corresponding image containing the latent variables describing the endocardium, atrial region and background. The
4.2. Bayesian Probability Maps

 contours of the endocardium and atrial regions are discretized in an equidistant manner. Let \( z^e_h \) be the \( h \)-th spatial coordinate in the \( xy \)-plane, taken from an equidistant set of \( n \) points from the boundary of the endocardium. For the atrial region the \( k \)-th spatial coordinate in a equidistant set of \( m \) points is denoted by \( z^a_k \). See Fig. 4.4 for illustration. Then \( v = (z^e_1, ..., z^e_n, z^a_1, ..., z^a_m) \) jointly provides the location of the endocardium and atrial region. Now let \( t_z \) and \( \phi \) be translation and rotation parameters, respectively, that take \( v \) to the mean shape \( \bar{v} \) of the training data. The parameters are collected in the vector

\[
\eta = (z^e_1, ..., z^e_n, z^a_1, ..., z^a_m, t_z, \phi) \in \mathbb{R}^{mn+3}.
\]

Now, let the shape prior be defined as

\[
P_{\text{shape} | \text{sd}}(u | \lambda_{\text{shape}}, \lambda_{\text{mean}}) \propto \prod \exp \left\{ -\lambda_{\text{shape}}(\eta_t - \bar{\eta}_t)^T((1 - \lambda_{\text{mean}})C_t^+ + \lambda_{\text{mean}}I)(\eta_t - \bar{\eta}_t) \right\},
\]

where \( \eta_t \) and \( C_t \) represent the joint shape of the endocardium and atrial region, and \( C_t^+ \) the Moore-Penrose pseudoinverse of the covariance matrix of the training shapes at the
time t, respectively. Increasing \( \lambda_{\text{shape}} \) makes the influence of the shape prior larger, while larger \( \lambda_{\text{mean}} \) increases the influence of the mean shape in the shape prior.

**Algorithm**

Our algorithm for generating Bayesian Probability Maps can be divided into three parts. First the censored Gamma mixture model parameters \( \Theta \) are estimated by an EM-MAP algorithm from our data; these parameters are used to compute the class posterior probabilities \( \gamma_i \) for each latent variable \( w_i \) after seeing the corresponding image values — these probabilities \( \gamma_i \) are used as an input in the following step in constructing the likelihood function that is further transformed to the posterior probabilities of in the second step. The posterior of the latent variables \( u_i \) is then sampled by Gibbs sampling \[71\] and the samples are used to compute the sample mean, which we refer to as a Bayesian probability map. The algorithm is summarized in Fig. 4.6.

**Estimation of Model Parameters**

The parameters \( \Theta \) of the model (4.25) are obtained as the solution to the MAP (Maximum a Posteriori) problem

\[
\hat{\Theta} = \arg\max_{\Theta} p(x | \Theta) P(\Theta).
\]  

(4.34)

We solve the MAP problem (4.34) by use of Expectation Maximization (EM) \[53\]. Thus the complete data likelihood of the model (4.25) is represented according to the latent variable model as

\[
p(x, w | \Theta) = \prod_{j=1}^{q} \prod_{i=1}^{n_j} \left( \alpha_j \int_{0}^{a+b_1} f_{\text{gam}}(s | k_j, \theta_j) ds \right)^{\delta(W_i - j)} \times \prod_{i=n_1+1}^{n} \left( \alpha_j f_{\text{gam}}(x_i + a | k_j, \theta_j) \right)^{\delta(W_i - j)} \times P(k_1, \theta_1, ..., k_q, \theta_q)
\]  

(4.35)

where \( x \) is censored B-mode data, and \( w = (w_1, ..., w_N) \) are latent variables which provide class membership assuming data independence, cf. Sec. 4.2.

On the E-step, we build the expected complete data log-likelihood, conditioned on the measured data and the previous parameter estimates. Specifically, index the censored intensities \( x \) s.t.

\[
i = \begin{cases} 
1, ..., n_1 & \text{if } x_i = b_1 \\
n_1 + 1, ..., N & \text{if } x_i > b_1
\end{cases}
\]  

(4.36)

where \( b_1 \) is the gray level intensity at the first bin.
The expected complete data log-likelihood is then
\[
\chi(\Theta, \Theta^{(n)}) = \mathbb{E}_{w|x, \hat{\Theta}^{(n-1)}}\{\log p(x, w | \Theta)\}
\]
\[
= \sum_{j=1}^{q} \sum_{i=1}^{n_1} \left[ \log \left( \alpha_j \int_0^a f_{\text{gam}}(s | k_j, \theta_j) ds \right) \right] 
\times P(W_i = j | \Theta(n), x_i = b_1) 
+ \sum_{j=1}^{q} \sum_{i=n_1+1}^{n} \left[ \log \left( \alpha_j f_{\text{gam}}(x_i + a | k_j, \theta_j) \right) \right] 
\times P(W_i = j | \Theta(n), x_j > b_1) 
+ \log P(k_1, \theta_1, ..., k_q, \theta_q) .
\] (4.37)

On the M-step, \(\chi(\Theta, \Theta^{(n)})\) is maximized for \(\Theta\) to obtain an update for the parameters. The steps are iterated until convergence.

Initial values for the EM algorithm is obtained by use of the \(k\)-means algorithm [130] using squared Euclidean distances. The data is divided into \(n\) groups using \(k\)-means, and the Maximum Likelihood Estimate (MLE) is computed for each group. The EM algorithm is run with 50 initializations, and the value of \(\Theta\) which produces the highest posterior value is adopted.

The weights \(\pi_{12}\) and \(\pi_3\) in (4.22) are estimated by the proportion of pixels inside and outside the region \(R_{12}\) in the training data, respectively.

The parameter \(\sigma_{\text{apex}}\) of the apex sampling weight \(A(z_i | \sigma_{\text{apex}})\) is estimated using leave-one-out cross-validation, see Sec. 4.2.

**Sampling of the Posterior**

The sampling of the posterior is sampled by alternating between conventional Gibbs sampling [71, 129] and sampling of latent variable image translations. On the Gibbs sampling step, we draw the elements of the sample latent variable vector \(u\) from the conditional distribution
\[
P(u_i | u_1^{(r)}, \ldots, u_{i-1}^{(r)}, u_{i+1}^{(r-1)}, \ldots, u_N^{(r-1)})
\]
\[
= \left\{ P(u_i = k | u_1^{(r)}, \ldots, u_{i-1}^{(r)}, u_{i+1}^{(r-1)}, \ldots, u_N^{(r-1)}) \right\}^3_{k=1}, \ i = 1, 2, \ldots, N,
\] (4.38)

where \(P\) is the posterior probability from Eq. 4.20. Then, to obtain sample vector \(u^{(r+1)}\), we sample the distribution of translations which spatially move the latent variable image \(I_u\). The details of the translation sampling step are as follows. We sample the conditional translation distribution
\[
P(t | u, x, \Delta) \equiv P(u' | u, x, \Delta) \propto L(x | u', \Theta)P(u' | \Lambda),
\] (4.39)
where the latent variable vector \( u' \) is obtained from \( u \) by spatially translating the latent variable image \( I_u \) by \( t \).

We sample (4.39) by the Metropolis Hastings algorithm [136] with proposal density

\[
P_{prop}(t | u, x, \Delta) \propto \left( \prod_{i=1}^{N} \gamma_i 1(u'_i \in R_{12}) (1 - \gamma_i) 1(u'_i \in R_3) \right) P(u' | \Lambda) \propto \prod_{i=1}^{N} \gamma_i 1(u'_i \in R_{12}) (1 - \gamma_i) 1(u'_i \in R_3),
\]

(4.40)

where we have used the fact that the conditional translation distribution is independent of the priors. The proposal distribution is an approximation of the conditional translation distribution (4.39), in that the apex sampling weight \( \Lambda(z_i | \sigma_{apex}) \) has been dropped in (4.40). It follows that

\[
\log P_{prop}(t | u, x, \Delta) = \sum_{i=1}^{N} 1(u'_i \in R_{12}) \log \gamma_i + 1(u'_i \in R_3) \log(1 - \gamma_i) + C,
\]

(4.41)

where \( C \) is a constant, related to the latent variable priors \( P(u' | \Lambda) \), which does not depend on the translation. The sum in (4.41) represents correlations between the translated latent variable image and log probability densities. Hence, the logarithms of the conditional translation probabilities can be computed efficiently in the Fourier domain by the correlation theorem. By the Metropolis-Hastings rule for independent proposals, we draw the translation sample \( t' \) from \( P_{prop}(t | u^{(r)}, x, \Delta) \), the sample is accepted if

\[
b = \frac{P(t' | u^{(r)})}{P(t = 0 | u^{(r)})} \cdot \frac{P_{prop}(t = 0)}{P_{prop}(t' = 0)} \geq 1,
\]

otherwise accept with probability \( b \). If the translation is accepted, we set \( u^{(r+1)} = T'_t(u^{(r)}) \), where \( T'_t \) denotes the translation operator.

**Sample Mean**

To characterize the posterior distribution of \( \hat{P}(u_{endo} | x, \Delta) \), we compute an estimate of the conditional mean of the endocardial latent variables \( u_{endo} \) over the posterior

\[
E\{u_{endo} | x, \Delta\} \approx \frac{1}{M} \sum_{r} u^{(r)}_{endo} = \left( \hat{P}(U_i \in \text{endocardium}) \right)_{i=1}^{N} \equiv \hat{u}_{CM} \tag{4.42}
\]

by the latent variable sample vectors \( u^{(r)}_{endo} \). By the strong law of large numbers \( \hat{u}_{CM} \rightarrow E\{u_{endo} | x, \Delta\} \) when \( M \rightarrow \infty \). The corresponding image \( I_{\hat{u}_{CM}} \) represents the Bayesian probability map.

**Estimation of Hyperparameters**

For each sequence manually segmented in its entirety, the data is divided into two sets corresponding to endocardium and atrial region, and background. ML estimates of distri-
4.2. Bayesian Probability Maps

Figure 4.7: Bayesian probability maps and ground truth (dashed white).

Bivariat distribution parameters $\hat{\Psi}_{12}$ and $\hat{\Psi}_3$ are computed for each region $R_{12}$ and $R_3$ by maximizing the likelihood

$$L_{\text{train}}(x) \doteq L_{\text{train}}^{12}(x; c_1, c) \cdot L_{\text{train}}^{3}(x; d_1, d)$$

(4.43)

where

$$L_{\text{train}}^{(i)}(x; c_1, c) = \prod_{i=1}^{c_1} \left( \int _{0}^{a} f(s \mid \hat{\Psi}^{(i)}) ds \right) \cdot \prod_{i=c_1+1}^{c} (f(x + a \mid \hat{\Psi}^{(i)}),$$

(4.44)

and number of observations are $c + d$, where $c$ observations are taken from the endocardium and atrial region, and $d$ observations from the background. The number of censored observations in each region is given by $c_1$ and $d_1$. Note that without the censoring limit $a$, the two likelihood terms in (4.43) would be maximized independently.

The hyperparameters $K_{12}$ and $K_3$ of the uniform distribution prior (4.29) are defined by

$$K_{12} = [\min_i (k_{12}^l), \max_i (k_{12}^l)] \times [\min_i (\hat{\theta}_{12}^l), \max_i (\hat{\theta}_{12}^l)]$$

(4.45)

and

$$K_3 = [\min_i (k_3^l), \max_i (k_3^l)] \times [\min_i (\hat{\theta}_3^l), \max_i (\hat{\theta}_3^l)],$$

(4.46)

where $k_{12}^l$ and $\hat{\theta}_{12}^l$ are the ML estimates of region $R_{12}$ obtained by maximizing the
### Table 4.1: Evaluated distribution models

<table>
<thead>
<tr>
<th>Model</th>
<th>Gamma</th>
<th>Nakagami</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution</td>
<td>$f_{\text{gam}}(x \mid k, \theta) = x^{k-1}e^{-x/\theta}/\Gamma(k)$, $x \geq 0$ and $k, \theta &gt; 0$</td>
<td>$f_{\text{nak}}(x \mid \mu, \omega) = \frac{2\mu^\mu}{\Gamma(\mu)}x^{2\mu-1}e^{-\mu x^2}$, $x \geq 0$ and $\mu, \omega &gt; 0$.</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>$f_{\text{logn}}(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$, $x \geq 0$ and $\sigma &gt; 0$</td>
<td>$f_{\text{rayl}}(x \mid \sigma) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$, $x &gt; 0$ and $\mu \in \mathbb{R}, \sigma &gt; 0$</td>
</tr>
<tr>
<td>Fisher-Tippett</td>
<td>$f_{\text{fish-tipp}}(x \mid \mu, \sigma) = \exp \left{ \frac{(x - \mu)}{\sigma} - \exp { (x - \mu)/\sigma } \right}$, $x \geq 0$ and $\mu \in \mathbb{R}, \sigma &gt; 0$</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.2: Mean AMDL score for distribution models

<table>
<thead>
<tr>
<th>Dist. model</th>
<th>Gamma</th>
<th>Nakagami</th>
<th>Rayleigh</th>
<th>Fisher-Tippett</th>
<th>Log-Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td># comp.</td>
<td>$q_c = 2$, $q_c = 2$, $q_c = 2$, $q_c = 3$, $q_c = 2$</td>
<td>$q_d = 2$</td>
<td>$q_d = 2$, $q_d = 1$, $q_d = 2$</td>
<td>$q_d = 2$</td>
<td></td>
</tr>
<tr>
<td>Mean AMDL</td>
<td>$5.10 \cdot 10^6$</td>
<td>$5.20 \cdot 10^6$</td>
<td>$5.38 \cdot 10^6$</td>
<td>$5.35 \cdot 10^6$</td>
<td>$5.38 \cdot 10^6$</td>
</tr>
</tbody>
</table>

Joint likelihood (4.43) for training sequence $l$, and $\hat{k}_l$ and $\hat{\theta}_l$ are the corresponding ML estimates for region $R_l$.

