

# **Non-rigid Image Registration: Regularization, Algorithms and Applications**

Andriy Myronenko

A dissertation submitted to the faculty of the  
Department of Science & Engineering  
School of Medicine  
at Oregon Health & Science University  
in partial fulfillment of the  
requirements for the degree  
Doctor of Philosophy  
in  
Electrical Engineering

June 2010

The dissertation “Non-rigid Image Registration: Regularization, Algorithms and Applications” by Andriy Myronenko has been examined and approved by the following Examination Committee:

---

Xubo Song  
Associate Professor  
Thesis Research Adviser

---

Todd K. Leen  
Professor

---

Izhak Shafran  
Assistant Professor

---

David J. Sahn  
Professor

---

Xenophon Papademetris  
Assistant Professor  
Yale University

# Dedication

To my parents for their constant love and support.

# Acknowledgments

I would like to thank the following people who all contributed in some form to the success of this endeavor:

I want to thank my adviser, Xubo Song, for her continuous support, patience and encouragement throughout my PhD studies. Her technical advice, knowledge and expertise were essential to the completion of this dissertation. She taught me the lessons of rigorous academic research and greatly supported me in all of my undertakings.

I would also like to thank David J. Sahn for his support and clinical expertise in the echocardiography research. I am grateful for his optimism and belief in the success of our research. Special thanks to Todd K. Leen for his insightful comments and constructive criticisms at different stages of my research. I also would like to thank Eric A. Wan, Deniz Erdogan, Miguel Á. Carreira-Perpiñán and Muhammad Ashraf for many helpful discussions and suggestions.

Finally, I want to thank the remaining members of my thesis committee, Izhak Shafran and Xenophon Papademetris, for taking the time to review this work.

My parents for their love and support. Andriy and Olga: Thank you for always believing in me.



# Contents

<b>Dedication</b> . . . . .	<b>iii</b>
<b>Acknowledgments</b> . . . . .	<b>iv</b>
<b>Abstract</b> . . . . .	<b>xvi</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 What is Image Registration? . . . . .	1
1.2 Contributions of This Work . . . . .	3
1.3 Thesis Outline . . . . .	4
1.4 Publications . . . . .	5
<b>2 Non-rigid Image Registration Framework</b> . . . . .	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Similarity Measure . . . . .	9
2.2.1 Feature-based Registration . . . . .	9
2.2.2 Intensity-based Registration . . . . .	10
2.3 Transformation Model . . . . .	16
2.3.1 Parametric Approach . . . . .	16
2.3.2 Non-parametric Approach . . . . .	18
2.4 Regularization . . . . .	19
2.4.1 Theory . . . . .	19
2.4.2 Regularization in Parametric Transformations . . . . .	21
2.4.3 Regularization in Non-parametric Transformations . . . . .	23
2.4.4 Invertibility of the Transformation . . . . .	24
2.4.5 Regularization in Similarity Measures . . . . .	25
2.5 Groupwise Image Registration . . . . .	26
2.6 Conclusion . . . . .	27

<b>3</b>	<b>Image Registration by Minimization of Residual Complexity</b>	<b>28</b>
3.1	Introduction	28
3.2	Background	31
3.3	Method	33
3.4	DCT basis	37
3.5	Analysis of Residual Complexity	38
3.6	Implementation	41
3.7	Results	41
3.7.1	Elastix Configuration	41
3.7.2	Synthetic Experiment 1	42
3.7.3	Synthetic Experiment 2	45
3.7.4	Retina Images	45
3.7.5	Iris Images	46
3.7.6	3D Echocardiography	47
3.8	Non-rigid Stabilization of Time Lapse Videos	48
3.9	Discussion and Conclusion	52
<b>4</b>	<b>Point Set Registration: Coherent Point Drift</b>	<b>55</b>
4.1	Introduction	56
4.2	Previous Work	58
4.2.1	Rigid Point Set Registration Methods	58
4.2.2	Non-rigid Point Set Registration Methods	60
4.3	General Methodology	61
4.4	Rigid & Affine Point Set Registration	63
4.4.1	Related Rigid Point Set Registration Methods	65
4.5	Non-Rigid Point Set Registration	67
4.5.1	Regularization of the Displacement Function	67
4.5.2	Variational Solution	68
4.5.3	The Coherent Point Drift (CPD) Algorithm	69
4.5.4	Related Non-rigid Point Set Registration Methods	71
4.6	Fast Implementation	73
4.7	Results	75
4.7.1	Rigid Registration Results	76
4.7.2	Non-rigid Registration Results	79
4.7.3	Image Registration Results	83
4.8	Discussion and Conclusion	84

<b>5</b>	<b>Motion Estimation from 3D Echocardiography</b>	<b>88</b>
5.1	Introduction	88
5.2	Motion Estimation Methods	91
5.2.1	Block Matching Methods	91
5.2.2	Boundary Tracking Methods	93
5.2.3	Dense Tracking Methods	95
5.2.4	Regularization of the Estimated Myocardial Deformation	96
5.3	Quantitative Characteristics	99
5.4	Similarity Measure Based on Physics of Speckle Noise	100
5.4.1	Speckle Correlation and Blurring	101
5.4.2	Maximum Likelihood Motion Estimation	101
5.4.3	Independent Rayleigh Noise Assumption	103
5.4.4	Correlated Rayleigh Noise Assumption	103
5.4.5	Blurred Speckle	105
5.4.6	Analysis	107
5.5	Shape Constraints	109
5.6	Transformation Model	110
5.7	Optimization	111
5.8	Sequential Registration	112
5.9	Dynamic Constrains	112
5.10	Spherical Coordinates	113
5.11	Validation by Sonomicrometry	115
5.12	Experimental Results	116
5.12.1	Animal Preparation	116
5.12.2	Data Acquisition	117
5.12.3	LV Strain Analysis	118
5.12.4	LV Torsion Analysis	121
5.13	Discussion and Conclusion	123
<b>6</b>	<b>Conclusions and Future Directions</b>	<b>125</b>
6.1	Summary	125
6.2	Possible Extensions & Future Work	126
6.3	Conclusion	130
	<b>Bibliography</b>	<b>131</b>
<b>A</b>	<b>Upper Bound of the GMM Log-likelihood Function</b>	<b>148</b>

<b>B</b>	<b>Optimal Rotation Matrix . . . . .</b>	<b>150</b>
<b>C</b>	<b>Green's function corresponding to <math>\phi_{RT}(v)</math> . . . . .</b>	<b>153</b>
<b>D</b>	<b>Green's function corresponding to <math>\phi_{MCT}(v)</math> . . . . .</b>	<b>155</b>
	<b>Biographical Note . . . . .</b>	<b>157</b>

# List of Tables

2.1	Common regularization operators of the displacement field and their variational derivatives. . . . .	23
4.1	The rigid CPD registration time for naive (no FGT) and FGT implementations. The FGT-based implementation is significantly faster. . . . .	79
4.2	Registration time required for non-rigid registration of 3D bunny point sets. The time is shown when using only FGT of vector-matrix products, only low-rank matrix approximation of Gaussian kernel matrix or both. . . . .	82
5.1	Algorithmically estimated peak stain (ES strain) for different segments between the sono crystals in the apical, basal and in apical-to-basal levels. The measurements were averaged among the combinations of sono crystal within a particular level and among all scans. The decrease in the peak strain caused by LAD ligation is clear especially in the apical region. Also the reduction of preload due to controlled IVC occlusion led to the slight increase of the overall strain. . . . .	119
5.2	Mean absolute error between algorithm and groundtruth sonomicrometry strain. The error is shown for different levels (apical, basal, apical-to-basal and overall) against different states (baseline, LAD ligation and IVC occlusion). . . . .	120
5.3	The average peak torsion as estimated by the algorithm and by the sonomicrometry. . . . .	121

# List of Figures

3.1	Gray stripe registration experiment. . . . .	29
3.2	Plot of several similarity measures (RC, SSD, CC, MI, L1 and L2 norm gradient magnitude) with respect to the translation of the smaller image over the larger one in Figure 3.1. Zero translation corresponds to the accurate registration. . . . .	30
3.3	A comparison of the distributions of the DCT coefficients for the different spatial translations (pixels) of the images shown in Figure 3.1. Here, we computed each histogram from $\log(\mathbf{x}^2 + 1)$ , where $\mathbf{x}$ are the 50 largest DCT coefficients of the residual image for the given translation. The correct alignment (zero translation) corresponds to the maximum number of zero coefficients. By enforcing sparseness on the DCT coefficients of the residual image, the RC similarity measure achieves optimum at the correct spatial alignment. . . . .	38
3.4	Synthetic experiment 1. We register the source image (b) onto the reference image (a). (f,g) Intensity distortion fields added to the reference and source images respectively. (d) The RC registration result. RC performs well. (d,e,i,j) Typical performance of global NMI, local CC, local MI and Preprocessing+SSD respectively. NMI shows unsatisfactory result. Local CC and local MI performs satisfactorily with slight misalignments, which we highlighted with contours (e,j). Preprocessing+SSD approach produces overall satisfactory result, but with several misalignments of the smaller structures. . . . .	43
3.5	Registration performances of RC, local CC, local MI and Preprocessing+SSD. Both images were corrupted by additive intensity field (simulated by a mixture of $K$ random Gaussians) and scaled. RC performs comparably well, often better, with the other tested similarity measures. Local CC and local MI produce similar satisfactory results. Preprocessing+SSD approach showed the worst performance. . . . .	44
3.6	Transformation error with respect to the different parameter $\alpha$ settings. . .	45

3.7	Registration performances of RC, local CC, local MI and Preprocessing+SSD. Both images were corrupted by multiplicative intensity field. (simulated by a mixture of $K$ random Gaussians) and scaled. Artificial spatial and intensity distortions were reinitialized at every run. RC performs comparably well, often better, with the other tested similarity measures in almost all cases. . . . .	46
3.8	Registration of retina images. (a) reference image; (b) source image (taken 2 years prior); (c) composite view before registration; (d) composite view after registration; (e) registered source image; (f) composite view through contour overlap after the registration. The RC registration result is accurate; vessel structures are aligned despite intensity distortions and artifacts.	47
3.9	Registration of two iris images. Notice the intensity variation across the images as well as intensity artifacts including moving cells and nonconsistent vessel reflections. Our method accurately registers the images without any additional preprocessing. . . . .	48
3.10	Estimated motion of the LV contour found during the cardiac cycle (24 consecutive volumes). LV achieves the maximum contraction at volume 13, then the LV dilates (diastolic phase) up to the volume 22 and starts contracting again (systolic phase). The last plot shows the displacement fields on LV endocardium between end diastole (ED) and end systole (ES) volumes. We validate the accuracy of RC registration with groundtruth motion of implanted sonomicrometers. The RC registration-based estimated motion is accurate. . . . .	49
3.11	Typical translation compensated image frames from the first video and the average image. Notice local intensity inconsistency, blurring and moving cells. The average image has blurred boundaries and a low contrast, which indicates non-rigid image misalignments. . . . .	50
3.12	The registered image frames from the first video after applying the transformation found using the groupwise non-rigid registration. The average image has much sharper boundaries than prior to registration, which indicates accurate image alignment. . . . .	51
3.13	Typical image frames from the second video and the average image. Notice local intensity inconsistency, blurring and moving cells. The average image has blurred boundaries and a low contrast, which indicates non-rigid image misalignment. . . . .	52

3.14	The registered image frames from the second video after applying the transformation found during the group-wise non-rigid registration. The average image has much sharper boundaries than prior to registration, which indicates accurate image alignment. . . . .	53
4.1	The point set registration problem: Given two sets of points, assign the correspondences and the transformation that maps one point set to the other.	57
4.2	Rigid point set registration algorithm. . . . .	65
4.3	Affine point set registration algorithm. . . . .	66
4.4	The Coherent Point Drift algorithm for non-rigid point set registration. . .	71
4.5	Matrix-vector products computation through FGT. . . . .	74
4.6	Fish data set, rigid registration examples. We align $\mathbf{Y}$ (blue circles) onto $\mathbf{X}$ (red stars). The columns show the iterative alignment progress. a) Registration of the point sets with missing non-overlapping parts ( $w = 0.5$ ); b) Registration of the point sets corrupted by random outliers ( $w = 0.5$ ); c) A challenging rigid registration example, where both point sets are corrupted by outliers and biased to different sides of the point sets. We have also deleted some parts from both point sets. We set $w = 0.8$ and fix scaling $s = 1$ . CPD registration is robust and accurate in all experiments. .	75
4.7	3D bunny point set rigid registration examples. We align $\mathbf{Y}$ (blue circles) onto $\mathbf{X}$ (red dots). The columns show the iterative alignment progress. We initialized one of the point sets with 50 degree rotation and scaling equal 2. a) Registration of the point sets with missing points ( $w = 0.5$ ); b) A challenging example of CPD rigid registration with missing points, outliers and noise. CPD shows robust and accurate registration result in all experiments. . . . .	76
4.8	A comparison of CPD and LM-ICP rigid registration performances with respect to noise in the $\mathbf{X}$ (first row) and the $\mathbf{Y}$ point sets (second row). We align $\mathbf{Y}$ (blue circles) onto $\mathbf{X}$ (red dots). The columns 2,3 and 4 show the examples of initial point sets for different random noise stds added to the point set positions. The first column shows the error in estimating the rotation matrix for CPD (blue) and TPS-RPM (red). CPD outperforms LM-ICP in all cases. . . . .	77



4.9	A comparison of CPD and LM-ICP rigid registration performances with respect to outliers in the $\mathbf{X}$ (first row) and the $\mathbf{Y}$ (second row) point sets. We align $\mathbf{Y}$ (blue circles) onto $\mathbf{X}$ (red dots). The columns 2,3 and 4 show the examples of initial point sets with different number of outliers added. The first column show the error in estimating the rotation matrix. CPD outperforms LM-ICP. . . . .	78
4.10	Registration error with respect to the initial rotation (degrees) of the fish point set registered to itself. CPD performs well if the angle of initial misalignment is less than 70 degrees, whereas ICP get trapped into a local minima with rotations beyond 40 degrees. . . . .	79
4.11	Non-rigid CPD registration of 2D fish point sets. a) Noiseless fish point sets registration ( $91 \times 2$ points, $w = 0$ ); b) Registration of 2D fish point set with missing points ( $w = 0.5$ ); c) Registration of 2D fish point set in presence of outliers ( $w = 0.5$ ). CPD registration is robust and accurate in all experiments. . . . .	80
4.12	A comparison of CPD and TPS-RPM on the 2D fish point sets with respect to a) Deformation level; b) Noise level; c) Outliers (x-axis is the ratio of the number of outliers to the number of clean data points). CPD shows more accurate registration performance compared to TPS-RPM, especially in presence of outliers and complex non-rigid deformations. d) Registration error of CPD with respect to the $\lambda$ and $\beta$ parameter settings. CPD performs well for $\lambda$ in a $[0.1..9]$ and $\beta$ in a $[1.5..3]$ intervals. . . . .	81
4.13	Non-rigid registration of 3D face point sets. a) Registration of clean point sets b) Registration of point sets with outliers. CPD shows accurate alignment. . . . .	82
4.14	Non-rigid registration of 3D left ventricle (LV) point sets. (a) two LV point sets at different time instances, (b) the registration result, (c) displacement field between the corresponding points. . . . .	83
4.15	Application of CPD non-rigid point set registration to image registration of histopathology (a) and ultrasound elastography (b) images. (c) Two point set were manually extracted from the images. (d) The aligned point sets using CPD non rigid point set registration. (e) The registered histopathology image. (f) Composite view of the images after the alignment. . . . .	86

4.16	Application of CPD non-rigid point set registration to image registration of histopathology (a) and ultrasound elastography (b) images. (c) Two point set were manually extracted from the images. (d) The aligned point sets using CPD non rigid point set registration. (e) The registered ultrasound elastography image. (f) Composite view of the images after the alignment. .	87
5.1	The plot of $CD2$ similarity function and its derivative versus L1 norm function. . . . .	108
5.2	The plot of $MS$ similarity function and its derivative for different values of correlation coefficient $\rho$ . When $\rho = 0$ , the $MS$ similarity measure simplifies to $CD_2$ . . . . .	109
5.3	Joint non-rigid image registration and segmentation schema. (a) Given two 3D ultrasound images $I$ and $J$ , and the initial position of the contour $C_t$ (b) We use FFD to parametrize the transformation $T$ by a mesh of control points (+). By manipulating the control points positions, we want to solve two tasks simultaneously: to deform the image $I$ to align it with $J$ , and to deform contour $C_t$ to align it to the high intensity gradient region of the image $J$ . . . . .	109
5.4	(a) An example of a 2D slice out of 3D spherical data displayed as a 2D array with $416 \times 62 \times 56$ resolution. (b) The spherical data is converted to the Cartesian data with $208 \times 297 \times 269$ resolution. . . . .	114
5.5	Placement of sono crystals on the LV. . . . .	115
5.6	The LV contour and sonomicrometer motion found during the cardiac cycle (24 consecutive volumes). LV achieves the maximum contraction at volume 13, then the LV dilates (diastolic phase) up to the volume 22 and starts contracting again (systolic phase). . . . .	118
5.7	Estimated motion of the LV contours and sonomicrometers superimposed on the 3D echo images in Cartesian coordinates. . . . .	119
5.8	Color-coded display of LV endocardium strain is visualized at 6 particular volumes (out of 24 total). LV achieves the maximum contraction at volume 13 (end diastolic phase). . . . .	120
5.9	Strain values against time (sec) from the apical region (between 2 particular sono crystals), derived from algorithm and sonomicrometry for the Baseline, IVC occlusion and LAD ligation states. The algorithm strain follows the sonomicrometry strains closely both in the amplitude and shape. . . . .	121
5.10	Color-coded display of LV endocardium ES-strain for LAD and Baseline states. . . . .	122

5.11	The displacement vector fields on LV endocardium between end diastole and end systole volumes at two different viewing angles. . . . .	122
5.12	The plots of LV torsion (degrees) against time (sec) in a cardiac cycle at different steady states. Torsion was increased with reduction in preload due to controlled IVC occlusion (green) and decreased with myocardial ischemia induced with ligation of LAD (red) as compared to baseline (blue). . . . .	123

# Abstract

## **Non-rigid Image Registration: Regularization, Algorithms and Applications**

**Andriy Myronenko**

Doctor of Philosophy  
Department of Science & Engineering  
School of Medicine  
Oregon Health & Science University  
June 2010

**Thesis Advisor: Xubo Song**

Images provide vital information about this world. Multiple images often share the same scene observed at different times, from different view angles or using different sensors. Image registration is a method of aligning two or more images into the same coordinate system, so that the aligned images can be directly compared, combined and analyzed. Correspondence identification between the images is usually a simple task for human visual system, but for the computer algorithm it represents a challenging problem. Automated estimation of the correspondences between the imaged objects and recovery of the underlying geometrical transformation is a fundamental goal of image registration. In medical imaging, images are often related through complex non-rigid deformations. Method to recover such non-rigid geometrical transformations are called non-rigid image registration methods.

In this thesis, we have developed several contributions to the field of non-rigid image

registration. These contributions are linked under the common theme of non-rigid image registration, but stand on their own as valuable components within image registration framework.

We have developed a new intensity-based similarity measure, called Residual Complexity (RC), to cope with images corrupted by spatially-varying intensity distortions. Such distortions are common in microscopy and magnetic resonance imaging, and represent many challenges for image registration. RC is optimized when the residual image can be sparsely coded using a few known basis functions, which explicitly account for spatially-varying distortions.

We have also developed a novel method for rigid and non-rigid point set registration, called Coherent Point Drift (CPD) algorithm. The algorithm simultaneously recovers the correspondences between two sets of multidimensional points as well as the underlying non-rigid transformation. CPD can be used as a key component in feature-based non-rigid image registration, but also has many applications in different computer vision areas.

Finally, we have developed an automated system for motion estimation from 3D+T echocardiography. The system is based on sequential non-rigid image registration, and includes several new contribution, such as ultrasound-specific similarity measure, shape and dynamic constraints. The system outputs the dense deformation field, which we use to derive myocardium quantitative characteristics, such as strain and torsion. We have validated the accuracy of our approach with the groundtruth measurements from implanted markers.

# Chapter 1

## Introduction

### 1.1 What is Image Registration?

Image registration is a method of transforming two images into the same coordinate system [21, 42]. Often, images of the same scene are obtained at different times, from different perspectives or using different modalities. Image registration is required to find the correspondences between the imaged objects and to recover the geometrical transformation that aligns the images, so that the aligned images can be directly compared, combined and analyzed. Main application domains of image registration include medical imaging [68, 42], remote sensing [48, 60] and computer vision [166]. In different applications, image registration can be also called image alignment, matching, stabilization, fusion or stitching. Historically, image registration has been *rigid*, where images need to be rotated and shifted with respect to each other to achieve correspondence, e.g. to correct for different head positions between scans or to register satellite images.

*Non-rigid* image registration is a class of methods where two image are related through non-rigid geometric transformation. Such non-rigid transformations often arise in medical imaging, where an imaged body organ undergoes soft-tissue type of deformation. Non-rigid image registration is one of the key technologies in medical image analysis following closely after image segmentation.

Applications of non-rigid image registration include: a) Motion compensation, which is required to correct for the anatomical body deformations, e.g for respiratory lung motion during radiosurgery or for iris jittering during microscopic imaging [159, 162]. b) Deformable tracking, which is required for motion analysis of the underlying objects. For

instance, in 3D echocardiography non-rigid image registration allows to recover the underlying myocardial motion, which provides insights into heart architecture and function [115, 116]. c) Multimodality fusion, which combines images of the same subject taken from scans with different modalities, e.g. X-ray and MRI, PET and CT [42, 175]. d) Monitoring of changes, e.g. due to disease progression, tumor growth or surgery [68]. e) Segmentation. If one of the images has apriory labeled structures or delineated segment boundaries, then the estimated geometrical transformation can be applied to these labels or boundaries to segment the second image [135, 7]. f) Atlas construction, which is used to describe anatomical variations of populations. Atlas construction requires simultaneous registration of several patient scans, e.g. brain MRIs, to a common coordinate system using groupwise image registration.

Despite large amount of work and relative success, non-rigid image registration is still not a solved problem [42, 71]. Many challenges remain in definition of correspondences between the image elements and of transformation models. Correspondence definition between the anatomical elements is difficult to formulate mathematically. In the simplest case, one can assume equal intensity levels for all the corresponding image elements. Such assumption rarely holds. Real-world images can be corrupted by noise, illumination changes and spatially-varying bias fields. Furthermore, if two images are from different modalities, then the same anatomical object can have completely different intensities. An image can be also corrupted by outliers, e.g. contrast agent, growing tumor or moving cells, which have no correspondences in the other image. The second main challenge in non-rigid image registration is the transformation model definition. The true underlying deformation is often unknown. One usually assumes a certain transformation model that produces physically realizable adequate approximation for a given application, e.g. smooth or locally rigid deformation. Comprehensive surveys on image registration and its applications can be found in works of Brown [21], Maintz and Viergever [100], Hill et al. [68], Crum et al. [42], Zitova and Flusser [197] and Szeliski [166].

## 1.2 Contributions of This Work

Most of the contributions of this thesis have been successfully completed and published during the course of the thesis research. A number of ancillary issues were addressed as needed within the larger scope of the thesis. In summary, the following concrete and substantial contributions to the body of knowledge in non-rigid image registration and its applications were made:

1. Similarity measure: We have developed a novel similarity measure, called Residual Complexity (RC), that is robust to spatially-varying intensity distortions. RC is minimized when the residual image can be sparsely coded using only a few basis functions. We have derived numerically efficient implementation of RC and its derivative and demonstrated the accuracy of RC in comparison with the other state-of-the-art similarity measures. We have also developed a system for non-rigid motion stabilization of microscopic time lapse videos with severe illumination and reflectance artifacts. The system is based on sequential and groupwise non-rigid image registration with the RC similarity measure. The system produces accurate video stabilization results.
2. Point set registration: We have developed a novel probabilistic method for rigid and non-rigid point set registration. The algorithm is called Coherent Point Drift and is currently the state-of-the-art in point set registration. The algorithm is of general interest in computer vision with many applications, including stereo matching, non-rigid image registration, 3D object registration, pattern recognition and shape correspondence estimation. Several key contributions made for CPD are: a) we derived a closed form solution of the optimal rotation matrix for the rigid case of CPD; b) we formulated the motion coherence constraint and derived a solution of the regularized maximum likelihood estimation through the variational approach in the non-rigid CPD case; c) we derived a closed form update for the non-rigid CPD kernel coefficients within the expectation maximization framework; d) we developed fast CPD implementations using low-rank matrix approximations and fast Gauss transform; e) we compared the CPD performance on multiple synthetic and real-world



examples with the other leading point set registration methods; CPD demonstrated the supreme performance in almost all tested cases.

3. Real-world application of non-rigid image registration in 3D+T echocardiography: We have develop an automated system for motion estimation from 3D+T echocardiography based on sequential non-rigid image registration with the following key contributions: a) we derived a new similarity measure that takes into account speckle formation, correlation and acquisition process. b) we developed an approach for registration in spherical coordinate system. c) we introduced shape and dynamic constraints for myocardial tracking. d) we validated the system in vivo on open-chest pigs with the groundtruth measurements from implanted sonomicrometers.
4. Matlab software libraries: We have developed and released the following toolboxes.
  - a) Medical Image Registration Toolbox (MIRT) for MATLAB. The toolbox allows 2D and 3D non-rigid image registration. Multiple similarity measures are implemented, including Residual Complexity and Mutual Information. Parametric (B-splines) and non-parametric transformation models are available. MIRT also supports pair-wise, sequential and groupwise image registration modes.
  - b) Point Set Registration Toolbox for MATLAB, based on the Coherent Point Drift (CPD) algorithm. The toolbox finds the correspondences between two multidimensional point sets, as well as rigid or non-rigid transformation to align the point sets. The toolbox has been downloaded thousands of times since 2007.

### 1.3 Thesis Outline

Chapter 2 overviews the image registration framework. It discusses different approaches to image registration including intensity- and feature-based ones, parametric and non-parametric ones. It also emphasizes the importance of regularization in all aspects of image registration. In Chapter 3, we introduce a novel intensity-based similarity measure, called Residual Complexity. We show the analytical derivation of RC, analyze its properties and test its performance on multiple synthetic and real-world examples. We also develop a system for non-rigid stabilization of microscopic time lapse videos, where the

RC similarity measure is the key component. In Chapter 4, we introduce a new probabilistic method for rigid and non-rigid point set registration called the Coherent Point Drift algorithm. We derive the method, analyze its properties and validate it against the other popular methods. We also demonstrate CPD application to the non-rigid registration of images. In Chapter 5, we introduce a system for an automated motion estimation from 3D+T echocardiography. The system is based on sequential non-rigid registration of the 3D ultrasound volumes. We validate the system against the groundtruth measurements provided by implanted sonomicrometers. Chapter 6 summarizes the thesis and suggests possible extensions and future directions.

## 1.4 Publications

A large amount of work contained in this thesis has already been published in the peer-reviewed literature and presented at numerous conferences and workshops [111, 113, 116, 115, 114, 160, 159]. Here is a list of those publications:

- [1] Andriy Myronenko and Xubo Song: “Intensity-based Image Registration by Minimizing Residual Complexity”, *Accepted to IEEE Trans. on Medical Imaging*.
- [2] Andriy Myronenko and Xubo Song: “Point-Set Registration: Coherent Point Drift”, *Accepted to IEEE Trans. on Pattern Analysis and Machine Intelligence*.
- [3] Andriy Myronenko and Xubo Song: “Image Registration by Minimization of Residual Complexity”, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 49-56, 2009.
- [4] Andriy Myronenko, Xubo Song and David J. Sahn: “Maximum Likelihood Motion Estimation in 3D Echocardiography through Non-rigid Registration in Spherical Coordinates”, *Functional Imaging and Modeling of the Heart (FIMH)*, pp. 427-436, LNCS vol. 5528, 2009.
- [5] Andriy Myronenko, Xubo Song and David J. Sahn: “LV Motion Tracking from 3D Echocardiography Using Textural and Structural Information”, *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*

- (*MICCAI*), LNCS vol. 4792, Springer, pp. 428-435, 2007.
- [6] Xubo Song, Andriy Myronenko and David J. Sahn: “Speckle Tracking in 3D Echocardiography with Motion Coherence”, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-7, 2007.
  - [7] Andriy Myronenko, Xubo Song and Miguel A. Carreira-Perpinan: “Non-rigid point set registration: Coherent Point Drift”, *Advances in Neural Information Processing Systems 19 (NIPS)*, pp. 1009-1016, 2006.
  - [8] Xubo Song, Andriy Myronenko, Stephen R. Plank and Stephen R. Plank: “Registration of Microscopic Iris Image Sequences Using Probabilistic Mesh”, *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS vol. 4191, Springer, pp. 553-560, 2006.
  - [9] Muhammad Ashraf, Andriy Myronenko, Thuan Nguyen, Akio Inage, Wayne Smith, Robert I. Lowe, Karl Thiele, Carol A. Gibbons Kroeker, John V. Tyberg, Jeffrey F. Smallhorn, David J. Sahn and Xubo Song: “Defining Left Ventricular Apex to Base Twist Mechanics Computed from High-Resolution 3D Echocardiography: Validation against Sonomicrometry”, *Journal of American College of Cardiology (JACC) Imaging*, vol. 3, pp. 227-234, 2010.

# Chapter 2

## Non-rigid Image Registration Framework

### 2.1 Introduction

Two fundamental approaches to image registration are the feature-based and intensity-based ones [42, 68]. Feature-based image registration methods extract a set of geometrical features from one image beforehand and match them with their counterparts in the second image. These features typically include locations of corner points, landmarks, line segments, curves and surfaces. In medical imaging, such features can represent distinct anatomical regions or fiducial markers, which ensure registration validity and allows the transformation to be consistent with the underlying anatomy [68]. The key advantage of the feature-based methods is their dimensionality reduction property, which significantly reduces the computational time and load. Once the geometrical features are identified, the image intensities are no longer required for the registration process, because the matching is based solely on geometrical features. The disadvantage of using features is that they heavily rely on the robust feature extraction and correspondence estimation pre-processing steps. Automatic feature extraction [94, 65] and correspondence estimation [13, 113] are themselves large areas of research in computer vision.

In contrast, intensity-based image registration aligns images directly based on their intensities and does not require any pre-processing steps [42, 68, 100]. Dense intensity information, which is available at each voxel, allows for more accurate estimation of local non-rigid deformations. However, the definition of the correspondences between the imaged areas based solely on intensities is non-trivial and currently remains a great challenge in computer vision. Another downside of intensity-based methods is their high computer

memory and computational time requirements especially in volumetric 3D registration tasks. Nevertheless, most of the research works in image registration are intensity-based, because the registration procedure uses all the available image information without relying on any preprocessing steps.

Aside from the source of information used (features or intensities levels), an image registration algorithm consists of three main components: the similarity measure, the transformation model and the optimization method. The first component of an image registration algorithm is the similarity measure. The similarity measure is a criterion, usually an objective function, that achieves its optimum when two images verify a certain relationship. For instance, sum-of-squared-differences remains a popular similarity measure with the identity intensity relationship assumption between the images. We discuss different similarity measures in Section 2.2.

The second component of an image registration algorithm is the transformation model. The transformation model defines the admissible set of transformations (e.g. physically realizable ones) that is required to align the images. For instance, rigid parametrization is adequate to align two images with different object orientations. Transformation models subdivide into rigid and non-rigid ones. A rigid transformation allows only rotation and translation. The simplest non-rigid transformation is affine, which also allows anisotropic scaling and skews. Non-rigid image registration is the most interesting and challenging work in registration today. For instance, medical images are often related through soft-tissue type of transformation due to the physical properties of body organs. Non-rigid transformations subdivide into parametric and non-parametric ones. Parametric methods assume a particular parametrization of the transformation, e.g. locally-affine or spline-based [71]. From this perspective, rigid image registration is parametric by definition. Non-parametric methods estimate a transformation as an unknown function within variational calculus [108]. Regularization of non-parametric transformations is always required to make the problem well-posed and to constrain the transformation, e.g. to be smooth and invertible. Such regularization defines the key properties and behavior of the corresponding non-parametric transformation model. We overview different transformation models in Section 2.3.

The last component of an image registration algorithm is the optimization method. The optimization method is an algorithm to find a set of parameters that optimize a given similarity measure. Usually, standard optimization methods are employed including Gradient Descent, Quasi-Newton, Conjugate Gradient, Levenberg-Marquardt, BFGS and Stochastic Gradient Descent methods [121, 85].

Finally, the vital mechanism that occurs throughout any image registration algorithm is *regularization*. Image registration is inherently ill-posed; the existence and the uniqueness of the solution is not guaranteed. Thus, regularization is essential. Non-parametric registration methods completely rely on the regularization, which fully defines admissible deformations. Parametric methods implicitly regularize the transformation by using low dimensional set of parameters. Many parameterizations, including splines, are the variational solutions of the corresponding regularized functionals. Not only the transformation model, but the similarity measure also takes advantage of the regularization. We reserve the detailed discussion of the regularization approaches in image registration for Section 2.4.

In many applications it is not a pair but a set of multiple images that we need to transform into a common coordinates system. Such methods are called groupwise image registration methods, and are usually build upon conventional image registration approaches with several specifics. We overview the groupwise image registration methods in Section 2.5

## 2.2 Similarity Measure

### 2.2.1 Feature-based Registration

Feature-based image registration methods define a similarity measure directly based on geometrical features extracted from the images beforehand. One of the simplest feature-based similarity measures is the  $L_2$  norm between the corresponding pairs of landmarks.

$$\mathbf{E}_{sim} = \sum_{k=1}^K \|\mathcal{T}(\mathbf{c}_k^J) - \mathbf{c}_k^I\|^2 \quad (2.1)$$

where  $\mathbf{c}_k^J$  and  $\mathbf{c}_k^I$  are the locations of feature points extracted from the images  $J$  and  $I$  respectively. Several robust distance measures, including  $L_1$  norm have been also used [143]. One of the main advantages of  $L_2$  norm is that the optimal transformation can be found in the closed form both for rigid [173] and some non-rigid [17] parameterizations. Equation 2.1 implies that both images have equal number of feature points with known correspondences. In real-world applications, the correspondences are not known in advance. Furthermore, a feature extraction method may detect different number of points in two images with several outliers and missing points. Automated point set matching methods are required. The methods to estimate point-wise correspondences are often referred to as point set registration methods [151, 32, 114, 113]. Some point set registration methods only estimate the unknown correspondences [151], whereas newer advanced methods simultaneously estimate the correspondences and the transformation [13, 113]. Methods based on simultaneous estimation usually alternate between the assignment of correspondences and the minimization of a similarity measure given the correspondences. In the simplest case, the correspondence assignment is based on the closest distance criterion, and then the familiar  $L_2$  norm is used as the similarity measure [13]. A more fundamental approach is to assign the correspondences between all possible combinations of points up to some probability. Then the similarity measure a generalized distance measure [113, 32]

$$\mathbf{E}_{sim} = \sum_{k=1}^K \sum_{l=1}^L P_{kl} \|\mathcal{T}(\mathbf{c}_k^J) - \mathbf{c}_l^I\|^2 \quad (2.2)$$

where  $P_{kl}$  is the probability of the point  $\mathbf{c}_k^J$  to correspond to the point  $\mathbf{c}_l^I$ . Such similarity measure analytically comes from the maximum likelihood approach to the probabilistic point set registration [113]. We thoroughly overview existing and introduce a novel point set registration method in Chapter 4.

### 2.2.2 Intensity-based Registration

Intensity-based registration methods define the similarity measure directly on image intensities and do not require any feature extraction or matching steps. These methods use dense voxel-wise similarity measures, which are suitable to estimate dense deformation

fields. Intensity based similarity measures are distinguished based on their application area: monomodal or multimodal.

One of the simplest intensity-based similarity measures is the sum-of-squared-differences (SSD)

$$E_{SSD}(\mathcal{T}) = \sum_{n=1}^N (I_n - J_n^T)^2 \quad (2.3)$$

SSD is the optimal similarity measure when two image only differ by Gaussian noise. SSD is widely used in image registration due to its simplicity [68], but is very sensitive to outliers and image artifacts. To reduce the effect of outliers, one can use sum-of-absolute-differences (SAD)

$$E_{SAD}(\mathcal{T}) = \sum_{n=1}^N |I_n - J_n^T| \quad (2.4)$$

Both SSD and SAD assume two images to have equal intensities at the correct alignment. In contrast, squared correlation coefficient (CC) is maximized when the images are linearly related

$$E_{CC}(\mathcal{T}) = \frac{(\sum_{n=1}^N (I_n - \bar{I})(J_n - \bar{J}^T))^2}{\sum_{n=1}^N (I_n - \bar{I})^2 \sum_{n=1}^N (J_n - \bar{J}^T)^2} \quad (2.5)$$

where  $\bar{I}$  is the intensity mean of image  $I$  and  $\bar{J}$  is the intensity mean of image  $J$ .

Mutual information (MI) [175] takes the most general assumption between the image intensities. It maximizes for the simultaneous low joint entropy  $\mathcal{H}(I, J^T)$  and high marginal entropies  $\mathcal{H}(I)$  and  $\mathcal{H}(J^T)$

$$E_{MI}(\mathcal{T}) = \mathcal{H}(I) + \mathcal{H}(J^T) - \mathcal{H}(I, J^T) = \sum_i \sum_j p^T(i, j) \log \frac{p^T(i, j)}{p^T(j)p(i)} \quad (2.6)$$

where  $p^T(i, j)$  is the joint probability of the intensities of the images, and  $p^T(j), p(i)$  are the marginal probabilities. MI can be seen as a measure of how well one image explains the other. MI is maximized when the value of a voxel in the first image is a good predictor of the corresponding voxel in the second image. Due to only the statistical relationship assumption, MI found a wide-spread application in multimodal image registration, e.g. MR-CT, MR-PET [133].

Mutual information depends on the amount of image overlap during the registration. Studholme et al. [163] proposed a normalized variant of mutual information (NMI) to



overcome the sensitivity of MI to the image overlap:

$$E_{NMI}(\mathcal{T}) = \frac{\mathcal{H}(I) + \mathcal{H}(J^{\mathcal{T}})}{\mathcal{H}(I, J^{\mathcal{T}})} \quad (2.7)$$

This version of normalized mutual information has been shown to be considerably more robust than standard mutual information for multimodal registration in which the overlap volume changes substantially [163, 68].

MI and NMI require an estimate of the joint probability density function of the intensities, which is practically done using histograms or kernel density estimation methods [133, 168].

### Maximum Likelihood Approach

Many intensity-based similarity measures can be formulated within the Bayesian framework, where the minimization of a similarity measure is equivalent to the maximum likelihood (ML) estimation of the transformation [141]. Popular image registration similarity measures, such as SSD, CC and MI, correspond to the ML estimation given the identity, affine or statistical relationship between the intensities of the images. ML approach allows to implement prior knowledge of image acquisition process and derive the efficient similarity measure for the given image modality. ML approach to register two images  $I$  and  $J$  is to maximize the conditional probability

$$\hat{\mathcal{T}} = \arg \max_{\mathcal{T}} P(I|J, \mathcal{T}) = \arg \max_{\mathcal{T}} P(I|J^{\mathcal{T}}) \quad (2.8)$$

where  $J^{\mathcal{T}}$  denotes the image  $J$  after the coordinate transformation  $\mathcal{T}$ . To simplify, we assume that conditional probabilities of image voxels are independent:

$$P(I|J, \mathcal{T}) = \prod_n P(I(\mathbf{x}_n)|J(\mathcal{T}(\mathbf{x}_n))) \quad (2.9)$$

For practical convenience, we minimize the negative log-likelihood function, which is equivalent to the maximization of the likelihood function:

$$\hat{\mathcal{T}} = \arg \min_{\mathcal{T}} \sum_n -\log P(I(\mathbf{x}_n)|J(\mathcal{T}(\mathbf{x}_n)), \mathcal{T}) \quad (2.10)$$

The conditional probabilities  $P(I(\mathbf{x}_n)|J(\mathcal{T}(\mathbf{x}_n)))$  may depend on a set of parameters  $\theta$ , which we can also estimate within the ML approach.

**Identity relationship:** If we assume that the image  $I$  is simply the image  $J$  corrupted by additive stationary Gaussian noise

$$I(\mathbf{x}_n) = J(\mathcal{T}(\mathbf{x}_n)) + \eta_n \quad (2.11)$$

then the conditional probabilities have the Gaussian form:

$$P(I(\mathbf{x}_n)|J(\mathcal{T}(\mathbf{x}_n))) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(I(\mathbf{x}_n) - J(\mathcal{T}(\mathbf{x}_n)))^2}{2\sigma^2}}. \quad (2.12)$$

And the negative log-likelihood function takes the form

$$-\log P(I|J, \mathcal{T}) = \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_n (I(\mathbf{x}_n) - J(\mathcal{T}(\mathbf{x}_n)))^2 \quad (2.13)$$

If the noise variance  $\sigma^2$  is fixed, then the negative loglikelihood function is the familiar SSD similarity measure.

**Linear relationship:** If we assume that the image  $I$  is the image  $J$  after some linear intensity transformation corrupted by additive stationary Gaussian noise

$$I(\mathbf{x}_n) = aJ(\mathcal{T}(\mathbf{x}_n)) + b + \eta_n \quad (2.14)$$

where  $a$  and  $b$  are unknown constants of linear intensity transformation, then we can obtain the correlation coefficient (CC) similarity measure. The negative log-likelihood function takes the form

$$E(a, b, \mathcal{T}) = -\log P(I|J, \mathcal{T}) = \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_n (I(\mathbf{x}_n) - aJ(\mathcal{T}(\mathbf{x}_n)) - b)^2 \quad (2.15)$$

We can estimate  $a$  and  $b$  by minimizing Equation 2.15 with respect to these parameters.

Taking the partial derivatives, we obtain:

$$\begin{aligned} \frac{\partial E(a, b, \mathcal{T})}{\partial a} &= -\frac{1}{\sigma^2} \sum_n J(\mathcal{T}(\mathbf{x}_n))(I(\mathbf{x}_n) - aJ(\mathcal{T}(\mathbf{x}_n)) - b) = 0 \\ \frac{\partial E(a, b, \mathcal{T})}{\partial b} &= -\frac{1}{\sigma^2} \sum_n (I(\mathbf{x}_n) - aJ(\mathcal{T}(\mathbf{x}_n)) - b) = 0 \end{aligned}$$

Simplifying for  $b$  and  $a$  we get

$$b = \frac{1}{N} \sum_n (I(\mathbf{x}_n) - aJ(\mathcal{T}(\mathbf{x}_n))) = \mu_I - a\mu_J \quad (2.16)$$

and

$$\begin{aligned} \sum_n J(\mathcal{T}(\mathbf{x}_n))(I(\mathbf{x}_n) - aJ(\mathcal{T}(\mathbf{x}_n)) - \mu_I + a\mu_J) &= 0 \\ \sum_n J(\mathcal{T}(\mathbf{x}_n))(I(\mathbf{x}_n) - \mu_I - a(J(\mathcal{T}(\mathbf{x}_n)) - \mu_J)) &= 0 \\ \sum_n J(\mathcal{T}(\mathbf{x}_n))(\hat{I}(\mathbf{x}_n) - a\hat{J}(\mathcal{T}(\mathbf{x}_n))) &= 0 \\ \sum_n \hat{J}(\mathcal{T}(\mathbf{x}_n))(\hat{I}(\mathbf{x}_n) - a\hat{J}(\mathcal{T}(\mathbf{x}_n))) &= 0 \\ a = \frac{\sum_n \hat{J}(\mathcal{T}(\mathbf{x}_n))\hat{I}(\mathbf{x}_n)}{\sum_n \hat{J}(\mathcal{T}(\mathbf{x}_n))\hat{J}(\mathcal{T}(\mathbf{x}_n))} &= \frac{\text{cov}(I, J^T)}{\text{var}(J^T)} \end{aligned}$$

Now, we can substitute  $a$  and  $b$  back in Equation 2.15

$$\begin{aligned} E(a, b, T) &= \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_n \left( \hat{I}(\mathbf{x}_n) - \frac{\text{cov}(I, J^T)}{\text{var}(J^T)} \hat{J}(\mathcal{T}(\mathbf{x}_n)) \right)^2 = \\ &= \frac{N}{2} \log(2\pi\sigma^2) + \frac{N}{2\sigma^2} \text{var}(I) - \frac{N}{2\sigma^2} \frac{\text{cov}(I, J^T)^2}{\text{var}(J^T)} \propto - \left( \frac{\text{cov}(I, J^T)}{\sqrt{\text{var}(J^T)}\sqrt{\text{var}(I)}} \right)^2 \end{aligned} \quad (2.17)$$

Thus the assumption of a linear intensity relationship between the images leads to the familiar squared CC similarity measure [21, 68].

**Functional relationship:** Roche et al. [141, 142] proposed to use a functional relationship between image intensities plus an additive Gaussian noise.

$$I(\mathbf{x}_n) = f(J(\mathcal{T}(\mathbf{x}_n))) + \eta_n \quad (2.18)$$

where  $f$  is an unknown function (parametric or nonparametric) that has to be also estimated withing ML framework. The similarity function, based on such image relationship depends on the functional form of  $f$  and is called Correlation Ratio (CR) [141, 142].

**Statistical relationship:** If we make no assumptions neither on the intensity relationship nor on the noise model, then the final similarity function is Mutual Information (MI) [175]. We can rewrite the sum over the voxels in Equation 2.10 as a summation over the areas of constant intensity.

$$E(\mathcal{T}) = \sum_{\mathbf{x}} -\log P(I(\mathbf{x})|J(\mathcal{T}(\mathbf{x})), \mathcal{T}) = -N \sum_i \sum_j p(i, j) \log P(i|j) \quad (2.19)$$

where  $p(i, j)$  is a total number voxels with intensity  $i$  in the image  $I$  that correspond to the voxels  $j$  in the image  $J^T$ , which is a joint probability of the intensities of the images. Now we have to find the  $P(i|j), \forall i, j$  withing the ML approach, subject to the constraint  $\sum_i P(i|j) = 1, \forall j$ . Introducing the Lagrange multipliers  $\lambda_j$  for the constraint, we can find the partial derivative of negative log-likelihood function and equate them to zero:

$$\begin{aligned} E(\mathcal{T}) &= -N \sum_i \sum_j p(i, j) \log P(i|j) + \sum_j \lambda_j (\sum_i P(i|j) - 1) \\ \frac{\partial E}{\partial P(i|j)} &= -\frac{Np(i, j)}{P(i|j)} + \lambda_j = 0, \Rightarrow \lambda_j = \sum_i Np(i, j) = Np(j) \\ P(i|j) &= \frac{p(i, j)}{p(j)} \end{aligned}$$

where  $P(j)$  is the marginal distribution of  $J^T$  voxels. Thus the negative log-likelihood function takes the form

$$\begin{aligned} E(\mathcal{T}) &= -N \sum_i \sum_j p(i, j) \log P(i|j) = -N \sum_i \sum_j p(i, j) \log \frac{p(i, j)}{p(j)} = \\ &\quad -N (\sum_i \sum_j p(i, j) \log p(i) - \sum_i \sum_j p(i, j) \log \frac{p(i, j)}{p(j)p(i)}) = \\ &\quad -N (\sum_i p(i) \log p(i) - \sum_i \sum_j p(i, j) \log \frac{p(i, j)}{p(j)p(i)}) = N(\mathcal{H}(I) - \mathcal{I}(I, J^T)) \propto -\mathcal{I}(I, J^T) \end{aligned} \quad (2.20)$$

where  $\mathcal{I}$  is mutual information. Thus, if we make no assumptions on conditional probability  $P(I|J, \mathcal{T})$ , except it is voxelwise independent and stationary, then ML estimation of the conditional probabilities and the transformation is equivalent to the maximization of MI [175].

## 2.3 Transformation Model

The transformation model defines the way to deform one image to match the other. Historically, the most common transformation model is rigid, which has a small number of parameters: rotation and translation parameters. The simplest non-rigid transformation is affine, which allows for skews and shearing. Affine transformation is often considered rigid due to its simplicity. Non-rigid transformations subdivide into parametric and non-parametric ones [71]. Parametric transformations depend on a set of parameters. For instance, B-spline parametrization uses a mesh of control points (parameters) and interpolates in between with the B-spline basis functions. The shape of the B-spline basis functions and the sparseness of control points limit the admissible transformations. Non-parametric image registration methods estimate the transformation as an unknown function within the variational calculus [108]. Non-parametric approach allows to model complex local deformations. This approach is ill posed and requires to use a regularization term to constrain the transformation, e.g. to be smooth or locally rigid.

### 2.3.1 Parametric Approach

Parametric image registration methods define the spatial transformation as a function of a set of parameters. Rigid image registration methods are parametric by definition, where unknown parameters are the rotation matrix and the translation vector. Popular non-rigid parametric transformations include piecewise affine, radial basis functions (RBF) and B-splines. Piecewise affine approach subdivides the image into rectangular blocks and assumes a simple transformation model, usually affine or rigid, for each block [66]. Such approach is fast, but introduces many approximation errors to the truly non-rigid deformation. RBF approach defines the transformation as a linear combination of the basis functions

$$\mathcal{T}(\mathbf{x}) = \sum_{n=1}^N w_n R(\|\mathbf{x} - \mathbf{x}_n\|) \quad (2.21)$$

where  $w_n$  are unknown parameters. The basis function  $R$  depends on the Euclidean distance from  $\mathbf{x}$  to the given control point  $\mathbf{x}_n$ . Examples of RBFs include Gaussian, multiquadric and thin plate spline (TPS). These functions approach zero asymptotically, but

have global support, which significantly increases the computation complexity with each additional control point. Among different RBFs, TPS has found many applications in medical image registration [17, 33]. In contrast to RBFs, B-splines define the transformation locally. B-spline control points can be shared among several connected segments to ensure the continuity. Cubic B-splines are probably the most popular parametric non-rigid transformation parametrization [145, 8, 79]. Cubic B-splines are twice differentiable and continuous at the joints. They can model localized deformations with low computational complexity. Due to its wide popularity in medical image analysis, we review the TPS and the cubic B-spline transformation models in more details in the following subsections.

### Thin Plate Spline

Thin Plate Spline (TPS) is an RBF defined as [17, 176]

$$\mathcal{T}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t} + \sum_n w_n U(\mathbf{x}_n, \mathbf{x}). \quad (2.22)$$

where  $\mathbf{A}$  is an affine transformation matrix,  $\mathbf{t}$  is a translation vector,  $w_n$  are the TPS coefficients and  $U$  is the TPS basis function. In 2D,  $U = r^2 \log r^2$  and in 3D,  $U = r$ , where  $r = \|\mathbf{x}_n - \mathbf{x}\|$  is the Euclidean distance from  $\mathbf{x}$  to the control point  $\mathbf{x}_n$ . In 4D or higher dimensions the TPS solution does not exist [156]. TPS arises from the Euler-Lagrange differential equation corresponding to minimization of bending energy (See Sec. 2.4.2). The function  $U$  is a solution to the squared Laplacian equation,  $\Delta^2 U \propto \delta_{0,0}$  [17].

Bookstein [17] applied TPS to interpolate the transformation from the given set of the corresponding landmarks in medical images. TPS transformation resembles the deformation of thin metal plate, which allows to approximate soft-tissue type of deformations. One of the advantages of TPS is its explicit decomposition into linear and non-linear parts. Among the disadvantages of TPS is its global support, that is each control point has a global influence on the overall transformation. Global support does not allow to model complex localized transformations and has high computational complexity [17, 176]. Nevertheless, TPS became popular in feature-based methods based on point set registration, where a set feature points is extracted from two image beforehand and matched subject to the TPS parameters [33, 76].

## Cubic B-splines

Cubic B-splines parametrization found a wide-spread attention in non-rigid medical image registration [145, 8, 79]. The transformation is often called Free Form Deformation (FFD) after it was first introduced to deform solid geometric models [152].

The basic idea of FFD is to deform a 3D object by manipulating a mesh of control points. The main advantage of FFD is that the complex non-rigid transformation is defined by a small number of parameters (control points positions).

We denote the 3D image volume as  $\Omega = \{(x, y, z) | 0 \leq x_n \leq N, 0 \leq y_m \leq M, 0 \leq z_k \leq K\}$ . We place a  $n_x \times n_y \times n_z$  mesh of equally spaced control points  $p_{i,j,l}$  over the image domain. The number of control points defines the complexity of the transformation. Then, the transformation  $\mathcal{T}$  is a 3D tensor product of the 1D cubic B-splines:

$$\mathcal{T}(x, y, z; p) = \sum_{k=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_k(u_x) B_m(v_y) B_n(w_z) p_{i+k, j+m, l+n} \quad (2.23)$$

where  $i = \lfloor x/n_x \rfloor - 1, j = \lfloor y/n_y \rfloor - 1, l = \lfloor z/n_z \rfloor - 1$ . B-splines are compactly supported, and thus are defined in terms of local coordinates  $(u_x, v_y, w_z)$ , where  $u_x = x/n_x - \lfloor x/n_x \rfloor, v_y = y/n_y - \lfloor y/n_y \rfloor, w_z = z/n_z - \lfloor z/n_z \rfloor$ .  $B_k$  are the  $k^{th}$  B-spline basis function:

$$\begin{aligned} B_1(u) &= (1-u)^3/6 & B_2(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_3(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 & B_4(u) &= u^3/6 \end{aligned}$$

The key advantage of cubic B-splines is their local support, which provides local transformation control and low computational complexity. They are also twice continuously differentiable at the control point joints. For digital images on a regular grid, the B-spline basis functions can be precomputed to allow fast transformation computations.

### 2.3.2 Non-parametric Approach

Non-parametric image registration estimates the transformation as an unknown function without explicit parametrization. Such problem is ill-posed, because the solution is not unique. To constrain the problem one can use a regularization term that penalizes some undesirable transformation properties, e.g. oscillations. The transformation is usually

defined as

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}) \quad (2.24)$$

where  $\mathbf{u}(\mathbf{x})$  is a unknown displacement function that optimizes the objective function:

$$E(\mathbf{u}) = E_{sim}(I, J(\mathbf{x} + \mathbf{u}(\mathbf{x}))) + \alpha E_{reg}(\mathbf{u}(\mathbf{x})) \quad (2.25)$$

where  $E_{sim}$  is the similarity measure term,  $E_{reg}(\mathbf{u})$  is the regularization term and  $\alpha$  is a trade-off parameter. The objective function is rather a functional, because the unknown  $\mathbf{u}$  is a function. One can think of the similarity measure as of the external energy function that moves the image, whereas the regularization term is an internal energy function that keeps the transformation coherent and constrained. We overview different regularization terms used in non-parametric image registration in Section 2.4.3

## 2.4 Regularization

Regularization plays an essential role in image registration. It enforces certain properties on a function, e.g. smoothness, sparsity or rigidity. For instance, consider estimation of a non-rigid transformation over the image domain given only a set of corresponding landmarks. The solution to such problem is not unique. There is an infinite number of transformations that will match the corresponding landmarks, but have totally different behavior in the remaining image parts. However, if we impose an additional requirement of using the smoothest transformation, as defined by the corresponding regularization term, then the problem can be uniquely solved. Alternatively, one can regularize the solution by using explicit parametrization, e.g. rigid, which also makes the solution unique. Regularization is essential not only in image registration; it is one of the key components in many machine learning and computer vision area. In this section, we overview the regularization framework and show its applications in image registration.

### 2.4.1 Theory

Regularization is a key tool to make a problem well posed. A problem is well posed in the Hadamard sense if it satisfies all of the following conditions: existence, uniqueness and



continuity. Otherwise the problem is called ill posed. Most of the engineering problems are ill posed including such areas as computer vision, system identification, reconstruction and density estimation [27].

The classical regularization theory originates from the works of Tikhonov [169]. A standard regularization approach is to augment the existing optimization problem with a regularization term and minimize the following functional

$$E(f) = S(f) + \alpha R(f) \quad (2.26)$$

where  $S(f)$  is some objective function,  $R(f)$  is the regularization term and  $\alpha$  is a trade-off parameter. One famous regularization example is to fit a function  $f$  to the given data points  $y_i$  specified at locations  $\mathbf{x}_i$  by minimizing the regularized loss function

$$E(f) = \sum_i (y_i - f(\mathbf{x}_i))^2 + \alpha \|\mathbf{P}f\|^2 \quad (2.27)$$

where  $\mathbf{P}$  is a linear operator, e.g. first or second order derivative operator.

In general an operator can be linear or nonlinear, finite- or infinite-dimensional. An operator extracts certain properties of the function for penalization, e.g. oscillations. In finite dimensions one can draw an analogy of an operator with a matrix, and the application of an operator to a function is equivalent to a matrix-vector product. A regularization term is usually a norm,  $R(f) = \|\mathbf{P}f\|^2 = \|f\|_{\mathbb{H}_m}^2$ , in the Hilbert space  $\mathbb{H}^m$ :

$$\|f\|_{\mathbb{H}^m}^2 = \sum_{k=0}^m \left\| \frac{\partial^k f}{\partial x^k} \right\|_2^2 \quad (2.28)$$

It reduces to the squared  $L_2$  norm for  $m = 0$ . Alternatively, we can define such norm in the Reproducing Kernel Hilbert Space (RKHS) [27, 55] as

$$\|f\|_{\mathbb{H}^m}^2 = \int_{\mathbb{R}^D} \frac{|\tilde{f}(\mathbf{s})|^2}{\tilde{G}(\mathbf{s})} d\mathbf{s} \quad (2.29)$$

where  $G$  is a unique kernel function associated with the RKHS, and  $\tilde{G}$  is its Fourier transform. Function  $\tilde{f}$  indicates the Fourier transform of the function  $f$  and  $\mathbf{s}$  is a frequency domain variable [55]. Regularization in the spectral (frequency) domain is essentially equivalent to the the spatial norm based approach [27]. Whereas spatial approach usually penalizes function derivatives to enforce smoothness, spectral approach penalizes high

frequency function content for the same purpose. Many regularization operators in spatial domain have their counterparts in the spectral domain [27].

Often regularization problems can be solved analytically. The solution involves using the Green's function, which is a continuous analog to the matrix inverse. The function  $G(\mathbf{x}, \xi)$  is the Green's function for a linear operator  $\hat{\mathbf{P}}\mathbf{P}$  if  $\hat{\mathbf{P}}\mathbf{P}G(\mathbf{x}, \xi) = \delta(\mathbf{x} - \xi)$  and  $G(\mathbf{x}, \xi)$  is differentiable everywhere but  $\mathbf{x} = \xi$ . One can obtain the analytical solution to Equation 2.27 using the Representer Theorem [83, 150], which states that the optimal functional form of  $f$  is a linear combination of the reproducing kernels ( $G$ ) in the space produced by the operator  $\mathbf{P}$ ; where the Green's function,  $G$ , of  $\hat{\mathbf{P}}\mathbf{P}$  is the reproducing kernel.

$$f(\mathbf{x}) = \sum_i w_i G(\mathbf{x}, \mathbf{x}_i) + \psi(\mathbf{x}) \quad (2.30)$$

Here, the coefficients  $w_i = (y_i - f(\mathbf{x}_i))/\alpha$  are obtained by evaluating Equation 2.30 at  $\mathbf{x}_i$  points. The term  $\psi(\mathbf{x})$  is a null space of the operator  $\mathbf{P}$ , which can be, e.g. constant or zero [83, 150]. Thus, even though the objective function in Equation 2.27 is being minimized over the whole function space, the optimal  $f$  is in the kernel form, which is analogous to RBF parametrization. An explicit function parametrization, e.g. RBFs, is thus equivalent to variational formulation with the corresponding regularization term. In practice, one can choose an explicit parametrization, e.g. B-splines, without prior knowledge of the corresponding regularization term, as soon as such parametrization sufficiently limits the admissible function space.

In the cases where the objective function,  $S(f)$ , is non-linear non-quadratic functional of  $f$  and/or the regularization term is non-linear then analytical solutions are not available; one usually has to discretize the function  $f$  and to use iterative optimization. In the following subsections, we show how the regularization approach is employed in image registration.

#### 2.4.2 Regularization in Parametric Transformations

Parametric transformation model explicitly regularizes the admissible set of transformations in image registration. For instance, rigid parametrization only allows for rotation

and translation transformations. B-spline parametrization limits the deformation by the control point spacing and by smoothness of the B-spline basis functions. The regularization is imposed implicitly by using a low dimensional set of parameters. Parametric and non-parametric transformation models are linked to each other. As we showed in the previous section, an RBF parametrization is a solution of the non-parametric problem with the corresponding regularization term.

**Thin Plate Spline:** TPS parametrization (Equation 2.22) is a variational solution to the following problem [17, 176]

$$E(f) = \sum_i (y_i - f(\mathbf{x}_i))^2 + \alpha \int_{\mathbb{R}^d} \sum_{m,n=1}^d \left( \frac{\partial^2 f}{\partial x_n \partial x_m} \right)^2 d\mathbf{x} \quad (2.31)$$

where the second term penalizes squared second order derivatives, which has a mechanical interpretation. If an infinite elastic flat thin plate is slightly deformed then the bending energy of the thin plate is to first order proportional to the integral over squared second order derivatives as in Equation 2.31. For this reason the solution (function) to the Equation 2.31 is called thin plate spline. The null space of such regularization includes affine transformations. Alternatively, one can obtain the TPS solution using the spectral form of regularization,  $\int_{\mathbb{R}^2} \|\mathbf{s}\|^4 |\tilde{f}(\mathbf{s})|^2 d\mathbf{s}$ , which is a special case of the regularization that leads to the Duchon splines [47].

**Gaussian Radial Basis Function:** Gaussian RBF parametrization is a variational solution to the following minimization problem

$$E(f) = \sum_i (y_i - f(\mathbf{x}_i))^2 + \alpha \int_{\mathbb{R}^d} \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} \|D^l v(\mathbf{x})\|^2 d\mathbf{x} \quad (2.32)$$

where  $\beta$  is a parameter and  $D$  is a derivative operator such that  $D^{2l}v = \nabla^{2l}v$  and  $D^{2l+1}v = \nabla(\nabla^{2l}v)$  [188]. Such regularization term penalizes all order derivatives. Its spectral analog is in the form  $\int_{\mathbb{R}^D} \frac{|\tilde{f}(\mathbf{s})|^2}{G(\mathbf{s})} d\mathbf{s}$ , where  $G$  is the Gaussian function with width  $\beta$ . The null space of such regularization operator is zero,  $\psi(\mathbf{x}) = 0$  [55]. In Section 4.5 we will show that Gaussian RBF is also a variational solution to the generalized distance measure in the context of non-rigid point set registration [113].

### 2.4.3 Regularization in Non-parametric Transformations

Non-parametric registration approach estimates the transformation function as an unknown function without assuming any parameterizations. Regularization term is always required to make the problem well posed. Standard regularization terms include the diffusion, curvature, elastic and fluid regularization of the displacement function  $\mathbf{u}$  [108, 52]:

	Regularization term ( $E_{reg}$ )	Variational derivative ( $\nabla E_{reg}$ )
Diffusion	$\frac{1}{2} \int \ \nabla \mathbf{u}\ ^2 d\mathbf{x}$	$\Delta \mathbf{u}$
Curvature	$\frac{1}{2} \int (\Delta \mathbf{u})^2 d\mathbf{x}$	$\Delta^2 \mathbf{u}$
Elastic	$\int \frac{\mu}{2} \sum_{i=1}^D \ \nabla \mathbf{u}^i\ ^2 + \frac{\lambda+\mu}{2} (\nabla \cdot \mathbf{u})^2 d\mathbf{x}$	$\mu \Delta \mathbf{u} + (\lambda + \mu) \nabla (\nabla \cdot \mathbf{u})$
Fluid	$\int \frac{\mu}{2} \sum_{i=1}^D \ \nabla \mathbf{v}^i\ ^2 + \frac{\lambda+\mu}{2} (\nabla \cdot \mathbf{v})^2 d\mathbf{x}$	$\mu \Delta \mathbf{v} + (\lambda + \mu) \nabla (\nabla \cdot \mathbf{v})$

Table 2.1: Common regularization operators of the displacement field and their variational derivatives.

where  $\mathbf{u}^i$  is the  $i^{th}$  component of  $\mathbf{u}$ ,  $\lambda$  and  $\mu$  are so-called the Lamé constants. The fluid regularization is similar to elastic, but applied to the velocity field  $\mathbf{v} = \frac{d}{dt} \mathbf{u}(\mathbf{x}, t)$ . The elastic regularization term corresponds to the image deformation as a linear elastic material, whereas fluid regularization allows deformation similar to idealistic behavior of fluids [71, 108]. One feature of the fluid regularization is that it allows large deformations and forces them to be continuous and invertible [29].

The similarity measure is usually a non-linear functional of the transformation. In this case the analytical closed form solution is not available. Optimization of Equation 2.25 requires solving the corresponding Euler-Lagrange differential equation

$$\nabla E_{sim}(\mathbf{u}) + \alpha \nabla E_{reg}(\mathbf{u}) = 0, \quad (2.33)$$

which is usually done iteratively. For instance, using the SSD similarity measure and the curvature regularization term, the corresponding Euler-Lagrange differential equation is

$$-(I - J(\mathbf{x} + \mathbf{u}(\mathbf{x}))) \nabla J(\mathbf{x} + \mathbf{u}(\mathbf{x})) + \alpha \Delta \mathbf{u} = 0 \quad (2.34)$$

Fast partial differential equation solvers include FFT-based and multigrid methods [108], which require only  $\mathcal{O}(N \log N)$  operations per iteration up to the multiplicative constant.

#### 2.4.4 Invertibility of the Transformation

Both parametric and non-parametric non-rigid registration methods do not guarantee the existence of the inverse transformation<sup>1</sup>. The inverse transformation,  $\mathcal{T}^{-1}$ , may never be used, however its existence is required to ensure that the transformation is physically meaningful. If the inverse transformation exists and  $\mathcal{T} \circ \mathcal{T}^{-1} = Id$ ,  $\mathcal{T}^{-1} \circ \mathcal{T} = Id$  then such transformation is called isomorphic. Such a structure-preserving transformation does not allow for any local foldings or twists, which are not physically realizable.