The prior parameters $\Lambda = \{\lambda_{\text{shape}}, \lambda_{\text{mean}}, \lambda_{\text{sd}}\}$ are estimated by leave-one-out cross-validation over the training data set, cf. Sec. 4.2.

### Estimation of Number of Mixture Components and Evaluation of Distribution Models

Several distribution models are considered for use in a mixture density, modeling endocardium and atrial region, and background gray level intensities in clinical B-mode US cardiac data, which have been reported as appropriate for modeling cardiac US gray levels in previous works [149, 204]. The ones considered here are the Gamma, Nakagami,
Rayleigh, Fisher-Tippett and Log-Normal distribution, given in Table 4.1. The relationship between Nakagami and Gamma distribution is detailed in Sec. 2.23 in Chp 2. The Rayleigh distribution is similar in shape to Gamma and Nakagami, and was first proposed in Burckhardt [33] for modeling US data. The log-normal distribution has been used for liver [236] and breast US [225]. In Dutt [60] the Fisher-Tippet, or double-exponential distribution, is shown to be most suitable for log-compressed envelope detected RF data with fully developed speckle. We consider using a maximum of two densities to model each of the region \( R_{12} \) and \( R_{3} \), thus six possible component density configurations for each distribution. In order to avoid densities which only model the first bin in the histogram, as mentioned for the case of the Gamma distribution in Sec. 4.2, the following constraints are imposed: \( \mu > 1 \) for the Nakagami distribution and \( \sigma < 0.5 \) for the Log-Normal distribution.

We compute the MAP solution, cf. (4.34), for each choice of component configuration and distribution. Uniform priors for the parameters of each distribution is constructed in an analogous manner to the case of the Gamma distribution, cf. Sec. 4.2. We select the configuration of components and distribution which produces the least mean Asymptotic MDL (AMDL) score, see e.g. [9], for taken over all training sequences. The Asymptotic MDL\(^1\) principle selects the mixture model with \( d \) components which minimizes the quantity

\[
-2 \log L(\hat{\omega}) + d \log n,
\]

where \( L \) signifies likelihood, \( \hat{\omega} \) the MLE of the model parameters and \( n \) the number of observations.

### Experiments

#### Material

The US data used consisted of 27 single beat cardiac cycles of two-chamber (2C) apical long-axis views of the heart, from the database described in Sec. 1.6.3. The length of the single beat sequences were normalized to length 16, which was found to be sufficient to capture the temporal variations for all sequences; the maximum length of an unnormalized sequence was 52.

The data was divided into two sets: training and test set. The training set consisted of 13 cardiac cycles. The training set was further divided into sets, corresponding to parts of the cardiac cycle. The test set consisted of 14 cardiac cycles, each acquired from 14 individual patients.

Significant left-censoring of data, more than 20 percent of total data observations in first histogram bin, was found in 16 out of 27 data sets. See Fig. 4.3 for an example of a left-censored US image.

\(^1\)The AMDL principle is equivalent to the Bayesian Information Criterion (BIC).
The training sets were segmented in their entirety, and for the test sets three frames at end-diastole and end-systole, respectively, were segmented. The atrial region was segmented by a trained non-expert in the training set. Since goal of the method is to determine the position of the endocardium, an approximation of the atrial region is sufficient for the shape model. The training sets were randomly selected among the 27 considered data sets in this paper.

Data with artifacts like substantial rib-shadows and those related to reverberations from pathological structures in lung were excluded from the study, since mixture modeling is invalid when large part of the data (falsely) exhibits tissue-like appearance.

Quantitative Validation

We evaluated the proposed method on 14 separate cardiac sequences. For each sequence 6 manually segmented frames were obtained, 3 end-diastolic and 3 end-systolic.

We employed validation measures of cardiac 2D+T US segmentation methods used in published quality studies [42, 28, 26, 142]: Contour Average Distance (Contour AD), Area Correlation, Area error and Corresponding Point Average Distance (CPAD). The ground truth segmentation was taken from a senior expert observer.

CPAD is the mean unsigned Euclidean distance between corresponding landmarks on ground-truth (see Fig. 4.8) and the segmentation obtained by the proposed method. In Bosch et al. [26] it is argued to be a more robust way of measuring automatic segmentation results than Contour AD, which measures minimal distance of points on ground-truth to algorithm segmentation. Contour AD may not register significant deviations along the border, since it only considers the distance to the closest point on the corresponding contour, and not in which order the points have been placed. In our case, two of the landmark points (1 and 37) were placed at the mitral valve and the third (19) at the apex of the endocardium, and the remaining were equidistantly placed between these three, see Fig. 4.8.

In addition to this score we determined the mean and standard deviation of each of 37 discretization point of endocardial contour. This was done to highlight the uncertainty of the segmentation at different part of the endocardium. For area measures the endocardial area was defined as area enclosed by the endocardial border.

The mean and standard deviation of the results of the proposed method in Fig. 4.9 and Table 4.4 were computed over all the samples, which is natural since the proposed method is stochastic.

Manual Inter- and Intraobserver Variability of Ground Truth Data

Manual segmentations of the endocardium were performed by three independent expert clinicians, using the software package ITK-Snap [229]. The most senior clinician provided segmentations of all sequences used in this paper, while the remaining clinicians segmented a selection of sequences. Naturally, it would have been preferable to obtain
4.2. Bayesian Probability Maps

Figure 4.8: Evaluation Landmarks.

Figure 4.9: Unsigned distances from sample points to corresponding expert points (mean ± std).

three expert segmentations of all sequences, but the manual segmentation work-load excluded this option.

The interobserver study was limited in scope, and so the numbers in Table 4.3 are necessarily approximate. Note that among other works concerning segmentation of LAX sequences, the number of experts used is either one [26, 142, 28] or two [105].
CHAPTER 4. Segmentation of the Endocardium - Bayesian MCMC

Figure 4.10: Expert contour (red), Contour of final sample using proposed method (black). Figures (a), (b), (c) and (d) are taken from two sequences where the proposed approach performs favorably. Figures (e), (f), (g) and (h) are taken from two sequences where the proposed approach has difficulties. Figures (a), (c), (e) and (g) depict Systolic frames, while figures (b), (d), (f) and (h) are Diastolic frames.

We selected 17 sequences randomly from the same database, from which our training and test data originates. In the interobserver study full sequences were segmented, and in the intra-observer study 3 end-diastolic and 3 end-systolic frames were segmented three
4.2. Bayesian Probability Maps

Figure 4.11: Ultrasound segmentation is a difficult task, prone to inaccuracies. Here we display three expert segmentations (green, red, black) together with the segmentation (final sample) from proposed approach (yellow).

Times by each clinician over a period of 6 months. The results of this study are presented in Table 4.3. In the calculations of deviations from groundtruth, the segmentation of the senior expert clinician was used as baseline.

Notice that the mean and standard deviation for inter- and intraobserver study presented in [142, 26], are lower than the ones we achieved in our study, which indicates that our data is more challenging. In Fig. 4.11 we highlight the difficulty of performing manual segmentation, and that it is not perfectly evident what is groundtruth. We also show the result of applying the proposed approach to the same sequence.

Results

Table 4.2 contains the least mean AMDL scores for the 5 considered distributions and 6 possible configurations of mixture components for the two regions $R_{12}$ and $R_{3}$. The Gamma model provides the lowest (and thus best) mean AMDL score, with $q_c = 2$ for endocardium and atrial region, and $q_d = 2$ for background. Thus we model the US data as a four component ($q = 4$) Gamma mixture model. The Nakagami distribution has results similar to Gamma, and thus one could employ this instead of Gamma with minor changes to the end result. In the case of Fisher-Tippett the EM algorithm did not converge in 5 instances. These results were not included in the calculation of the mean, since we do not wish to punish the Fisher-Tippett distribution inordinately. It may be possible that convergence could be obtained using a different initialization scheme.

When sampling the posterior, 60 discretization points for the endocardium and atrium ($n = m = 60$), cf. Sec. 4.2, was found to be sufficient to capture small sample variations. Leave-out-one cross-validation was performed on the training set, and the mean parameter estimate was computed over all cross-validation rounds, for parameters $\lambda_{\text{shape}}$, $\lambda_{\text{mean}}$, $\lambda_{\text{sd}}$ and $\sigma_{\text{apex}}$. The estimates $\bar{\lambda}_{\text{shape}} = 1.15$, $\bar{\lambda}_{\text{mean}} = 0.16$, $\bar{\lambda}_{\text{sd}} = 0.6$ and $\bar{\sigma}_{\text{apex}} = 1.22$ were obtained.

Leave-out-one cross-validation is performed as follows. Each training sequence of images has a corresponding groundtruth segmentation sequence. In each round of cross-
validation the covariance matrix of shapes is computed, cf. (4.33), from groundtruth segmentations, leaving out one training sequence. Using the estimated covariance matrix the algorithm is applied to the left out training sequence, and the optimal parameters are determined w.r.t. Average Distance (AD) from the algorithm segmentation to the groundtruth segmentation. This procedure is repeated for all samples, and the mean is taken as the parameter estimate.

The simplices $K_{12}$ and $K_3$, constructed from shape and scale parameters estimated in (4.43), were found to be $K_{12} = [1, 1.2] \times [0.022, 0.109]$ and $K_3 = [1.50, 7.12] \times [0.0562, 0.273]$. The small values for scale parameters $\theta$ is due to normalization of the data. The prior weights on the relative size of regions, cf. Sec. 4.2, are $\pi_{12} = 0.32$ and thus $\pi_3 = 1 - \pi_{12} = 0.68$. Initialization of latent variables $u$ is given by the mean of the training shapes for each time instant. The positioning of the initial shape is achieved by drawing a sample from the translation distribution of the latent variables, cf. Sec. 4.2, hence no manual initialization is needed.

In Fig. 4.7 we display Bayesian Probability Maps formed from 50 samples; with a burn-in of 100 samples. The probability map spans colors from red to blue with degree of probability, of area being within the endocardium. Hence, red indicates the highest probability while no color signifies zero probability. It follows that the more varied in color the map is, the more uncertain it is. This is seen in the right-hand plots in Fig. 4.7, which in places have colorful border regions.

The obtained results have lower mean and standard deviation than those obtained through our interobserver study. A Welch two-sample t-test, hence assumption of unequal sample-size and unknown variance in both groups, at significance level 0.05 reveals that the proposed approach produces segmentations significantly better than an expert (p-value of this test is $7.41 \times 10^{-28}$). However, since the interobserver study is small, one needs to be careful with drawing inference from this result. The result indicates that the proposed method performs as well or better than expert segmentation.

Quantitative results for the proposed method and other comparable works for 2D+T segmentation are summarized in Table 4.4. The cited works are chosen, since they all have varying degrees of systematic validation. In Fig. 4.9 the mean and standard deviation of distances from the samples of the contour, to each expert point along the endocardium is displayed.

Running time on a standard laptop (Intel (R) Core Duo, 2.40 GHz, 4 GB RAM) is 3.2 sec/sample. Total running time for a 16 slice sequence, where 150 samples are obtained of each slice, is approximately 2.5 hours. The code is written in MATLAB with some elements in C++.

**Discussion**

We have not had access to the US data against which the other algorithms were validated, so comparison against these models is only approximate. To check the validity of our
algorithm we compare our results against those produced on sequences from our database. When the proposed method is compared against the results of experts on material from our hospital database, the method performs quite well, see Table 4.4 and 4.3. Our inter- and intraobserver study does not achieve the same low mean and standard deviations as in Mitchell et al. [142], which may indicate that our data material is more challenging. Note that same data material is used in Bosch et al. [26] as in Mitchell et al. [142].

The left endocardial walls (point 1 – 18) are easier for the proposed algorithm to properly segment, as seen in Fig. 4.9, while there is an increase in mean distance of the segmentation to groundtruth in the right wall 19 – 28. It is also evident that the variation of points is larger on the right wall. This is typically the most difficult part of the endocardium to segment, as it is the region most prone to signal dropout.

Fig. 4.10 highlights situations where the proposed approach works well and less so. The most important factor in how well the algorithm will perform is how severe signal dropout is in the right chamber wall. If the dropout is severe, then the algorithm will have to rely more on priors which necessarily will cause the result to deviate more from the true chamber wall, as seen in Fig. 4.10g and 4.10f. One should also note that not only the algorithm is adversely affected by severe signal dropout; manual segmentation by expert will also be adversely affected. The extent of how much a single expert is affected by signal dropout is difficult to quantify, barring an extensive intra-expert study.

The parameters of the model, determined by cross-validation over 13 training samples, are not optimal for every setting. Having an adaptive scheme, where the parameters are set according to e.g. the current distribution of the sample, would be more desirable and would potentially increase the flexibility of the model. Increasing the training set may also yield a similar result, although one must take into account the effect of specific diseases on the ultrasound echo, which may make a cross-validated parameter estimate, even over a large sample, too approximate.

We have made our study with censored clinical B-mode US data, by using a censored mixture model, which does not assume any additional information about the signal, such as knowledge of dynamic compression parameter [191]. The use of a censored mixture model is easily extendible to other statistical segmentation paradigms, such as variational formulations. Furthermore the study is performed on the most challenging of views of the heart, namely apical long axis view, as opposed to the more commonly examined short-axis view.

The proposed method assumes that it is possible to separate US data into components corresponding to blood and background. If artifacts, e.g. rib-shadows, are very substantial, this assumption will be violated. It would be possible to include this data in the training data set, but since these data sets have substantially different appearance, it would skew parameter estimation in an unwanted manner; rib-shadow would require very large shape prior parameters.
4.3 Conclusion

We have proposed a Bayesian MCMC approach for describing the position distribution of the endocardium. A left-censored Gamma mixture model was introduced, to address the quantization error in the B-mode US data. The problem of determining the position distribution was cast as a latent variable model, which represents the three regions endocardium, atrial region and background, for which the posterior density is estimated. The data likelihood of the endocardium and atrial region, and background was described as a mixture of components of the censored gamma mixture model. The model was refined by incorporating priors for spatial and temporal smoothness, and preferred shapes, by using the principal components of manually segmented training shapes. Misclassification of the apex of the endocardium, due to foreshortening related to a highly curved apex of the endocardium, was handled by down-weighting of the likelihood in this region. The posterior density was sampled by a Gibbs method to estimate the expected latent variable image of the endocardium, which we call the Bayesian Probability Map, since it describes the probability of pixels being classified as being within the endocardium. The algorithm was initialized by sampling the translation distribution of the latent variables, improving the convergence rate of the algorithm. The regularization parameters of the model were estimated by cross-validation, and the results are validated using a variety of validation measures, in order to compare with expert results and previously published works.

Our experiments show the usefulness of the Bayesian Probability Maps for the clinician since, instead of producing a single segmenting curve, it highlights the uncertain areas and suggests possible segmentations. Furthermore, an interobserver study shows that the results produced by the algorithm agree well with expert segmentations.

4.4 Future work

The parameters of the model, determined by cross-validation over 13 training samples, are not optimal for every setting. Having an adaptive scheme, where the parameters are set according to e.g. the current distribution of the sample, would be more desirable and would potentially increase the flexibility of the model. Increasing the training set may also yield a similar result, although one must take into account the effect of specific diseases on the ultrasound echo, which may make a cross-validated parameter estimate, even over a large sample, too approximate.
### Table 4.3:
Manual inter- and intraobserver errors (mean ± std)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Interobserver</th>
<th>Intraobserver</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPAD</td>
<td>4.61 ± 3.67 mm (7.47 ± 5.68 pxl)</td>
<td>1.75 ± 2.51 mm (2.80 ± 4.05 pxl)</td>
</tr>
<tr>
<td>Contour AD</td>
<td>3.11 ± 1.04 mm (5.11 ± 1.72 pxl)</td>
<td>2.02 ± 1.03 mm (3.24 ± 1.65 pxl)</td>
</tr>
<tr>
<td>Area Correlation coeff.</td>
<td>0.87</td>
<td>0.94</td>
</tr>
<tr>
<td>Area Error (pxl²)</td>
<td>-1.04 ± 16.62 %</td>
<td>1.05 ± 13.58 %</td>
</tr>
<tr>
<td>CPAD (Bosch [26], Mitchell [142])</td>
<td>3.82 ± 1.44 mm</td>
<td>2.32 ± 0.75 mm</td>
</tr>
<tr>
<td>Area Error (pxl²) (Bosch [26], Mitchell [142])</td>
<td>-4.39 ± 10.3 %</td>
<td>0.92 ± 6.19 %</td>
</tr>
</tbody>
</table>
## Table 4.4: Results of the proposed method and other comparable works on 2D+T cardiac segmentation

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Modality</th>
<th>Ad hoc parameters</th>
<th>Interaction</th>
<th>Evaluation</th>
<th>Performance</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen [42]</td>
<td>2004</td>
<td>LAX</td>
<td>(-)</td>
<td>(-)</td>
<td>Ellipse</td>
<td>initial QN(m) Contour AD (end)</td>
<td>1.97 pxl 3(3) OB(1)</td>
</tr>
<tr>
<td>Boukerroui [28]</td>
<td>2003</td>
<td>LAX</td>
<td>(+/-)</td>
<td>(+)</td>
<td>Area error (pxl^2)</td>
<td>-7.59 ± 6.71 %</td>
<td></td>
</tr>
<tr>
<td>Bosch [26]</td>
<td>2002</td>
<td>4C</td>
<td>(-)</td>
<td>(+)</td>
<td>Manual Training (AAM) QN(m) CPAD</td>
<td>3.35 ± 1.22 mm 64(1024) OB(1)</td>
<td></td>
</tr>
<tr>
<td>Mitchell [142]</td>
<td>2002</td>
<td>4C</td>
<td>(+)</td>
<td>(+)</td>
<td>Manual Training (AAM) QN(m) CPAD</td>
<td>3.90 ± 1.38 mm 64(1024) OB(1)</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE KEY:**
- **Modality:** LAX = Long Axis, 2C = Two chamber, 4C = Four chamber.
- **Evaluation:** QN(m) = quantitative (manual delineations).
- **Adhoc parameters:** (-) none or sensitivity analysis is done, (+/-) yes and some attempt at sensitivity analysis, (+) yes and no sensitivity done.
- **Validation number:** Number of subjects (number of images).
- **Automation:** (+) full, (-) interactive guidance/correction.
- **Validation type:** OB(m) = human observer (number of observers).

Note: We employ the same notation for properties of models as used in Noble et al. [151].
Chapter 5

Feature Descriptor for and Registration of Envelope Detected Radio Frequency Ultrasound

5.1 Background

Recently, modeling textures using MRFs has become quite popular. Especially in the domain of synthetic aperture radar (SAR) image analysis they enjoy wide-spread use. For MRF texture modeling as well as synthesizing there exist numerous approaches for formalizing the underlying distributions, ranging from parametric to non-parametric. Inherent to all of them is the idea of combining the notion of distribution and neighborhood dependency. Besag [20] proposed the so called auto-models, allowing description of interaction patterns within the statistical environment of exponential family distributions. An ultrasound specific auto-model was proposed by Bouhlel et al. [27] by embedding the Nakagami distribution into a Markov random field (MRF) facilitating the classification of cancerous breast tissue. In Sec. 5.6 we formulate and analyse a MRF-based feature descriptor for RF ultrasound data. Among the distributions described in Chp. 2 the, comparably simple but nevertheless versatile, Nakagami distribution [188, 146] can account for varying scattering conditions. The closest work, which shares analogy with our method but, however, does not formulate a global feature descriptor, see [27]. There have been attempts at using texture for segmentation of ultrasound, see [152] for an overview. However, these models do not use Besag’s auto-model, which is essential to our work; instead they employ simpler statistics like local mean and variance as feature descriptors.
Registration of images is a fundamental problem in several fields, particularly in the domain of medical imaging. However, issues such as noise and artifacts complicate this process and often make automatic registration processes intractable. Those issues are in particular prevalent in ultrasound (US) imaging due to its complex physical nature. Nevertheless, there exist a multitude of applications in US registration such as in elastography, speckle tracking and motion recovery. Hence, the development of similarity metrics in US registration is an active field of research. A common approach in modeling this problem is to employ statistical methods, employing the distributional models described in Chp. 2. Unlike in the conventional vision domain, where the Gaussian intensity distribution has been successfully applied in a variety scenarios, in US other models are more appropriate. This is particularly the case for envelope detected radio frequency (RF) US data. The statistical properties of the echo envelope of US data depend on numerous factors. Among them is, in particular, the density and spatial distribution of scatterers in the medium. As different types of biological tissue exhibit various characteristics w.r.t. density and scatterer distribution, this can be utilized for registration purposes. In Sec. 5.8 we propose a hybrid approach combining global statistics (in the form of distribution matching) with local statistics (texture patterns). We refer to this methods as hybrid local binary patterns (HLBP).

As described in Chp. 1, ultrasound is highly view dependent. In Section 5.8 we extend the classical US RF envelope modeling by employing a Nakagami FMM-based approach for 3D US freehand data. This allows us to embed the view-dependent property of US in a statistical formulation. FMMs have already been successfully applied to US data for segmentation of the carotid artery [56]. However, our target applications for this model are registration and reconstruction. To our knowledge we are the first to apply a statistical parametric approach in conjunction with similarity measures (in our case the J-divergence distance metric) to registration of 3D US data. Previous works [100] have applied similarity metrics, e.g. Kullback-Leibler (KL), Hellinger and Bhattacharyya, directly without any parametric distributional assumptions on the data. Performance and use of the model described in Sec. 5.8 is show-cased on transcranial US (TCUS) brain data, where 3D freehand image sequences are taken through a narrow bone window at the temporal lobe. The domain of neuro-US is quite relevant for the proposed approach. This is because of the need of high registration accuracy for brain data for applications such as electrode implantation in the brain. Also, TCUS has been recently shown to be appropriate for early diagnosis of PD [217]. The process is associated with the agglomeration of ferrite deposits, which form hyperechogenic areas in the SN [12] that are visible in US. In this regard, generating 3D data from 2D images by means of a tracked transducer can help in reducing subjectivity in diagnosis [161]. Furthermore, accurate registration of 3D data is required to perform continuous staging of hyperechogenic SN regions.
5.2 Markov Random Fields

We here consider Markov Random Fields defined over spatially regular lattices, which is need for the formulation of the feature descriptor in Sec. 5.6. Let $\mathcal{S} = \{(i,j) \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ be a regular lattice of sites for the image of size $m \times n$. The labels of the sites can be continuous or discrete. In our work we deal with regularly spaced sites with continuous labels. We denote a labeling of the sites on the lattice by

$$f = (f_1, \ldots, f_{m \cdot n}) \quad (5.1)$$

where $f_i : \mathcal{S} \rightarrow \mathcal{L}$. Here $\mathcal{L}$ is a discrete or continuous label, e.g. for gray-scale images, $\mathcal{L} = [0, 256]$. The possible combinations of labels of sites are contained in the configuration space

$$\mathbb{F} = \mathcal{L} \times \ldots \times \mathcal{L} = \mathcal{L}^{m \cdot n}, \quad (5.2)$$

thus $f \in \mathbb{F}$.

The sites in $\mathcal{S}$ are related by neighborhood system $\mathcal{N}$

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{S}\} \quad (5.3)$$

where $\mathcal{N}_i$ is set of sites neighboring $i$.

The system needs to fulfill two fundamental conditions,

1. A site cannot be a neighbor to itself: $i \notin \mathcal{N}_i$
2. The neighbour relationship is mutual: $i \in \mathcal{N}_j \iff j \in \mathcal{N}_i$.
The most common neighbourhood sets are first and second order neighbourhoods. A first order neighbourhood has 4 elements, and two horizontal and vertical elements neighbouring the site. A second order neighbourhood on the other hand has 8 elements surrounding the site.

A clique \( c \) is a ordered set of sites in a neighbourhood, as described in e.g. Li [124]. A pair-wise clique as \( c = \{i, i'\} \), and triple-site clique as \( c = \{i, i', i''\} \). Traditionally one assumes the cliques to be ordered, that is that \( \{i, i'\} \) and \( \{i', i\} \) are distinct. However, in our work with MRFs in Sec. 5.6 we assume the cliques to be non-ordered.

The set of collections of pair-wise cliques is
\[
C_2 = \{ \{i, i'\} | i' \in N_i, i \in S \} \tag{5.4}
\]
and triple-site cliques
\[
C_3 = \{ \{i, i', i''\} | i, i', i'' \in S \text{ are pair-wise neighbours} \} \tag{5.5}
\]
See Fig. 5.1 for examples of elements in \( C_2 \) and \( C_3 \).

Let \( F = \{F_1, ..., F_m\} \) be a family of random variables on the set \( S \), in which random variable \( F_i \) takes a value in the state space \( E \); this amounts to letting the labels \( L \) in Eq. 5.2 be stochastic. The family is called a random field. In this context \( F = f \) denotes joint event, \( f = \{f_1, ..., f_m\} \) is a configuration \( F \) corresponding to a realization of the field. The set of configurations is \( F \), which is the same as (5.2) but in this case stochastic.

### 5.2.1 Markov Random Fields

**Definition 5.2.1.** \( F \) is a Markov random field on \( S \) w.r.t. a neighbourhood system \( N \) iff
\[
P(f) > 0, \forall f \in F \tag{5.6}
\]
\[
P(f_i | f_{S \setminus \{i\}}) = P(f_i | f_{N_i}) \tag{5.7}
\]
where
\[
f_{N_i} = \{f_j | j \in N_i\} \tag{5.8}
\]
is the set of labels at sites neighboring \( i \).