One indicator of the invertibility of the transformation is the determinant of the Jacobian matrix (or simply Jacobian). In 3D the Jacobian is

$$\mathcal{J} = \begin{vmatrix} \frac{\partial \mathcal{T}_x}{\partial x} & \frac{\partial \mathcal{T}_x}{\partial y} & \frac{\partial \mathcal{T}_x}{\partial z} \\ \frac{\partial \mathcal{T}_y}{\partial x} & \frac{\partial \mathcal{T}_y}{\partial y} & \frac{\partial \mathcal{T}_y}{\partial z} \\ \frac{\partial \mathcal{T}_z}{\partial x} & \frac{\partial \mathcal{T}_z}{\partial y} & \frac{\partial \mathcal{T}_z}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 + \frac{\partial \mathbf{u}_x}{\partial x} & \frac{\partial \mathbf{u}_x}{\partial y} & \frac{\partial \mathbf{u}_x}{\partial z} \\ \frac{\partial \mathbf{u}_y}{\partial x} & 1 + \frac{\partial \mathbf{u}_y}{\partial y} & \frac{\partial \mathbf{u}_y}{\partial z} \\ \frac{\partial \mathbf{u}_z}{\partial x} & \frac{\partial \mathbf{u}_z}{\partial y} & 1 + \frac{\partial \mathbf{u}_z}{\partial z} \end{vmatrix} \quad (2.35)$$

The Jacobian equals 1 if the volume remains constant after the transformation. The Jacobian is smaller than 1 in the presence of local volume shrinkage and is greater than 1 in the presence of local volume expansion. Negative Jacobian implies that local foldings and twists had occurred, which is physically not realizable and mathematically not invertible.

There are several ways to enforce the existence of the inverse transformation. If the final transformation has local foldings, one can re-run the registration algorithm with a higher regularization weight  $\alpha$  or simply post-smooth the transformation [2]. A more fundamental approach is to constrain the Jacobian to be non-negative [64] or to impose an additional regularization term on the Jacobian using  $L_2$  norm [28] or KL-divergence regularization [183]. Implementation of the additional regularization on the Jacobian can significantly increase the computation load of non-parametric registration algorithms.

**Diffeomorphic Approach:** An alternative idea to ensure the invertibility of the transformation is to consider the transformation as a solution of the ordinary differential equation [11]

$$\frac{d\mathcal{T}_t(\mathbf{x})}{dt} = \mathbf{v}_t(\mathcal{T}_t(\mathbf{x})) \quad (2.36)$$

---

<sup>1</sup>with the exception of Fluid-based regularizer.

The function  $\mathcal{T}_t(\mathbf{x})$  describes a diffeomorphic transformation through space and time. A diffeomorphic transformation ensures that the inverse transformation and its derivatives exist and are both smooth functions, which is a stronger property than isomorphism. The initial transformation at  $t = 0$  is the identity transformation:  $\mathcal{T}_0(\mathbf{x}) = Id(\mathbf{x}) = \mathbf{x}$ . The final transformation at  $t = 1$  is the diffeomorphic transformation that aligns the images.

Typical objective functional of diffeomorphic methods is

$$E(\mathbf{v}) = E_{sim}(I, J(\mathcal{T}_1(\mathbf{x}))) + \alpha \int_0^1 \|\mathbf{v}_t(\mathbf{x})\|_V^2 dt \quad (2.37)$$

where  $E_{sim}(I, J^T)$  is a similarity measure (typically SSD) and  $\|\mathbf{v}_t\|_V$  is a Sobolev norm on the velocity field (typically  $\|\mathbf{v}_t\|_V = \|\mathbf{P}\mathbf{v}_t\|_2^2$ , where  $\mathbf{P}$  is a linear differential operator). Thus, by minimizing Equation 2.37, one seeks a time-dependent velocity vector field  $\mathbf{v}$  that when integrated from Equation 2.36 generates the diffeomorphic transformation  $\mathcal{T}_1$  that aligns the images [11].

Iterative computation of the final transformation  $\mathcal{T}_1$  by integration from Equation 2.36 is required at each registration iteration, which is extremely computationally expensive. Alternatively, one can assume a time-independent velocity field  $\mathbf{v}$ , which allows to integrate from Equation 2.36 fast through recursion [4, 6]. The downside of using the time-independent velocity field is that some transformations can be impossible to reach or the transformation flow can require a very windy trajectories. In summary, the diffeomorphic image registration methods explicitly account for invertible transformation, but suffer from a large computational complexity.

#### 2.4.5 Regularization in Similarity Measures

Regularization is often present in the similarity measure definition. For instance, the CC similarity measure can be seen as the SSD similarity measure between two images, where one of the images is intensity normalized by the global scaling and shift. Such linear parametrization of the underlying intensity normalization function implements implicit regularization of the similarity measure. Optimal parameters of the intensity normalization can be determined analytically, which leads to the CC similarity measure (see Sec. 2.2.2). Mutual information also can be seen to implement regularization by assuming

voxels of equal intensities to belong to the same class. Historically, transformation was the only unknown function of an image registration algorithm. In many situations, the similarity measure itself includes several unknowns that need to be estimated simultaneously with the transformation. For instance, images often require intensity correction prior to registration. One can define an intensity correction function as an unknown component of the similarity measure and estimate it simultaneously with the transformation. Depending on the application, parametric or non-parametric models are used for the intensity correction function. In non-parametric case an additional regularization term of the correction function is required [111, 109, 110]. In Chapter 3 we will develop a new similarity measure that uses non-parametric estimation of the correction field.

## 2.5 Groupwise Image Registration

Groupwise image registration methods register several images to a common coordinate system. It is not a single geometric transformation, but a group of transformations that has to be found [198, 172]. There are several approaches to groupwise image registration. In the simplest case, one can pick the first (or any other) image as the reference image, and then register the rest of the images to the reference one using conventional (pairwise) image registration. Such approach is very sensitive to the choice of the reference image.

When images are acquired at consecutive time instance, such as in time lapse video imaging, then two consecutive images should be close to each other in terms of the underlying transformation. Thus, one can sequentially register next image frame onto the previous and track the deformation field over time. Often two neighboring image frames have a large overlap, whereas some distant frames have small or zero overlap due to the progressive object or camera shift, which is common in such areas as remote sensing or time lapse video microscopy. Sequential image registration not only significantly improves the processing speed, but also benefits from a large area overlap between the consecutive frames, in contrast to the fixed-reference-image approach.

Another variation of sequential image registration, is to register the next image onto the average of previously aligned images. The main benefit of averaging is to remove noise

and artifacts (e.g. from moving cells). The disadvantage can be in the reduction of the texture statistics (e.g. speckle patterns in ultrasound images).

Finally, the so-called “true” groupwise non-rigid image registration registers all images simultaneously [198, 102, 172, 78]. Such formulation eliminates the bias of choosing any particular reference image. Twining et al. [172, 102] proposed to use a Minimum Description Length (MDL) criterion to align the images. The MDL principle states that the transmission of a model of the data, together with the parameters of that model, should be as short as possible. Zollei et al. [198] investigated the *congealing* approach for groupwise registration. Their objective function is the total element-wise entropy of the input image sequence. The entropies are computed at each coordinate location and then added together. Vedaldi et al. [174] improved the congealing approach by minimizing the amount of complexity of the data set together with the amount of image distortion. Joshi et al. [78] used a sum of SSD’s between all images and the reference one, which is estimated simultaneously with registration. By choosing the SSD similarity measure the reference image can be found analytically as an arithmetical average of the data set.

We would like to emphasize that most of the groupwise registration methods register several images to build a statistical model of the population, while in this thesis we will employ groupwise image registration for a different purpose. In Chapter 3, we will use the groupwise approach to stabilize the time lapse video microscopy, which is necessary for the analysis of moving cells. In Chapter 5, we will sequentially estimate the non-rigid transformation between 3D ultrasound volumes in order to track the myocardial deformation.

## 2.6 Conclusion

In this chapter, we overviewed the general framework to image registration. We showed that two fundamental approaches to image registration are feature- and intensity-based ones. We overviewed the key components of image registration, including popular similarity measures and transformation models, and showed the essential role of the regularization theory in image registration.



# Chapter 3

## Image Registration by Minimization of Residual Complexity

### 3.1 Introduction

Image registration is a method of aligning two images into the same coordinate system, so that the aligned images can be directly compared, combined and analyzed. One of the key components of image registration is the similarity measure between the images that is optimized at the correct spatial alignment. Intensity-based similarity measures are defined directly on image intensities, and often rely on the assumption of independence and stationarity of the intensities from pixel to pixel [141, 68]. Such similarity measures, including Sum-of-Squared-Differences (SSD), Correlation Coefficient (CC), Correlation Ratio (CR) and Mutual Information (MI), are defined only between the corresponding pixels without considering their spatial dependencies. Further, the intensity relationship is assumed to be spatially stationary<sup>1</sup>. As a result, such measures are less robust and tend to fail when registering two images corrupted by spatially-varying intensity distortion.

Real-world images often have spatially-varying intensity distortions. For instance, brain MRI images often can be corrupted by slow-varying intensity bias fields [162]; visual-band images can have illumination non-homogeneity and reflectance artifacts [68]. These complex distortions do not obey the pixel-wise independence or stationarity assumption and cannot be captured by simplistic intensity relationships. To illustrate this argument,

---

<sup>1</sup>Spatial stationarity implies the equal form of the probability density function regardless of any shift in image spatial dimension.

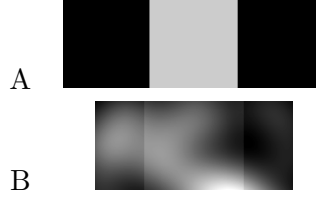


Figure 3.1: Gray stripe registration experiment.

consider aligning the two images in Figure 3.1. The smaller image (B) is a crop of the larger image (A) corrupted by an additive spatially-varying intensity field (simulated by a sum of random Gaussians). Figure 3.2 plots the values of several similarity measures, including our new similarity measure called Residual Complexity (RC), with respect to the horizontal shift of image B over A. Global similarity measures, including SSD, CC and MI, do not give an optimum at the correct image alignment (zero translation), whereas RC achieves a distinct optimum with a wide convergence range (Fig. 3.2). We also applied the  $L_1$  and  $L_2$  norms to the difference of the gradient magnitude images, which provides a correct optimum position, however optimization of such similarity measures is challenging due to its uniformity even in a small vicinity from the optimum. In this example, global similarity measures are not adequate to register the images. It would require our method or local similarity measures to cope with the given spatially-varying intensity distortion.

The effect of slow-varying intensity distortions becomes even more apparent with dense non-rigid registration, which estimates the transformation (e.g. displacement vector) at every image pixel. Such methods require high accuracy and consistency of the similarity measure at every pixel. Dense non-rigid registration using global criteria, e.g. MI or normalized MI, often provides solutions with significant local errors in the estimated transformation, as the similarity criteria attempts to reduce the size of locally corrupted regions to increase the overall image similarity [162].

Among current methods to deal with spatially-varying intensity distortion are methods based on simultaneous intensity correction and registration [54, 134, 109], on modeling higher order pixel interdependencies [49, 180, 191], and on employing local similarity criteria [2, 162, 184]. Such methods perform relatively well with spatially-varying intensity

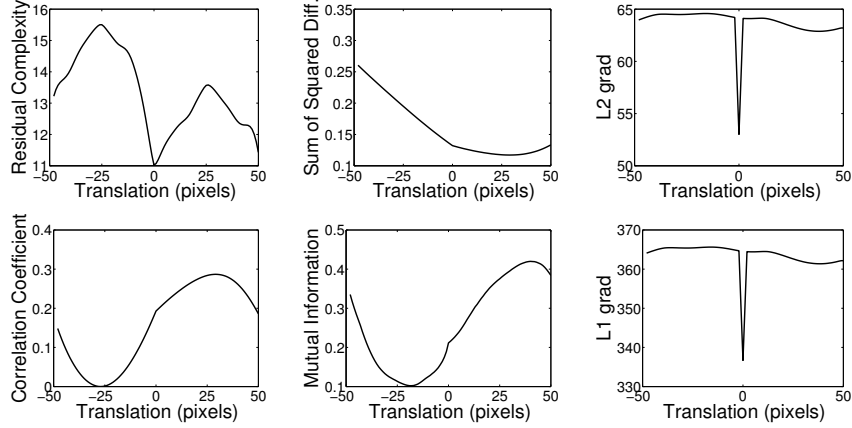


Figure 3.2: Plot of several similarity measures (RC, SSD, CC, MI, L1 and L2 norm gradient magnitude) with respect to the translation of the smaller image over the larger one in Figure 3.1. Zero translation corresponds to the accurate registration.

distorted images, but have disadvantages in their high computation complexity or presence of multiple local minima of its similarity function.

We propose a new intensity-based similarity measure to deal with complex spatially-varying intensity distortions. We start deriving the similarity measure by introducing an intensity correction field that brings images into agreement in the intensity space. We learn the adaptive regularization for the correction field. Analytically solving for the correction field and adaptive regularization allows us to derive a similarity measure that is robust to spatial intensity distortions. Interestingly, the final form of our similarity measure has many analogies in several computer vision areas, such as image compression, sparse coding and topographic learning. Our similarity measure can be interpreted as one that favors a registration with minimum complexity of the residual (difference) image between the two registered images. Thus, we name it the Residual Complexity (RC) similarity measure.

Intuitively, the residual image is expected to achieve the minimal complexity at the correct spatial alignment. RC measures the sparseness of the residual image in terms of the given basis functions. From this perspective, one can consider the SSD similarity measure as the one that favors all zeros residual regardless of the spatial positions of misaligned structures; a random simultaneous shuffle of pixel positions in both images

does not change the SSD measure. In contrast to SSD, RC is minimized when the residual image can be sparsely coded using only a few basis functions (e.g. favoring smooth or piece-wise smooth residual). Such properties of RC allow robust alignment of the images distorted by the spatially-varying intensity distortions.

## 3.2 Background

Definition of proper similarity measure is a challenging task in the presence of non-stationary intensity distortions [68]. Many intensity-based similarity measures can be unified within a maximum likelihood (ML) framework [141], which is to maximize the joint probability of two images given the transformation. To simplify the likelihood function, it is commonly assumed that pixel-wise probabilities are independent and stationary, which allows likelihood factorization, and the final intensity similarity measure is defined only between the corresponding pixels. Random shuffle of all the corresponding image pixels in both images does not change the value of such similarity functions, because it is only pixel-wise defined and evaluated.

Spatial stationarity and pixel-wise independence are often not realistic in medical imaging, especially in the presence of spatially-varying intensity distortions. To overcome intensity non-stationarities, current registration methods include ones based on 1) simultaneous intensity correction and registration, 2) on modeling higher order pixel interdependencies and, 3) on employing local similarity estimation.

The first group of methods correct for intensity distortions simultaneously with image registration [54, 134, 109]. Friston et al. [54] proposed to align the images using SSD, but assumed that one of the images has to be intensity corrected with non-linear intensity transformation and a convolution filter. The intensity correction function was defined as a linear combination of some basis functions with spatially smooth-varying coefficients. The convolution filter needed to be chosen manually for a specific problem or estimated from the images. Modersitzki and Wirtz [109] used a similar approach, but defined a multiplicative intensity correction function with a total variation regularizer. The authors had to iteratively solve a non-linear system of partial differential equations simultaneously

with respect to the transformation and intensity correction. In more recent work, Ashburner and Friston [7] proposed a probabilistic framework for joint registration, intensity normalization and segmentation, using alternating optimization of corresponding parameters. Such combined approaches require ad hoc definition of the intensity normalization function and can be more time consuming.

The second group of methods address the problem of non-stationary distortions from a Bayesian perspective by using more complex probabilistic models, beyond simplistic assumptions of pixel-wise independence and stationarity [49, 180, 191]. El-Baz et al. [49] proposed to learn a prior appearance model of the first image using a Markov-Gibbs random field with pairwise interaction, and then to transform the second image to maximize its probability under the learned appearance model. Wyatt and Noble [180] proposed to use Markov random fields (MRF) to iteratively segment the images and register the class labels. Zheng and Zhang [191] proposed a MAP-MRF framework, which includes some standard similarity measures and allows the definition of new ones by taking advantage of local pairwise intensity interactions. Such MRF-based methods heavily rely on the initialization of parameters [180] and definition of local intensity interactions [191].

The last and the largest group of methods use local similarity measures or similarities defined only on a small pixel neighborhood averaged over the image domain. The intuitive idea behind such approaches is that a spatially-varying intensity distortion is constant within a small pixel neighborhood. Ardecani et al. [2] used a local correlation coefficient similarity measure within a small window centered at every pixel and averaged over the image domain. To overcome multiple local minima, the registration is carried at increasing image sizes with block-matching strategy for every pixel. Hermosillo et al. [67] proposed a framework for local similarity measures, by localizing the correlation ratio and mutual information similarity criteria. Studholme et al. [162] considered the spatial location as third channel in estimation of mutual information. Their similarity measure was a sum of regional MI criteria evaluated within small overlapping regions, centered on a sparse regular grid of points. The regional MI similarity functions were weighted proportional to the size of the corresponding region. Regional MI showed improved registration performance compared to the global MI in the presence of slow-varying intensity distortions. Similar

to [162], Loeckx et al. [93] proposed to use the probability distributions conditioned on spatial location, which resulted in a conditional MI similarity measure. Klein et al. [86] also used local MI, defined as sum of MIs evaluated over small overlapping regions. The overlapping regions were reinitialized randomly at every iterations, in contrast to pre-defined region centers on a regular grid. They also adopted stochastic gradient descent optimization algorithm, which often has faster convergence and is less likely to get stuck in a local minima. Yi and Soatto [184] used local neighborhood NMI at each pixel (averaged over the image domain), but combine it with the global NMI. The influence of local NMI versus global NMI is weighted at every pixel proportional to the confidence of local image agreement (given by the local NMI). Whereas global NMI has less local optima, local NMI is less sensitive to spatially-varying intensity distortions.

Overall, the methods based on local similarity measures showed better performance compared to global ones in the presence of non-stationary intensity distortions. The problem of such local approaches is that local statistics are only effective within a small image region, which created numerous local minima of the objective function. Also such local estimates are much more sensitive to noise and outliers than global measures.

We introduce a new similarity measure that is closely related to the first group of above-mentioned methods, of simultaneous image registration and intensity correction. The unique feature of our approach is that we analytically solve for the intensity correction field and eliminate it from the similarity function. Thus, the intensity correction field does not appear in the final registration algorithm. We use the definition of the correction field mostly to set up an elegant theoretical framework. From a Bayesian perspective, the pixel-wise probabilities, by our method, are independent only conditioned on the intensity correction field.

### 3.3 Method

Consider two images  $I$  and  $J$  to be aligned, assuming the following intensity relationship:

$$I = J(\mathcal{T}) + S + \eta \quad (3.1)$$

where  $S$  is an intensity correction field and  $\eta$  is zero mean Gaussian noise (note that for any two images there always exists a correction field  $S$ , such that Equation 3.1 holds, if the shape of  $S$  is not restricted).  $\mathcal{T}$  is the geometric transformation that aligns  $I$  and  $J$ . The maximum a posteriori (MAP) approach to estimate  $S$  and  $\mathcal{T}$  is to maximize the probability

$$P(\mathcal{T}, S|I, J) \propto P(I, J|\mathcal{T}, S)P(\mathcal{T})P(S), \quad (3.2)$$

where we assume the independence of  $\mathcal{T}$  and  $S$ . The term  $P(I, J|\mathcal{T}, S)$  is a joint likelihood of the images, which leads to the familiar similarity measure of SSD [141];  $P(\mathcal{T})$  is a prior used to regularize the transformation and  $P(S) \propto e^{-\beta\|\mathbf{P}S\|^2}$  is a prior on  $S$  that reflects our assumption on spatial intensity interactions [27]. Now, we also assume that pixel-wise probabilities are i.i.d., but only given the correction field. Maximization of the posterior probability in Equation 3.2 (excluding the transformation regularization<sup>2</sup>) is equivalent to minimization of the following objective function:

$$E(\mathbf{S}, \mathcal{T}) = \|\mathbf{I} - \mathbf{J}(\mathcal{T}) - \mathbf{S}\|^2 + \beta \|\mathbf{P}\mathbf{S}\|^2 \quad (3.3)$$

where  $\mathbf{I}, \mathbf{J}$  and  $\mathbf{S}$  are in column-vector form,  $\|\cdot\|$  is the Euclidean norm and  $\mathbf{P}$  is a regularization operator for  $\mathbf{S}$  (we have not yet specified the form of  $\mathbf{P}$ ). We can analytically solve for  $\mathbf{S}$ , and thus eliminate it from the objective function. Equating the gradient to zero, we solve for  $\mathbf{S}$ :

$$-2(\mathbf{I} - \mathbf{J}(\mathcal{T}) - \mathbf{S}) + 2\beta\mathbf{P}^T\mathbf{P}\mathbf{S} = 0 \quad (3.4)$$

$$\mathbf{S} = (\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\mathbf{r} \quad (3.5)$$

where  $\mathbf{Id}$  is the identity matrix and  $\mathbf{r} = \mathbf{I} - \mathbf{J}(\mathcal{T})$  is the residual vector (difference image) that explicitly depends on the transformation  $\mathcal{T}$ . The inverse always exists, because  $\mathbf{P}^T\mathbf{P}$  is positive semidefinite. Defining a new square symmetric matrix  $\mathbf{V} = (\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}$

---

<sup>2</sup>Depending on the transformation model  $\mathcal{T}$ , Equation 3.3 can also include the regularization on  $\mathcal{T}$  term, which comes from a prior  $P(\mathcal{T})$ .

and substituting  $\mathbf{S}$  back into the objective function (3.3), we obtain:

$$\begin{aligned} E(\mathcal{T}) &= \|\mathbf{r} - \mathbf{V}\mathbf{r}\|^2 + \beta \|\mathbf{P}\mathbf{V}\mathbf{r}\|^2 = \mathbf{r}^T ((\mathbf{Id} - \mathbf{V})^T (\mathbf{Id} - \mathbf{V})) \mathbf{r} + \beta \mathbf{r}^T (\mathbf{V}^T \mathbf{P}^T \mathbf{P} \mathbf{V}) \mathbf{r} = \\ &= \mathbf{r}^T (\mathbf{Id} - 2\mathbf{V} + \mathbf{V}^T (\mathbf{Id} + \beta \mathbf{P}^T \mathbf{P}) \mathbf{V}) \mathbf{r} = \mathbf{r}^T (\mathbf{Id} - 2\mathbf{V} + \mathbf{V}) \mathbf{r} = \mathbf{r}^T (\mathbf{Id} - \mathbf{V}) \mathbf{r} = \\ &= \mathbf{r}^T (\mathbf{Id} - (\mathbf{Id} + \beta \mathbf{P}^T \mathbf{P})^{-1}) \mathbf{r} \quad (3.6) \end{aligned}$$

To simplify further, consider the matrix  $\mathbf{P}^T \mathbf{P}$ ; it is square, symmetric and positive semidefinite. Thus, it allows spectral decomposition:

$$\mathbf{P}^T \mathbf{P} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T, \quad \mathbf{\Lambda} = \text{d}[\lambda_1, \dots, \lambda_N], \quad \lambda_i \geq 0 \quad (3.7)$$

where  $\mathbf{\Lambda}$  is the diagonal matrix of eigenvalues and  $\mathbf{Q}$  is the eigenvector matrix. Substituting Equation 3.7 in Equation 3.6, we can simplify the objective function greatly, because  $\mathbf{Q}$  is orthogonal, and the majority of multiplications and inversions are only with diagonal matrices within.

$$E(\mathcal{T}) = \mathbf{r}^T \mathbf{Q} \text{d} \left( 1 - \frac{1}{1 + \beta \lambda_i} \right) \mathbf{Q}^T \mathbf{r} = \mathbf{r}^T \mathbf{Q} \text{d} \left( \frac{\beta \lambda_i}{1 + \beta \lambda_i} \right) \mathbf{Q}^T \mathbf{r} = \mathbf{r}^T \mathbf{Q} \mathbf{L} \mathbf{Q}^T \mathbf{r} = \mathbf{r}^T \mathbf{A} \mathbf{r}. \quad (3.8)$$

where  $\text{d}()$  denotes a diagonal matrix. With  $\mathbf{A} = \mathbf{Q} \mathbf{L} \mathbf{Q}^T$  we defined a new square, symmetric and positive semidefinite matrix with eigenvalues  $\mathbf{L}$ :

$$\mathbf{L} = \text{d}(l_1, \dots, l_N) = \text{d} \left( \frac{\beta \lambda_i}{1 + \beta \lambda_i} \right), \quad 1 \geq l_i \geq 0 \quad (3.9)$$

Note that eigenvalues of  $\mathbf{A}$  are all nonnegative and bounded between  $[0, 1]$ . Now, if we choose a particular regularization operator  $\mathbf{P}$ , then  $\mathbf{A}$  is known and we can minimize

$$E(\mathcal{T}) = \mathbf{r}^T \mathbf{A} \mathbf{r} = (\mathbf{I} - \mathbf{J}(\mathcal{T}))^T \mathbf{A} (\mathbf{I} - \mathbf{J}(\mathcal{T})) \quad (3.10)$$

with respect to the transformation  $\mathcal{T}$ . The matrix  $\mathbf{A}$  is of size  $N^2$ , where  $N$  is a number of image pixels. This equation can be interpreted as the squared Mahalanobis distance between the two random vectors  $I$  and  $J$  of the same distribution with the inverse covariance  $\mathbf{A}$ . Alternatively, Equation 3.10 can be seen as the generalized distance measure, with the distance metric  $\mathbf{A}$ .



So far we have not specified the eigenvectors and eigenvalues of the operator  $\mathbf{P}^T\mathbf{P} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ . A large class of operators  $\mathbf{P}^T\mathbf{P}$  have the same eigenvector basis  $\mathbf{Q}$  and differ only in eigenvalues. In our method, we will choose a particular form of  $\mathbf{Q}$  (see Section 3.4) without specifying the eigenvalues, which are absorbed in a matrix  $\mathbf{L}$ . Instead, we estimate  $\mathbf{L}$  within the optimization framework. Such adaptive regularization allows greater flexibility with respect to the choice of  $\mathbf{P}$ . Consider solving for the eigenvalues matrix  $\mathbf{L}$ :

$$E(\mathbf{L}, \mathcal{T}) = \mathbf{r}^T \mathbf{A} \mathbf{r} = (\mathbf{Q}^T \mathbf{r})^T \mathbf{L} (\mathbf{Q}^T \mathbf{r}), \quad 1 \geq l_i \geq 0 \quad (3.11)$$

A trivial, but not interesting, minimum for  $E(\mathbf{L})$  is with an all-zeros matrix  $\mathbf{L}$ . To avoid this, we impose a regularization on the eigenvalues  $\mathbf{L}$  that prefers  $\mathbf{L}$  to be flat (has relatively equal elements) and large but still bounded in  $[0, 1]$ . The motivation is as follows. Notice that if  $\mathbf{L}$  is identity (or multiples of identity), then the inverse covariance  $\mathbf{A}$  is also identity, and the objective function simplifies to regular SSD:

$$E(\mathbf{L}, \mathcal{T}) = (\mathbf{Q}^T \mathbf{r})^T \mathbf{I} (\mathbf{Q}^T \mathbf{r}) = \mathbf{r}^T \mathbf{r} = \|\mathbf{r}\|^2 \quad (3.12)$$

which is equivalent to no intensity correction ( $\mathbf{S} = 0$ ). SSD is a valid and optimal objective function if the images are corrupted by i.i.d. Gaussian noise. The more the eigenspectra ( $\mathbf{L}$ ) deviate from flat, the more off-diagonal elements appear in  $\mathbf{A}$ , which means that the noise is more correlated. As far as the true noise covariance is unknown, we want to assume the least required correlation possible and thus a flat  $\mathbf{L}$ . We define the regularization term on  $\mathbf{L}$  as a Kullback-Leibler (KL) divergence between a uniform distribution  $p$  and a distribution given by eigenvalues  $\mathbf{L}$

$$R(\mathbf{L}) = KL(p||\mathbf{L}) = \sum_i p_i \log \frac{p_i}{l_i} + l_i - p_i \quad (3.13)$$

where the last two terms is a correction factor, so that it applies to unnormalized distributions [106, 192]. KL-based regularization enforces closeness of the data to satisfy an information measure rather than some distance measure (associated with its function space) [137]. Also it guarantees positivity for the regularized solution, in contrast to other regularizers, such as the  $L_2$  norm. Including the regularization term, our objective function takes the form

$$E(\mathbf{L}, \mathcal{T}) = (\mathbf{Q}^T \mathbf{r})^T \mathbf{L} (\mathbf{Q}^T \mathbf{r}) + \alpha R(\mathbf{L}) \quad (3.14)$$

where  $\alpha$  is a trade-off parameter. Differentiating the equation with respect to all  $l_i$  and equating the derivative to zero, we obtain

$$(\mathbf{q}_i^T \mathbf{r})^2 - \alpha \frac{1}{l_i} + \alpha = 0 \Rightarrow l_i = \frac{1}{(\mathbf{q}_i^T \mathbf{r})^2 / \alpha + 1} \quad (3.15)$$

where  $\mathbf{q}_i$  are eigenvectors in  $\mathbf{Q}$  ( $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_N]$ ). As a check we see that  $l_i$  are indeed bounded within  $[0, 1]$ . Substituting this result back into Equation 3.14, we obtain

$$E(\mathcal{T}) = \sum_{n=1}^N \left( \frac{(\mathbf{q}_n^T \mathbf{r})^2}{(\mathbf{q}_n^T \mathbf{r})^2 / \alpha + 1} - \alpha \log \frac{1}{(\mathbf{q}_n^T \mathbf{r})^2 / \alpha + 1} + \frac{\alpha}{(\mathbf{q}_n^T \mathbf{r})^2 / \alpha + 1} \right) = \alpha N - \alpha \sum_{n=1}^N \log \frac{1}{(\mathbf{q}_n^T \mathbf{r})^2 / \alpha + 1} \quad (3.16)$$

Ignoring the terms independent of  $\mathcal{T}$ , we obtain a novel similarity measure

$$E(\mathcal{T}) = \sum_{n=1}^N \log((\mathbf{q}_n^T \mathbf{r})^2 / \alpha + 1); \quad \mathbf{r} = (\mathbf{I} - \mathbf{J}(\mathcal{T})) \quad (3.17)$$

Note that even though we have derived it starting from solving for the correction field and its regularizer, they are not explicitly present in the final similarity function form. However, we still need to define the basis eigenvectors  $\mathbf{Q}$ .

### 3.4 DCT basis

We proceed to specify  $\mathbf{Q}$ . Recall that the functional form of basis  $\mathbf{Q}$  initially comes from the eigenvectors of  $\mathbf{P}^T \mathbf{P} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$ , where  $\mathbf{P}$  is a regularization operator of the correction field. We choose the discrete cosine transform (DCT) basis [161] as a functional form of  $\mathbf{Q}$ , because it corresponds to the discrete derivative-based regularizer  $\mathbf{P}$ . In this case, the matrix multiplication  $\mathbf{Q}^T \mathbf{r}$  is just a discrete cosine transform (DCT) of  $\mathbf{r}$ , which can be computed through FFT in  $\mathcal{O}(N \log N)$  [161].

Consider the first order derivative regularizer (finite differences matrix)  $\mathbf{P}_1$ , the corresponding self-adjoint operator  $\mathbf{P}_2 = \mathbf{P}_1^T \mathbf{P}_1$ , which appears after taking the gradient of  $\|\mathbf{P}_1 \mathbf{S}\|^2$ , is a second order derivative matrix (discrete Laplacian), which decomposes as [161]:

$$\mathbf{P}_2 = \mathbf{P}_1^T \mathbf{P}_1 = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \quad (3.18)$$

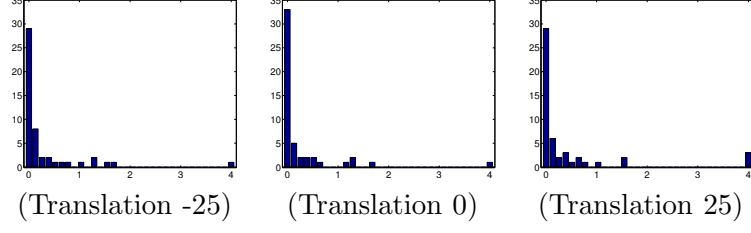


Figure 3.3: A comparison of the distributions of the DCT coefficients for the different spatial translations (pixels) of the images shown in Figure 3.1. Here, we computed each histogram from  $\log(\mathbf{x}^2 + 1)$ , where  $\mathbf{x}$  are the 50 largest DCT coefficients of the residual image for the given translation. The correct alignment (zero translation) corresponds to the maximum number of zero coefficients. By enforcing sparseness on the DCT coefficients of the residual image, the RC similarity measure achieves optimum at the correct spatial alignment.

where  $\mathbf{Q}$  is the DCT basis (real and orthogonal). Depending on a choice of boundary conditions and boundary approximations, there are several DCT choices. The most common in image processing is DCT-2, which corresponds to Neumann midpoint boundary conditions [161]. For instance, in 1D the DCT basis functions are

$$\mathbf{q}_n(k) = \frac{w_n}{\sqrt{N}} \cos\left(\frac{\pi(2k-1)(n-1)}{2N}\right), \quad (3.19)$$

for  $k = 1..N$ ,  $n = 1..N$ , and

$$w_n = \begin{cases} 1 & \text{for } n = 1, \\ \sqrt{2} & \text{for } n = 2..N. \end{cases} \quad (3.20)$$

If we use the second order regularizer  $\mathbf{P}_2$ , the corresponding self-adjoint operator,  $\mathbf{P}_2^T \mathbf{P}_2$ , is the fourth order discrete derivative operator  $\mathbf{P}_4 = \mathbf{P}_2^T \mathbf{P}_2 = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T = \mathbf{Q} \mathbf{\Lambda}^2 \mathbf{Q}^T$ , with same set of eigenvectors as  $\mathbf{P}_2$ . Similarly, derivative operators of higher orders and their linear combinations lead to the same eigenvectors  $\mathbf{Q}$  and differ only in eigenvalues.

### 3.5 Analysis of Residual Complexity

Low complexity of a signal expressed in the DCT basis corresponds to a small number of non-zero coefficients. Our similarity measure penalizes small DCT coefficients disproportionately stronger than the large ones. Thus, it encourages the presence of few large

coefficients opposed to many small ones. Let  $\mathbf{x} = \mathbf{Q}^T \mathbf{r}$  be the DCT coefficients of the residual image. Intuitively, the image can be highly compressed if only a few coefficients are non-zero. The term  $\sum \log(\mathbf{x}^2 + 1)$  indeed forces sparseness of the DCT coefficients of the residual image, because the logarithm function decreases quickly to zero compared to its increase for larger values of the DCT coefficients.

Figure 3.3 shows the histogram of the DCT coefficients of the residual image for the different spatial translation positions from the example in Figure 3.1. At the correct spatial alignment, zero translation, the residual image has the maximum number of (almost) zero DCT coefficients. By enforcing sparseness of the DCT coefficients, RC reaches the optimum that corresponds to the correct alignment.

Several other sparseness measures have been proposed for image and speech analysis, neural networks, independent component analysis, sparse coding, etc. [81, 125]. Popular sparseness measures include:

- $|\mathbf{x}|^b$ , for  $b \leq 1$
- $\log(\mathbf{x}^2 + 1)$
- $\tanh(|a\mathbf{x}|^b)$ , for  $b > 1$

In our case the signal  $\mathbf{x}$  represents the DCT coefficient of the residual image. Karvanen and Chichocki showed no consistent advantage of one sparseness measure over the others [81]. The L1 norm  $|\mathbf{x}|$  perhaps is the most popular sparseness measure due to its simple representation. The tanh function has a nice property of saturating to 1 when  $|x| \rightarrow \infty$ . Our sparseness measure  $\log(\mathbf{x}^2 + 1)$  comes from the formulation and analytical derivation described in Section 3.3.

Some other similarity measures, including SSD and entropy of the difference (EOD) image [132], also define an objective function on the residual image. The SSD similarity measure favors all zeros residual regardless of the spatial positions of misaligned structures. EOD is optimized when the residual image is clustered into a small number of intensity levels, regardless of the spatial position. In contrast to SSD and EOD, the RC similarity measure is minimized when the residual image can be sparsely represented using only a

few basis functions (e.g. favoring spatially smooth or piece-wise smooth residual). Such properties of RC allow robust alignment of the images distorted by the spatially-varying intensity distortions.

**Analogy with sparse coding:** In sparse coding [125], one tries to decompose images in terms of basis functions  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]$  and sparse coefficients  $\mathbf{c} = [c_1, \dots, c_N]^T$ , by minimizing

$$E(\mathbf{W}, \mathbf{c}) = \sum_i \|\mathbf{x}^i - \mathbf{W}\mathbf{c}\|^2 + \lambda \sum_n \log(1 + (\frac{c_n}{\sigma})^2) \quad (3.21)$$

where  $\mathbf{x}^i$  are vectorized image patches, and the bases ( $\mathbf{W}$ ) are usually overcomplete and nonorthogonal. The last term measures the sparseness of coefficients  $\mathbf{c}$  [125], which is similar to our objective function in Equation 3.17. Variations of this formulation lead to principal components, independent components, or more specialized filters [178].

Welling et al. [178] proposed to learn the overcomplete set of filters  $\mathbf{J} = [\mathbf{J}_1, \dots, \mathbf{J}_N]$ , using the product of experts (PoE) framework. PoE is an energy based method, which defines a probability of  $\mathbf{x}$  as a normalized product of all the distributions represented by the individual experts. Unnormalized experts are chosen to be Student-t distributions, because this distribution has heavy tails, which makes it a suitable candidate for modeling constraints that are found in natural images. The energy of the PoE model is

$$E(\alpha, \mathbf{J}) = - \sum_{i=1}^M \alpha_i \log \frac{1}{(1 + \frac{1}{2}(\mathbf{J}_i^T \mathbf{x})^2)} + \log(Z) \quad (3.22)$$

which is again similar to our objective function in Equation 3.17, but for a different purpose. The estimated filters  $\mathbf{J}_i$  are further applied for image denoising and inpainting tasks [178].

Whereas in sparse coding, the task is to search for sparse bases and coefficients to represent the image, in image registration we are given the bases and looking for the alignment where the residual image has sparse representation by the bases.

### 3.6 Implementation

We model the transformation  $\mathcal{T}$  using the Free Form Deformation (FFD) transformation with 3 hierarchical levels of B-spline control points [146]. We use the gradient descent optimization method to iteratively update the transformation parameters  $\mathcal{T}$ . The pseudo-code to compute the objective function and its gradient is

$$\mathbf{r} = \mathbf{I} - \mathbf{J}(\mathcal{T}); \mathbf{c} = \text{dctn}(\mathbf{r}); E = \sum \log(\mathbf{c}^2/\alpha + 1);$$

$$\nabla E = -\text{idctn}\left(\frac{2\mathbf{c}/\alpha}{\mathbf{c}^2/\alpha + 1}\right) \nabla \mathbf{J}(\mathcal{T}) \frac{\partial \mathcal{T}}{\partial \theta}$$

where *dctn* and *idctn* are the forward and inverse multidimensional DCTs,  $\nabla \mathbf{J}$  is the intensity image gradient and  $\theta$  represents the transformation parameters. We set the parameter  $\alpha = 0.05$ . We have also used an addition regularization term for FFD; we penalized the Euclidean distance between all the neighboring displacements of B-spline control points to prevent unnatural wraps. The weight of the regularization term was set to 0.01, which we empirically determined to give the best performance.

### 3.7 Results

We have implemented the algorithm in Matlab, and tested it on a AMD Opteron CPU 2GHz Linux machine with 4GB RAM. We show the performance of the Residual Complexity similarity measure on several challenging synthetic and real-life examples. For the synthetic experiments, we simulated additive and multiplicative spatially varying intensity distortion fields. Such distortion fields are consistent with the common assumptions in medical imaging, e.g. MRI can be corrupted a multiplicative bias field. We compare our similarity measure with the local CC and local MI similarity measures implemented in Elastix [84, 86].

#### 3.7.1 Elastix Configuration

Elastix is an image registration software toolbox. It provides a broad range of intensity-based registration options and settings. We used B-spline parametrization of non-rigid

transformation with 3 hierarchical levels. We used mesh size of 48, 24 and 12 pixels respectively for both our and Elastix implementations.

Elastix includes local similarity measure implementations (local CC and local MI), which are well suited to cope with image intensity inhomogeneity and non-stationarity. At each iteration Elastix randomly selects one voxel, and then selects a set of random samples in a local square region around that voxel [84, 86]. From the local region, the local similarity measure is computed and used to update the transformation using the Gradient Descent optimization method. Empirically, we found a  $15 \times 15$  region size with 1000 random samples<sup>3</sup> to produce the best results in our synthetic experiments. We set a maximum number of iterations of 2000 for each of Elastix hierarchical levels. The maximum number of iteration was also the only stopping condition available in the Elastix implementation. We have empirically found the 2000 iterations to be sufficient to achieve an accurate registration result with no further improvements.

### 3.7.2 Synthetic Experiment 1

We used the BrainWeb T1-weighted MRI images [35]. We selected a 2D slice ( $216 \times 180$ ) for the experiments. We normalized image intensities to the  $[0, 1]$  interval before registration. We generated synthetic examples by introducing both geometric and intensity distortions to the images. To simulate the geometric distortion, we perturbed a uniform grid of points followed by the thin-plate spline (TPS) interpolation of the source image according to the grid deformation to create the target images. The grid size was  $7 \times 7$  and its random perturbation was drawn from a zero-mean Gaussian with 3 pixels standard deviation (std). To simulate the spatially-varying intensity distortion we corrupted both images according to the formula

- $I(x, y) = I(x, y) + \frac{1}{K} \sum_{k=1}^K e^{-\frac{\| [x; y] - \mu_k \|^2}{2(30)^2}};$
- Rescale  $I$  to  $[0, 1]$ .

The last term models locally-varying intensity field with mixture of  $K$  randomly centered Gaussians.

---

<sup>3</sup>A large number of samples has been chosen to ensure the full information use from the neighborhood.

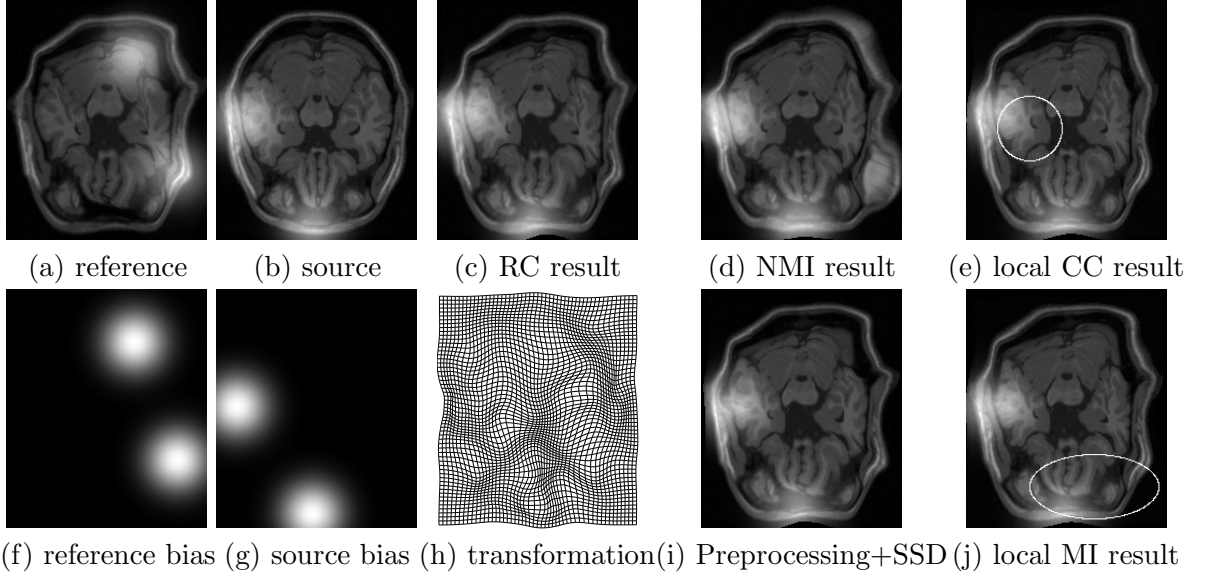


Figure 3.4: Synthetic experiment 1. We register the source image (b) onto the reference image (a). (f,g) Intensity distortion fields added to the reference and source images respectively. (d) The RC registration result. RC performs well. (d,e,i,j) Typical performance of global NMI, local CC, local MI and Preprocessing+SSD respectively. NMI shows unsatisfactory result. Local CC and local MI performs satisfactorily with slight misalignments, which we highlighted with contours (e,j). Preprocessing+SSD approach produces overall satisfactory result, but with several misalignments of the smaller structures.

Figure 3.4 shows typical registration results for  $K = 2$  (two Gaussians). The RC similarity measure showed accurate performance. Global normalized MI showed poor performance with significant misalignment. We have also tested some other global measures including CC, CR and SSD, all of which performed poorly. Local similarity measures performed well with a slight advantage of local CC over the local MI. Both local CC and local MI accurately aligns most of the image regions. We also tested the performance of SSD after both images were Gaussian high pass filtered with 10 std, which led to an overall satisfactory alignment, except for some smaller structures that still remained misaligned.

To quantify the registration performance, we used two different error measures. The first one is the transformation root mean square error (RMSE) between the true and estimated transformations:  $\varepsilon_{RMSE} = \sqrt{\frac{1}{N} \sum \|\mathcal{T}_{true} - \mathcal{T}_{estimated}\|^2}$ . We did not include the area outside the skull (roughly found by thresholding) for the transformation  $RMSE$  computation to avoid the boundary error influence. The second error is the intensity



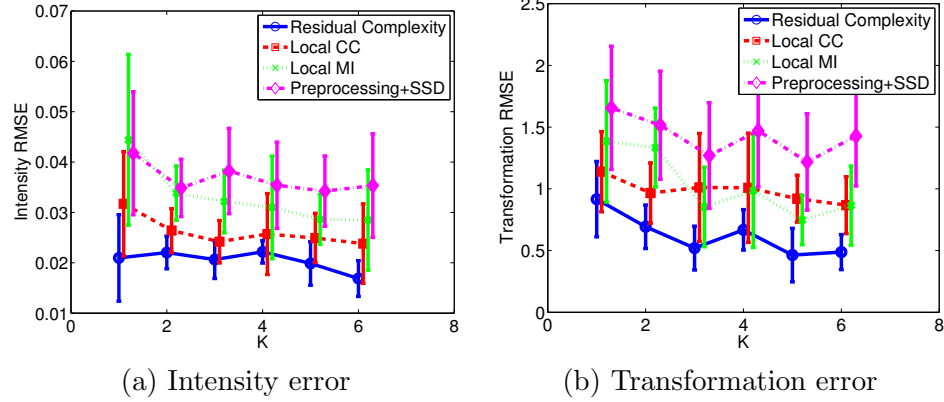


Figure 3.5: Registration performances of RC, local CC, local MI and Preprocessing+SSD. Both images were corrupted by additive intensity field (simulated by a mixture of  $K$  random Gaussians) and scaled. RC performs comparably well, often better, with the other tested similarity measures. Local CC and local MI produce similar satisfactory results. Preprocessing+SSD approach showed the worst performance.

RMSE between the reference and the clean registered images, where the clean registered image was obtained by applying the estimated transformation  $\mathcal{T}$  to the source image without intensity distortion.

We performed 120 automatic registrations for  $K = [1..6]$ . The random intensity and spatial distortions were reinitialized at every run. Figure 3.5 demonstrates the registration performances of RC, local CC, local MI and Preprocessing+SSD. For the Preprocessing+SSD, we preprocessed the images with a Gaussian high pass filter (10 std) followed by SSD-based registration. RC performed well and outperformed the local CC and local MI similarity measures. All tested similarity measures produced satisfactory results with subpixel accuracy. Nevertheless, visually local CC and local MI tend to align all image regions except for one or two problematic regions, which remained misaligned. The average processing time with RC was 20 sec for a pair of images, whereas local MI and local CC Elastix implementations took around 60 sec.

We also tested the sensitivity of RC with respect to the choice of the  $\alpha$  parameter. We conducted 50 runs with random deformation and intensity distortion ( $K = 3$ ) initializations for each of  $\alpha = [0.01, 3]$ . Figure 3.6 shows the mean and standard deviation of the Transformation RMSE obtained for various  $\alpha$ . Overall, the values of  $\alpha$  in a range  $[0.01, 3]$

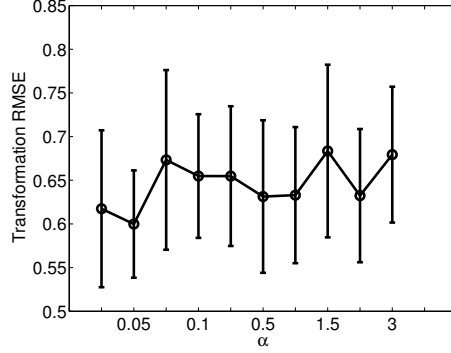


Figure 3.6: Transformation error with respect to the different parameter  $\alpha$  settings.

produce similar accurate registration results with a small advantage when  $\alpha = 0.05$ .

### 3.7.3 Synthetic Experiment 2

We corrupted both images with a multiplicative intensity field, otherwise using similar set-up as in the synthetic experiment 1. We simulated the spatially-varying multiplicative intensity distortion according to the formula

- $I(x, y) = I(x, y) \cdot \left( 0.2 + \frac{1}{K} \sum_{k=1}^K e^{-\frac{\| [x; y] - \mu_k \|^2}{2(30)^2}} \right)$ ;
- Rescale  $I$  to  $[0, 1]$ .

Figure 3.7 demonstrates the registration performances of RC, local CC, local MI and Preprocessing+SSD. The registration performances are comparable with a slight advantage of RC, and definite disadvantage of Preprocessing+SSD. Interestingly local MI shows a comparable performance to RC in terms of transformation RMSE, but larger error in terms of intensity RMSE between the clean reference and the final registered images. Visually both RC, local CC and local MI produced accurate results.

### 3.7.4 Retina Images

We registered retina images taken 2 years apart [189]. Retinal images are used in ophthalmology to assess the evolution of illness, e.g. diabetic retinopathy [148, 123]. The images show the vascular tree of the eye, which permits to determine the areas where blood flow seems occluded or leaking. For this, the images have to be aligned first.

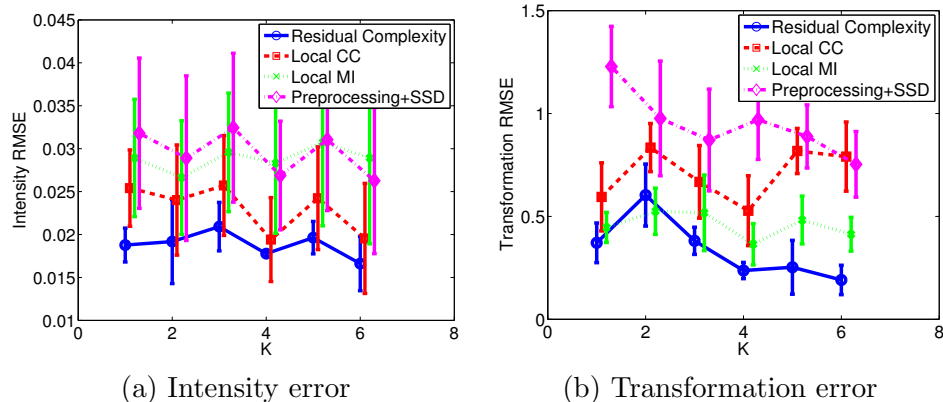


Figure 3.7: Registration performances of RC, local CC, local MI and Preprocessing+SSD. Both images were corrupted by multiplicative intensity field. (simulated by a mixture of  $K$  random Gaussians) and scaled. Artificial spatial and intensity distortions were reinitialized at every run. RC performs comparably well, often better, with the other tested similarity measures in almost all cases.

Retina images are challenging to register due to multiple intensity artifacts including non-uniform background and blood vessels (with intensity variations and changed structure) [123]. Most of the retina registration methods are feature-based [148, 189]; they extract the vascular structure or landmarks and align them. Figure 3.8 demonstrates the RC performance on 2 retina images ( $200 \times 250$  pixels). We achieved accurate registration results as demonstrated through composite views. We did not do image preprocessing or normalization and did not excluded the background outside the retina circle. We used the images as they appear in the figure and did no rigid pre-registration beforehand. Both global and local CC and MI similarity measures (see Elastix configuration in Section 3.7.1) produced similar registration results, but only after we defined the region of interest to exclude the area outside the retina circle and after image filtering and heavy regularization of the transformation (to the extent where the transformation is almost rigid).

### 3.7.5 Iris Images

We stabilize a video sequence of microscopic iris images through frame-by-frame registration. This was necessary to remove the severe jitter and deformation across frames in order to be able to track leukocyte motion [159]. The deformation between frames is highly nonlinear. Notice the intensity variation across the images as well as intensity

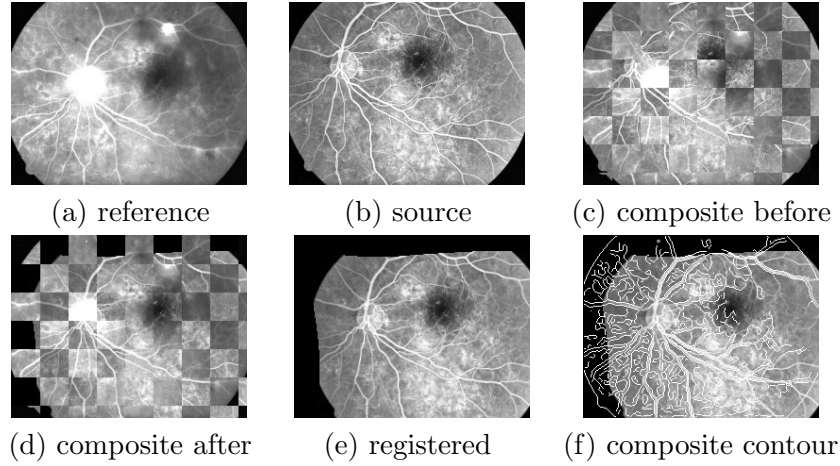


Figure 3.8: Registration of retina images. (a) reference image; (b) source image (taken 2 years prior); (c) composite view before registration; (d) composite view after registration; (e) registered source image; (f) composite view through contour overlap after the registration. The RC registration result is accurate; vessel structures are aligned despite intensity distortions and artifacts.

artifacts from moving cells, nonconsistent vessel reflections, etc. Our algorithm proved accurate and effective for these images (Figure 3.9). Other similarity measures, including local CC and local MI showed poor performance on iris images without preprocessing, even though different combinations of parameters were tried.

### 3.7.6 3D Echocardiography

We sequentially registered a set of 3D echo images (24 frames  $192 \times 274 \times 248$  pixels) to find a displacement field of imaged area through time [115]. The idea is to register the next 3D volume onto the previous, sequentially, in order to track the underlying myocardial motion. We have used the similar set up for each pair of 3D volumes as described in Section 3.6. The estimated transformation field can be used to study the myocardial deformation. We acquired the 3D echo sequences using a Philips iE33 with EKG gating from openchest piglets. Figure 3.10 shows the estimated motion of LV superimposed on the 3D echo images and the displacement vector field visualized at the LV endocardium position between end diastole and end systole volumes. We used a groundtruth sonomicrometer motion

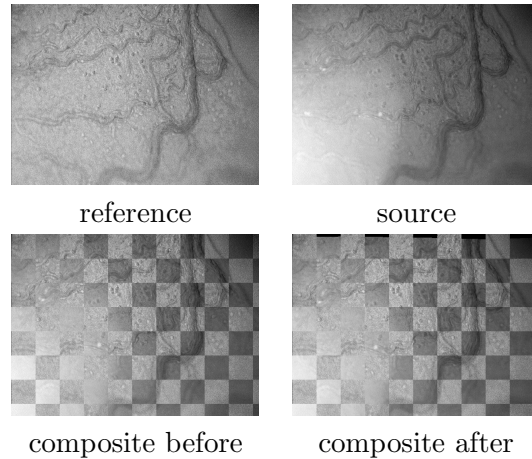


Figure 3.9: Registration of two iris images. Notice the intensity variation across the images as well as intensity artifacts including moving cells and nonconsistent vessel reflections. Our method accurately registers the images without any additional preprocessing.

to validate the algorithm. The estimated motion showed a high correlation with the groundtruth one, which shows the high accuracy of our method [115].

### 3.8 Non-rigid Stabilization of Time Lapse Videos

In this section we introduce a system for non-rigid stabilization of time lapse videos. The RC similarity measure plays a key role in stabilization of such videos, which have many illumination artifacts and slow-varying intensity distortions.

The videos were taken of sedated murine eyes infected with uveitis (inflammation of the uveal tract), with static microscopic camera looking at a portion of the iris. Videos were taken at 3 frames per minute rate for 30 to 90 minutes. The images are monotone of size  $720 \times 480$ . The motion artifacts are caused by wandering of the eye, dilation and contraction of the pupil, head and respiration motion, and sometimes refocusing of the camera during imaging. The net result is jitter and distortion (both spatial and intensity) in the image plane, which subsequently obscure the T-cell motion.

The motion patterns of leukocytes, specifically T-cells, are directly related to the cellular and chemical environment at the site of eye inflammation and can reveal underlying

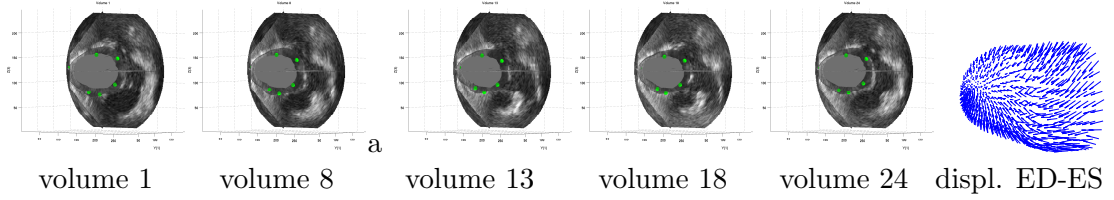


Figure 3.10: Estimated motion of the LV contour found during the cardiac cycle (24 consecutive volumes). LV achieves the maximum contraction at volume 13, then the LV dilates (diastolic phase) up to the volume 22 and starts contracting again (systolic phase). The last plot shows the displacement fields on LV endocardium between end diastole (ED) and end systole (ES) volumes. We validate the accuracy of RC registration with groundtruth motion of implanted sonomicrometers. The RC registration-based estimated motion is accurate.

disease mechanisms [10, 9]. Visualization of T-cell behavior in disease models is made possible by the state-of-the-art intravital microscopy technology which can record cellular activity in the eye in vivo without resorting to any surgical trauma [144]. These records (videos) have revealed patterns of T-cell migration in the ocular uveal tract, indicating a complexity in immune responses that has not been closely examined before. By studying these microscopic videos, we can expect to characterize the migration of T-cells within the iris stroma in disease models.

Groupwise image registration (see Sec. 2.5) is needed to stabilize and rectify the image sequences, which will pave the way for subsequent T-cell tracking. The deformation across frames are locally non-linear, and it is not feasible to obtain an accurate closed-form deformation model by examining the image formation process. In addition, the motion artifacts can cause the region on the iris being imaged to go in and out of depth of field of the camera, resulting in local blurring and local intensity instability.

The first step in video stabilization is to compensate for global motion. In particular we compensate for the translation, because our videos have significant view-area shifts between the frames. We find the translation between every two consecutive pair of video frames and then transform all image frames into the same coordinate system. The translation vector for a particular image is given by a cumulative sum of all inter-frame translations prior to the given frame. To find the translation between two images we compute the normalized cross-correlation function, which is the normalized CC computed

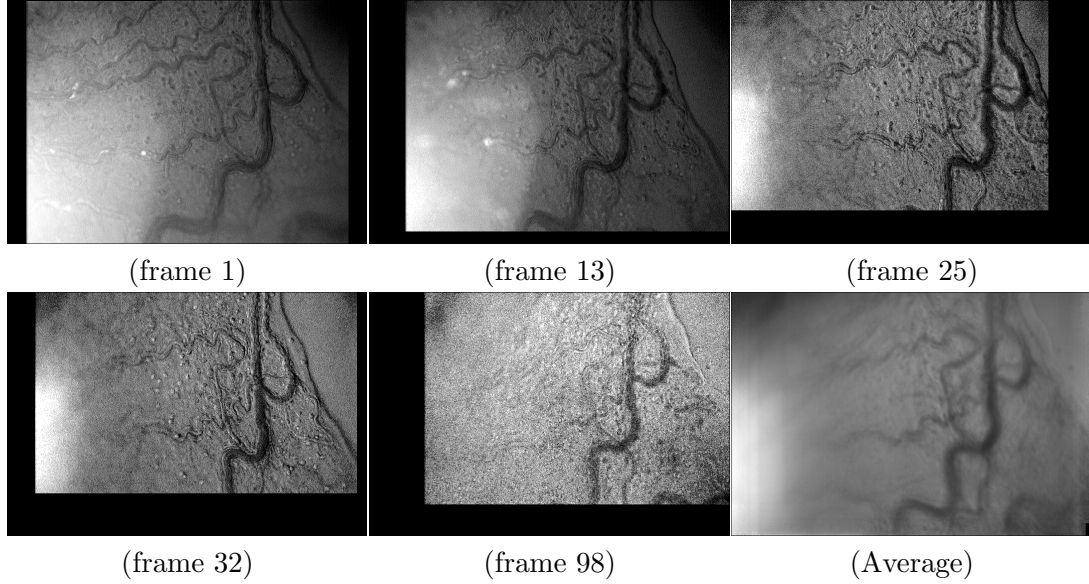


Figure 3.11: Typical translation compensated image frames from the first video and the average image. Notice local intensity inconsistency, blurring and moving cells. The average image has blurred boundaries and a low contrast, which indicates non-rigid image misalignments.

for all  $(x, y)$  image shifts. We use the coordinates of the maximum of the computed cross-correlation function as the translation vector. Such an exhaustive search is much more reliable compared to the gradient based optimization. As far the parameter space is small ( $x$  and  $y$  shift in  $2D$  plane), we can compensate for the translation fast. For a video of a 100 frames, the translation compensation takes around 1 minute. Figure 3.11 shows several translation compensated video frames of the example video 1.

The second step in video stabilization is to compensate for local deformations. We use the sequential image registration approach. We first register and transform the second image frame into the coordinate system of the first one. Then, we use the obtained transformation as the initialization to search for the transformation that aligns the third frame to the (registered) second frame. Similarly, we can sequentially map all image frames into the same coordinate system of the first frame. We emphasize that we do not register all image frames to the first one, we register the consecutive frames and then map the result into the coordinate system of the first one. For the inter-frame (pair-wise) registration, we use FFD parametrization with 3 hierarchical levels and RC similarity

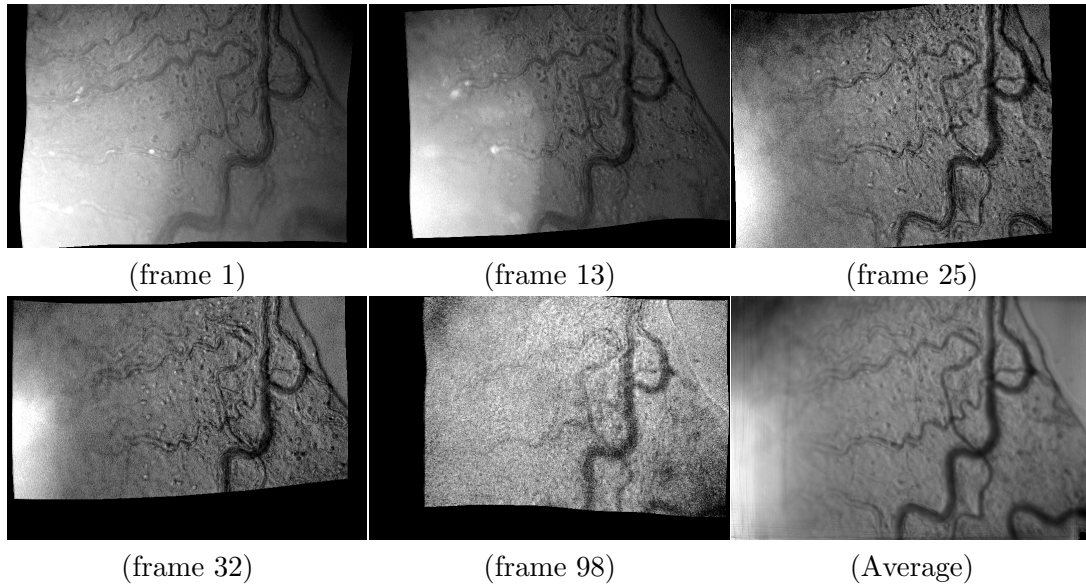


Figure 3.12: The registered image frames from the first video after applying the transformation found using the groupwise non-rigid registration. The average image has much sharper boundaries than prior to registration, which indicates accurate image alignment.

measure. Because the image frames are large,  $720 \times 480$ , registration of the 100 frames sequence takes around 1 hour. Often the final stabilization result is sufficiently accurate. However, in some cases an additional fine tuning is required.

To fine tune the non-rigid stabilization result, we proceed to register all image frames onto the average (mean) image. We compute the arithmetical mean for each of the image pixels. Only the images whose domain (visible area) exist for a given pixel are used during averaging. We then register all image frames onto the average image, update the average and continue in this fashion until convergence. We use the RC similarity measure. Such groupwise non-rigid registration is accurate because all image frames have full overlap with the average image. Also the average image has less noise and little artifacts, such as moving cells, due to averaging. Such an “all-to-the-mean” registration can be very computationally expensive, as it can take many cycles for the average image to converge. Without preliminary sequential non-rigid registration, such groupwise approach takes around 10 hours. After pre-registration, it takes only around 1 hour for a video of 100 frames.

Figure 3.12 demonstrates the non-rigidly registered image frames from the first video,



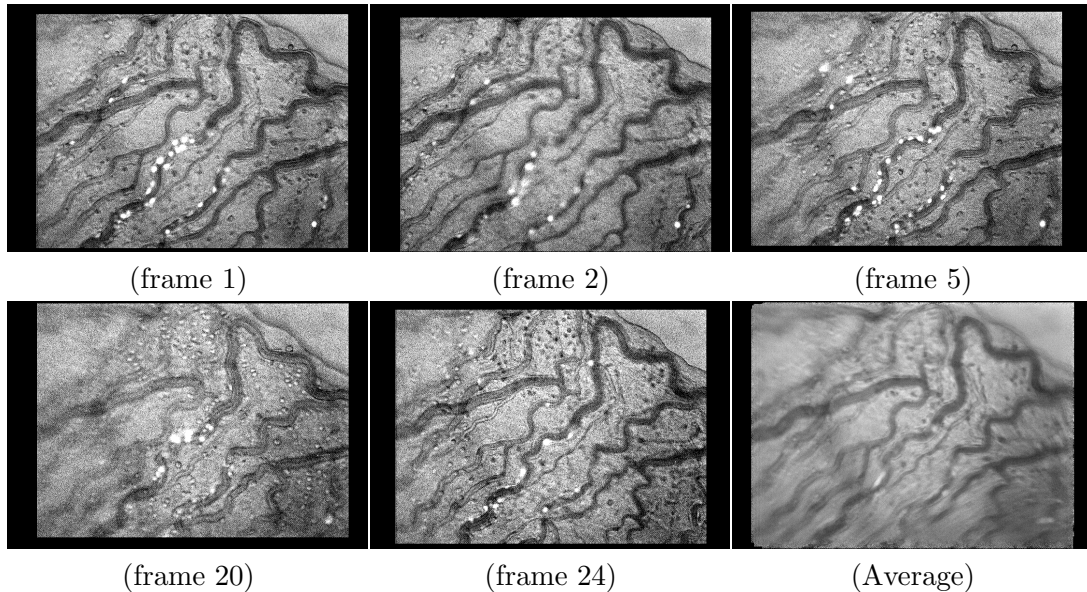


Figure 3.13: Typical image frames from the second video and the average image. Notice local intensity inconsistency, blurring and moving cells. The average image has blurred boundaries and a low contrast, which indicates non-rigid image misalignment.

as well as the average image. Notice that the average image after the registration have sharp vessel structures with little blurring, which shows the accuracy of the registration.