### 5.2.2 Gibbs Random Fields

**Definition 5.2.2.** The set of random variables \( F \) is a Gibbs Random Field (GRF) on \( S \) w.r.t. \( N \) iff its configurations follow a Gibbs distribution.

The Gibbs distribution is defined as
\[
P(f) = Z^{-1} \exp\{-\frac{1}{T} U(f)\} \tag{5.9}
\]
where
\[ Z = \sum_{f \in \mathcal{F}} \exp\{-\frac{1}{T} U(f)\} \] (5.10)

Here \( Z \) is a normalizing constant, \( T \) a temperature constant, usually assumed to be 1, and lastly \( U(f) \) is an energy function.

The energy function is defined by
\[ U(f) = \sum_{c \in \mathcal{C}} V_c(f) \] (5.11)
and is the sum of clique potentials \( V_c(f) \) over all possible cliques \( \mathcal{C} \).

The value of the clique potential \( V_c(f) \) is determined by the configuration of the clique \( C \). Thus when one considers that \( P(f) \) depends on the configurations of the cliques, it is clear that \( P(f) \) is in fact the probability of the occurrence of a particular pattern.

Following 5.11 case of cliques of size up to two the energy \( U(f) \) becomes,
\[ U(f) = \sum_{i \in S} V_i(f) + \sum_{i \in S, j \in N_i} V_{ij}(f_i, f_j) \] (5.12)

Probability of configuration \( f_i \) conditioned on neighbourhood \( f_{N_i} \) is now
\[ P(f_i|f_{N_i}) = \frac{\exp\left\{-\left[ V_1(f_i) + \sum_{j \in N_i} V_2(f_i, f_j) \right]\right\}}{\sum_{f_i \in S} \exp\left\{-\left[ V_1(f_i) + \sum_{j \in N_i} V_2(f_i, f_j) \right]\right\}} \] (5.13)

Markov Random Fields and Gibbs Random fields are fundamentally related, as is clear by the following theorem.

**Theorem 5.2.1.** ([20]) \( F \) is an Markov Random Field w.r.t. \( S \) iff \( F \) is a Gibbs Random Field w.r.t. \( \mathcal{N} \).

Thus a Markov Random Field may be expressed using an equivalent Gibbs Random Field, something which is used for e.g. auto-models.

### 5.2.3 Auto-models

Assume that \( X \) Markov random field on \( S = \{1, 2, \ldots, n\} \) with pair potentials
\[ P(f) = Z \exp\left\{\sum_{i \in S} V_i(f_i) + \sum_{i \in N_i} V_{ij}(f_i, f_j)\right\}, \] (5.14)
where \( A \) is a constant. In the following we can assume without loss of generality that \( Z = 1 \).
Theorem 5.2.2. Suppose that each conditional distribution \( P_i(\cdot | N_i) \) of \( P \) belongs to the exponential family:
\[
\log P_i(f_i|N_i) = A_i(N_i)B_i(f_i) + C_i(f_i) + D_i(N_i),
\]
where \( B_i(0) = C_i(0) = 0 \). Then:
For any \( i, j \in S, i \neq j \), there exists \( \alpha_i \) and \( \beta_{ij} = \beta_{ji} \) such that:
\[
A_i(N_i) = \alpha_i + \sum_{j \neq i} \beta_{ij}B_j(f_j)
\]
\[
\Phi_i(f_i) = \alpha_iB_i(f_i) + C_i(f_i), \quad \Phi_{ij}(f_i, f_j) = \beta_{ij}B_i(f_i)B_j(f_j)
\]

Using Thm. 5.2.2 we can model both local and global statistics. Thus if we assume that the intensities in an image are modeled by e.g. a Nakagami distribution,
\[
f_{\text{nak}}(x|\mu, \omega) = \frac{2\mu^\mu \omega^{2\mu-1}}{\Gamma(\mu)\omega^\mu} \exp\left(-\frac{\mu}{\omega}x^2\right), \forall x \in \mathbb{R}_+
\]

Then
\[
\log f_{\text{nak}}(x|\mu, \omega) = \log 2 + \mu \log \mu + (2\mu - 1) \log x - \left(\frac{\mu}{\omega}\right)x^2 - \log \Gamma(\mu) - \mu \log \omega
\]

It follows from Eq. 5.19 and 5.15 that
\[
A_i(N_i) = \alpha_i + \sum_{j \neq i} \beta_{ij}B_j(f_j) = 2\mu - 1
\]
and
\[
\mu = \frac{1}{2} \left( \alpha_i + 1 + \sum_{j \neq i} \beta_{ij}B_j(f_j) \right).
\]

Then using estimates of \( \mu \) in local regions obtained by e.g. Expectation Maximization, we can connect the global statistics of an image, here provided by the distributional parameters, with local characteristics provided by interaction potentials.
5.3 Local Binary Patterns

Texture classification is a wide domain in which numerous approaches exist. A relatively simple, but powerful and popular technique in computer vision is the so called local binary patterns (LBP), which were originally introduced by Ojala et al. in [154]. In its standard formulation it encodes a second order neighborhood of a pixel into a $2^8$ bits code. Encoding is based on the inequality relationship between the central site intensity and its 8 neighbors. A pixel position with lower intensity than the central one is attributed the binary code 0, otherwise 1. From those individual binary digits a code for the neighborhood and finally a histogram of codes, serving as descriptor for a region, is built.

The neighbourhood formulation above is a special case of the circular neighbourhood formulation: Let $I(x, y)$ be an image and $f_c$ a gray level value of an arbitrary pixel situated at $(x_c, y_c)$, surrounded by peripheral pixels at positions $(x_j, y_j)$. The circular neighbourhood of radius $R$ consisting $P$ points is given by

$$f_j = I(x_j, y_j),$$

$$x_j = x + R \cos(2\pi j / P),$$

$$y_j = y - R \sin(2\pi j / P)$$

for $j = 0, ..., P - 1$.

Define the LBP operator

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{P-1} H(f_p - f_c)2^p.$$
A texture is described as a $2^p$-bin discrete distribution of LBP codes. For an example of an LBP histogram, see Fig. 5.4. LBPs have been mainly applied to natural scene images. However, there also exist adaptations to cope with images that are inherently subject to noise. Extensions of LBP are in rotation invariance [3], uniform patterns [154] limiting the number of codes generated, multi-scale [154] and probabilistic LBP [122]. A specific extension of interest to us is the so called Fuzzy LBP (FLBP) for ultrasound proposed by Iakovidis et al. [99], in which not only a single LBP but various codes can represent a region, therefore being more robust in noisy image domains. A threshold specifies an intensity range in which fuzziness is assumed and a ramp (membership function) $m_0$, and its complement $m_1$, associates the probability for the binary class encoding. The membership functions are defined by

$$m_0(i) = \begin{cases} 
0, & \text{if } f_i - f_c > T \\
\frac{T-(f_i-f_c)}{2T}, & \text{if } |f_i - f_c| < T \\
1, & \text{if } f_i - f_c < -T .
\end{cases}$$

(5.26)

and

$$m_1(i) = 1 - m_0(i) .$$

(5.27)

The parameter $T \in [0, 255]$ controls the degree of fuzziness, or vagueness. Instead of a unit contribution of each LBP code, the contribution under FLBP is flexible. In the example of a $3 \times 3$ neighbourhood the contribution of each LBP code to the FLBP histogram is

$$C_{\text{LBP}} = \prod_{i=0}^{7} m_{H(f_i-f_c)}(i) .$$

(5.28)

A normalization over each neighbourhood is imposed in FLBP s.t.

$$\sum_{LBP=0}^{255} C_{\text{LBP}} = 1 .$$

(5.29)
See Fig. 5.4 for a comparison of LBP and FLBP histograms. For an exhaustive overview of variations of LBP, we refer the reader to [160]. For measuring similarity between LBP histograms, several methods have been proposed such as Histogram intersection, Log-likelihood statistic and Chi-square statistic - see [3] for a comparative study.

In Sec. 5.7 we propose an change to original FLBP approach, which defines the membership function as a ramp. Instead we propose a non-linear function associated with the underlying statistical properties of data. This method is then combined with Hellinger divergence to create a similarity measure, HLBP, Hybrid Local Binary Patterns.

5.4 Texture Based US Models

In the field of texture modeling in ultrasound a variety of approaches exist, we here provide a limited overview: Bleck et al. [25] uses texture classification method based on autoregressive periodic random field models used in discriminating between liver with and without microfocal lesions. A combination of fuzzy local binary patterns and fuzzy grey-level histogram features are used, in Iakovidis et al. [99], to model thyroid textures in B-mode ultrasound. For classification of thyroid a polynomial kernel support vector machine is utilized. Bouhlel et al. [27] estimates parameters of a model combining the Nakagami distribution, describing the envelope of RF-data, and a Markov Random Field model, describing spatial interaction. The parameters are used to model textures, specifically used to distinguish between normal and abnormal tissue in the case of choroidal malignant melanoma. A statistical feature vector is used in conjunction with probabilistic neural network is proposed in Huang et al. [97], to differentiate fatty liver B-mode ultrasound image from normal liver images.

5.5 Registration Methods

The background on registration methods we give here essentially follows the one description given in Klein [113].

There have been a number of works regarding registration of ultrasound images [201, 118, 45, 224, 29, 173, 235, 78, 10, 63, 65, 145]. These works can be roughly divided into four categories: (i) motion measurements in echocardiography for detecting and characterizing abnormalities, (ii) breast deformation analysis to assess the elastic properties of tissues, (iii) assessment of tissue strain with elastography, and (iv) multi-view compounding.

Block matching is a common strategy for registration, and there are several works exploring this approach. Incorporating statistical noise-models is a commonly applied strategy. Strintzis and Kokkinidis [201] observes that Gaussian statistics are poorly suited for ultrasound data, and proposes the use of the Rayleigh distribution to model data in block matching. In Cohen and Dinstein [45] noise is modeled in both moving and fixed
image, as opposed to just the moving image, and logarithmic transform integrated, in order to model log-compression (recall that this is a specific characteristic on B-mode data). The resulting similarity metric is applied for motion estimation employing a block matching approach in [29, 173]. Correlated speckle in adjacent frames in log-compressed US images sequences is modeled in Myronenko et al. [145] by use of bivariate Rayleigh and bivariate Nakagami.

Block matching by use of normalized cross-correlation (NCC) is used in Basarab et al. [10] for flow estimation and tracking tissue motion in elasticity imaging.

Block matching is applied in [118, 164] to the registration of multiple compounded data sets. Sum of squared differences (SSD) is in Krucker et al. [118] found to be suitable only in imagery that is subject to low noise levels.

It follows that SSD is less suited for ultrasound-to-ultrasound registration where speckle contamination is significant. Poon and Rohling [164] describes a compounding optimization algorithm using NCC.

Common to all the presented methods is that the parameters of the distributions are set heuristically on a global basis, which is at odds with the underlying local data variation. For instance, the variance of the Rayleigh distribution is set to $\frac{2}{\pi}$ in Cohen and Dinstein [45], or the shape and correlation parameters are set to $m = 0.5, \rho = 0.8$ in Myronenko et al. [145].

Further registration approaches: Registration by mutual information is proposed in [128, 234], which in turn apply a joint density formulation in local regions, or a combination of local and global estimates. These strategies attempt to deal with the deficiency of mutual information w.r.t. images which contain high intensity non-uniformity due to the bias field (a description of this issue can be found in Loeckx et al. [128]).

Lastly there exist various metric learning approaches, where similarity metrics are constructed by boosting in Babenko et al. [7], Random Forest [30] classification in [226] and supervised learning on previously registered data to train an appropriate similarity function in [121, 31].
5.5. Registration Methods

Figure 5.4: Top: patch of envelope detected neck US (log-transformed for display purposes). Middle: corresponding (8,1) LBP histogram. Bottom: corresponding (8,1) FLBP histogram. (8,1) denotes eight peripheral points and one central point.
CHAPTER 5. Feature Descriptor and Registration

5.6 Spatial Statistics Based Feature Descriptor for RF Ultrasound Data

This section is based on the paper Spatial Statistics Based Feature Descriptor For RF Ultrasound Data [115].

The theoretical prerequisites are provided in:
- Sec. 1.2 (Physical Background)
- Sec. 1.3 (Transmission to Display of Ultrasound Echo)
- Sec. 2.1 (Statistical Modeling of Ultrasound Data)
- Sec. 5.2 (Markov Random Fields)

5.6.1 General Background

Speckle in ultrasound is the result of the wavefront interference characterized by the spacing and organization of scatterers in tissue, [33, 216]. The resulting stochastic patterns are distinctive for organs. Backscatter intensities within an organ are of stochastic nature, however, follow a certain distribution and therefore are subject to a spatial interaction relationship within a neighborhood. This spatial interaction within a statistical framework can be nicely modeled with a MRF texture model. Following Bouhlel et al. [27] we propose a feature descriptor employing the descriptive nature of the interaction parameters of auto-models. This contrasts with the conventional MRF based SAR image analysis, where one set of interaction parameters is sought providing the best classification results. In the sequel we assume all intensities $x \in \Omega$ in the RF envelope image $\Omega$ to follow a Nakagami distribution

$$f_{\text{nak}}(x, \mu, \omega) = \frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\omega^\mu} \exp\left(-\frac{\mu}{\omega} x^2\right), \forall x \in \mathbb{R}_+ \quad (5.30)$$

with $\mu, \omega$ the shape and scale parameters respectively. When the data follows the Nakagami distribution, it can be readily embedded into an auto-model, since the Nakagami belongs to the exponential family. This in turn facilitates the decomposition of the log-likelihood as:

$$\log f_{\text{nak}}(x_i, \mu, \omega) = A(x_{N_i})B(x_i) + C(x_i) + D(x_{N_i}) , \quad (5.31)$$

where $A(.)$ and $D(.)$ depend on the intensities $x_{N_i}$ in the neighborhood $N_i$ of the central pixel $x_i$, whereas $B(.)$ and $C(.)$ depend only on $x_i$.