We also show the non-rigid stabilization of the second video example. Figure 3.13 shows several frames from the second video after translation compensation. Notice, high local area deformations, which results in significant average image blur. Figure 3.14 demonstrates the accurate non-rigid stabilization result, because the average image has sharp structures with little blur. We emphasize that utilization of the RC similarity measure in stabilization of such microscopic videos was vital. We have also tested this stabilization approach with other similarity measures, including SSD, SAD and CC, after advanced preprocessing to remove image artifacts and normalize the intensity. Only the RC allows to achieve stable and accurate registration result.

### 3.9 Discussion and Conclusion

We have derived a novel similarity measure for image registration, which accounts for spatially-varying intensity distortions. The similarity measure intuitively measures the

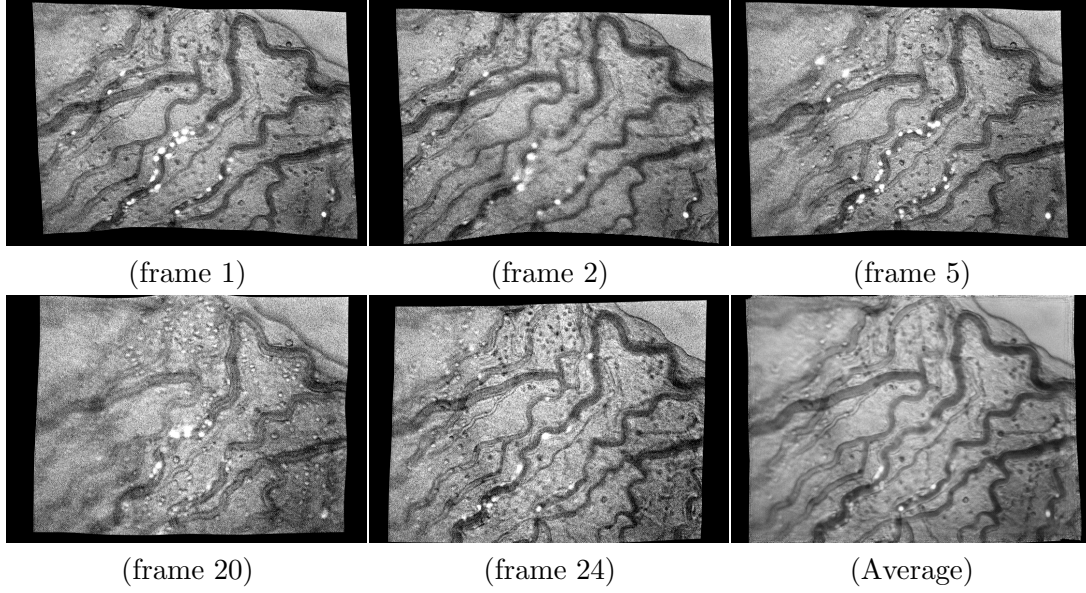


Figure 3.14: The registered image frames from the second video after applying the transformation found during the group-wise non-rigid registration. The average image has much sharper boundaries than prior to registration, which indicates accurate image alignment.

coding complexity of the residual image. This measure produces accurate registration results on both artificial and real-world problems that we have tested, and outperforms other state-of-the-art similarity measures, in term of accuracy, in these cases.

The key advantage of our new similarity measure is its simplicity in terms of both computational complexity ( $\mathcal{O}(N \log N)$ ) and implementation. The similarity measure and its pixel-wise gradient can be implemented within several Matlab lines. Nevertheless, our new similarity measure shows robust and accurate performance on real-word tested examples corrupted by severe intensity distotion fields. The limitation of our approach is that it is only applicable to areas where images are from the same modality.

Our similarity measure requires initialization of the coefficient  $\alpha$ . For the examples we tested, we found its value in a range  $[0.01, 3]$  to give accurate results with no significant difference in performance. Generally smaller values of  $\alpha$  forced sparser coefficients, which resulted in more accurate registration, but also lead to more local minima of the similarity measure.

Our underlying assumption on the correction field was a simple additive one. Nevertheless, our similarity measure showed robust performance for both synthetic and real examples, where the intensity correction function can be complex. Our explanation for the accurate performance of RC in such cases is that even when the real residual image is complex, its form still has to be simpler (represented sparsely by the basis functions) than the one with inaccurate registration. Thus, RC will still work with some complex intensity distortions. We discuss the future directions to improve the RC similarity measure in Section 6.2.

## Chapter 4

# Point Set Registration: Coherent Point Drift

Point set registration is a key component in many computer vision tasks. The goal of point set registration is to assign correspondences between two sets of points and to recover the transformation that maps one point set to the other. Multiple factors, including an unknown non-rigid spatial transformation, large dimensionality of point set, noise and outliers, make the point set registration a challenging problem. We introduce a probabilistic method, called the Coherent Point Drift (CPD) algorithm, for both rigid and non-rigid point set registration. We consider the alignment of two point sets as a probability density estimation problem. We fit the Gaussian mixture model (GMM) centroids (representing the first point set) to the data (the second point set) by maximizing the likelihood. We force the GMM centroids to move *coherently* as a group to preserve the topological structure of the point sets. In the rigid case, we impose the coherence constraint by re-parametrization of GMM centroid locations with rigid parameters and derive a closed form solution of the maximization step of the EM algorithm in arbitrary dimensions. In the non-rigid case, we impose the coherence constraint by regularizing the displacement field and using the variational calculus to derive the optimal transformation. We also introduce a fast algorithm that reduces the method computation complexity to linear. We test the CPD algorithm for both rigid and non-rigid transformations in the presence of noise, outliers and missing points, where CPD shows accurate results and outperforms current state-of-the-art methods.

## 4.1 Introduction

Registration of point sets is a key component in many computer vision tasks including stereo matching, content-based image retrieval, image registration and shape recognition. The goal of point set registration is to assign correspondences between two sets of points and/or to recover the transformation that maps one point set to the other. For example, in stereo matching, in order to recover depth and infer structure from a pair of stereo images, it is necessary to first define a set of points in each image and find the correspondence between them. An example of point set registration problem is shown in Fig. 4.1. The “points” in a point set are often features extracted from an image, such as the locations of corners, boundary points or salient regions. The points can represent both geometric and intensity characteristics of an image.

Practical point set registration algorithms should have several desirable properties: (1) Ability to accurately model the transformation required to align the point sets with tractable computational complexity; (2) Ability to handle possibly high dimensionality of the point sets; (3) Robustness to degradations such as noise, outliers and missing points that occur due to imperfect image acquisition and feature extraction.

The transformation usually falls into two categories: rigid or non-rigid. A rigid transformation allows only for translation, rotation and scaling. The simplest non-rigid transformation is affine, which also allows anisotropic scaling and skews. Non-rigid transformation occurs in many real-world problems including deformable motion tracking, shape recognition and medical image registration. The true underlying non-rigid transformation model is often unknown and challenging to model. Simplistic approximations of the true non-rigid transformation, including piece-wise affine and polynomial models, are often inadequate for correct alignment and can produce erroneous correspondences. Due to the usually large number of transformation parameters, the non-rigid point set registration methods tend to be sensitive to noise and outliers and are likely to converge into local minima. They also tend to have a high computational complexity. A practical non-rigid point set registration method should be able to accurately model the non-rigid transformation with tractable computational complexity.

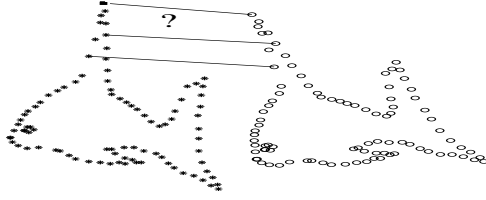


Figure 4.1: The point set registration problem: Given two sets of points, assign the correspondences and the transformation that maps one point set to the other.

Multidimensional point sets are common in many real world problems. Most current rigid and non-rigid point sets registration algorithm are well suited for 2D and 3D cases, but their generalization to higher dimensions are not always trivial. Furthermore, degradations such as noise, outliers and missing points significantly complicates the problem. Outliers are the points that are incorrectly extracted from the image; outliers have no correspondences in the other point set. Missing points are the features that are not found in the image due to occlusion or inaccurate feature extraction. A point set registration method should be robust to these degradations.

We present a robust probabilistic multidimensional point set registration algorithm for both rigid and non-rigid transforms. We consider the alignment of two point sets as a probability density estimation problem, where one point set represents the Gaussian Mixture Model (GMM) centroids, and the other one represents the data points. We fit the GMM centroids to the data by maximizing the likelihood. At the optimum, the point sets become aligned and the correspondence is obtained using the posterior probabilities of the GMM components. Core to our method is to force GMM centroids to move *coherently* as a group, which preserves the topological structure of the point sets. We impose the coherence constraint by explicit re-parametrization of GMM centroid locations (for rigid and affine transformations) or by regularization of the displacement field (for smooth non-rigid transformation). We also show how the computational complexity of the method can be reduced to linear, which makes it applicable for large data sets. The rest of the chapter is organized as follows. In Section 4.2, we overview the current rigid and non-rigid point set registration methods and state our contributions. In Section 4.3, we formulate a probabilistic point set registration framework. In Sections 4.4 and 4.5, we

describe our algorithms for rigid, affine and non-rigid registration cases, and relate them to existing works. In Section 4.6, we discuss the computational complexity of the method and introduce its fast implementation. In Section 4.7, we evaluate the performance of our algorithm. Section 4.8 concludes with some discussions.

## 4.2 Previous Work

Many algorithms exist for rigid and for non-rigid point set registration. They aim to recover the correspondence or the transformation required to align the point sets or both. Many algorithms involve a dual-step update, iteratively alternating between the correspondence and the transformation estimation. Here, we briefly overview the rigid and non-rigid point set registration methods and state our contributions.

### 4.2.1 Rigid Point Set Registration Methods

Iterative Closest Point (ICP) algorithm, introduced by Besl and McKay [13] and Zhang [190], is the most popular method for rigid point set registration due to its simplicity and low computational complexity. ICP iteratively assigns correspondences based on the closest distance criterion and finds the least-squares rigid transformation relating the two point sets. The algorithm then redetermines the correspondences and continues until it reaches the local minimum. Many variants of ICP have been proposed that affect all phases of the algorithm from the selection and matching of points to the minimization strategy [53, 147]. ICP requires that the initial position of the two point sets be adequately close.

To overcome the ICP limitations, many probabilistic methods were developed [136, 96]. These methods use soft-assignment of correspondences that establishes correspondences between all combinations of points according to some probability, which generalizes the binary assignment of correspondences in ICP. Among these methods are Robust Point Matching (RPM) algorithm introduced by Gold et al. [56], and its later variants [136, 32]. In [31] it was shown that in RPM alternating soft-assignment of correspondences and transformation is equivalent to the Expectation Maximization (EM) algorithm for GMM, where one point sets is treated as GMM centroids with equal isotropic covariances and the

other point set is treated as data points. In fact, several rigid point set methods, including Joshi and Lee [77], Wells [179], Cross and Hancock [41], Luo and Hancock [96, 98], McNeill and Vijayakumar [103] explicitly formulate point set registration as a maximum likelihood (ML) estimation problem, to fit the GMM centroids to the data points. These methods re-parameterize GMM centroids by a set of rigid transformation parameters (translation and rotation). The EM algorithm used for optimization of the likelihood function consists of two steps: E-step to compute the probabilities, M-step to update the transformation. Common to such probabilistic methods is the inclusion of an extra distribution term to account for outliers (e.g. large Gaussian [136] or uniform distribution [179]) and deterministic annealing on the Gaussian width to avoid poor local minima. These probabilistic methods perform better than conventional ICP, especially in presence of noise and outliers.

Another class of rigid point sets registration methods are the spectral methods. Scott and Longuet-Higgins [151] introduced a non-iterative algorithm to associate points of two arbitrary patterns, exploiting some properties of Gaussian proximity matrix (Gram matrix) of point sets. The algorithm works well with translation, shearing and scaling deformations, but performs poorly with non-rigid transformations. Li and Hartley showed that correspondence and transformation are two factors of Gram matrices, and can be found iteratively using Newton-Schulz factorization [89]. This method performs well for moderate linear transformations. In spite of its elegance, the large computational effort required for spectral methods prohibits its wide applicability. There are a few other nonspectral methods worth mentioning. Ho et al. [70] proposed an elegant non-iterative algorithm for 2D affine registration by searching for the roots of the associated polynomials. Unfortunately this method does not generalize to higher dimensions. Belongie et al. [12] introduced the “shape context” descriptor, which incorporates the neighborhood structure of the point set and thus helps to recover the correspondence between the point sets.

Our approach to the rigid point set registration is probabilistic and most closely related to the works of Rangarajan et al. [136], Wells [179] and Luo and Hancock [98]. Despite extensive work in rigid probabilistic registration, none of the methods, to our best knowledge, provides a closed form solution to the maximization step (M-step) of the EM algorithm for a general multidimensional case. The fact that the rotation matrix should



be constrained to be orthogonal and to have a positive determinant further complicates its estimation. Rangarajan and collaborators [136] showed the solution for 2D case only, where rotation is parametrized by a single angle. In higher dimensions the closed form solution with Euler angles parametrization is not feasible. Luo and Hancock [96, 98] find the rotation matrix through singular value decomposition (SVD). They ignored some terms of the objective function, which leads to only an approximate solution. We shall derive the exact closed form solution (M-step) for the rigid point set registration and discuss its difference from the related methods in Section 4.4.

#### 4.2.2 Non-rigid Point Set Registration Methods

Earlier works on non-rigid point set registration include Hinton et al. [69, 139], who used the probabilistic GMM formulation. The GMM centroids are uniformly positioned along the contour (modeled using splines), which allows for non-rigid transformations. In practice, the method is applicable only to contour-like point sets. One of the most popular non-rigid point set registration method is by Chui and Rangarajan [32]. They proposed to use Thin Plate Spline (TPS) [176, 17] parametrization of the transformation, following RPM, which results into the TPS-RPM method. Similar to the rigid case, they use deterministic annealing and alternate updates for soft-assignment and TPS parameters estimation. They also showed that TPS-RPM is equivalent (with several modifications) to EM for GMM [31]. Tsin and Kanade [170] proposed a correlation-based approach to point set registration, which was later improved by Jian and Vemuri [76]. The method considers the registration as an alignment between two distributions, where each of the point sets represents the GMM centroids. One of the point sets is parametrized by rigid/affine parameters (in rigid/affine case) or TPS (in non-rigid case). The transformation parameters are estimated to minimize the  $L_2$  norm between the distributions. These methods all use explicit TPS parametrization, which is equivalent to a regularization of second order derivatives of the transformation [176, 17]. The TPS parametrization does not exist when the dimension of points is higher than three, which limits the applicability of such methods.

Huang et al. [73] proposed to implicitly embed the shape (or point sets in our case)

into distance transform space, and align them similar to non-rigid image registration algorithms. The authors use sum-of-squared-differences similarity measure between the embedded spaces and incremental free form deformation (FFD) to parameterize the transformation. The method performs well on relatively simple data sets.

Finally, we introduced the Coherent Point Drift (CPD) algorithm [113, 114] for non-rigid point sets registration. The algorithm regularizes the displacement (velocity) field between the point sets following the motion coherence theory (MCT) [187, 188]. Using variational calculus, we obtained that the optimal displacement field has an elegant kernel form in multiple dimensions. We shall discuss and compare our method to the works of Chui and Rangarajan [32], and Jian and Vemuri [76] in Section 4.5.

### 4.3 General Methodology

We consider the alignment of two point sets as a probability density estimation problem, where one point set represents the Gaussian mixture model (GMM) centroids, and the other one represents the data points. At the optimum, two point sets become aligned and the correspondence is obtained using the maximum of the GMM posterior probability for a given data point. Core to our method is to force GMM centroids to move coherently as a group to preserve the topological structure of the point sets. Throughout the chapter we use the following notations:

- $D$  - dimension of the point sets,
- $N, M$  - number of points in the point sets,
- $\mathbf{X}_{N \times D} = (\mathbf{x}_1, \dots, \mathbf{x}_N)^T$  - the first point set (the data points),
- $\mathbf{Y}_{M \times D} = (\mathbf{y}_1, \dots, \mathbf{y}_M)^T$  - the second point set (the GMM centroids),
- $\mathcal{T}(\mathbf{Y}, \theta)$  - Transformation  $\mathcal{T}$  applied to  $\mathbf{Y}$ , where  $\theta$  is a set of the transformation parameters,
- $\mathbf{I}$  - identity matrix,
- $\mathbf{1}$  - column vector of all ones,

- $\mathbf{d}(\mathbf{a})$  - diagonal matrix formed from the vector  $\mathbf{a}$ .

We consider the points in  $\mathbf{Y}$  as the GMM centroids, and the points in  $\mathbf{X}$  as the data points generated by the GMM. The GMM probability density function is

$$p(\mathbf{x}) = \sum_{m=1}^{M+1} P(m)p(\mathbf{x}|m) \quad (4.1)$$

where  $p(\mathbf{x}|m) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp^{-\frac{\|\mathbf{x}-\mathbf{y}_m\|^2}{2\sigma^2}}$ . We also added an additional uniform distribution  $p(\mathbf{x}|M+1) = \frac{1}{N}$  to the mixture model to account for noise and outliers. We use equal isotropic covariances  $\sigma^2$  and equal membership probabilities  $P(m) = \frac{1}{M}$  for all GMM components ( $m = 1, \dots, M$ ). Denoting the weight of the uniform distribution as  $w$ ,  $0 \leq w \leq 1$ , the mixture model takes the form

$$p(\mathbf{x}) = w\frac{1}{N} + (1-w) \sum_{m=1}^M \frac{1}{M}p(\mathbf{x}|m) \quad (4.2)$$

We re-parameterize the GMM centroid locations by a set of parameters  $\theta$  and estimate them by maximizing the likelihood, or equivalently by minimizing the negative log-likelihood function

$$E(\theta, \sigma^2) = - \sum_{n=1}^N \log \sum_{m=1}^{M+1} P(m)p(\mathbf{x}_n|m) \quad (4.3)$$

where we make the i.i.d. data assumption. We define the correspondence probability between two points  $\mathbf{y}_m$  and  $\mathbf{x}_n$  as the posterior probability of the GMM centroid given the data point:  $P(m|\mathbf{x}_n) = P(m)p(\mathbf{x}_n|m)/p(\mathbf{x}_n)$ .

We use the Expectation Maximization (EM) algorithm [44, 14] to find  $\theta$  and  $\sigma^2$ . The idea of EM is first to guess the values of parameters (“old” parameter values) and then use the Bayes’ theorem to compute *a posteriori* probability distributions  $P^{old}(m|\mathbf{x}_n)$  of mixture components, which is the expectation or E-step of the algorithm. The “new” parameter values are then found by minimizing the expectation of the complete negative log-likelihood function [14]

$$Q = - \sum_{n=1}^N \sum_{m=1}^{M+1} P^{old}(m|\mathbf{x}_n) \log(P^{new}(m)p^{new}(\mathbf{x}_n|m)) \quad (4.4)$$

with respect to the “new” parameters, which is called the maximization or M-step of the algorithm. The  $Q$  function, which we call *the objective function*, is also an upper bound of the negative log-likelihood function in (4.3). The EM algorithm proceeds by alternating between E- and M-steps until convergence. Ignoring the constants independent of  $\theta$  and  $\sigma^2$ , we rewrite (4.4) as

$$Q(\theta, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M P^{\text{old}}(m|\mathbf{x}_n) \|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \theta)\|^2 + \frac{N_{\mathbf{P}} D}{2} \log \sigma^2 \quad (4.5)$$

where  $N_{\mathbf{P}} = \sum_{n=1}^N \sum_{m=1}^M P^{\text{old}}(m|\mathbf{x}_n) \leq N$  (with  $N = N_{\mathbf{P}}$  only if  $w = 0$ ) and  $P^{\text{old}}$  denotes the posterior probabilities of GMM components calculated using the previous parameter values:

$$P^{\text{old}}(m|\mathbf{x}_n) = \frac{\exp^{-\frac{1}{2} \left\| \frac{\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \theta^{\text{old}})}{\sigma^{\text{old}}} \right\|^2}}{\sum_{k=1}^M \exp^{-\frac{1}{2} \left\| \frac{\mathbf{x}_n - \mathcal{T}(\mathbf{y}_k, \theta^{\text{old}})}{\sigma^{\text{old}}} \right\|^2} + c} \quad (4.6)$$

where  $c = (2\pi\sigma^2)^{D/2} \frac{w}{1-w} \frac{M}{N}$ . Minimizing the function  $Q$ , we necessarily decrease the negative log-likelihood function  $E$ , unless it is already at a local minimum. To proceed, we specify the transformation  $\mathcal{T}$  for rigid, affine and non-rigid point set registration cases separately.

## 4.4 Rigid & Affine Point Set Registration

For rigid point set registration, we define the transformation of the GMM centroid locations as  $\mathcal{T}(\mathbf{y}_m; \mathbf{R}, \mathbf{t}, s) = s\mathbf{R}\mathbf{y}_m + \mathbf{t}$ , where  $\mathbf{R}_{D \times D}$  is a rotation matrix,  $\mathbf{t}_{D \times 1}$  is a translation vector and  $s$  is a scaling parameter. The objective function (4.5) takes the form:

$$Q(\mathbf{R}, \mathbf{t}, s, \sigma^2) = \frac{1}{2\sigma^2} \sum_{m,n=1}^{M,N} P^{\text{old}}(m|\mathbf{x}_n) \|\mathbf{x}_n - s\mathbf{R}\mathbf{y}_m - \mathbf{t}\|^2 + \frac{N_{\mathbf{P}} D}{2} \log \sigma^2, \quad \text{s.t. } \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1 \quad (4.7)$$

Note that the first term is similar to the one in the absolute orientation problem [5, 173], which is defined as  $\min \sum_{n=1}^N \|\mathbf{x}_n - (s\mathbf{R}\mathbf{y}_n + \mathbf{t})\|^2$  in our notations. Equation (4.7) can be seen as a generalized weighted absolute orientation problem, because it includes weighted differences between all combinations of points. The exact minimization solution of the

objective function (4.7) is complicated due to the constraints on  $\mathbf{R}$ . To obtain the closed form solution we shall use Lemma 1 [112]:

**Lemma 1.** *Let  $\mathbf{R}_{D \times D}$  be an unknown rotation matrix and  $\mathbf{A}_{D \times D}$  be a known real square matrix. Let  $\mathbf{USV}^T$  be a Singular Value Decomposition (SVD) of  $\mathbf{A}$ , where  $\mathbf{UU}^T = \mathbf{VV}^T = \mathbf{I}$  and  $\mathbf{S} = \text{d}(s_i)$  with  $s_1 \geq s_2 \geq \dots \geq s_D \geq 0$ . Then the optimal rotation matrix  $\mathbf{R}$  that maximizes  $\text{tr}(\mathbf{A}^T \mathbf{R})$  is  $\mathbf{R} = \mathbf{UCV}^T$ , where  $\mathbf{C} = \text{d}(1, 1, \dots, 1, \det(\mathbf{UV}^T))$ .*

To apply this lemma, we need to simplify the  $Q$  function to a form equivalent to  $\text{tr}(\mathbf{A}^T \mathbf{R})$ . First, we eliminate translation  $\mathbf{t}$  from  $Q$ . Taking partial derivative of  $Q$  with respect to  $\mathbf{t}$  and equate it to zero, we obtain:

$$\mathbf{t} = \frac{1}{N_P} \mathbf{X}^T \mathbf{P}^T \mathbf{1} - s \mathbf{R} \frac{1}{N_P} \mathbf{Y}^T \mathbf{P} \mathbf{1} = \mu_{\mathbf{x}} - s \mathbf{R} \mu_{\mathbf{y}},$$

where the matrix  $\mathbf{P}$  has elements  $p_{mn} = P^{old}(m|\mathbf{x}_n)$  in (4.6) and the mean vectors  $\mu_{\mathbf{x}}$  and  $\mu_{\mathbf{y}}$  are defined as:

$$\mu_{\mathbf{x}} = \mathbf{E}(\mathbf{X}) = \frac{1}{N} \mathbf{X}^T \mathbf{P}^T \mathbf{1}, \quad \mu_{\mathbf{y}} = \mathbf{E}(\mathbf{Y}) = \frac{1}{N} \mathbf{Y}^T \mathbf{P} \mathbf{1}.$$

Substituting  $\mathbf{t}$  back into the objective function and rewriting it in matrix form, we obtain

$$Q = \frac{1}{2\sigma^2} [\text{tr}(\hat{\mathbf{X}}^T \text{d}(\mathbf{P}^T \mathbf{1}) \hat{\mathbf{X}}) - 2s \text{tr}(\hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}} \mathbf{R}^T) + s^2 \text{tr}(\hat{\mathbf{Y}}^T \text{d}(\mathbf{P} \mathbf{1}) \hat{\mathbf{Y}})] + \frac{N_P D}{2} \log \sigma^2 \quad (4.8)$$

where  $\hat{\mathbf{X}} = \mathbf{X} - \mathbf{1}\mu_{\mathbf{x}}^T$  and  $\hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{1}\mu_{\mathbf{y}}^T$  are the centered point set matrices. We use the fact that trace is invariant under cyclic matrix permutations and  $\mathbf{R}$  is orthogonal. We can rewrite (4.8) as  $Q = -c_1 \text{tr}((\hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}})^T \mathbf{R}) + c_2$ , where  $c_1, c_2$  are constants independent of  $\mathbf{R}$  and  $c_1 > 0$ . Thus minimization of  $Q$  with respect to  $\mathbf{R}$  is equivalent to maximization of

$$\max \text{tr}(\mathbf{A}^T \mathbf{R}), \mathbf{A} = \hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}}, \quad \text{s.t. } \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1.$$

Now we are ready to use Lemma 1, and the optimal  $\mathbf{R}$  is in the form

$$\mathbf{R} = \mathbf{UCV}^T, \text{ where } \mathbf{USV}^T = \text{svd}(\hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}}) \quad (4.9)$$

and  $\mathbf{C} = \text{d}(1, \dots, 1, \det(\mathbf{UV}^T))$ . To solve for  $s$  and  $\sigma^2$ , we equate the corresponding partial derivative of (4.8) to zero. Solving for  $\mathbf{R}, s, \mathbf{t}, \sigma^2$  is the M-step of the EM algorithm. We summarize the rigid point sets registration algorithm in Fig. 4.2.

**Rigid point set registration algorithm:**

- Initialization:  $\mathbf{R} = \mathbf{I}, \mathbf{t} = 0, s = 1, 0 \leq w \leq 1$   
 $\sigma^2 = \frac{1}{DNM} \sum_{n=1}^N \sum_{m=1}^M \|\mathbf{x}_n - \mathbf{y}_m\|^2$
- EM optimization, repeat until convergence:
  - E-step: Compute  $\mathbf{P}$ ,  

$$p_{mn} = \frac{\exp^{-\frac{1}{2\sigma^2} \|\mathbf{x}_n - (s\mathbf{R}\mathbf{y}_m + \mathbf{t})\|^2}}{\sum_{k=1}^M \exp^{-\frac{1}{2\sigma^2} \|\mathbf{x}_n - (s\mathbf{R}\mathbf{y}_k + \mathbf{t})\|^2} + (2\pi\sigma^2)^{D/2} \frac{w}{1-w} \frac{M}{N}}$$
  - M-step: Solve for  $\mathbf{R}, s, \mathbf{t}, \sigma^2$ :
    - $N_{\mathbf{P}} = \mathbf{1}^T \mathbf{P} \mathbf{1}, \mu_{\mathbf{x}} = \frac{1}{N_{\mathbf{P}}} \mathbf{X}^T \mathbf{P}^T \mathbf{1}, \mu_{\mathbf{y}} = \frac{1}{N_{\mathbf{P}}} \mathbf{Y}^T \mathbf{P} \mathbf{1},$
    - $\hat{\mathbf{X}} = \mathbf{X} - \mathbf{1} \mu_{\mathbf{x}}^T, \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{1} \mu_{\mathbf{y}}^T,$
    - $\mathbf{A} = \hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}},$  compute SVD of  $\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T,$
    - $\mathbf{R} = \mathbf{U} \mathbf{C} \mathbf{V}^T,$  where  $\mathbf{C} = \text{d}(1, \dots, 1, \det(\mathbf{U} \mathbf{V}^T)),$
    - $s = \frac{\text{tr}(\mathbf{A}^T \mathbf{R})}{\text{tr}(\hat{\mathbf{Y}}^T \text{d}(\mathbf{P} \mathbf{1}) \hat{\mathbf{Y}})},$
    - $\mathbf{t} = \mu_{\mathbf{x}} - s \mathbf{R} \mu_{\mathbf{y}},$
    - $\sigma^2 = \frac{1}{N_{\mathbf{P}} D} (\text{tr}(\hat{\mathbf{X}}^T \text{d}(\mathbf{P}^T \mathbf{1}) \hat{\mathbf{X}}) - s \text{tr}(\mathbf{A}^T \mathbf{R})).$
  - The aligned point set is  $\mathcal{T}(\mathbf{Y}) = s \mathbf{Y} \mathbf{R}^T + \mathbf{1} \mathbf{t}^T,$
  - The probability of correspondence is given by  $\mathbf{P}.$

Figure 4.2: Rigid point set registration algorithm.

The algorithm has one free parameter,  $w$  ( $0 \leq w \leq 1$ ), which reflects our assumption on the amount of noise in the point sets. The solution for the rotation matrix is general  $D$ -dimensional.

**Affine point set registration:** The affine registration case is simpler compared to the rigid case, because the optimization is unconstrained. Affine transformation is defined as  $\mathcal{T}(\mathbf{y}_m; \mathbf{B}, \mathbf{t}, s) = \mathbf{B} \mathbf{y}_m + \mathbf{t}$ , where  $\mathbf{B}_{D \times D}$  is an affine transformation matrix,  $\mathbf{t}_{D \times 1}$  is translation vector. The objective function takes the form:

$$Q(\mathbf{B}, \mathbf{t}, \sigma^2) = \frac{1}{2\sigma^2} \sum_{m,n=1}^{M,N} P^{\text{old}}(m|\mathbf{x}_n) \|\mathbf{x}_n - (\mathbf{B} \mathbf{y}_m + \mathbf{t})\|^2 + \frac{N_{\mathbf{P}} D}{2} \log \sigma^2 \quad (4.10)$$

We can directly take the partial derivatives of  $Q$ , equate them to zero, and solve the resulting linear system of equations. The solution is straightforward and similar to the rigid case. We summarize the affine point set registration algorithm in Fig. 4.3.

#### 4.4.1 Related Rigid Point Set Registration Methods

Here, we discuss the probabilistic rigid point set registration methods most closely related to ours. Rangarajan et al. [136] presented the RPM method for rigid point set registration.

**Affine point set registration algorithm:**

- Initialization:  $\mathbf{B} = \mathbf{I}, \mathbf{t} = 0, 0 \leq w \leq 1$   
 $\sigma^2 = \frac{1}{DNM} \sum_{n=1}^N \sum_{m=1}^M \|\mathbf{x}_n - \mathbf{y}_m\|^2$
- EM optimization, repeat until convergence:
  - E-step: Compute  $\mathbf{P}$ ,  

$$p_{mn} = \frac{\exp^{-\frac{1}{2\sigma^2} \|\mathbf{x}_n - (\mathbf{B}\mathbf{y}_m + \mathbf{t})\|^2}}{\sum_{k=1}^M \exp^{-\frac{1}{2\sigma^2} \|\mathbf{x}_n - (\mathbf{B}\mathbf{y}_k + \mathbf{t})\|^2} + (2\pi\sigma^2)^{D/2} \frac{w}{1-w} \frac{M}{N}}$$
  - M-step: Solve for  $\mathbf{B}, \mathbf{t}, \sigma^2$ :
    - $N_{\mathbf{P}} = \mathbf{1}^T \mathbf{P} \mathbf{1}, \mu_{\mathbf{x}} = \frac{1}{N_{\mathbf{P}}} \mathbf{X}^T \mathbf{P}^T \mathbf{1}, \mu_{\mathbf{y}} = \frac{1}{N_{\mathbf{P}}} \mathbf{Y}^T \mathbf{P} \mathbf{1},$
    - $\hat{\mathbf{X}} = \mathbf{X} - \mathbf{1} \mu_{\mathbf{x}}^T, \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{1} \mu_{\mathbf{y}}^T,$
    - $\mathbf{B} = (\hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}})(\hat{\mathbf{Y}}^T \mathbf{d}(\mathbf{P} \mathbf{1}) \hat{\mathbf{Y}})^{-1},$
    - $\mathbf{t} = \mu_{\mathbf{x}} - \mathbf{B} \mu_{\mathbf{y}},$
    - $\sigma^2 = \frac{1}{N_{\mathbf{P}} D} (\text{tr}(\hat{\mathbf{X}}^T \mathbf{d}(\mathbf{P}^T \mathbf{1}) \hat{\mathbf{X}}) - \text{tr}(\hat{\mathbf{X}}^T \mathbf{P}^T \hat{\mathbf{Y}} \mathbf{B}^T)).$
- The aligned point set is  $\mathcal{T}(\mathbf{Y}) = \mathbf{Y} \mathbf{B}^T + \mathbf{1} \mathbf{t}^T,$
- The probability of correspondence is given by  $\mathbf{P}$ .