Following [20, 27] the neighborhood depending term $A(.)$ can be written as linear combination of intensities,

$$A(x_{N_i}) = \alpha_i + \sum_{i \neq i'} \beta_{i,i'} B_i(x_{i'}) . \quad (5.32)$$
such that $\alpha_i$ is the weight of the central site $x_i$ and $\beta_{i,i'}$ are the interaction weights between neighboring pixels. From Eq. 5.32 and 5.31 we then obtain

$$\mu_i = \frac{1}{2} \left( \alpha_i + \sum_{i \neq i'} \beta_{i,i'} \log x_{i'} \right). \quad (5.33)$$

### 5.6.2 Method

The goal of our approach is to derive a descriptor for a small region of the image referred to as patch. As a first step, in order to decompose the image, a sparse grid is placed on the entire image $\Omega$ subdividing the data into patches $P_i \subset \Omega$, s.t. $\bigcup_i P_i = \Omega$. Subsequently, in order to estimate the finite mixture model, a Nakagami MLE is operated on a patch containing area (or Gamma MLE, making use of the close Nakagami-Gamma relationship stated in Eq. 2.23) Having obtained the MLE for the Nakagami distribution, another more dense grid is placed on the patch, with $S \subset P_i$ denoting the set of all grid points on a specific patch. This is followed by an instantiation of the MRF interaction model centered on each sub-grid point $s \in S$. We define the cliques to be spatially opposite segments of an annulus of varying radius (depending on the degree of neighborhood) centered on sub-grid points, similar to Bouhlel et al. [27]. In the following we assume a second-order neighborhood - see Fig. 5.6.
CHAPTER 5. Feature Descriptor and Registration

5.6.3 Optimization

At the final stage we compute the interaction parameters for all patches $P_i$, using either conditional-least-squares (CLS) or maximum-pseudo-likelihood (MPL). The interaction feature vector

$$\psi = (\alpha, \beta_1, \ldots, \beta_n) \in \mathbb{R}^{n+1},$$  \hspace{1cm} (5.34)

employing a neighborhood system with $n$ cliques, is computed for each patch separately. It is comprised of the $\alpha$ and the $\beta_i$ parameters. The former can be related to the density and amplitudes of scatterers. In contrast to that, the latter type of parameters represent the interaction of the diametrically opposite neighbor pixels influencing the center pixel and therefore are closely related to scatterer spacing [27]. They have positive values when the respective neighbors and the central site tend to have similar intensities, otherwise they are negative.

**Conditional-Least-Squares**

Following a conditional-least-squares (CLS) approach - the variance between observed and expected intensities is minimized,

$$\hat{\psi} = \arg\min_{\theta} \sum_i (x_i^2 - E[X_i^2])^2, \theta_i = \{\alpha_i, \beta_{i,i'}\}$$  \hspace{1cm} (5.35)

$$E[X_i^2] = \frac{1}{2\omega_i} (\alpha_i + \sum_{i \neq i'} \beta_{i,i'} \log x_i),$$  \hspace{1cm} (5.36)

with $E[\cdot]$ being the expected value. It should be noted that CLS implicitly assumes a Gaussian variation around $x_i^2$.

The interaction feature vector $\psi$ (see Eq. 5.34) is obtained by least-squares employing Eqns. 5.37, with $C_i$ denoting the cliques

$$\theta_s = \left(1, \sum_{i \in C_1} \log x_i, \ldots, \sum_{i \in C_n} \log x_i\right)$$

$$\hat{\psi} = \left[\sum_{s \in S} \theta_s^T \theta_s\right]^{-1} \sum_{s \in S} \theta_s \left(2\frac{\omega}{\mu} x_s^2\right)$$  \hspace{1cm} (5.37)
5.6. Spatial Statistics Based Feature Descriptor for RF Ultrasound Data

Figure 5.6: Scheme of US MRF model - from right to left (a) Sparse grid placed on ultrasound image; Example patch $\mathcal{P}$ with sub-grid and grid points $\mathcal{S}$ (filled ♦ symbols); second-order neighborhood of a sub-grid point (b) Interaction parameters and the associated neighborhood relation.

Figure 5.7: US human neck; left training image, right test image.

Maximum-Pseudo-Likelihood

Another possibility of deriving the MRF parameters is by computing the pseudo-likelihood [21], which is the product of the conditional likelihood terms,

$$ P(\theta) \approx \prod_{i \in \mathcal{S}} P(x_i \mid x_{\mathcal{N}_i}; \theta) $$

$$ \hat{\psi} = \arg\max_{\theta} P(\theta), \theta_i = \{\alpha_i, \beta_{i,i'}\} $$

However, the pseudo-likelihood is just an approximation of the real likelihood. The actual optimization, computation of the interaction parameters, is then performed using a direct search algorithm e.g. simplex.

5.6.4 Training Distance Function

To cope with the different variation statistic among the components we computed a distance based on the learned distance - DistBoost algorithm [89, 90, 88]. For this purpose,
out of a sequence of ultrasound images, one was selected and used for supervised-learning. For testing, another image from the sequence was taken - see Fig. 5.7. Specifically, \( N \) patches of size \([w, h]\) are extracted from the training image, and from these patches feature vectors \( x \in \mathbb{R}^{w+h+n+1} \) are constructed from \( w+h \) MRF parameters vectors \( \psi \in \mathbb{R}^{n+1} \). These form a set of feature vectors \( X = \{x_i\}_{i=1}^N \). Various texture patches were manually segmented in the training image and labeled \(-1, 1, *\). The labels form a label set \( Y = \{y_i\}_{i=1}^N \) corresponding to \( X \), where \( y_i \in \{-1, 1, *\} \). The numbers 1 and \(-1\) denote two distinct classes, while * denotes an unassigned point. Thus two equivalence sets are formed of \( N_p \) pairs of positive labeled points \( \{(p_1^j, p_2^j)\}_{j=1}^{N_p} \), and \( N_n \) pairs of negative labeled points \( \{(n_1^j, n_2^j)\}_{j=1}^{N_n} \).

5.6.5 DistBoost

DistBoost, introduced in Hertz et al. [89] and used in various applications in [90] and Hertz [88], is an algorithm which can learn highly non-linear distance functions. DistBoost is a binary margin-based classification algorithm: if pairs of points are labeled 0 they are from the same class and 1 signifies that they are from different classes, then the margin (a positive scalar) between classes can be interpreted as a distance function. These are equivalence constraints on the product space \( A \times A \), where \( A \) is the original data domain. DistBoost also, in contrast to Support Vector Machines (SVM) and boosting methods, uses unlabeled data.

The DistBoost algorithm iteratively trains weak learners using of the constrained Expectation Maximization (cEM) algorithm [192]. These weak learners are then linearly combined to form a strong learner. The weak learner finds plausible partitions in the product space which comply with the equivalence constraints. Unlabeled data, denoted *, is incorporated in the learning process by a modified version of the AdaBoost with intervals algorithm [185].

A Gaussian Mixture Model (GMM) is given by \( p(x \mid \Psi) = \sum_{l=1}^{M} \alpha_l p(x \mid \theta_l) \), where \( \Psi = \{\alpha_k, \theta_k; k = 1, \ldots M\} \), \( \alpha_l \) is the mixture weight of component \( l \), \( \theta_l \) is the parameter of component \( l \) and \( M \) the number of components. Let \( X = \{x_i\}_{i=1}^N \) be a set of iid sample points. The constraints on \( X \) are indexed by \( \{(p_1^j, p_2^j)\}_{j=1}^{N_p} \) for data points labeled 1 and \( \{(n_1^j, n_2^j)\}_{j=1}^{N_n} \) for points labeled \(-1\).

The GMM is now estimated by constrained EM (cEM). Assume that here is a hidden variable \( h_i \) for each \( x_i \), denoting the class assignment of each data point. Let \( p(X, H \mid \Psi, \Omega) = \frac{1}{Z} \prod_{i=1}^n \alpha_{h_i} p(x_i \mid \theta_{h_i}) \prod_{j=1}^{N_p} \delta_{h_1, h_2}^{r_j} \prod_{k=1}^{N_n} (1 - \delta_{h_1, h_2})^{k_k} \), \((5.40)\)

where \( H \) are is the set of hidden variables and \( \Omega \) are the set of equivalence constraints. The conditional EM algorithm maximizes the data likelihood, which is the marginal distribution of \((5.40)\) w.r.t. \( H \). The sample size in the cEM estimation is modified in
5.6. Spatial Statistics Based Feature Descriptor for RF Ultrasound Data

order to take into account point weights: each points is replicated \( w_i \cdot N_v \) times in the sample size, and thus the virtual sample size is \( N_v \).

From the fitted mixture model, obtained by cEM estimation, a weak distance function \( h_t : X \rightarrow [-1, 1] \) is constructed. Let the MAP assignment of point \( x_i \) be

\[
p_{\text{MAP}}(x_i) = \arg\max_m p(h_i = m | x_i, \Theta). \tag{5.41}
\]

The weak distance function \( h_t : X \rightarrow [0, 1] \) is defined as

\[
h_t(x_i, x_j) = \frac{1}{2} \left( 1 - \tilde{h}_t(x_i, x_j) \right), \tag{5.42}
\]

where

\[
\tilde{h}_t(x_1, x_2) = \begin{cases} +p_{\text{MAP}}(x_1) \cdot p_{\text{MAP}}(x_2) & \text{if } MAP(x_1) = MAP(x_2) \\ -p_{\text{MAP}}(x_1) \cdot p_{\text{MAP}}(x_2) & \text{if } MAP(x_1) \neq MAP(x_2) \end{cases}. \tag{5.43}
\]

Next, the scaling parameter \( \alpha_t \) for the weak distance function is calculated

\[
\alpha_t = \frac{1}{2} \log \left( \frac{1 + r_t}{1 - r_t} \right) \tag{5.44}
\]

where

\[
r_t = \sum_{x_{i1}, x_{i2}, y_i = \pm 1} W_{i1i2} y_i \tilde{h}_t(x_{i1}, x_{i2}) > 0 \tag{5.45}
\]

Update weighting

\[
W_{i1i2}^{t+1} = \begin{cases} W_{i1i2}^t \exp(-\alpha_t y_i \tilde{h}_t(x_{i1}, x_{i2}), & y_i \in \{-1, 1\} \\ W_{i1i2}^t \exp(-\alpha_t), & y_i = * \end{cases} \tag{5.46}
\]

Normalize weighting

\[
W_{i1i2}^{t+1} = \frac{W_{i1i2}^{t+1}}{\sum_{i1i2} W_{i1i2}^{t+1}} \tag{5.47}
\]

Weights for cEM are updated through marginalization of \( W_{i1i2}^{t+1} \):

\[
w_{k}^{t+1} = \sum_j W_{k_j}^{t+1} \tag{5.48}
\]

The final output is a distance function \( D(x_i, x_j) = \sum_{t=1}^{T} \alpha_t h_t(x_i, x_j) \).

Summarizing the algorithm:
Algorithm 5 DistBoost Algorithm

1: **Input:** Data points \((x_1, ..., x_n), x_k \in A\)
2: **Set of equivalence constraints** \((x_{i_1}, x_{i_2}, y_i), y_i \in \{-1, 1\}\).
3: Unlabeled pairs of points: \((x_{i_1}, x_{i_2}, y_i = *)\).
4: **Initialize:** \(W_{i_1, i_2}^1 = 1/(n^2), i_1, i_2 = 1, ..., n\) (pair-wise data points weights)
5: \(w_k = 1/n, k = 1, ..., n\) (single data point weights)
6: **for** \(t = 1 \rightarrow T \) **do**
7: Perform cEM over data points in A weighted with \(w_k\) using equivalence constraints.
8: Generate \(h_t : A \times A \rightarrow [0, 1]\) by Eq. 5.42
9: Generate scaling parameter \(\alpha_t\) for \(h_t\) by Eq. 5.44
10: Update and normalize weights \(W_{i_1, i_2}^{t+1}\) by Eq. 5.46 and Eq. 5.47
11: Compute weight \(w_k^{t+1}\) by Eq. 5.48
12: **end for**
13: **Output:** \(D(x_i, x_j) = \sum_{t=1}^{T} \alpha_t h_t(x_i, x_j)\)
5.6. Spatial Statistics Based Feature Descriptor for RF Ultrasound Data

Figure 5.8: Texture data for training DistBoost - consisting of synthesized pattern from real US data.