Figure 4.3: Affine point set registration algorithm.

The method is shown for 2D case, where rotation matrix is parametrized by a single rotation angle, which allows to find an explicit update. Such Euler's angles approach is not feasible in multidimensional cases. RPM also uses deterministic annealing on  $\sigma^2$ , which requires to set the starting and stopping criteria as well as the annealing rate. The EM iterations has to be repeated at each annealing step, which can be slow. We argue that it is preferable to estimate  $\sigma^2$  instead of using deterministic annealing. The deterministic annealing helps to overcome poor local minima, but for the rigid point set registration problem the rigid parametrization is a strong constraint that alleviates the advantages of the deterministic annealing.

Luo and Hancock [97, 98] introduced the rigid point sets registration algorithm that is the most similar to ours. The authors optimize the objective function rather intuitively than rigorously, which leads to several assumptions and approximate minimization. They ignore a few terms of the objective function (see Eqs.10,11 in [97]), where the last term does depend on transformation parameters, and must not be ignored. If such optimization converge, the M-step of the EM algorithm is only approximate. Among other differences, we want to mention that the authors use *structural editing*, a technique to remove some undesirable points, instead of using an additional uniform distribution to account for these

points. Some other authors [103] also follow the rigid parameters estimation steps of Luo and Hancock [97].

## 4.5 Non-Rigid Point Set Registration

Non-rigid point set registration remains a challenging problem in computer vision. The transformation that aligns the point sets is assumed to be unknown and non-rigid, which is generally broad class of transformations that can lead to an ill-posed problem. In order to deal with the problem we use Tikhonov regularization framework [169, 27, 150]. We define the transformation as the initial position plus a displacement function  $v$ :

$$\mathcal{T}(\mathbf{Y}, v) = \mathbf{Y} + v(\mathbf{Y}), \quad (4.11)$$

We regularize the norm of  $v$  to enforce the smoothness of the function [27]. Such approach is also supported by the Motion Coherence Theory (MCT) [187, 188], which states that points close to one another tend to move coherently, and thus, the displacement function between the point sets should be smooth. This is mathematically formulated as a regularization on the displacement (also called velocity) function.

Adding a regularization term to the negative log-likelihood function we obtain

$$f(v, \sigma^2) = E(v, \sigma^2) + \frac{\lambda}{2} \phi(v) \quad (4.12)$$

where  $E$  is the negative log-likelihood function (4.3),  $\phi(v)$  is a regularization term and  $\lambda$  is a trade-off parameter. Such regularization is well formulated in Bayesian approach, where the regularization term comes from a prior on displacement field:  $p(v) = \exp^{-\frac{\lambda}{2}\phi(v)}$ .

We estimate the displacement function  $v$  using variational calculus. We shall define the regularization term  $\phi(v)$  in different but equivalent forms and show that the optimal functional form of  $v$  is a linear combination of particular kernel functions. A particular choice of the regularization will lead to our non-rigid point set registration method.

### 4.5.1 Regularization of the Displacement Function

A norm of  $v$  in the Hilbert space  $\mathbb{H}^m$  is defined as:

$$\|v\|_{\mathbb{H}^m}^2 = \int_{\mathbb{R}} \sum_{k=0}^m \left\| \frac{\partial^k v}{\partial x^k} \right\|^2 dx. \quad (4.13)$$



Alternatively, we can define the norm in the Reproducing Kernel Hilbert Space (RKHS) [27, 55] as

$$\|v\|_{\mathbb{H}^m}^2 = \int_{\mathbb{R}^D} \frac{|\tilde{v}(\mathbf{s})|^2}{\tilde{G}(\mathbf{s})} d\mathbf{s} \quad (4.14)$$

where  $G$  is a unique kernel function associated with the RKHS, and  $\tilde{G}$  is its Fourier transform. Function  $\tilde{v}$  indicates the Fourier transform of the function  $v$  and  $\mathbf{s}$  is a frequency domain variable. The Fourier domain norm definition has been used in the Regularization Theory (RT) [55] to regularize the smoothness of a function. Regularization theory defines smoothness as a measure of the “oscillatory” behavior of a function. Within the class of differentiable functions, one function is said to be smoother than another if it oscillates less; in other words, if it has less energy at high frequency. The high frequency content of a function can be measured by first high-pass filtering the function, and then measuring the resulting power. This can be represented by (4.14), where  $\tilde{G}$  represents a symmetric positive definite low-pass filter, which approaches zero as  $\|\mathbf{s}\| \rightarrow \infty$ . For convenience, we shall write the regularization term as

$$\phi(v) = \|v\|_{\mathbb{H}^m}^2 = \|Lv\|^2 \quad (4.15)$$

where an operator  $L$  “extracts” a part of the function for regularization, in our case, the high frequency content part [27, 150].

#### 4.5.2 Variational Solution

We find the functional form of  $v$  using calculus of variation. Minimization of regularized negative log-likelihood function in (4.12) boils down to minimization of the following objective function (M-step):

$$Q(v, \sigma^2) = \frac{1}{2\sigma^2} \sum_{m,n=1}^{M,N} P^{\text{old}}(m|\mathbf{x}_n) \|\mathbf{x}_n - (\mathbf{y}_m + v(\mathbf{y}_m))\|^2 + \frac{N_{\mathbf{P}}D}{2} \log \sigma^2 + \frac{\lambda}{2} \|Lv\|^2 \quad (4.16)$$

A function  $v$  that minimizes (4.16) must satisfy the Euler-Lagrange differential equation

$$\frac{1}{\sigma^2\lambda} \sum_{n=1}^N \sum_{m=1}^M P^{\text{old}}(m|\mathbf{x}_n) (\mathbf{x}_n - (\mathbf{y}_m + v(\mathbf{y}_m))) \delta(\mathbf{z} - \mathbf{y}_m) = \hat{L}Lv(\mathbf{z}) \quad (4.17)$$

for all vectors  $\mathbf{z}$ , where  $\hat{L}$  is the adjoint operator to  $L$ . The solution to such partial differential equation can be written as the integral transformation of its left side with a Green's function  $G$  of the self-adjoint operator  $\hat{L}L$ .

$$v(\mathbf{z}) = \frac{1}{\sigma^2\lambda} \sum_{m,n=1}^{M,N} P^{\text{old}}(m|\mathbf{x}_n)(\mathbf{x}_n - (\mathbf{y}_m + v(\mathbf{y}_m)))G(\mathbf{z}, \mathbf{y}_m) = \sum_{m=1}^M \mathbf{w}_m G(\mathbf{z}, \mathbf{y}_m) \quad (4.18)$$

where  $\mathbf{w}_m = \frac{1}{\sigma^2\lambda} \sum_{n=1}^N P^{\text{old}}(m|\mathbf{x}_n)(\mathbf{x}_n - (\mathbf{y}_m + v(\mathbf{y}_m)))$ . Note that this solution is incomplete. In general, the solution also includes the term  $\psi(\mathbf{z})$  that lies in the null space of  $L$  [55, 83]. Thus, we achieve Lemma 2.

**Lemma 2.** *The optimal displacement function that minimizes (4.16) is given by linear combination of the particular kernel functions centered at the points  $\mathbf{Y}$  plus the term  $\psi(\mathbf{z})$  in the null space of  $L$ :*

$$v(\mathbf{z}) = \sum_{m=1}^M \mathbf{w}_m G(\mathbf{z}, \mathbf{y}_m) + \psi(\mathbf{z}) \quad (4.19)$$

where the kernel function is a Green's function of the self-adjoint operator  $\hat{L}L$ .

### 4.5.3 The Coherent Point Drift (CPD) Algorithm

We choose the regularization term according to (4.14):

$$\phi(v) = \int_{\mathbb{R}^D} \frac{|\tilde{v}(\mathbf{s})|^2}{\tilde{G}(\mathbf{s})} d\mathbf{s} \quad (4.20)$$

where  $G$  is a Gaussian (note it is not related to the Gaussian form of the distribution chosen for the mixture model). There are several motivations for such a Gaussian choice: First, the Green's function (the kernel) corresponding to the regularization term in (4.20) is also a Gaussian (and remains a Gaussian for an arbitrary dimensional case); the Gaussian kernel is positive definite and the null space term  $\psi(\mathbf{z}) = 0$  [55]. Second, by choosing an appropriately sized Gaussian function we have the flexibility to control the range of filtered frequencies and thus the amount of spatial smoothness. Third, the choice of the Gaussian makes our regularization term equivalent to the one in the Motion Coherence Theory (MCT) [188]:

$$\phi_{MCT}(v) = \int_{\mathbb{R}^d} \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} \left\| D^l v(\mathbf{x}) \right\|^2 d\mathbf{x} \quad (4.21)$$

where  $D$  is a derivative operator so that  $D^{2l}v = \nabla^{2l}v$  and  $D^{2l+1}v = \nabla(\nabla^{2l}v)$ , where  $\nabla$  is the gradient operator and  $\nabla^2$  is the Laplacian operator.

**Lemma 3.** *The regularization term in (4.20) with a Gaussian choice of low-pass filter  $G$  is equivalent to the regularization term in (4.21). Both terms represent the norm of the function  $v$ , after applying the operator  $L$ , and the corresponding Green's function is a Gaussian in both cases [27].*

The equivalence of our regularization term with that of the Motion Coherence Theory implies that we are imposing motion coherence among the points and thus we call the non-rigid point set registration method the Coherent Point Drift (CPD) algorithm.

We can obtain the coefficients  $\mathbf{w}_m$  by evaluating (4.19) at  $\mathbf{y}_m$  points

$$(\mathbf{G} + \lambda\sigma^2 d(\mathbf{P}\mathbf{1})^{-1})\mathbf{W} = d(\mathbf{P}\mathbf{1})^{-1}\mathbf{P}\mathbf{X} - \mathbf{Y} \quad (4.22)$$

where  $\mathbf{W}_{M \times D} = (\mathbf{w}_1, \dots, \mathbf{w}_M)^T$  is a matrix of coefficients,  $\mathbf{G}_{M \times M}$  is a kernel matrix with elements  $g_{ij} = G(\mathbf{y}_i, \mathbf{y}_j) = e^{-\frac{1}{2}\left\|\frac{\mathbf{y}_i - \mathbf{y}_j}{\beta}\right\|^2}$  and  $d^{-1}(\cdot)$  is the inverse diagonal matrix. The transformed position of  $\mathbf{y}_m$  are found according to (4.11) as  $\mathbf{T} = \mathcal{T}(\mathbf{Y}, \mathbf{W}) = \mathbf{Y} + \mathbf{G}\mathbf{W}$ . We obtain  $\sigma^2$  by equating the corresponding derivative of  $Q$  to zero

$$\begin{aligned} \sigma^2 = \frac{1}{N_{\mathbf{P}}D} \sum_{n=1}^N \sum_{m=1}^M \|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \mathbf{W})\|^2 = \\ \frac{1}{N_{\mathbf{P}}D} (\text{tr}(\mathbf{X}^T d(\mathbf{P}^T \mathbf{1})\mathbf{X}) - 2\text{tr}((\mathbf{P}\mathbf{X})^T \mathbf{T}) + \text{tr}(\mathbf{T}^T d(\mathbf{P}\mathbf{1})\mathbf{T})) \end{aligned} \quad (4.23)$$

We summarize the CPD non-rigid point set registration algorithm in Fig. 4.4.

**Analysis:** The CPD algorithm includes three free parameters:  $w, \lambda$  and  $\beta$ . Parameter  $w$  ( $0 \leq w \leq 1$ ) reflects our assumption on the amount of noise in the point sets. Parameters  $\lambda$  and  $\beta$  both reflect the amount of smoothness regularization. A discussion on the difference between  $\lambda$  and  $\beta$  can be found in [187, 188]. Briefly speaking, parameter  $\beta$  defines the model of the smoothness regularizer (width of smoothing Gaussian filter in (4.20)). Parameter  $\lambda$  represents the trade-off between the goodness of maximum likelihood fit and regularization.

We note that solution of (4.22) gives the exact minimum of  $Q$  (4.16), if  $\sigma^2$  is assumed fixed. As far as we are estimating  $\sigma^2$ , (4.22) and (4.23) should be solved simultaneously.

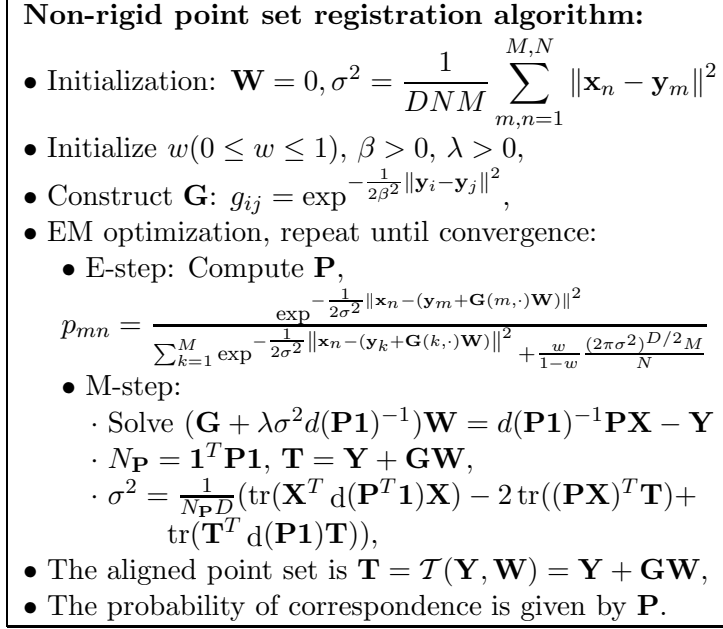


Figure 4.4: The Coherent Point Drift algorithm for non-rigid point set registration.

The non-linear dependency of  $\sigma^2$  on  $\mathbf{W}$  and vice-versa does not allow for simultaneous analytical solution. Iterative exact solution can be obtained by performing a few cyclic iterations on (4.22) and (4.23) within a single EM step. Practically, a single iteration, given by (4.22) and (4.23), decrease the  $Q$  function almost to the exact minimum. Such an iterative procedure, which decreases the  $Q$  function but not to exact minimum, is often called the generalized EM algorithm [44, 119].

#### 4.5.4 Related Non-rigid Point Set Registration Methods

The CPD algorithm follows our previous work [114] on non-rigid point set registration. However, previously we have used deterministic annealing on  $\sigma^2$ , whereas here, we estimate the Gaussian width  $\sigma^2$  within ML framework. This allows us to significantly speed up the algorithm, alleviating the repeated EM-iterations for every single annealing step. We have not observed any decrease in accuracy of the method related to this change. In [114], we used a slightly different notation for the GMM centroid locations: we called  $\mathbf{Y}_0$  the initial centroids position (which we call  $\mathbf{Y}$  here), and  $\mathbf{Y}$  for the final GMM centroid position (which we call  $\mathcal{T}(\mathbf{Y})$  here).

The most relevant non-rigid point sets registration algorithm to ours is TPS-RPM, more precisely its GMM formulation [31]. TPS-RPM uses Thin Plate Spline (TPS) [176, 17] parametrization of the transformation, which can be obtained by adding the regularization term that penalizes second order derivatives of the transformation. For instance, in 2D such regularization term is

$$\|L\mathcal{T}\|^2 = \int \int [(\frac{\partial^2 \mathcal{T}}{\partial x^2})^2 + 2(\frac{\partial^2 \mathcal{T}}{\partial x \partial y})^2 + (\frac{\partial^2 \mathcal{T}}{\partial y^2})^2] dx dy \quad (4.24)$$

This term can be equivalently formulated in the Fourier space as:

$$\|L\mathcal{T}\|^2 = \int_{\mathbb{R}^2} \|\mathbf{s}\|^4 |\tilde{\mathcal{T}}(\mathbf{s})|^2 d\mathbf{s} \quad (4.25)$$

which is a special case of the Duchon splines [47]. The null space of such regularization includes affine transformations. Using the variational approach we can show that the optimal transformation  $\mathcal{T}$  for such regularization is in the form  $\mathcal{T}(\mathbf{Y}) = \mathbf{Y}\mathbf{A} + \mathbf{K}\mathbf{C}$ , where  $\mathbf{A}$  is a matrix of affine transformation coefficients,  $\mathbf{C}$  is a matrix of non-rigid coefficients. For 2D case, matrix  $\mathbf{K}_{M \times M}$  is the kernel matrix with elements  $k_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|^2 \log \|\mathbf{y}_i - \mathbf{y}_j\|$ . For 3D case, matrix  $\mathbf{K}$  has elements  $k_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|$ . For 4D or higher dimensions the TPS kernel solution does not exist [156]. Finally, to link such regularization to our non-rigid registration framework, we note that the regularization of the displacement field  $v$ , instead of the transformation itself, is exactly the same, because, (4.24) is invariant under affine transformations, in other words  $\|L\mathcal{T}(\mathbf{z})\|^2 = \|L(\mathbf{z} + v(\mathbf{z}))\|^2 = \|Lv(\mathbf{z})\|^2$ . This means that both CPD and TPS-RPM regularizes the displacement function, but using different regularization terms.

The advantage of CPD regularization (as given by (4.20) or (4.21)) comparing to TPS ((4.24) or (4.25)), is that it easily generalizes to  $N$  dimensions. Also we can control the locality of spatial smoothness by changing the Gaussian filter width  $\beta$ , whereas TPS does not have such flexibility. Among other differences, TPS-RPM approximates the M-step solution of the EM algorithm [31] for simplicity and use deterministic annealing on  $\sigma^2$ .

Finally, Jian and Vemuri [76] consider the registration as an alignment between the distributions of two point sets, where a separate GMMs are used to model the distribution for the point sets. One of the point sets is parametrized by TPS. The transformation

parameters are estimated to minimize the  $L_2$  norm between the distributions. In our case, the CPD method maximizes the likelihood function, which is equivalent to KL divergence minimization between two mixture distributions: GMM and mixture of delta functions. KL divergence is more appropriate similarity measure for the densities than  $L_2$  norm, because it weights the error according to its probability.

## 4.6 Fast Implementation

Here we show that CPD computational complexity can be reduced to a linear up to a multiplicative constant. We use the fast Gauss transform (FGT) [62] to compute the matrix-vector products  $\mathbf{P}\mathbf{1}$ ,  $\mathbf{P}^T\mathbf{1}$ ,  $\mathbf{P}\mathbf{X}$ , which are the bottlenecks for both rigid and non-rigid cases. We use low-rank matrix approximation to speed-up the solution of the linear system of equations (4.22) for the non-rigid case.

**The fast Gauss transform:** Greengard and Strain [62] introduced the fast Gauss transform (FGT) for fast computation of the sum of exponentials:

$$f(\mathbf{y}_m) = \sum_{n=1}^N z_n \exp^{-\frac{1}{2\sigma^2} \|\mathbf{x}_n - \mathbf{y}_m\|^2}, \quad \forall \mathbf{y}_m, \quad m = 1, \dots, M. \quad (4.26)$$

The naive approach takes  $\mathcal{O}(MN)$  operations, while FGT takes only  $\mathcal{O}(M + N)$ . The basic idea of FGT is to expand the Gaussians in terms of truncated Hermit expansion, and approximate (4.26) up to the predefined accuracy. Rewriting (4.26) in matrix form, we obtain  $\mathbf{f} = \mathbf{K}\mathbf{z}$ , where  $\mathbf{z}$  is some vector and  $\mathbf{K}_{M \times N}$  is a Gaussian affinity matrix with elements:  $k_{mn} = \exp^{-\frac{1}{2\sigma^2} \|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m)\|^2}$ , which we have already used in our notations. We simplify the matrix-vector products  $\mathbf{P}\mathbf{1}$ ,  $\mathbf{P}^T\mathbf{1}$  and  $\mathbf{P}\mathbf{X}$ , to the form of  $\mathbf{K}\mathbf{z}$  and apply FGT. Matrix  $\mathbf{P}$  (4.6) can be partitioned into

$$\mathbf{P} = \mathbf{K} \mathbf{d}(\mathbf{a}), \quad \mathbf{a} = \mathbf{1} / (\mathbf{K}^T \mathbf{1} + c \mathbf{1}) \quad (4.27)$$

where  $\mathbf{d}(\mathbf{a})$  is diagonal matrix with a vector  $\mathbf{a}$  along the diagonal. Here, we use Matlab element-wise division  $(./)$  and element-wise multiplication  $(.*)$  notations. We show the algorithm to compute the bottleneck matrix-vector products  $\mathbf{P}\mathbf{1}$ ,  $\mathbf{P}^T\mathbf{1}$  and  $\mathbf{P}\mathbf{X}$  using FGT in Fig. 4.5. We note that for dimensions higher than three, we can use the improved fast

<p><b>Compute <math>\mathbf{P}^T \mathbf{1}</math>, <math>\mathbf{P1}</math> and <math>\mathbf{PX}</math>:</b></p> <ul style="list-style-type: none"> <li>• Compute <math>\mathbf{K}^T \mathbf{1}</math> (using FGT),</li> <li>• <math>\mathbf{a} = 1./(\mathbf{K}^T \mathbf{1} + c\mathbf{1})</math>,</li> <li>• <math>\mathbf{P}^T \mathbf{1} = \mathbf{1} - c\mathbf{a}</math>,</li> <li>• <math>\mathbf{P1} = \mathbf{Ka}</math> (using FGT),</li> <li>• <math>\mathbf{PX} = \mathbf{K(a * X)}</math> (using FGT),</li> </ul>
---

Figure 4.5: Matrix-vector products computation through FGT.

Gauss transform (IFGT) method [181], which is a faster alternative to FGT for higher dimensions.

During the final EM iterations, the width of the Gaussians  $\sigma^2$  becomes small. The Hermitian expansion thus requires many terms to approximate highly multimodal Gaussian distribution for a given precision. At the final iterations, the Gaussian becomes very narrow, and we can switch to the truncated Gaussian approximation (set zeros outside a predefined box).

**Low-rank matrix approximation:** In the non-rigid case, we need to solve the linear system (4.22), which is  $\mathcal{O}(M^3)$  using direct matrix inversion. We note that the left hand side matrix of (4.22) is symmetric and positive definite.

We use low-rank matrix approximation of  $\mathbf{G}$ , where  $\mathbf{G}$  is a Gaussian affinity matrix with elements  $g_{ij} = \exp^{-\frac{1}{2\beta^2}\|\mathbf{y}_i - \mathbf{y}_j\|^2}$ . We approximate the matrix  $\mathbf{G}$  as

$$\hat{\mathbf{G}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \quad (4.28)$$

where  $\mathbf{\Lambda}_{K \times K}$  is a diagonal matrix with  $K$  largest eigenvalues and the matrix  $\mathbf{Q}_{M \times K}$  is formed from the corresponding eigenvectors.  $\hat{\mathbf{G}}$  is the closest  $K$ -rank matrix approximation to  $\mathbf{G}$  both in  $L_2$  and Frobenius norms [57]. To solve the linear system in (4.22) we use the Woodbury identity and invert the first term as

$$(\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T + \lambda\sigma^2 \mathbf{d}(\mathbf{P1})^{-1})^{-1} = \frac{1}{\lambda\sigma^2} \mathbf{d}(\mathbf{P1}) - \frac{1}{(\lambda\sigma^2)^2} \mathbf{d}(\mathbf{P1})\mathbf{Q}(\mathbf{\Lambda}^{-1} + \frac{1}{\lambda\sigma^2}\mathbf{Q}^T \mathbf{d}(\mathbf{P1})\mathbf{Q})^{-1}\mathbf{Q}^T \mathbf{d}(\mathbf{P1}) \quad (4.29)$$

The inside matrix inversion is of  $\mathcal{O}(K^3)$ , where  $K \ll M$ . For instance choosing  $K = M^{1/3}$  largest eigenvalues, we reduce the computational complexity to linear. We can pre-compute  $K$  largest eigenvalues and eigenvectors of  $\mathbf{G}$  using deflation techniques [43]. It

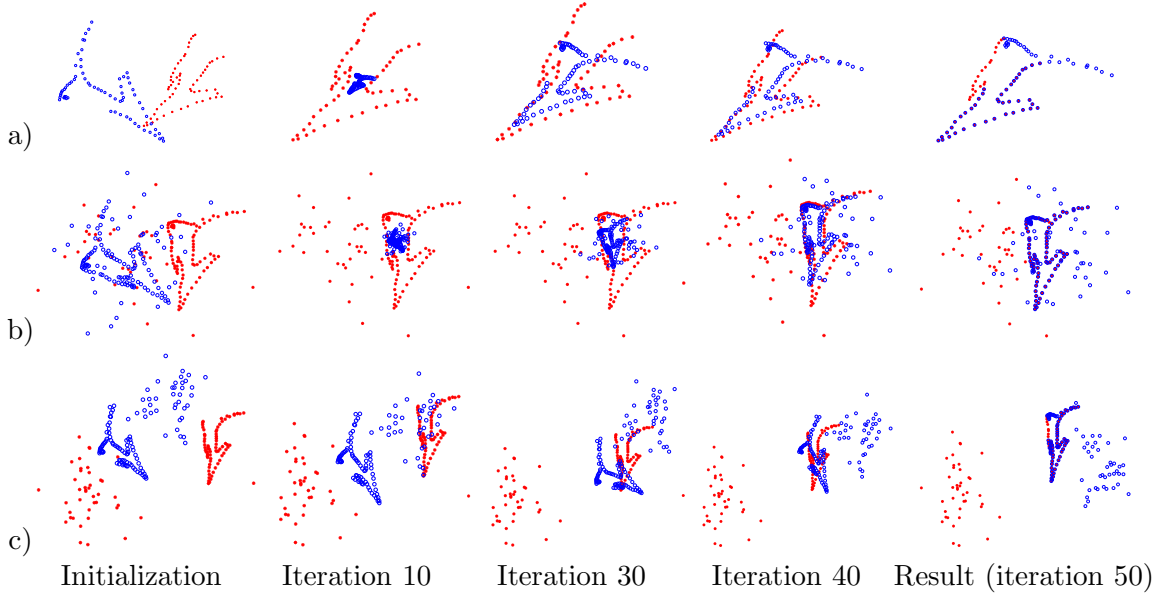


Figure 4.6: Fish data set, rigid registration examples. We align  $\mathbf{Y}$  (blue circles) onto  $\mathbf{X}$  (red stars). The columns show the iterative alignment progress. a) Registration of the point sets with missing non-overlapping parts ( $w = 0.5$ ); b) Registration of the point sets corrupted by random outliers ( $w = 0.5$ ); c) A challenging rigid registration example, where both point sets are corrupted by outliers and biased to different sides of the point sets. We have also deleted some parts from both point sets. We set  $w = 0.8$  and fix scaling  $s = 1$ . CPD registration is robust and accurate in all experiments.

requires several iterations with the matrix-vector product  $\mathbf{G}\mathbf{z}$ , which can be implemented explicitly or through FGT.

The low-rank matrix approximation intuitively constraints the space of the non-rigid transformations, and can be even desirable to further constrain the non-rigid transformation. If the number of points is large and well clustered, then an extremely small percent of eigenvalues will be sufficient for an accurate approximation.

## 4.7 Results

We implemented the algorithm in Matlab, and tested it on a Pentium4 CPU 3GHz with 4GB RAM. We implemented the matrix-vector products in C as a Matlab mex functions to avoid the storage of  $\mathbf{P}$ . The code is available at [www.bme.ogi.edu/~myron/matlab/cpd](http://www.bme.ogi.edu/~myron/matlab/cpd). We shall refer to our method as Coherent Point Drift (CPD) both for rigid and non-rigid



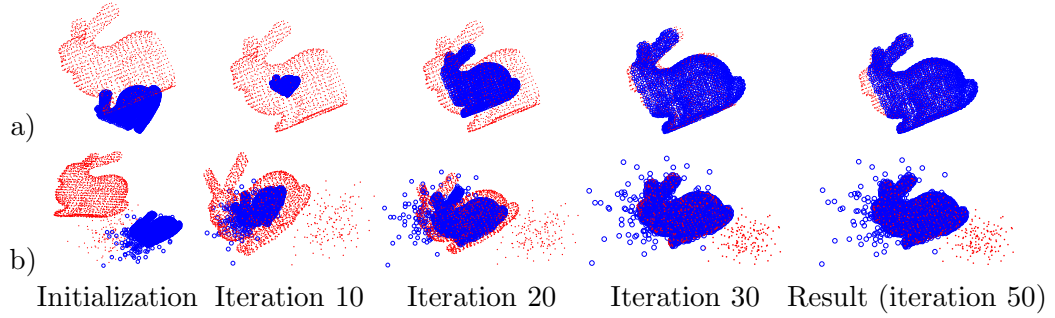


Figure 4.7: 3D bunny point set rigid registration examples. We align  $\mathbf{Y}$  (blue circles) onto  $\mathbf{X}$  (red dots). The columns show the iterative alignment progress. We initialized one of the point sets with 50 degree rotation and scaling equal 2. a) Registration of the point sets with missing points ( $w = 0.5$ ); b) A challenging example of CPD rigid registration with missing points, outliers and noise. CPD shows robust and accurate registration result in all experiments.

point sets registration methods. We have also implemented the matrix-vector products through FGT using the Matlab FGT implementation by Sebastien Paris [131].

We consider rigid and non-rigid experiments separately below. We shall always pre-align both point sets to zero mean and unit variance before the registration.

#### 4.7.1 Rigid Registration Results

We show the CPD rigid registration on several examples, test the fast CPD implementation and evaluate its performance in comparison with LM-ICP [53], which is one of the most popular robust rigid point set registration methods.

**Rigid fish point set registration:** Fig. 4.6 shows several rigid registration tests on 2D fish point sets. In Fig. 4.6a we deleted non-overlapping parts in both point sets and set  $w = 0.5$ , where  $w$  is a weight of the uniform distribution that accounts for noise and outliers. In Fig. 4.6b we corrupted the point sets by outliers. We generate outliers randomly from a normal zero-mean distribution. CPD demonstrates robust and accurate performance in all examples. Fig. 4.6c demonstrates a challenging example, where both point sets have missing points and are corrupted by outliers. The most challenging here is that we biased the outliers to the different sides of fish point sets. We were able to register such point sets only by fixing the scaling to be constant (estimating rotation and translation only). CPD demonstrates accurate and robust registration performance.

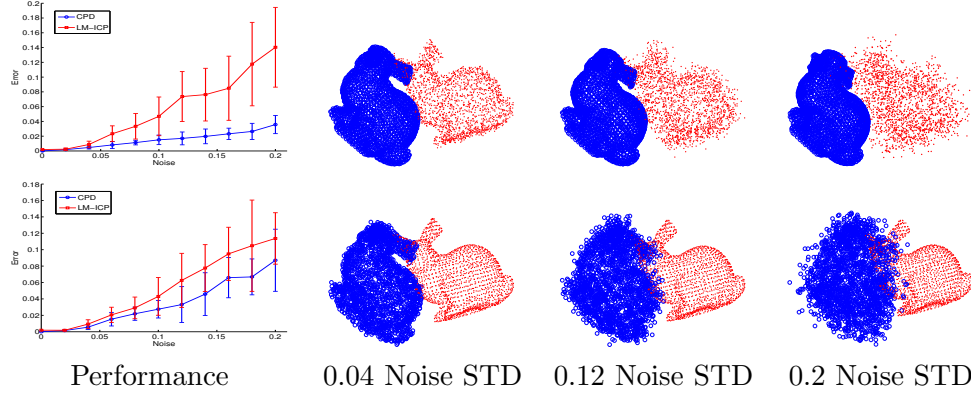


Figure 4.8: A comparison of CPD and LM-ICP rigid registration performances with respect to noise in the  $\mathbf{X}$  (first row) and the  $\mathbf{Y}$  point sets (second row). We align  $\mathbf{Y}$  (blue circles) onto  $\mathbf{X}$  (red dots). The columns 2,3 and 4 show the examples of initial point sets for different random noise stds added to the point set positions. The first column shows the error in estimating the rotation matrix for CPD (blue) and TPS-RPM (red). CPD outperforms LM-ICP in all cases.

We also test the CPD algorithm with respect to different initializations of point sets. Fig. 4.10 shows the registration error under varying degrees of initial rotation when registering the clean fish point set to itself. CPD performs well if the angle of initial misalignment is less than 70 degrees, whereas ICP get trapped into a local minima with rotations beyond 40 degrees. Empirically, CPD is not sensitive to the initial translation and scaling as these parameters are compensated after the first iteration.

**Rigid bunny point set registration:** We test 3D rigid point sets registration on the Stanford “bunny” data set [171]. We use a subsampled bunny version of  $1889 \times 3$  points. In Fig. 4.7a, we have deleted the front and back parts of the bunny point sets. In Fig. 4.7b, we have added random outliers to different sides of the point sets. We set  $w = 0.7$ . CPD registration is accurate and robust in all examples.

We compare the CPD rigid algorithm to the LM-ICP method [53], a robust version of ICP. Fig. 4.8 shows the performance of CPD and LM-ICP with respect to noise in the point sets. We align the  $\mathbf{Y}$  point set (blue circles) onto the  $\mathbf{X}$  point set (red dots). We set  $w = 0.5$ . The known initial rotation discrepancy between the point sets is 50 degrees. The first and second rows shows the alignment performance when a random noise is added to the  $\mathbf{X}$  and  $\mathbf{Y}$  point set positions respectively. We use a norm of the difference between the

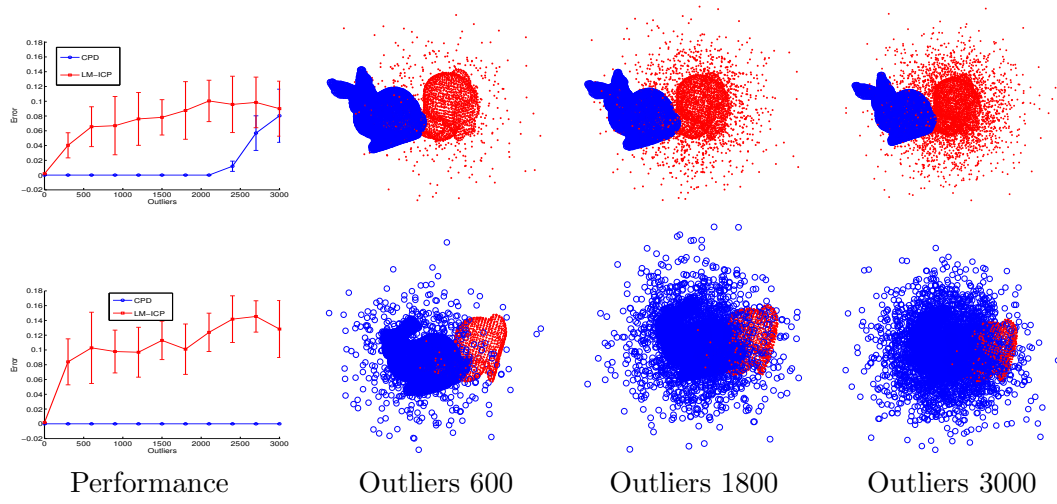


Figure 4.9: A comparison of CPD and LM-ICP rigid registration performances with respect to outliers in the  $\mathbf{X}$  (first row) and the  $\mathbf{Y}$  (second row) point sets. We align  $\mathbf{Y}$  (blue circles) onto  $\mathbf{X}$  (red dots). The columns 2,3 and 4 show the examples of initial point sets with different number of outliers added. The first column show the error in estimating the rotation matrix. CPD outperforms LM-ICP.

true and estimated rotation matrix as an error measure. A few initial point sets examples with different noise std are shown in the columns 2, 3 and 4 of Fig. 4.8. For each level of the noise stds we made 25 independent runs. The first column plots the error values (mean and standard deviation) in the estimated rotation matrix as a function of noise levels. On average CPD required 26 iterations (11 sec total), whereas LM-ICP required 40 iterations (10 sec total). The CPD rigid algorithm outperforms the robust LM-ICP method, especially when the noise is present in the  $\mathbf{X}$  point set.

Fig. 4.9 shows the performance of CPD and LM-ICP with respect to the outliers in the point sets. We add different number of outliers (irrelevant random points) to the point sets. An examples of such initial point sets are shown in columns 2, 3 and 4 of Fig. 4.9 for 600, 1800 and 3000 outlier points added respectively. The first and second row show the cases of outliers present in the  $\mathbf{X}$  and  $\mathbf{Y}$  point sets respectively. CPD performs well in all experiments, whereas LM-ICP performance is less accurate.

**Fast rigid CPD implementation:** We also test the CPD performance with FGT implementation of the matrix-vector products. We use four Stanford bunny sets of sizes:  $453 \times 3$ ,  $1889 \times 3$ ,  $8171 \times 3$  and  $35947 \times 3$ . For each of the cases we add a small amount

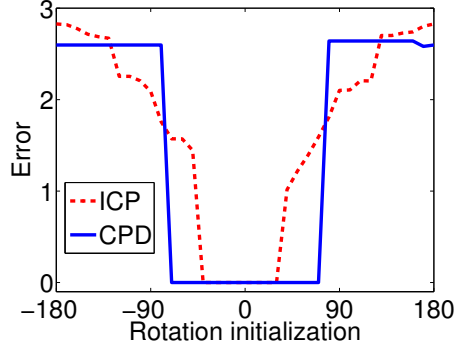


Figure 4.10: Registration error with respect to the initial rotation (degrees) of the fish point set registered to itself. CPD performs well if the angle of initial misalignment is less than 70 degrees, whereas ICP get trapped into a local minima with rotations beyond 40 degrees.

$N, M$	Naive	FGT
$453 \times 3$	0.6s	0.7s
$1889 \times 3$	11s	3s
$8171 \times 3$	4m	10s
$35947 \times 3$	3.5hr	51s

Table 4.1: The rigid CPD registration time for naive (no FGT) and FGT implementations. The FGT-based implementation is significantly faster.

of noise and outliers to both point sets, initialized them with 50 degrees rotation and set  $w = 0.3$ . For the FGT parameters, we used “ratio of far field”=8, “number of centers”=80, “order of truncation”=5. Table. 4.1 shows the registration time with and without FGT. The FGT implementation is significantly faster. We note that there are several downsides of using the FGT: a) FGT requires its own parameter initialization; b) CPD (with FGT) aligns the point sets to 0.1 degree error rotation and then starts being unstable. This is because  $\sigma^2$  becomes small and the FGT approximation error becomes significant. At this point one can either stop (the alignment already is reasonably accurate) or proceed with ICP or truncated Gaussian CPD.

#### 4.7.2 Non-rigid Registration Results

We show CPD non-rigid registration on several examples, test the fast CPD implementation and evaluate CPD performance in comparison to TPS-RPM [32], which is one of the

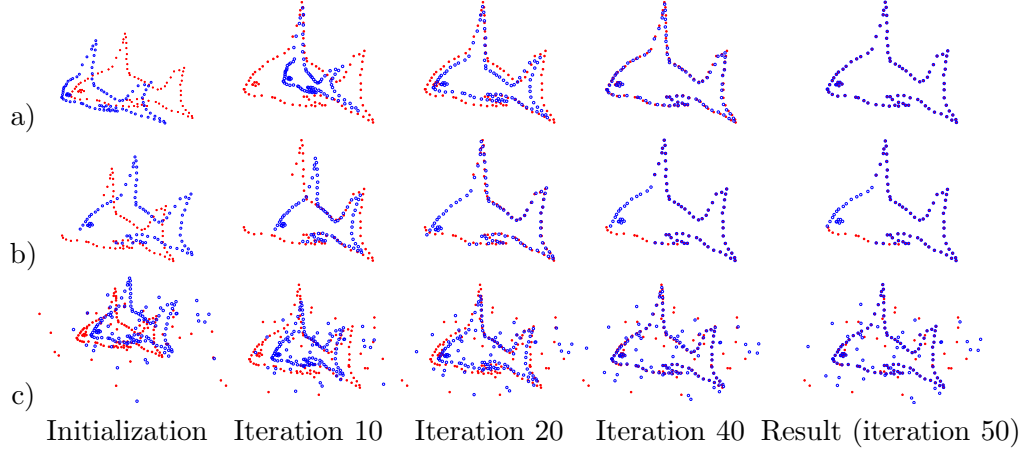


Figure 4.11: Non-rigid CPD registration of 2D fish point sets. a) Noiseless fish point sets registration ( $91 \times 2$  points,  $w = 0$ ); b) Registration of 2D fish point set with missing points ( $w = 0.5$ ); c) Registration of 2D fish point set in presence of outliers ( $w = 0.5$ ). CPD registration is robust and accurate in all experiments.

best performing non-rigid point set registration methods. We set  $\lambda = 2$ ,  $\beta = 2$ .

**Non-rigid fish point set registration:** Fig. 4.11a shows non-rigid CPD registration of two fish point sets with clean data. Fig. 4.11b is with missing points ( $w = 0.5$ ). Fig. 4.11c is with both point sets are corrupted by outliers ( $w = 0.5$ ). The non-rigid CPD registration results are accurate in all experiments.

We test CPD against TPS-RPM [32] on synthetically generated 2D fish non-rigid examples with respect to a) level of non-rigid deformation, b) amount of noise in the point sets locations c) number of outliers. We set  $w = 0.3$  in all experiments. Since we know the true correspondences, we use the mean squared distance between the corresponding points after the registration as an error measure. For each set of parameters we have conducted 25 runs. The computational time of CPD was 3 sec with 31 iterations on average, whereas TPS-RPM required 12 sec with 79 iterations. Fig. 4.12a shows the methods performances with respect to the level of initial non-rigid deformation between the point sets. To generate the non-rigid transformation, we parameterize the point sets domain by a mesh of control points, perturb the points and use splines to interpolate the deformation. The higher level of mesh point perturbations produce the higher deformation. CPD shows accurate registration performance and outperforms the TPS-RPM. Fig. 4.12b shows the

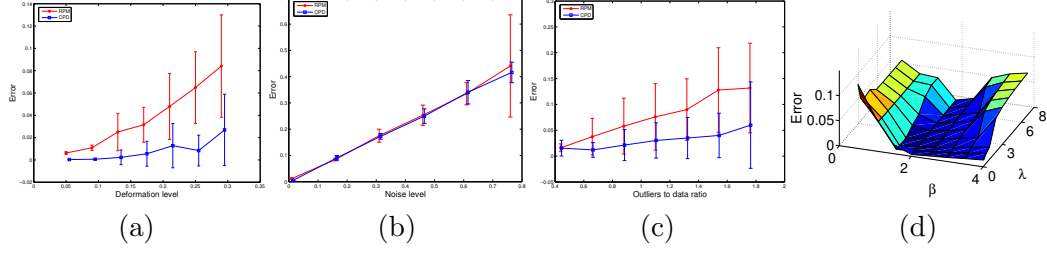


Figure 4.12: A comparison of CPD and TPS-RPM on the 2D fish point sets with respect to a) Deformation level; b) Noise level; c) Outliers (x-axis is the ratio of the number of outliers to the number of clean data points). CPD shows more accurate registration performance compared to TPS-RPM, especially in presence of outliers and complex non-rigid deformations. d) Registration error of CPD with respect to the  $\lambda$  and  $\beta$  parameter settings. CPD performs well for  $\lambda$  in a  $[0.1..9]$  and  $\beta$  in a  $[1.5..3]$  intervals.

methods performances with respect to the amount of noise. We add a zero-mean white noise with increasing levels of stds to the point sets. Both CPD and TPS-RPM show accurate performances. Fig. 4.12c shows the methods performances with respect to the number of outliers. We add random outliers to the point sets and plot the registration error with respect to the ration of number of outliers to the number of data points. At most we have added almost twice as many outliers as the actual data points. CPD shows robust registration performance and outperforms the TPS-RPM.

We also evaluate CPD with respect to the values of parameters  $\lambda$  and  $\beta$  (Fig. 4.12d) on the fish data with small amount of noise and outliers added. CPD performs well for  $\lambda$  in a  $[0.1..9]$  and  $\beta$  in a  $[1.5..3]$  intervals. CPD is more sensitive to the choice of the parameter  $\beta$ , because it defines the model of the non-rigid transformation.

**Non-rigid 3D face registration:** We show the CPD performance on 3D face point sets. Fig. 4.13a shows two 3D face point sets related through non-rigid deformation. Fig. 4.13b shows two 3D face point sets point sets with added outliers and non-rigid deformation. Non-rigid CPD registration is accurate in all experiments.

**Non-rigid 3D LV point set registration:** Finally, we demonstrate the CPD performance on non-rigid a 3D left ventricle (LV) contours segmented from 3D ultrasound images, using active contour based segmentation [115]. Fig. 4.14 shows (a) two LV point sets at different time instances, (b) the registration result, (d) the displacement field required for CPD alignment. The registration result is accurate.

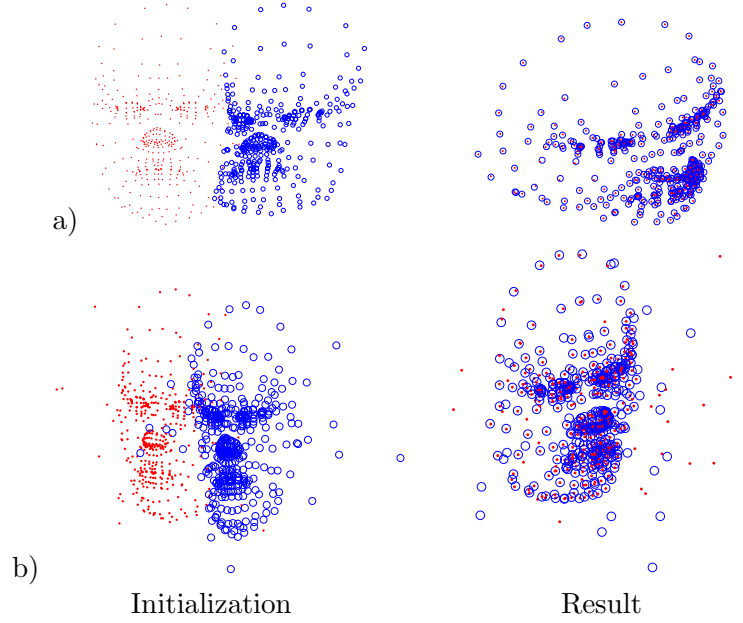


Figure 4.13: Non-rigid registration of 3D face point sets. a) Registration of clean point sets b) Registration of point sets with outliers. CPD shows accurate alignment.

$N, M \times D$	Naive	FGT	Low-rank	FGT & Low-rank
$453 \times 3$	2s	2.3s	1.7s	1.8s
$1889 \times 3$	1m22s	1m16s	19s	11s
$8171 \times 3$	3hr	2hr26m	10m20s	1m37s
$35947 \times 3$	—	—	40m	10m

Table 4.2: Registration time required for non-rigid registration of 3D bunny point sets. The time is shown when using only FGT of vector-matrix products, only low-rank matrix approximation of Gaussian kernel matrix or both.

**Fast non-rigid CPD implementation:** We test the computational time of the fast CPD non-rigid implementation on several subsampled 3D Stanford bunny point sets. We use FGT of the matrix-vector products, the low-rank matrix approximations of the kernel matrix, or both. We applied a moderate non-rigid deformation to the bunny point sets. The registration time of the non-rigid CPD is shown in Table 4.2.

We were unable to run the test without the low-rank matrix approximation for the largest bunny set ( $35947 \times 3$ ), because of the large RAM requirements to construct the kernel matrix  $\mathbf{G}$ . We used only 100 leading eigenvalues and eigenvectors in all cases. Table 4.2 shows that the main computational bottleneck is in solving the linear system

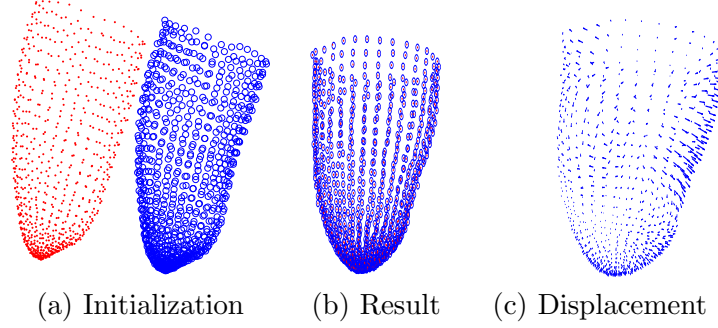


Figure 4.14: Non-rigid registration of 3D left ventricle (LV) point sets. (a) two LV point sets at different time instances, (b) the registration result, (c) displacement field between the corresponding points.

of equations (4.22), because the low-rank matrix approximation alone can reduce the computational time significantly. Both FGT and low-rank approximations provide further speed-up with only moderate loss of accuracy. We note that almost 60% of the time required to complete the CPD registration using the low-rank matrix approximation were required to pre-compute the eigenvalues and eigenvectors of the kernel matrix  $\mathbf{G}$ .

### 4.7.3 Image Registration Results

We show the application of CPD to non-rigid image registration problem. In particular, we register ultrasound (US) elastography images with histopathology images [30]. Ultrasound elastography is a relatively new image modality that is derived from an ultrasound signal by measuring local deformations under the application of an external force. Ultrasound elastography provides the measurements of elastic properties of tissues and is useful to localize a malignant tissue that is usually harder than benign tissue. Histopathology image is an optical, visual band image of the tissue section. Non-rigid registration of such image modalities is required to better understand the information provided by US elastography for tumor detection in comparison to histopathology information.

Figure 4.15(a,b) shows two images to be registered: histopathology and US elastography images of the breast cancer. Boundaries of tumors and fibers were selected as feature points and manually extracted from both images (Fig. 4.15(c)). Given the locations of two



point sets we applied CPD to estimate the non-rigid transformation. The estimated non-rigid transformation has been used to transform the histopathology image (Fig. 4.15(e)) into the coordinate system of the US elastography image. Figure 4.15(f) demonstrates the combined view of the aligned images. Figure 4.15 shows a similar example of feature-based non-rigid registration using CPD. In this example we registered the US elastography image into the coordinate system of the histopathology image. CPD performs well and demonstrates visually accurate image alignment.

## 4.8 Discussion and Conclusion

We introduce a probabilistic method for rigid and non-rigid point set registration, called the Coherent Point Drift algorithm. We consider the alignment of two point sets as a probability density estimation, where one point set represents the Gaussian Mixture Model centroids, and the other represents the data points. We iteratively fit the GMM centroids by maximizing the likelihood and find the posterior probabilities of centroids, which provide the correspondence probability. Core to our method is to force the GMM centroids to move coherently as a group, which preserves their topological structure.

Our contribution includes the following aspects. For the rigid case, we derived the closed form multidimensional solution (of the M-step of the EM algorithm), which has not been derived exactly before. For the non-rigid point set registration, we formulate the motion coherence constraint and derive a solution of the regularized ML estimation through the variational approach, which leads to an elegant kernel form. CPD simultaneously finds both the transformation and the correspondence between two point sets without making any prior assumption on the non-rigid transformation model except that of motion coherence. Finally, we introduced the fast CPD implementation using fast Gauss transform and low-rank matrix approximation to reduce the computational complexity of the method to as low as linear. On top of the computational advantage, the low-rank kernel approximation provides more stable solutions in cases where the matrix  $\mathbf{G}$  is poorly conditioned. To our best knowledge, CPD is the only method capable of non-rigid registration of large data sets. Both rigid and non-rigid CPD registration methods can be applied to arbitrary

dimensional data sets.

We estimate the GMM width,  $\sigma^2$ , within the ML formulation. We have not observed any decrease in performance compared to the deterministic annealing approach. Estimation  $\sigma^2$  allows to reduce the number of free parameters and, most importantly, to significantly reduce the number of iterations and the processing time.

We have used an addition uniform distribution to account for noise and outliers. The weight,  $w$ , of this distribution provides a flexible control over the method robustness and allows accurate CPD performance, especially in presence of severe outliers and missing points.

We have tested CPD on various synthetic and real examples and compare it to LM-ICP (in rigid case) and TPS-RPM (in non-rigid case). CPD shows robust and accurate performance with respect to noise, outliers and missing points. Our method is of general interest with numerous computer vision applications. We provide the Matlab code of the CPD algorithm free for academic research.

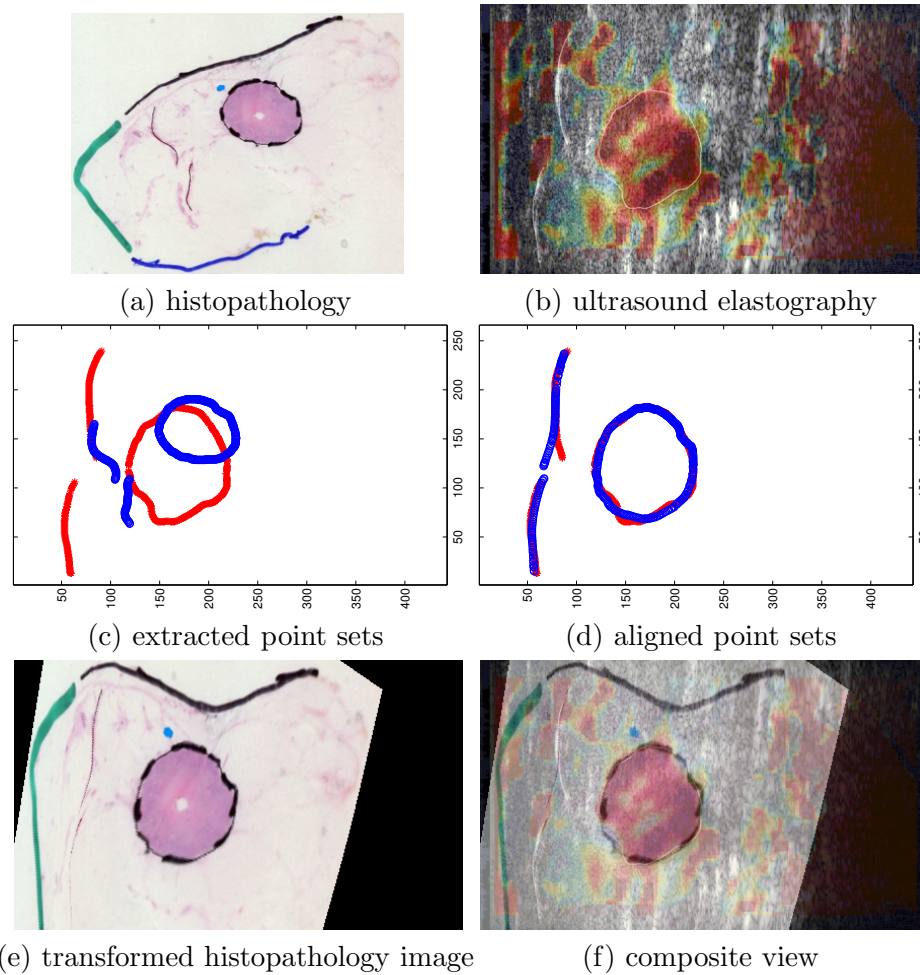


Figure 4.15: Application of CPD non-rigid point set registration to image registration of histopathology (a) and ultrasound elastography (b) images. (c) Two point set were manually extracted from the images. (d) The aligned point sets using CPD non rigid point set registration. (e) The registered histopathology image. (f) Composite view of the images after the alignment.

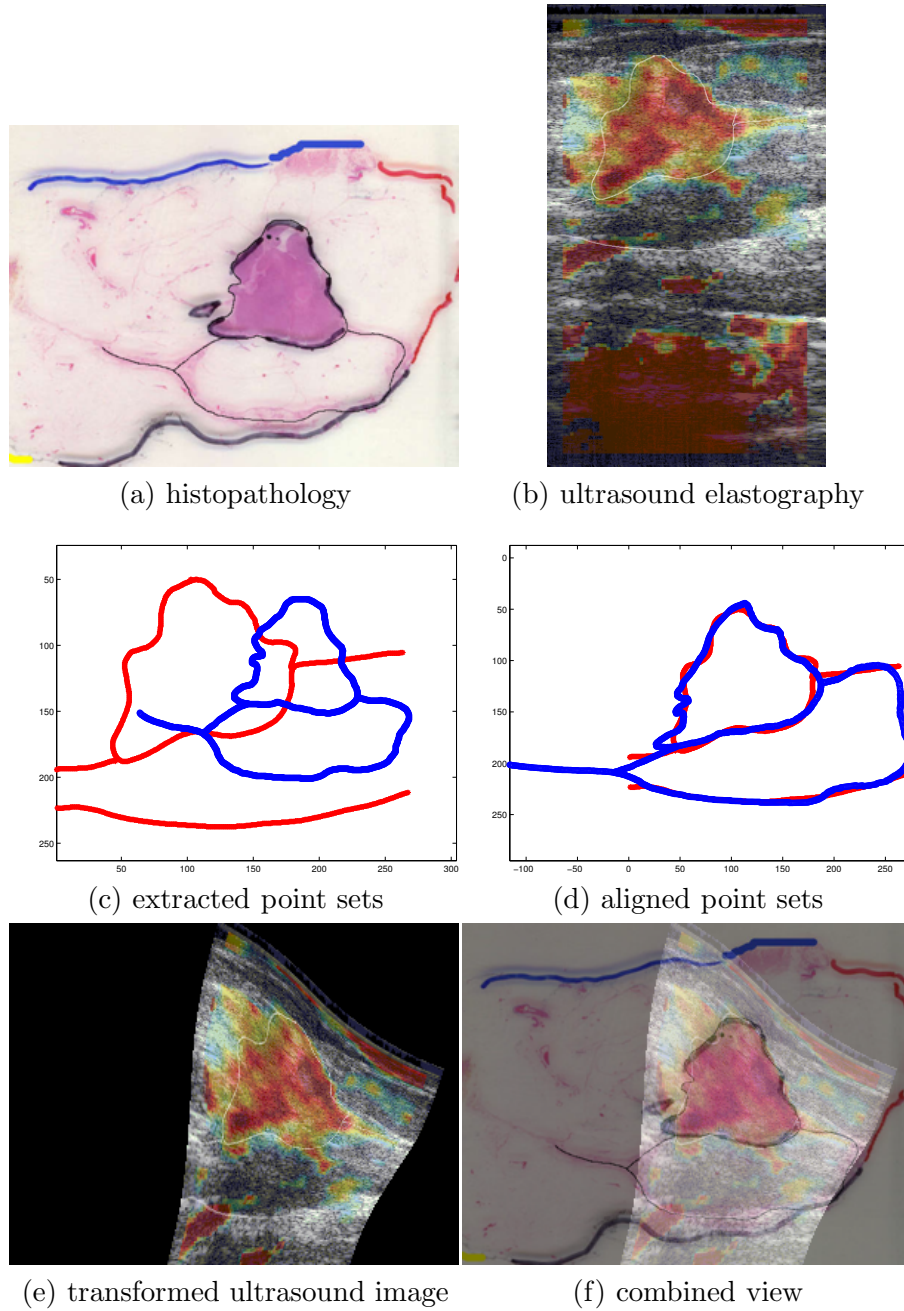


Figure 4.16: Application of CPD non-rigid point set registration to image registration of histopathology (a) and ultrasound elastography (b) images. (c) Two point set were manually extracted from the images. (d) The aligned point sets using CPD non rigid point set registration. (e) The registered ultrasound elastography image. (f) Composite view of the images after the alignment.

# Chapter 5

## Motion Estimation from 3D Echocardiography

### 5.1 Introduction

Cardiovascular disease is a major cause of death in the United States, claiming 876,000 lives each year [1], as many as cancer, chronic lower respiratory diseases, accidents and diabetes mellitus combined. Motion of myocardial tissue conveys the presence of disease processes, and provides insight as to the mechanism by which compensatory processes manifest themselves in pathologic settings. A long sought goal in imaging in cardiology and cardiovascular physiology has been to develop tools by which both global and regional ventricular tissue motion and tissue deformation can be quantitatively analyzed and characterized. Tools for accurate quantification of heart deformation are not well developed yet, despite the efforts from both commercial vendors and the research community. The methods available for characterizing myocardial motion are insufficient to examine the complex array of heart disorders that exist (adult congenital heart disease, cardiomyopathies with dysynergic contraction patterns, ischemic heart disease with regional variations in wall thickening, etc). Even in the normal heart, the complex myocardial motion is inadequately analyzed by current commercial non-invasive imaging modalities.

Current studies of regional myocardium motion include invasive methods, such as opaque markers and sonomicrometers, and non-invasive methods, including MRI and ultrasound based approaches. In sonomicrometry, an array of sono crystals is implanted into the myocardium to provide the groundtruth pairwise distance measurements. Such

invasive methods, while being potentially accurate, provides only sparse measurements (pairwise distances between sono crystals) and incapable to recover the spatially dense deformation maps of the heart motion. MRI based methods, including MR tissue tagging and MR velocity imaging may be the most accurate non-invasive methods [61]. However, high cost and low portability of MRI often prohibits its extensive use in clinical settings.

Ultrasound offers significant advantage over MR based imaging in terms of its availability, portability, cost, minimal adverse secondary effects and high temporal resolution. It can go to the patient, into the interventional or electrophysiology catheterization lab, or into the operating room. Traditionally tissue Doppler imaging has been used to measure the heart motion velocities [122]. However, Doppler-based imaging suffers from angle dependency. It can only detect velocity components parallel to the direction of the ultrasound wave propagation. To overcome such limitation, speckle tracking methods has been proposed [26]. Speckle is a random pattern formed from the interference of the backscattered echoes in myocardium tissue. In contrast to other random noise model, speckle pattern depends on the underlying imaging area, which allows to track speckle (and the myocardium) from echo images. Thus a regional speckle pattern may be considered as natural acoustic tags, equivalent to magnetic tags in MRI.

Historically, ultrasound echocardiography images were two dimensional. Recently developed three dimensional (3D) echocardiography is as a natural evolution of 2D echocardiography [87, 74]. It provides realistic and comprehensive views of cardiac valves and congenital abnormalities; It is extremely useful in the intraoperative and postoperative settings for allowing immediate feedback on the effectiveness of surgical interventions; And, it allows accurate evaluation of cardiac chamber volumes.

Quantitative analysis of 3D echo images can provide mechanical measurements such as strain, wall thickening, torsion, volume and ejection fraction, which can be used to evaluate the elasticity and contractility of the myocardium. For instance, an ischemic or infarcted segments of the heart are typically associated with reduced regional elasticity and contractility. Such measurements may also serve as earlier sub-clinical markers for ventricular dysfunction and myocardial disease [87, 128].

In this chapter, we introduce an integrated automated system for 3D dense speckle

tracking from 4D echocardiography. Our speckle tracking approach consists of several novel components that take into account specifics of ultrasound image formation, speckle pattern statistics, shape constraints and motion dynamic. The key idea is to consecutively align 3D images one onto another and track the motion field using non-rigid image registration [115, 116]. The algorithm outputs a spatially-dense displacement field of heart deformation, which allows to compute quantitative measurements such as strain and torsion. Our novelties include: a) a new similarity measure for ultrasound image registration; we derive it from a maximum likelihood perspective taking into account the physical properties of ultrasound image acquisition and formation; b) an approach to use envelope-detected 3D echo images in the raw *spherical coordinates*, which preserves speckle statistics and represents a compromise between signal detail and data complexity; c) we incorporate shape and dynamic constraints to add robustness to the registration procedure; and d) we validate the algorithm-derived measures to those derived from sonomicrometry in vivo on open chest pigs.

The rest of the chapter is organized as follows. In Section 5.2 we overview current approaches to speckle and boundary tracking. In Section 5.3 we describe the common quantitative characteristics used for regional myocardial analysis. In Section 5.4, we develop a new similarity measure for ultrasound images using physical properties of speckle. In Sections 5.5, we introduce the shape constraint to assist the registration-based tracking. In Sections 5.6, 5.7 and 5.8, we overview the transformation model, the optimization method and the sequential approach that are integrated into the tracking system. In Sections 5.9, we introduce a dynamic model to constrain and smooth the dense displacement field through time. In Section 5.10, we introduce the approach for image registration using envelope-detected 3D echo image in the raw spherical coordinates. In Section 5.11, we explain our validation procedure using implanted sonomicrometers. In Section 5.12 we present our experimental results. This section includes descriptions of animal preparation, data acquisition, automated processing using our system and analysis of the results. Section 5.13 concludes this chapter with discussions.

## 5.2 Motion Estimation Methods

Motion estimation from echocardiography is essential for regional deformation analysis. Initial studies of myocardial deformation relied on Doppler techniques, which provide the myocardial velocity only along the beam lines. To overcome the angular dependency of Doppler techniques, automated tracking methods have appeared; these methods derive the myocardium deformation from the ultrasound signal itself. Myocardium tracking approaches subdivide into boundary tracking and speckle tracking methods. Boundary tracking methods estimate the location of the heart boundaries, usually LV endocardium, throughout the ultrasound sequence. Thus, motion information is only obtained at the boundaries.

Speckle tracking methods track several myocardial regions, which are called speckle patterns. Speckle arises from the effects of interferences caused by microscopic scatterers from the imaged region. Speckle pattern characterizes the underlying myocardial tissue acoustically and serves as a fingerprint of the myocardial region within the ultrasound image. Speckle is just noise, however for the tracking purposes, speckle represents valuable information, because it follows the myocardium motion. Tracking such speckle patterns during the cardiac cycle thus allows to follow the motion of the myocardium. Speckled tracking methods subdivide into block matching methods, optical flow and image registration methods. Block matching methods track one or several square image patches, whereas optical flow and image registration methods tracks the whole imaged area and are referred to as dense speckle tracking methods.

Current tracking methods use 3D echo images to estimate the myocardial motion. Such methods naturally evolved from 2D methods [87, 74]. 3D methods do not suffer from out of plane regional motion and speckle decorrelation, but have a low temporal and spatial resolution due to the bigger number of beams required to cover an imaged 3D sector.

### 5.2.1 Block Matching Methods

The majority of speckle tracking methods are block matching approaches [36, 185, 186, 20, 138, 91]. The idea is to define a square image block (speckle patch) in one image and



then find its best match in the next image. The matching criterion is usually based on a standard similarity measure, such as SSD, SAD, CC or MI. The block is usually matched only within some neighborhood of its initial position to reduce the computational time. The size of each block has to be chosen sufficiently big to avoid multiple erroneous matches, but sufficiently small for the block-motion approximation to be adequate.