Figure 5.9: Distribution of the interaction parameters for the second-order neighborhood within a homogenous texture.

5.6.6 Results

In order to analyze the feature descriptor we computed a feature image (size: 206 × 40 × 5) for an ultrasound image of the human neck\(^1\) (size: 2080 × 256) - see Fig. 5.6 & 5.7. The image was recorded using a linear transducer, avoiding the issue of locally varying neighborhood, e.g. prominent in convex or phased array transducers. The descriptive power of the feature description is visualized using distance maps. Thereby a designated feature is chosen in a characteristic area and compared to all other feature vectors - see Fig. 5.5 and 5.7.

5.6.7 Discussion

We learned our metric on synthesized patches with distributions acquired from the image data, see Fig. 5.8. Although sufficient for our purposes, it would have been desirable to train our metric directly on the image. This would require a refinement of our MRF

---
\(^1\) courtesy of Ultrasonix Medical Corporation, Canada
model, to take into account the highly heterogeneous nature of real RF-data. From the comparably low computational complexity, calculating the interaction parameters using the CLS method seems attractive. The underlying local Gaussian variation assumption of CLS is generally not fulfilled due to e.g. heavy tailed distributions. In our case, however, estimates provided by CLS proved to be sufficiently regular in homogeneous areas. Parameters derived using MPL tend to showed even more regularity in homogeneous areas. It is this property that facilitates the learning of distance metric and the computation of distance maps. However, MPL optimization was quite time consuming. Employing other metrics such as Mahalanobis was considered, but was deemed inappropriate due to the non-Gaussian distribution of the parameters - see Fig. 5.9. An example distance map is shown in Fig. 5.5. It indicates high similarity in the proximity to the neighborhood of the comparison feature (marked with X). With increasing distance and growing pattern dissimilarity the feature distance grows. On the border regions, one can also find regions of similar distributions, which is correctly expressed by our feature descriptor with low distance. Furthermore, dissimilarity to the feature of interest was captured correctly as can be seen in regions of high intensity streaks, resulting in high distance peaks in the map.

It should be noted that we are not assuming that our proposed procedure is independent of specific machine settings (gain, frequency etc.). In fact, it is quite evident that there is a strong inherent dependency between machine settings and features. Therefore we have formulated a procedure for computing a feature descriptor by which one can learn tissue specific similarities for images obtained using a specific machine setting. For our future work, improvement of the generality of the model is envisaged.
5.7 Registration of RF Ultrasound Data Using Hybrid Local Binary Patterns

This section is based on the paper Registration of RF Ultrasound Data using Hybrid Local Binary Patterns[114].

The theoretical prerequisites are provided in:

- Sec. 1.2 (Physical Background)
- Sec. 1.3 (Transmission to Display of Ultrasound Echo)
- Sec. 2.1 (Statistical Modeling of Ultrasound Data)
- Sec. 5.3 (Local Binary Patterns)
- Sec. 5.5 (Registration Methods)

Given the noisy nature of US, using statistical measures for the sake of robustness seems evident. In order to improve the registration accuracy a hybrid approach is proposed that couples the global concept of distribution matching with a local one measuring texture patterns. We refer to this in the following as hybrid local binary pattern (HLBP). The two components of HLBP are based on the Gamma Hellinger distance metric, cf. Sec. 5.7.1, and Local Binary Patterns, cf. Sec. 5.3, resp. Additionally, we make use of US confidence maps for parameter estimation.

5.7.1 Gamma Hellinger Distance Metric

We use the Nakagami distribution [188, 146] for modeling the speckle distribution in envelope detected RF US data. The Nakagami distribution is closely related to the Gamma distribution, and so may be used in its stead, as seen in Eq. 2.23.

One option in distribution matching is the Hellinger distance metric. The Hellinger distance $H_{\text{gam}}$, cf. Eq. 2.8.1, between two probability Gamma distributions $F$ and $G$ is expressed by their associated density functions $f \sim f_{\text{gam}}(x|m_1, k_1)$ and $g \sim f_{\text{gam}}(x|m_2, k_2)$ as

$$H_{\text{gam}}(f, g) = 1 - \frac{\Gamma\left(\frac{m_1+m_2}{2}\right)}{\Gamma(m_1)\Gamma(m_2)} \cdot \left(\frac{k_1+k_2}{2}\right)^{-\frac{m_1+m_2}{2}} \cdot \left[\frac{m_1}{k_1}, \frac{m_2}{k_2}\right]^\frac{1}{2},$$

with $H_{\text{gam}}(f, g) \in [0, 1]$. We apply Hellinger distance alone and in conjunction with Local Binary Patterns, cf. Sec. 5.7.3, in registration of envelope detected US data, which to our knowledge, is a novel approach.
5.7.2 Statistics-based Membership Function

Unlike the original FLBP [99], which defines the membership function as a ramp, we propose a non-linear function associated with the underlying statistical properties of data. Let

\[ f_{nak}(x \mid \mu, \omega) = \frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\omega^\mu} \exp\left(-\frac{\mu}{\omega}x^2\right), \forall x \in \mathbb{R}_+ \]  \hspace{1cm} (5.50)

be the pdf of the Nakagami distribution with \( \mu, \omega \) the shape and scale parameters, resp. Its corresponding cdf is

\[ F_{nak}(x \mid \theta) = \frac{\gamma(\mu, \frac{x^2}{\omega})}{\Gamma(\mu)}, \]  \hspace{1cm} (5.51)

with \( \theta = \{\mu, \omega\} \) and \( \gamma \) being the incomplete gamma function. Following the idea of FLBP, a membership function \( m_j(x) \) is defined, denoting the confidence of a class association \( j \in \{0, 1\} \) given intensity \( x \). Membership symmetry is assumed, i.e., \( m_0(x) = 1 - m_1(x) \). However, as opposed to FLBP, the membership function \( m_1 \) is non-linear, given by

\[
m_1(x_i) = \begin{cases} 
1, & x_i > F_{nak}^{-1}(K + \epsilon|\theta), \\
\frac{F_{nak}(x_i|\theta) - K}{2\epsilon}, & x_i \in (x_{centr}, F_{nak}^{-1}(K + \epsilon|\theta)], \\
\frac{K - F_{nak}(x_i|\theta)}{2\epsilon}, & x_i \in [F_{nak}^{-1}(K - \epsilon|\theta), x_{centr}], \\
0, & x_i < F_{nak}^{-1}(K - \epsilon|\theta) 
\end{cases} \]  \hspace{1cm} (5.52)

where \( K = F_{nak}(x_{centr}|\theta) \), and \( x_{centr} \) and \( x_i \) denotes the intensity at the center site and site \( i \), corresp. In the fuzzy region \([F_{nak}^{-1}(K - \epsilon|\theta), F_{nak}^{-1}(K + \epsilon|\theta)]\), as illustrated in

**Figure 5.10:** The threshold parameter \( \epsilon \) defining a probability interval around the central site intensity value.
Fig. 5.10, the membership function is the normalized cumulative within class probability. As $\epsilon$ increases, more fuzziness is assumed, and thus more noise is compensated for. Generally, the influence of $\epsilon$ is quite data dependent. One possibility is to define it as a combination of a probability threshold $T$ with a confidence value $C(x)$, s.t. $\epsilon = T \cdot C(x)$ and $x$ is a spatial index. To approximate $C$ we employ a Confidence Map (see Sec. 5.7.3) corresponding to the US image. The underlying idea is that low confidence regions should be compensated in terms of increased fuzziness. See Fig. 5.2 for a visualization how the magnitude of the threshold $\epsilon$ affects the feature histogram.

5.7.3 Confidence Maps

We formulate pixel-wise confidence of an US image as a random walk problem. The solution to this is based on the algorithm proposed for image segmentation [77]. There a multi-label image segmentation is obtained by the analytic computation of the probabilities for random walks reaching user-defined image labels. For the confidence estimation we are interested in the probabilities of random walks reaching the transducer elements under US specific constraints. More specifically, the random walks behavior is adjusted by modifying the graph Laplacian [109] to model US transmission, beam-width, and depth-dependent attenuation. Subsequently, the analytic solution expresses the probability of US energy reaching a point in the RF data domain.

In Karamalis et al. [109] confidence maps $C$ are computed for Intravascular Ultrasound RF to emphasize uncertainty and detect lack of acoustic energy in regions of interest. In this work they are utilized for a different application, namely, a new similarity measure for mono-modal US registration.

We now give a more detailed description of the Random-Walk algorithm, and the method proposed by Karamalis et al. [109]. In the original Random-Walk approach the image is modeled as an undirected graph $G = (V, E)$, where $V$ are the set of pixels and $E$ the set of edges between pixels. Thus in this context the pixels are viewed as nodes. $e_{ij}$ is the edge between pixel $v_i$ and pixel $v_j$. To each edge $e_{ij}$ a weight $w_{ij} > 0$ is assigned; this weight is based on intensities in a local region of the image. The weight gives the likelihood of a random walker crossing the specific edge.

In [77], a Gaussian weighting is used, i.e.,

$$w_{ij} = \exp \left[ -\beta (g_i - g_j)^2 \right],$$

where $g_k$ denotes pixel intensity at node $k$ and $\beta$ is scalar weight.

Each unlabeled node, i.e. all nodes which are not seed points, is assigned a $K$-dimensional vector. The $k$:th element in the vector denotes the probability of a random walker, starting at the unlabeled node, reaching the $k$:th seed point. The segmentation is achieved by assigning each pixel with the seed class of highest probability.

As outlined in [77], the solution to the random walk problem is equivalent to the combinatorial Dirichlet problem, which can be solved analytically.
The random walk problem is solved using a combinatorial Laplacian matrix \([59]\) or weighted Graph Laplacian \([200]\).

The Laplacian matrix is defined as

\[
L_{ij} = \begin{cases} 
  d_i & \text{if } i = j \\
  -w_{ij} & \text{if } v_i \text{ and } v_j \text{ adjacent nodes} \\
  0 & \text{otherwise}
\end{cases}, \tag{5.54}
\]

where \(d_i = \sum_j w_{ij}\).

The elements on the diagonal correspond to the accumulated weights of all connected neighbors. Adjacent nodes are associated with negative weight, therefore giving affinity to the neighbor. Non-connected nodes are represented with a zero.

An alternative construction of the Laplacian matrix is possible as \(L = A^T C A\), via the common graphic incidence matrix \(A\) in combination with diagonal matrix \(C\), consisting of the edge weights, with the elements of \(A\) defined as

\[
A_{e_{ij} v_k} = \begin{cases} 
  +1 & \text{if } i = k \\
  -1 & \text{if } i = j \\
  0 & \text{otherwise,}
\end{cases} \tag{5.55}
\]

for every node \(v_k\) and edge \(e_{ij}\).

In the process of solving for the K-tuple probabilities, the matrix \(L\) is decomposed in blocks, followed by an rearrangement, yielding

\[
L = \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}, \tag{5.56}
\]

where \(M\) and \(U\) correspond to marked (seed points) and unmarked nodes.

Finally, the desired probabilities are obtained by solving the system of linear equations given as

\[
L_U x_U = -B^T x_M, \tag{5.57}
\]

where \(x_U\) denotes the unknown probabilities for the unmarked nodes and \(x_M\) the known unit probabilities of the seeds.

For the confidence/uncertainty estimation of ultrasound imagery, the random walk problem is modified, in \([109]\), in order to incorporate ultrasound specific constraints. Thereby the notion of seed points as virtual transducer elements at the beginning of each scanline is employed. Consequently, the random walk formulation models the problem of obtaining the probability of a random walker starting from a pixel arriving at the
transducer. However, the weights of the Graph Laplacian $L$ are fundamentally different than in the original approach due to ultrasound specific properties.

For computing confidence estimates for ultrasound images, virtual transducer elements are defined at the beginning of the scanline, thus typically at the top of the ultrasound image. The virtual transducer elements are the seed points in the random walker formulation. The weights $w_{ij}$ need to be reformulated in order to take into the specific characteristics of the ultrasound signal.

Image intensity is not constant, it dissipates/attenuates as the signal moves away from the transducer elements, see Sec. 1.2. Thus the likelihood of pixel reaching the virtual transducer element should decrease with increasing distance, to model attenuation.

A transmitted ultrasound signal does not travel in a straight line through the medium, due to e.g. reflection and scattering. Ultrasound travels along a narrow beam, and thus this should be taken into account in the model, while allowing deviations in horizontal and vertical direction. The width of the scanline is represented by a decrease in likelihood as the pixel moves away from the center of the scanline. In vertical direction the likelihood decreases as the pixel moves away from the virtual transducer element. To achieve this the weights are defined as

$$w_{ij} = \begin{cases} 
w_{ij}^H & \text{if } i, j \text{ adjacent and } e_{ij} \in E_H \\
w_{ij}^V & \text{if } i, j \text{ adjacent and } e_{ij} \in E_V \\
0 & \text{otherwise} 
\end{cases}$$

(5.58)

$$w_{ij}^H = \exp(-\beta((g_i \exp(-\alpha x_i) - g_j \exp(-\alpha x_j))^2 + \gamma))$$

(5.59)

$$w_{ij}^V = \exp(-\beta((g_i \exp(-\alpha x_i) - g_j \exp(-\alpha x_j))^2)),$$

(5.60)

where $E_H, E_V$ are the edges along the horizontal and vertical graph direction, respectively.

The adaptation of the model with respect to ultrasound physics becomes more intuitive when discussed in terms of potential theory (as described in [77]). In this respect, the beam-width is represented by the $\gamma$ parameter that simultaneously acts as a penalty on the random walks, reducing the probability of a random walk moving in scanline perpendicular direction.

Specifically, as mentioned above the horizontal move probability decays with distance to scanline is implicitly encoded into this formulation due to the inherent accumulative increase in circuit resistance, whereby the additive term acts as resistor.

In the model high intensity gradients represented by $(g_i \exp(-\alpha x_i) - g_j \exp(-\alpha x_j))^2$ resulting in proportionally reduced energy transmission depending on the echo amplitude. This modified random walk approach [109] yields a confidence map for a given ultrasound image.
CHAPTER 5. Feature Descriptor and Registration

Figure 5.11: Confidence map (left) and the corresp. RF image (right).

Hybrid Similarity Measure

A common approach to measuring the similarity between LBPs is histogram intersection

\[ D_{X,Y} = \sum_i \min(X_i, Y_i) , \tag{5.61} \]

where \( X_i \) is the count of the \( i \)-th bin of histogram \( X \). However, standard histogram intersection is prone to yield several local minima. In order to avoid this, we endow the standard histogram intersection with a component measuring the statistical similarity of distributions. This follows the notion that patterns, in the two patches of intensities being considered, should be considered as relevant where the underlying distributions significantly exhibit high similarity. Statistical similarity alone on the other hand is globally precise (rough scale), but is locally (fine scale) imprecise. However, in combination with, the proposed statistics-based FLBP in Sec. 5.7.2, which has diametrically opposite behaviour, i.e. globally imprecise and locally precise (see Fig. 5.14), we achieve overall high reliability. The hybrid Hellinger weighted histogram intersection is defined as

\[ D_{\text{Hybrid}}(\Omega_1, T(\Omega_2)) = D_{X,Y} \cdot \exp(-\Delta \bar{H}_{\text{gam}}(f, h)) + \bar{H}_{\text{gam}}(f, h) , \tag{5.62} \]

where \( \Omega_1 \) is a patch in image 1, and \( T(\Omega_2) \) is a translated (by translation \( T \)) patch in image 2. \( D_{X,Y} \) is computed over the patches, that is the LBP feature histogram (see Sec. 5.7.2) \( X \) is computed on image 1 in the patch \( \Omega_1 \), and the same is done for \( Y \) over \( T(\Omega_2) \) in image 2. \( \bar{H}_{\text{gam}} \) is the average value of Hellinger distance \( H_{\text{gam}} \) computations over subpatches on \( \Omega_1 \) and \( T(\Omega_2) \), thus the pdf’s \( f \) are computed over subpatches on \( \Omega_1 \) and the pdf’s \( g \) are computed over \( T(\Omega_2) \). The influence radius, of the patch-averaged distribution distance \( \bar{H}_{\text{gam}} \), is denoted \( \Delta \).

In the measure \( D_{\text{Hybrid}} \), the LBP term \( D_{X,Y} \) is primarily used to enhance the accuracy of Hellinger, which determines a robust rough approximation of the location of the global optima.
Experiments

Experiments were performed on 5 human neck datasets each acquired with a 10 MHz linear array probe yielding RF envelope images of size $2048 \times 256$ pixels. Images were recorded with an Ultrasonix MDP machine. The registration dataset consists of pairs of images constituting a moving and a fixed image for the registration (translation, 2 degrees of freedom). Consequently, block matching (block size $20 \times 40$ pixels) was performed at 28 regular spaced points on the image domain, yielding similarity maps. See Fig. 5.12 for a schematic visualization of the block matching process. For each block 20 registration
runs were performed with random initial start points to estimate the susceptibility of metrics towards building local minima. Manual alignment of each dataset pair served as ground truth. The following methods were used for the registration test: SSD (Sum-Of-Squared Differences), SSDNAK (SSD on the shape parameter of Nakagami images), NCC (Normalized Cross-Correlation), FLBP (threshold $\epsilon = 30$, confidence weighted), Hellinger and HLBP (threshold $\epsilon = 0.35$, confidence weighted). Finally, the global similarity was computed by accumulating estimates from all blocks on each dataset - see Table 5.1 for median pixel errors. For registration performance evaluation, comparison results from all datasets are combined - see Fig. 5.13. Additionally, in order to assess the different similarity measures we extracted a patch in the moving image, and compute corresp. similarity maps - see Fig. 5.14 for an example.

Table 5.1: Median errors (pixels) of random registration study for various datasets and similarity metrics.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SSD</th>
<th>SSDNAK</th>
<th>NCC</th>
<th>FLBP</th>
<th>HEL</th>
<th>HLBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>20.1</td>
<td>23.0</td>
<td>26.1</td>
<td>28.1</td>
<td>4.7</td>
<td>3.5</td>
</tr>
<tr>
<td>#2</td>
<td>16.1</td>
<td>19.8</td>
<td>26.3</td>
<td>28.2</td>
<td>7.4</td>
<td>6.3</td>
</tr>
<tr>
<td>#3</td>
<td>20.1</td>
<td>12.1</td>
<td>26.2</td>
<td>27.2</td>
<td>7.9</td>
<td>7.4</td>
</tr>
<tr>
<td>#4</td>
<td>17.9</td>
<td>16.5</td>
<td>25.0</td>
<td>24.0</td>
<td>8.5</td>
<td>7.4</td>
</tr>
<tr>
<td>#5</td>
<td>24.4</td>
<td>19.9</td>
<td>27.2</td>
<td>3.3</td>
<td>1.4</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Figure 5.14: Similarity maps from left to right: SSD, SSDNAK, NCC, FLBP, HEL, and HLBP. Circle: ground truth optimum; Cross: optimum in similarity map.
5.8 Modeling of Multi-View 3D Freehand Radio Frequency Ultrasound

This section is based on the paper *Modeling of Multi-View 3D Freehand Radio Frequency Ultrasound* [116].

The theoretical prerequisites are provided in:
- Sec. 1.2 (Physical Background)
- Sec. 1.3 (Transmission to Display of Ultrasound Echo)
- Sec. 1.4 (3D Freehand Ultrasound)
- Sec. 1.5 (Reconstruction of 3D Data)
- Sec. 1.7 (Applications in Neurology)
- Sec. 2.1 (Statistical Modeling of Ultrasound Data)
- Sec. 2.2 (Mixture Models)
- Sec. 2.4 (Maximum Likelihood Estimators)
- Sec. 2.5 (EM Algorithm)
- Sec. 2.6 (EM-Gamma)
- Sec. 2.8 (Similarity Measures)
- Sec. 5.5 (Registration Methods)

5.8.1 Method

Freehand 3D RF Data

Due to the spatial relationship of the data, the RF image requires disintegration into individual scanlines such that the reconstruction process, following [221], becomes ray-based - see Fig. 5.15. Beside the intensity data we also record geometric information such as viewpoint and direction. This additional data is used in a follow-up processing step. The distribution of the envelope of the RF signal, resulting from backscattered tissue echo, has been shown to be modeled, in a simple and versatile way, by the Nakagami distribution [188]. Thus we assume all intensities in the RF envelope image to follow a Nakagami distribution

\[
    f_{\text{nak}}(x|\mu, \omega) = \frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\omega^\mu} \exp\left(-\frac{\mu}{\omega}x^2\right) \quad \text{s.t.} \quad \forall x \in \mathbb{R}_+, \quad (5.63)
\]

with \(\mu, \omega\) denoting the shape and scale parameters, respectively.
Mixture Model Motivation

Since ultrasound is highly view-dependent [87], it is desirable to incorporate this property when modeling back-scatter. However, we are not interested in the individual backscatter intensities, but in the distribution within a small finite volumetric element (voxel) w.r.t. views of a data point \( x \), i.e.

\[
p(x) = \int_{\mathcal{D}(x)} p(x|\phi)p(\phi) d\phi , \tag{5.64}
\]

where \( \mathcal{D}(x) \) is the set of all possible viewing cones \( \phi \) of \( x \). The distribution \( p(x) \) is approximated by a FMM of \( K \) Nakagami densities [56],

\[
p(x) \approx \sum_{k=1}^{K} p(\phi_k)p(x|\phi_k) = \sum_{k=1}^{K} w_k f_{\text{nak}}(x|\mu_k, \omega_k) \quad \text{s.t.} \quad \sum_{k=1}^{K} w_k = 1 , \tag{5.65}
\]

where the distribution \( p(\phi_k) \) of the \( k^{th} \) cone is represented by a mixture weight. The \( K \) viewing cones are assumed approx. evenly spaced around the object of interest. We use 2 cones, i.e. a bilateral view, in our experiments, see Sec. 5.8.2. See Fig. 5.15 for illustration of viewing cones of the midbrain with beams originating at different skull positions. A popular choice for FMM estimation is the Expectation-Maximization (EM) algorithm [53]. However, as the EM algorithm can potentially overfit the data and is also quite flexible in component modeling, we do not instantiate it on the pooled data from all views. Rather, we instantiate individual mixture estimations within geometrical subspaces obtained from each view (each containing viewing cones) that were recorded during the acquisition process. Altogether this yields robust component estimation, s.t.

\[
p(x) = \sum_{i=1}^{N} \sum_{k=1}^{K_i} w_{ik} p(x|\phi_{i,k}) = \sum_{i=1}^{N} \sum_{k=1}^{K_i} w_{ik} f_{\text{nak}}(x|\mu_{i,k}, \omega_{i,k}) . \tag{5.66}
\]
Specifically, within each view \( i \in N \) we determine the number of components \( K \leq K_{\text{max}} \), following the approach of Frayley and Raftery [69]. Here the Asymptotic Minimum Description Length (AMDL / BIC) principle selects the FMM with \( d \) free parameters, which minimizes the quantity \(-2 \log L + d \log n\), where \( L \) is the likelihood of data given model parameters and \( n \) the number of observations. This estimation process is performed voxelwise in the entire volume, yielding a mixture model representation for each voxel, where the number of components naturally varies from voxel to voxel.

**Registration**

For registration, we perform a voxel-wise distribution matching employing J-divergence in conjunction with a data fidelity term. This provides higher robustness compared to a pure intensity-based model, as was also observed in Ijaz et al. [100]. Considering a fixed volume (A) and moving volume (B), we seek the rigid transformation \( \hat{T} \) that yields optimal alignment between the two, s.t. \( \hat{T} = \arg \min_T D_{\text{PJD}}(A, T(B)) \) for the pseudo-distance

\[
D_{\text{PJD}}(C, D) = \arg \min_{i, j} \sum_{k=1}^Z J(f_C^k, f_D^j) \cdot e^{\lambda ((1-w_C^k)+(1-w_D^j))}.
\]

(5.67)

We refer to this pseudo-distance as *Pseudo-J-Divergence*. For each voxel \( k \in Z \) it takes the mixture components pair \( f_C^k \) and \( f_D^j \) from the two volumes \( C \) and \( D \), resp., to be registered with least J-divergence [107] times the exponentially scaled sum of corresp. mixture weights \( w_C^k, w_D^j \geq \tau \), where \( \tau \) is the min. weight. Lastly, \( \lambda \) is a parameter.

Jeffreys (J) divergence is also known as symmetric KL distance; KL distance is defined formally in Sec. 2.8. The pseudo-distance \( D_{\text{PJD}} \) does not satisfy the triangle inequality, but inherits symmetry and uniqueness from the J-divergence \( J(f, g) \). The exponential weight in (5.67) punishes distances formed from components with low mixture weights, since these components are assumed to be less descriptive of the underlying distribution. The J-divergence between two Gamma distributions \( f, g \in \mathcal{G}A \) is the sum of two non-symmetric KL distances with switched arguments, s.t.

\[
J(f, g) = \int \log \frac{f(x)}{g(x)} (f(x) - g(x)) dx = (\mu_a - 1)\Psi(\mu_a) - \log \omega_a - \mu_a - \log \frac{\Gamma(\mu_a)}{\Gamma(\mu_b)} + \mu_b \log \omega_b - (\mu_b - 1)(\Psi(\mu_a) + \log \omega_a) + \frac{\omega_a \mu_a}{\omega_b}.
\]

(5.68)

As seen by Eq. 2.23, the Gamma distribution may be used instead of Nakagami.

**Reconstruction**

Given the voxel FMM representation we can perform a novel type of reconstruction. See Sec. 1.5 for a review of methods for reconstruction of 3D data from 2D data. There-
fore a reference component is chosen from the mixture for each voxel, on basis of maximum mean intensity, although other approaches are conceivable. However, this criterion guarantees that no high intensity backscatter is missed during reconstruction. Note that artifacts, such as shadows, might require specific treatment, which is beyond the scope of this work. By optimizing the reference component parameters, i.e. minimizing the sum of upper bounds of geodesic distances to neighbours component parameters, on the manifold $G$ of Gamma model parameters, smoothness of the reconstructed volume is achieved. Thus, given a point $\theta_a = (\mu_a, \omega_a) \in G$, the geodesic distance $D_{geo}$ to a locally neighbouring point $\theta_b = (\mu_b, \omega_b) \in G$ is bounded s.t.

$$D_{geo}(\theta_a, \theta_b) \leq \left| \frac{d^2 \log \Gamma}{d \mu^2}(\mu_b) - \frac{d^2 \log \Gamma}{d \mu^2}(\mu_a) \right| + |\mu_a \cdot \log \frac{\mu_a \omega_a}{\mu_b \omega_b} - \mu_a \cdot \log \frac{\mu_a \omega_a}{\mu_b \omega_b}| = D_{geo}^{\text{bnd}}(\theta_a, \theta_b),$$

(5.69)

$\text{cf. Ch. 7, [5] and [58].}$ Applying (5.69), the reference distribution $\theta_{\text{ref}}$ is optimized by minimizing the sum of geodesic distances to all of its neighbors s.t.

$$\hat{\theta}_{\text{ref}} = \arg\min_{\theta_{\text{ref}}} \sum_{k \in N_{\text{ref}}} D_{geo}^{\text{bnd}}(\theta_{\text{ref}}, \theta_k),$$

(5.70)

keeping the neighbours $\theta_k \in \mathcal{N}_{\text{ref}}$ fixed, where $\mathcal{N}_{\text{ref}}$ defines the neighborhood of a reference voxel. This yields a spatially consistent image without over-smoothing or loss of detail in terms of highlights. Typically a few optimization steps are sufficient and allow for a fast reconstruction.

For data reconstruction within a voxel we apply a Gaussian-weighted (GW) [221] reconstruction scheme in order to increase homogeneity,

$$y_j = \frac{1}{Z} \sum_{i=1}^{N} x_i e^{-d_i^2/\sigma^2} \quad \text{s.t.} \quad Z = \sum_{i=1}^{N} e^{-d_i^2/\sigma^2},$$

(5.71)

yielding the reconstructed intensity $y_j$ at voxel position $j$. Here the intensities $x_i$ are sampled from the reference distribution, where $d_i$ are the distances from the voxel centroid that can be obtained by regression from the measured data.

5.8.2 Results

Registration

The 3D US RF freehand data used to validate our method was obtained using the optical tracking system NDI Spectra in conjunction with an Ultrasonix MDP US machine. For testing the registration performance multiple transcranial 3D volumes for several patients were acquired. An acquisition consists of bilateral scans; in doing so various sweeps from numerous possible views were obtained. Furthermore, each patient was recorded with a reference target rigidly attached to the head, which allows establishment of groundtruth
position between numerous volumes. The three-dimensional RF datasets were acquired using a phased-array probe with a frequency of 3.3 MHz and depth 14 cm. RF data is sampled at 40 MHz. Each 3D RF data set was built from approximately 4000 2D RF images (2000 images from each side of the skull), each having a resolution of 3648×96 pixels. Acquisition time is 2-4 min (1800 images/min unilateral scan). Volumes were reconstructed with isotropic voxel size of 0.65 mm (mean 1400 samples/voxel). For the mixture modeling we assumed \( N = 2 \) views (bilateral) as well as \( K_{max} = 2 \). This yielded a maximum of four mixture components per voxel and sufficiently modeled the data while avoiding overfits.

For evaluating the quality of the proposed approach, we performed registration by block matching, which is commonly used for US [164, 118]. For each patient two multi-view volumes were constructed for distinct US data. These two distinct volumes were then rigidly registered by taking 27 equally distributed blocks (each of size 6 × 6 × 6 voxels) for matching, within each multi-view volume. For each of 10 runs we randomly displace, with initial deviation of ±6 cm, the moving subvolume/block from the ground truth position (obtained via head target), in each spatial direction from its ground truth position. This was followed by registration by block-matching at each position using state of the art similarity metrics for US, aligning the moving and the fixed volume. In the case of global registration, the result from each block is accumulated. Parameters for our distance metric were set to \( \lambda = 2 \) and \( \tau = 0.3 \). In spite of rigidity, registration of transcranial brain US brain data is quite challenging due to the relative low SNR as a result of variable transmission through skull bone and the low transducer frequency required. Nevertheless, the proposed pseudo-J-Divergence (PJD) yields up to 15% better registration results compared to Normalized Cross-Correlation (NCC) and Sum of Squared Differences (SSD) [164, 118], as can be seen in Tab. 5.2. Intensity volumes were created using GW [221]. The difference in median between methods is statistically significant. The Hodges-Lehmann 95% confidence interval for the difference in median error be-

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SSD</th>
<th>NCC</th>
<th>PJD</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>3.2 ± 1.3</td>
<td>3.0 ± 1.3</td>
<td>2.7 ± 1.1</td>
</tr>
<tr>
<td>#2</td>
<td>2.7 ± 1.3</td>
<td>2.7 ± 1.2</td>
<td>2.4 ± 1.2</td>
</tr>
<tr>
<td>#3</td>
<td>3.2 ± 1.2</td>
<td>3.3 ± 1.1</td>
<td>2.9 ± 1.1</td>
</tr>
<tr>
<td>#4</td>
<td>3.7 ± 1.1</td>
<td>3.4 ± 1.1</td>
<td>3.0 ± 1.1</td>
</tr>
<tr>
<td>#5</td>
<td>3.4 ± 1.1</td>
<td>3.2 ± 1.1</td>
<td>2.9 ± 1.1</td>
</tr>
<tr>
<td>#6</td>
<td>3.3 ± 1.2</td>
<td>3.4 ± 1.2</td>
<td>2.9 ± 1.1</td>
</tr>
</tbody>
</table>
between PJD and NCC is (-0.25, -0.08), and (-0.3, -0.12) for PJD and SSD. The corresp. Mann-Whitney U-test yields p-values $0.14 \cdot 10^{-3}$ and $0.2 \cdot 10^{-6}$, resp. Obtained results are quite close to technical possible limits due to the errors that are propagated through the processing chain. In this respect, US calibration affects the accuracy significantly, depending largely on the approach used. In our setup, calibration was performed using a single wall phantom, yielding an error between 1-2 mm [95, 135].

Reconstruction

As reconstruction methods are difficult to compare due to the highly subjective nature of US image analysis, we will elaborate on the obvious differences between a state of the art method (GW - increased smoothness favourable for segmentation and registration [221]) and our approach as well as discuss the potential implications. Reconstructions following the geodesic approach yielded more coherent and homogeneous images - see Fig. 5.17, 5.18, 5.19 and 5.16. Note that for Gaussian reconstruction we chose $\sigma = 1$ in all cases. Checker-board and radial artifacts disappear, which were visible in the GW approach due to the multi-view mixture nature of the signal. Additionally, the mid-brain area exhibits sharper edges using the proposed approach. Regions that are clinically relevant for classification such as the SN, where the level of hyperechogenicity is a risk-assessment
5.9. Conclusion

1) Feature Descriptor

We have presented a feature descriptor, based on a Markov random field (MRF) texture model. The proposed approach combines global data statistics in terms of a maximum-likelihood-estimated (MLE) distribution with local pattern characteristics employing MRF interaction parameters. This combining approach facilitates the encoding of the underlying nature of the ultrasound envelope data and therefore represents a powerful feature descriptor. Applicability and performance is showcased on RF data from a human neck.
2) Registration Methods

We have studied two approaches for registration of ultrasound data. First we have examined registration of 2D ultrasound data, by proposing similarity metrics Nakagami Hellinger and Hybrid Local Binary Patterns (HLBP). Results of the experiments show that HLBP greatly outperforms standard methods, such SSD and NCC, and also improves upon the result obtained by use of Hellinger distance metric. The robustness of HLBP is due to its hybrid local and global strategy. On the one hand, LBP and its variants are able to take into account very fine textural features in data, which can discriminate small local differences between images, and are thus suitable for small-scale alignments. On the other hand, when differences are large between images, fine textural differences are not relevant. Here a purely statistical (dis-)similarity measure like e.g. Hellinger distance, is a better candidate for detecting global differences between images.

We have also explored registration of 3D freehand ultrasound data. There exist a
multitude of applications where doctors could benefit from three dimensional ultrasound providing better judgment, due to the extended spatial view. 3D freehand US allows acquisition of images by means of a tracking device attached to the ultrasound transducer. Unfortunately, view dependency makes the 3D representation of ultrasound a non-trivial task. To address this we model speckle statistics, in envelope-detected radio frequency (RF) data, using a finite mixture model (FMM), assuming a parametric representation of data, in which the multiple views are treated as components of the FMM. The proposed model is show-cased with registration, using an ultrasound specific distribution based pseudo-distance, and reconstruction tasks, performed on the manifold of Gamma model parameters. Example field of application is neurology using transcranial US, as this domain requires high accuracy and data systematically features low SNR, making intensity based registration difficult. In particular, 3D US can be specifically used to improve differential diagnosis of Parkinson’s disease (PD) compared to conventional approaches and is therefore of high relevance for future application.
5.10 Future Work

1) Feature Descriptor
The influence of different neighborhood systems on the descriptive power should be investigated. The global part of the formulation is provided by an MLE, which is suboptimal, since a mixture model has the potential of providing much better fits to the underlying data. Substituting the MLE for a mixture representation is however non-trivial as this would violate the conditions of the auto-model, which require the distribution to be part of the exponential family, which is not the case for mixture models. Integrating a mixture model into the texture representation would be highly desirable. Potential fields of applications of this feature descriptor could be registration and segmentation; the specific application requires further study.

2) Registration Methods
Optimization of the neighborhood system of HLBP is a worthwhile venue of investigation, as initial experiments to be a source of potential major improvement. Extending the backscatter model to a mixture representation has the potential of greater accuracy.

In our work on 3D freehand US, the use of view-dependency has potential to facilitate several applications; specifically, results from a block-matching based rigid registration study suggest improvements in terms of accuracy compared to conventional similarity metrics. Moreover, image reconstruction promises to be an interesting domain of application. Further applications will be studied e.g. multi-view bone imaging where, as a result of varying reflectance properties of bone, muscle, and ligaments, different backscatter scenarios are encountered.

The choice of the maximum mean component, in our mixture representation of a voxel, for use in reconstruction was made to capture the highly echogenic area of the substantia nigra, in our application of Parkinson’s disease. In general for Parkinson's disease as well as other applications this choice certainly requires further evaluation. Other applications may warrant other choices of representative component for a voxel.

The similarity measure employed in this work, Pseudo-J-Divergence (PJD), does not fulfill the triangle inequality, which certainly is not theoretically desirable. Further effort may be put into modifying PJD so that it becomes a true metric, and also investigation on the detrimental effects of this lacking propriety in further validation.

The use of EM may be too restrictive in terms of computational complexity, when working with large sets of 3D data. Here implementations employing the GPU (Graphical Processing Unit) could radically decrease these costs.
Final Remarks

In segmentation of the LV endocardium, our method *Bayesian Probability Maps* has some interesting properties: sampling from the position distribution provides a stability which we have not found in our active contour models. Being Bayesian it is also very flexible in would easily admit adjustments in the prior structure. In addition the model requires little in the form of initialization, which is not the case in the active contour models. It is also our most extensively validated model. These benefits come at the cost of speed however, but we are confident that in an optimized version this will be a minor issue.

In registration our focus has been on similarity measures. Having first explored registration methods in 2D ultrasound registration and feature description, we most recently developed a method for 3D free US registration. Using geodesic interpolation on the Gamma manifold and then a statistical registration measure we achieved registration results which outperform the standard methods of NCC and SSD. These results are especially exciting since improvements in 3D freehand US registration are crucial for providing tools in aiding diagnosis of degenerative conditions such as Parkinson's disease.
Bibliography


Bibliography


178
Bibliography

[120] R. M. Lang, M. Bierig, R. B. Devereux, F. A. Flachskampf, E. Foster, P. A. Pel-
likka, M. H. Picard, M. J. Roman, J. Seward, J. S. Shanewise, S. D. Solomon,
K. T. Spencer, M. S. Sutton, and W. J. Stewart. Recommendations for Cham-
ber Quantification: A Report from the American Society of Echocardiography's
Guidelines and Standards Committee and the Chamber Quantification Writing
Group, Developed in Conjunction with the European Association of Echocardi-
ography, a Branch of the European Society of Cardiology. *J Am Soc Echocardiogr*,

[121] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N.D. Cahill, and B. Scholkopf. Learn-

for Face Recognition. In *Image Processing (ICIP), 2010 17th IEEE International

2000.


[125] J. Lie, M. Lysaker, and X.-C. Tai. A binary level set model and some applications
to Mumford-Shah image segmentation. *Image Processing, IEEE Transactions on*,

[126] J. Lie, O.M. Lysaker, and X.-C. Tai. A variant of the level set method and applica-
tions to image segmentation. *Mathematics for Computation*, 75(255):1155–1174,

for Echocardiographic Image Segmentation. *Medical Image Analysis*, 7:529–537,
2003.

[128] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid Im-
age Registration Using Conditional Mutual Information. *Medical Imaging, IEEE


[130] J. B. MacQueen. Some Methods for Classification and Analysis of MultiVariate
Bibliography


Bibliography


183


Bibliography


186