Speckle tracking methods can be defined on different signal types: radio frequency (RF) or B-mode. B-mode signal is obtained from the RF signal after filtering, envelope detection and log compression. B-mode signal preserves only signal amplitude and has lower spatial resolution. The benefits of B-mode signal is its lower storage and processing load requirements. Interestingly, Yu et al. [186] showed that the correlation-based block matching of B-mode signals outperforms RF signal matching for the deformations larger than 2% compression, which is a typical inter-frame deformation in 3D echocardiography. They showed that B-mode signal is more robust to higher levels of speckle decorrelation caused by larger deformations, because RF finer structure makes it more vulnerable in the analysis of large deformations.

Tracking individual speckle patches are often prone to errors, which motivated researches to track multiple speckles simultaneously [160, 40]. We proposed to track multiple speckle patches as a coherent group to obey smooth deformation of the myocardium [160]. We used the  $CD_2$  similarity criteria [37] and constrained the groupwise motion to obey the coherence constraint, which enforces spatial smoothness. We also employed the dynamic constraint to enforce temporal continuity and smoothness of speckle motion throughout the sequence [160]. Our method results showed a good agreement with manual expert tracking. Gosby et al. [40] proposed to track each speckle block individually with SAD similarity criteria (followed by CC fine tuning) and post-smooth the individual displacements based on the displacements of its neighbors, which is an ad-hoc way to enforce coherence of the group motion. The authors validated the method on synthetic examples with a good agreement of strain and rotation measurements. They also applied the method to real patient data and showed the reduced strain in the infarcted areas, as was expected theoretically. Duan et al. [45] used a similar block matching strategy; the initial blocks were manually allocated along the endocardial surface and tracked with the

correlation similarity measure followed by local averaging. The ultrasound images were pre-filtered with anisotropic diffusion, which reduces the amount of noise, but preserves the boundaries. The downside of such approach is a loss of potentially useful speckle structure. The endocardial surface tracking method has been validated against manual segmentation of each volume on 40 open-chest dog data sets in terms of surface geometry and volume agreement with satisfactory performance [45]. Groupwise speckle tracking is more accurate compared to the individual speckle tracking. However, even with coherence of block motion, such methods suffer from pure translational motion approximation, which is truly non-rigid. Finally, such methods provide only sparse measurement of the displacement fields for the preselected blocks, which have to be allocated in advance.

### 5.2.2 Boundary Tracking Methods

Boundary tracking methods sequentially segment 3D echo images, typically endo- and epi-cardial surfaces. Such methods integrate ultrasound segmentation approaches into the sequential tracking algorithm by take advantage of the shape and dynamic priors [128, 194, 24]. Segmentation of the boundaries can be based on gradient intensity information [38], phase-based acoustic feature [149], region-based [90] and global and local statistics [19, 195]. A good review of the ultrasound segmentation techniques can be found in [120].

Papademetris et al. [129, 128] decoupled the boundary tracking method into three steps: a) segmentation of the endocardial and/or epicardial surfaces; b) estimation of the initial correspondences between the surfaces; c) estimation of the deformation between the surfaces, which reflects the motion of the myocardium [127]. The key advantage of their method is the use of biomechanical model at the final deformation estimation step. The authors assumed a transversely isotropic linear-elastic model, which accounts for the muscle fiber directions in the LV. Such biomechanical model was implemented as a quadratic regularization term on the components of the strain tensor [128]. Specifically, regularization of the strain tensor components along the fiber direction was set several times greater than the gross-fiber ones. It corresponds to the physical model of fiber stiffness to be greater than gross-fiber stiffness [128]. The model of fiber orientations was taken from the work of Guccione and McCulloch [63]. Papademetris et al. [129, 128]

demonstrated a high correlation of such boundary tracking method with groundtruth implanted markers on open-chest dogs.

Jacob et al. [75] used a Kalman filter tracking of the heart boundaries based on the dynamic active contour tracking approach of Blake and Isard [16]. Similarly, Orderud and Rabben [126] tracked the deformable left ventricle surface with extended Kalman filter. Bosch et al. [18] used an extension of the active appearance model (AAM) approach to represent the shape and the appearance of the endocardium, as well as its motion. Manual delineations were compared with the algorithm contours with an average error of 3.3 mm, which is comparable with intra observer variability. Mitchell et al. [107] included the time dimension into the AAM. The model was learned from manually segmented training set and produces an average error of 3.9 mm between the true and estimated contours.

Yang et al. [182] used a novel one step forward position prediction to generate the motion prior. Their prediction model depends not only on the most recent boundary position, but also on all the previous ones, and is estimated from a training set. Such prediction model was used as an initialization position for the boundary detection based on steerable features.

Zhu et al. [195] tracked the endo- and epi-cardium surfaces by allocating the boundaries so that the intensity distribution of the myocardium, of the blood pool and the rest of the image follow the different parametric Nakagami distributions within the maximum likelihood framework. The authors also used the incompressibility of the myocardium constraint to further regularize the problem. Results from 80 sets of synthetic data and 286 sets of real canine data were evaluated against the ground truth and against outlines from three independent observers, respectively. Improving this work, Zhu et al. [194] built a shape prior model by accounting for both inter subject variability and temporal dynamics of cardiac deformation. This prior model was used to predict the next boundary position by first identifying the closest subject and then predicting the segmentation. Such prediction was used to regularize the estimated boundary position.

Boundary tracking methods are much faster than intensity-based speckle tracking methods, but provide only sparse displacement estimates (along the boundary) and can suffer from the aperture problem.

### 5.2.3 Dense Tracking Methods

Dense speckle tracking methods have emerged based on optical flow [164, 46, 104] and non-rigid image registration [115, 50, 116, 88]. The idea is to track the deformation of the whole 3D volume densely. Such methods utilize the full image information in contrast to block matching or boundary tracking methods.

**Optical Flow Methods:** Optical flow speckle tracking methods follow the classical computer vision optical flow algorithms of Horn and Schunk [72] and Lucas and Kanade [95]. Such approaches assume that the gray values do not change over time. A change of gray value at any voxel is thus caused only by the motion of this voxel to another location.

Mailloux et al. [99] extended the the Horn-Schunck [72] optical flow method by adding linearity constraint to the motion field. Zini et al. [196] also added an additional incompressibility constraint. Suhling et al. [164] extended the Lucas-Kanade [95] optical flow method by using local-affine model for the velocity in space and a linear model in time. Duan et al. [46] validated optical flow speckle tracking with groundtruth implanted sonomicrometers. The assumption of constant gray values is often violated between echo images, especially when interframe displacements are relatively large. Thus, optical flow methods work well only for small interframe displacements, which requires high frame rate. In contrast to optical flow methodology, non-rigid image registration allows to specify the similarity measure between the images, which can account for speckle variability and allows to capture larger deformations.

**Non-rigid Image Registration Methods:** Non-rigid image registration has been proved to be a powerful tool for dense speckle tracking from 2D [88, 101] and recently from 3D echo images [115, 50, 116, 39]. The idea is to sequentially register the next time image onto the previous and track the underlying transformation [115]. Some methods estimate the transformations between all image volumes simultaneously by registering all image volumes onto the first one [39]; such approach can suffer from large speckle decorrelation. Most of the 3D methods use B-splines (FFD) transformation parametrization to be able to cope with large 3D data in a reasonable time.

The key component of the image registration is the similarity measure, that reaches its optimum at the correct image alignment. Elen et al. [50] and Craene et al. [39] used mutual information similarity measure, which is popular for multi-model image registration [175]. While making the least assumption on the image intensity relationship, such similarity measure may not be the most appropriate for ultrasound images - it does not account for speckle variability, which makes it less robust in speckle tracking. In Section 5.4 we introduce a new similarity measure derived specifically for ultrasound images by taking into account physical image acquisition and speckle formation process [116].

While accounting for the dense speckle information, image registration does not account for the shape information, e.g., those of endo- and epi-cardium boundaries. We have proposed a hybrid approach to combine non-rigid image registration and model-based boundary segmentation [115]. Whereas non-rigid image registration accounts for intensity variations (speckle statistics), model-based segmentation adds shape constraints to the tracking procedure. Currently motion estimation from 3D echocardiography through non-rigid image registration is the state-of-the-art in the speckle tracking, especially when prior information, such as shape and dynamic, is also employed.

#### 5.2.4 Regularization of the Estimated Myocardial Deformation

Regardless of the tracking approach used, estimation of the myocardial motion is an ill-posed problem, which requires regularization. Estimated deformations can be regularized in spatial and temporal domains.

**Spatial regularization:** Heart is a connected organ and the motion of the adjacent areas is expected to be coherent. One usually regularizes the estimated displacement field to be spatially smooth. In block matching speckle tracking methods, the regularization is imposed by either penalization of spatial derivatives [160] or an ad hoc smoothing of the adjacent speckle blocks [40, 45] to enforce coherence. In dense speckle tracking methods, myocardium is usually modeled as an elastic material. Standard regularization terms, based on penalization of spatial derivatives of the displacement field, are employed (See Sec. 2.4.3). Such regularization does not account for fiber orientation, which is equivalent

to the isotropic linear elastic mechanical model of the heart. For instance, optical flow speckle tracking method penalize first order derivatives of the displacement field [72, 164]. In non-rigid image registration one also regularizes the displacement field either explicitly (non-parametric approach) or by using parametric approach (e.g. B-splines) that limits admissible deformations [115].

Boundary tracking method also impose smoothness regularization. As far as the initial boundary segmentation is available, such methods can take advantage of the myocardium biomechanical model. Papademetris et al. [128] assumed a transversely isotropic model, which is essentially equivalent to penalization of the displacement field derivatives taking into account fiber orientations. This model is more accurate compared to isotropic linear elastic model, because it takes into account disproportional stiffness of the myocardium along and across fibers.

Myocardium is often considered to be incompressible, which can be used as an addition constraint. This assumption is only approximately true due to the blood flow into the wall. Zhu et al. [193, 195] regularized the volume between the estimated endo- and epi-cardium contours to be constant by penalizing  $L_2$  norm between the initial and current volume estimates. Incompressibility constraint has been also used in non-rigid image registration based methods [50, 15, 92]. For instance, Elen et al. [50] penalized the deviation of the Jacobian determinant from unity in terms of  $L_2$  norm. Incompressibility constraint imposed on the entire image (not only the myocardium) is somewhat questionable as the blood pool undergoes significant volume changes.

Another spatial regularization approach is to take into account the shape constraints. Endo- and epi-cardial boundaries should persist and should remain continuous after the transformation. In fact, boundary tracking method explicitly regularize the heart boundaries using either surface parametric model [8] or active shape model (ASM) and its extensions [194, 120]. Block matching methods can benefit from adaptive smoothing by putting more importance on the speckle blocks located along the myocardium boundaries. In non-rigid image registration, one can add an additional regularization term to account for the boundary shape constraints, e.g. by ensuring that boundaries correspond to the high intensity gradient regions [115].

**Temporal regularization:** Temporal regularization ensures that the displacement field of the myocardium is smooth and continuous in time. Prior motion knowledge, such as distinct pseudo-periodic dynamics of the myocardium motion, can be also utilized to regularize tracking in time. Typically, one uses recursive Bayesian estimation [177] to smooth the estimated displacement field trajectory in time according to a specified dynamic model [24, 182].

We have used a simple random walk model to smooth the estimated speckle trajectories over time [160]. Such model does not require any learning steps, can be implemented fast using the Kalman filter and provides sufficient temporal smoothing of speckle trajectories. Such a simple model have a low computational complexity, but does not include any heart specific motion priors.

More sophisticated dynamic models are possible only given the location of heart (e.g. LV endocardium and epicardium), because one can apply the physical dynamic model only to the organ, and not to the whole image domain, as in the dense speckle tracking methods. Such models are usually non-linear and require to use particle filtering [177]. In notations of tracking methods, the position of heart boundaries can represent the state of a system. One is interested in modeling the state transition model, which describes the motion of the heart between two time instances [24]. Sun et al. [165] described cardiac dynamics using second-order non-linear transition model, which is learned from the training data. Similarly, Yang et al. [182] used a transition model that depends not only on a single previous state, but also on all the other previous states. The transition model was automatically learned from the training data and represented as a manifold that describes the motion pattern of the heart. Nascimento and Marques [118] proposed to use mixture of two transition models describing the systole (shrinking) and diastole (dilation) cardiac cycles, without the need of having an extensive training set. Carneiro and Nascimento [24] further improved this work by additionally learning the model from the data.

Zhu et al. [194] built a shape prior model by accounting for both inter subject variability and temporal dynamics of cardiac deformation. They used multilinear PCA and ICA to reduces the dimensionally, which is similar to the Active Shape Model (ASM) approach but also includes the temporal variability. This prior model was used to predict the

next boundary position by first identifying the closest subject and then predicting the segmentation, which is similar to recursive Bayesian tracking. Their objective function consists of data fitness term, which depends on the image and the regularization term, which is the  $L_2$  norm between the the estimated and predicted boundary positions [194]. This is a similar approach to the previous work of Zhu et al. [193, 195], but regularizes the boundary segmentation based on dynamic model rather than using the incompressibility constraint.

One downside of the dynamic model learned from the training data is its potential danger when applying to the cases not described by the training set, e.g. infarcted heart. In such cases, the dynamic model can significantly bias the deformation estimate toward the expected normal motion of the myocardium, whereas the actual motion pattern may include a few dysfunctional regions.

### 5.3 Quantitative Characteristics

From the estimated dense transformation field, we can derive quantitative characteristics to analyze the cardiac mechanic and function. Strain is one of the fundamental quantitative characteristics that is used to analyze myocardial deformations. Strain is a relative change of an object length. Linear strain is defined as

$$\varepsilon = \frac{\mathbf{L} - \mathbf{L}_0}{\mathbf{L}_0} \quad (5.1)$$

where  $\mathbf{L}_0$  is the initial and  $\mathbf{L}$  is the current distance between two points. Such strain is often called Lagrangian strain, which is an object lengthening relative to its original length. Negative strain corresponds to object shortening. Such strain can be also called longitudinal strain, but this definition is not advised because of the possible confusion with longitudinal heart orientation. Strain is dimensionless ratio, which can be also expressed in percent. Another important quantitative characteristic is strain rate, which is the strain change per time instance

$$\varepsilon' = \frac{d\varepsilon}{dt} \quad (5.2)$$

Strain rate indicates the rate of deformation change. Lagrangian strain is unknown if the original length is unknown. A more natural approach may be to define instantaneous



strain in relation to the instantaneous length and integrate over the time of observation

$$\varepsilon_N = \int_{t_0}^{t_1} \frac{d\mathbf{L}}{\mathbf{L}} = \int_{t_0}^{t_1} \frac{\mathbf{L} - \mathbf{L}_0}{\mathbf{L}} = \ln \frac{\mathbf{L}}{\mathbf{L}_0} \quad (5.3)$$

Such strain is called natural or Eulerian strain; it is related to the Lagrangian strain through the equality

$$\varepsilon_N = \ln(\varepsilon + 1) \quad (5.4)$$

Myocardial deformation is truly three dimensional, that is any myocardial region, e.g. speckle patch, undergoes deformation in three dimensions. 3D strain can be described through the strain tensor, which is a  $3 \times 3$  matrix with 3 diagonal elements corresponding to normal strain along the coordinate axes and 6 off-diagonal shear components, which measure angular changes between the axes. The orientation of the coordinate axes can be global, e.g. relative to the transducer probe location. For LV imaging, a local coordinate system is often used relative to the LV boundary, so that normal strain components correspond to longitudinal (along the heart main axis), transmural (perpendicular to the heart surface) and circumferential (around the heart) directions. Local coordinate system is always relative to the placement of its origin on the LV; its orientation varies spatially. Only dense speckle tracking allows to calculate the strain tensor, which provides more comprehensive evaluation of myocardial deformation.

Another useful quantitative characteristic is torsion. Torsion is defined as a base to apex angle of rotation along the longitudinal axis of the LV. Usually, one defines the torsion as a rotational angle between two parallel planes located at 20% and 80% from the apex level. Other quantitative characteristics include untwisting velocity, LV and RV volumes and ejection fraction [87, 74].

## 5.4 Similarity Measure Based on Physics of Speckle Noise

Speckle arises from the effects of interferences caused by microscopic scatterers from the imaged region. Speckle is just noise, which corrupts the image. For the tracking purposes, speckle represents a valuable information, because it follows the myocardium motion.

Registration of 3D echocardiogram can take advantage of the physical properties of the ultrasound images. In this section we derive a new similarity measure from the Maximum

likelihood (ML) perspective by taking into account speckle formation and acquisition process.

#### 5.4.1 Speckle Correlation and Blurring

The speckle noise is a random signal, but has implicit dependence on the underlying imaging object structure. If the static object is imaged twice, the speckle remains the same [58, 34]. This is in contrast to other noise cases, e.g. electrical noise, which always has a random nature. In the seminal paper on ultrasound speckle formation, Burckhardt [22] investigated the speckle correlation for a static object with respect to the rigid translation of a transducer. He theoretically derived the correlation coefficient to be  $\rho = (a-b)^2/a^2$  for  $|b| \leq a$  and zero otherwise, where  $a$  is a transducer aperture length and  $b$  is a translation amount. Which means that if we move the transducer, e.g. on 10% of its aperture length, the correlation coefficient will be  $\rho = 0.81$ . Other researches also showed the dependence of speckle correlation on object translation [124], rotation [80], linear transformations [104]. The myocardium motion is complex nonrigid deformation, but if two consecutive images are taken with sufficiently high frame rate, which is the case for modern ultrasound devices, the speckle noise formation between the consecutive frames is similar, and 2 speckle noise variables are correlated [34]. Another important aspect to take into account is that the actual measured speckle is *blurred*. In an experiment, speckle can not be measured at an ideal point, but rather is spatial-integrated over some finite area of a detector element. In addition, even with an extremely tiny detector, a speckle is moving and we are measuring again a speckle time-integrated pattern [59]. The spatial and temporal speckle integration both take places in echocardiography and it is important to use statistics of blurred speckle for more accurate motion estimation.

#### 5.4.2 Maximum Likelihood Motion Estimation

Consider two 3D ultrasound images  $I$  and  $J$  acquired at consecutive time instances. The maximum likelihood (ML) approach to estimate the transformation  $\mathcal{T}$  between the images is to maximize the conditional probability [37, 141],  $p(J(\mathcal{T})|I, \mathcal{T})$ , or equivalently minimize

the following energy function

$$E(\mathcal{T}) = \sum_{n=1}^N -\log p(J_n(\mathcal{T})|I_n, \mathcal{T}) \quad (5.5)$$

where we assumed that all voxelwise conditional probabilities are i.i.d. and  $J_n(\mathcal{T})$  denotes the intensity values of  $n^{th}$  voxel after applying the transformation  $\mathcal{T}$ . The negative log-likelihood function defines the (dis-) similarity function between the image voxels, which is minimized to align the images.

In ultrasound imaging, the image intensities are the log-compressed values of the envelope amplitudes (denoted as  $a$  and  $b$ ) [34]:

$$I_n = D \log(a) + G; \quad J_n(\mathcal{T}) = D \log(b) + G \quad (5.6)$$

where  $G$  and  $D$  are linear gain and scaling constants of the dynamic range.

When the speckle is fully developed the noise is multiplicative and follows the Rayleigh pdf [34, 59]:

$$p(n) = \frac{n}{\sigma^2} \exp \frac{-n^2}{2\sigma^2}, \quad n \geq 0 \quad (5.7)$$

Thus  $a = sn_1$  and  $b = sn_2$  can be seen as a realization of the informative signal  $s$  corrupted by multiplicative Rayleigh noise  $n_1$  and  $n_2$  respectively. Simplifying Eq. 5.6, we achieve

$$I_n = J_n(\mathcal{T}) + D \log\left(\frac{n_1}{n_2}\right) \quad (5.8)$$

Defining the division of two random variable as  $\eta = \frac{n_1}{n_2}$ , we can compute the conditional probability in Eq. 5.5 as

$$p(J_n(\mathcal{T})|I_n, \mathcal{T}) = \frac{1}{D} \eta p(\eta) \quad (5.9)$$

The probability density function of the division of two random variables can be found as [130]

$$p(\eta) = \int_{-\infty}^{\infty} n_2 p(\eta n_2, n_2) dn_2 \quad (5.10)$$

which defines the conditional probability as

$$p(J_n(\mathcal{T})|I_n, \mathcal{T}) = \frac{\eta}{D} p(\eta) = \frac{\eta}{D} \int n_2 p(\eta n_2, n_2) dn_2 \quad (5.11)$$

### 5.4.3 Independent Rayleigh Noise Assumption

Cohen and Dinstein [36, 37] assumed the noise at consecutive ultrasound images to be independent with equal variances, then Eq. 5.11 simplifies to

$$p(J_n(\mathcal{T})|I_n, \mathcal{T}) = \frac{\eta}{D} \int_{-\infty}^{\infty} n_2 p(\eta n_2) p(n_2) dn_2 = \frac{\eta}{D} \int_{-\infty}^{\infty} \frac{\eta n_2^3}{\sigma^4} e^{-n_2^2 \left( \frac{\eta^2 + 1}{2\sigma^2} \right)} dn_2 = \frac{2\eta^2}{D(\eta^2 + 1)^2} \quad (5.12)$$

From the Eq. 5.8, we find that<sup>1</sup>

$$\eta = \frac{n_1}{n_2} = e^{\frac{J_n(\mathcal{T}) - I_n}{D}} = e^{d_n}, \quad (5.13)$$

where by  $d_n$  we define a scaled difference between the intensities of corresponding image voxels:  $d_n = \frac{J_n(\mathcal{T}) - I_n}{D}$ . The similarity function (Eq 5.5) takes the form

$$E(\mathcal{T}) = \sum_{n=1}^N -\log \frac{e^{2d_n}}{(e^{2d_n} + 1)^2} - \log \frac{2}{D}; \quad d_n = \frac{J_n(\mathcal{T}) - I_n}{D} \quad (5.14)$$

where the last term  $\log \frac{2}{D}$  is a constant, which can be ignored. Cohen and Dinstein derived this similarity function and called it  $CD_2$  similarity measure [36, 37]. We also note, that Cohen and Dinstein in its original derivation do not account for scaling constant  $D$  as in Eq. 5.6, which is equivalent to assume  $D = 1$ .  $CD_2$  similarity measure showed an improved performance comparing to other similarity measure in ultrasound tracking, and later found a widespread attention in block matching algorithms [20, 138, 160, 91].

### 5.4.4 Correlated Rayleigh Noise Assumption

In this section we build upon the work of Cohen and Dinstein [36, 37] by assuming that Rayleigh noise  $n_1$  and  $n_2$  are not independent. If two consecutive images  $I$  and  $J$  are taken with sufficiently high frame rate, which is the case for modern ultrasound devices, the speckle noise formation between the consecutive frames is similar, and the random variables  $n_1$  and  $n_2$  are correlated [34].

---

<sup>1</sup>Usually, the ultrasound log-compression is done with decimal logarithm function, which results in  $10^{2d_n} = e^{2d_n \ln 10}$  terms instead of  $e^{2d_n}$ . We absorb the scaling  $\ln 10$  into the constant  $D$  for convenience.

The bivariate Rayleigh distribution [59] of two random variable  $n_1$  and  $n_2$  with equal variance  $\sigma^2$  is

$$p(n_1, n_2) = \frac{n_1 n_2}{(1 - \rho)\sigma^4} e^{-\frac{n_1^2 + n_2^2}{2\sigma^2(1-\rho)}} I_0 \left[ \frac{\sqrt{\rho} n_1 n_2}{(1 - \rho)\sigma^2} \right] \quad (5.15)$$

where  $I_0$  is a modified Bessel function of the first kind with zero order

$$I_0(z) = \sum_{k=0}^{\infty} \frac{z^{2k}}{2^{2k} (k!)^2} \quad (5.16)$$

and  $0 \leq \rho < 1$  is correlation coefficient and  $n_1 \geq 0$ ,  $n_2 \geq 0$ . Substitution Eq. 5.15 in Eq. 5.11, we have to take a non-trivial integral

$$\begin{aligned} p(J_n(T)|I_n, T) &= \frac{\eta}{D} \int_0^{\infty} n_2 p(\eta n_2, n_2) dn_2 = \frac{\eta}{D} \int_0^{\infty} \frac{\eta n_2^3}{(1 - \rho)\sigma^4} e^{-\frac{n_2^2(1+\eta^2)}{2\sigma^2(1-\rho)}} \cdot I_0 \left( \frac{\sqrt{\rho} \eta n_2^2}{(1 - \rho)\sigma^2} \right) dn_2 \\ &= \frac{\eta}{D} \int_0^{\infty} \frac{\eta n_2^3}{(1 - \rho)\sigma^4} e^{-\frac{n_2^2(1+\eta^2)}{2\sigma^2(1-\rho)}} \sum_{k=0}^{\infty} \frac{\left( \frac{\sqrt{\rho} \eta n_2^2}{(1 - \rho)\sigma^2} \right)^{2k}}{2^{2k} (k!)^2} dn_2 \\ &= \frac{\eta}{D} \sum_{k=0}^{\infty} \frac{\eta^{2k+1} \rho^k}{\sigma^{4k+4} (1 - \rho)^{2k+1} (k!)^2 2^{2k}} \int_0^{\infty} n_2^{3+4k} e^{-n_2^2 \frac{(1+\eta^2)}{2\sigma^2(1-\rho)}}. \quad (5.17) \end{aligned}$$

Using the formula

$$\int_0^{\infty} x^m e^{-ax^2} = \frac{\Gamma(\frac{m+1}{2})}{2a^{\frac{m+1}{2}}} \quad (5.18)$$

we achieve

$$\begin{aligned} p(J_n(T)|I_n, T) &= \frac{\eta}{D} \sum_{k=0}^{\infty} \frac{\eta^{2k+1} \rho^k}{\sigma^{4k+4} (1 - \rho)^{2k+1} (k!)^2 2^{2k}} \frac{\Gamma(2k+2)}{2 \left( \frac{(1+\eta^2)}{2\sigma^2(1-\rho)} \right)^{2k+2}} = \\ &= \frac{2(1 - \rho)\eta^2}{D(1 + \eta^2)^2} \sum_{k=0}^{\infty} \left( \frac{\eta^2 \rho}{(1 + \eta^2)^2} \right)^k \frac{(2k+1)!}{(k!)^2}. \quad (5.19) \end{aligned}$$

Finally, using the formula for infinite series [157]

$$\sum_{k=0}^{\infty} z^k \frac{(2k+1)!}{(k!)^2} = \frac{1}{\sqrt{(1-4z)^3}} \quad (5.20)$$

we obtain

$$p(J_n(T)|I_n, T) = \frac{2(1 - \rho)\eta^2}{D(1 + \eta^2)^2} \left( 1 - \frac{4\rho\eta^2}{(1 + \eta^2)^2} \right)^{-\frac{3}{2}} \quad (5.21)$$

The similarity function (Eq 5.5) takes the form

$$E(\mathcal{T}) = \sum_{n=1}^N -\log \frac{e^{2d_n}}{(e^{2d_n} + 1)^2} + \frac{3}{2} \log(1 - \frac{4\rho e^{2d_n}}{(e^{2d_n} + 1)^2}) - \log \frac{2(1 - \rho)}{D} =$$

$$\sum_{n=1}^N \frac{3}{2} \log(\cosh^2(d_n) - \rho) - \frac{1}{2} \log(\cosh^2(d_n)) + \text{const} \quad (5.22)$$

where  $d_n$  is a scaled difference between the intensities of corresponding image voxels:  $d_n = \frac{J_n(\mathcal{T}) - I_n}{D}$ . By convention, we shall call this similarity function the *MS* similarity measure. As a check on this result, if the correlation coefficient  $\rho = 0$ , then *MS* simplifies to *CD*<sub>2</sub> in Equation 5.14.

#### 5.4.5 Blurred Speckle

The probability density function of speckle noise at an ideal point in space is quite idealistic. During data acquisition, the speckle is not measured at an ideal point, but rather integrated over some area of detector element. Also, the speckle pattern motion, caused by imaging area motion, makes the measurement quantity integrated over some part of ideal speckle pattern. This results in *blurred speckle* effect, which modifies the density function [58, 59].

To deal with the same problem with optical speckle, Goodman [58], following the work of Rice [140], proposed to approximate the continuous intensity pattern falling on the detector by a two-dimensional boxcar function, consisting of  $m$  rectangular pillboxes with different height. The intensity values represented by pillboxes are assumed to be iid with negative exponential distribution. Then the probability density function of intensity is a Gamma density

$$p(w) = \frac{1}{\Gamma(m)} \left( \frac{m}{2\sigma^2} \right)^m w^{m-1} e^{-\frac{mw}{2\sigma^2}}, \quad w > 0 \quad (5.23)$$

In ultrasound imaging, the detector is pressure sensitive, which means that we measure the amplitude (square root of intensity). Using a change of variable for the Gamma pdf ( $n = \sqrt{w}$ ), we achieved that the blurred speckle has a Nakagami distribution [117]

$$p(n) = \frac{m^m}{2^{m-1}\sigma^{2m}\Gamma(m)} n^{2m-1} e^{-\frac{mn^2}{2\sigma^2}}, \quad n > 0 \quad (5.24)$$

which is also called a generalized Rayleigh distribution [105]. For  $m = 1$  the distribution reduces to the Rayleigh distribution (Eq. 5.7). The application of the Nakagami distribution in ultrasound imaging has been popularized by Shankar [154]. He showed that the Nakagami distribution fits the real speckle distribution superior to Rayleigh and K-distributions. There is no relationship between Shankar's [154] application of the Nakagami distribution and Goodman's [58, 59] approach to blurred speckle statistics. Both works, however, show that Nakagami distribution is superior to other common distributions to model speckle statistics.

To derive the similarity measure, assuming that two speckle variable are correlated, we use the bivariate Nakagami distribution [117, 167]

$$p(n_1, n_2) = \frac{m^{m+1}(n_1 n_2)^m}{2^{m-1} \sigma^{2(m+1)} (1-\rho) \rho^{\frac{m-1}{2}} \Gamma(m)} e^{-\frac{m(n_1^2 + n_2^2)}{2(1-\rho)\sigma^2}} I_{m-1} \left( \frac{m\sqrt{\rho} n_1 n_2}{(1-\rho)\sigma^2} \right) \quad (5.25)$$

Here, we assumed that the distribution widths  $\sigma^2$  are equal,  $\rho$  is a squared correlation coefficient  $0 \leq \rho < 1$ ,  $m > 0.5$  and  $I_{m-1}$  is a modified Bessel function of the first kind of order  $m - 1$ :

$$I_{m-1}(z) = \sum_{k=0}^{\infty} \frac{z^{2k+m-1}}{2^{2k+m-1} \Gamma(k+m) k!} \quad (5.26)$$

As a check, for  $m = 1$  the bivariate Nakagami distribution simplifies to bivariate Rayleigh distribution in Eq. 5.15.

Substitution Eq. 5.25 in Eq. 5.11, we take the integral

$$\begin{aligned} p(J_n(T)|I_n, T) &= \frac{\eta}{D} \int_0^{\infty} n_2 p(\eta n_2, n_2) dn_2 = \\ &= \frac{\eta}{D} \int_0^{\infty} \frac{m^{m+1} \eta^m n_2^{2m+1}}{2^{m-1} \sigma^{2(m+1)} (1-\rho) \rho^{\frac{m-1}{2}} \Gamma(m)} e^{-\frac{m(\eta^2 + 1)n_2^2}{2(1-\rho)\sigma^2}} I_{m-1} \left( \frac{m\sqrt{\rho} \eta n_2^2}{(1-\rho)\sigma^2} \right) dn_2 = \\ &= \frac{\eta}{D} \sum_{k=0}^{\infty} \frac{m^{2m+2k} \eta^{2k+2m-1} \rho^k}{2^{2k+2m-2} (1-\rho)^{2k+m} \sigma^{4k+4m} \Gamma(k+m) \Gamma(m) k!} \int_0^{\infty} n_2^{4k+4m-1} e^{-\frac{m n_2^2 (1+\eta^2)}{2\sigma^2(1-\rho)}}. \end{aligned} \quad (5.27)$$

Using the formula from Eq. 5.18, we achieve

$$p(J_n(T)|I_n, T) = \frac{2(1-\rho)^m \eta^{2m}}{D(\eta^2 + 1)^{2m} \Gamma(m)} \sum_{k=0}^{\infty} \left( \frac{\rho \eta^2}{(\eta^2 + 1)^2} \right)^k \frac{(2k + 2m - 1)!}{(k + m - 1)! k!} \quad (5.28)$$

Finally, using the formula for infinite series [158]

$$\sum_{k=0}^{\infty} z^k \frac{(2k + 2m - 1)!}{(k + m - 1)! k!} = \frac{(2m - 1)!}{(m - 1)!} (1 - 4z)^{-\frac{2m+1}{2}} \quad (5.29)$$

we obtain

$$p(J_n(\mathcal{T})|I_n, \mathcal{T}) = 2(1 - \rho)^m \frac{\Gamma(2m)}{\Gamma(m)^2} \frac{\eta^{2m}}{D(\eta^2 + 1)^{2m}} \left(1 - \frac{4\rho\eta^2}{(1 + \eta^2)^2}\right)^{-\frac{2m+1}{2}} \quad (5.30)$$

By definition  $\beta(m) = \frac{\Gamma(m)^2}{\Gamma(2m)}$ , where  $\beta(m)$  is a central beta function. The similarity function (Eq 5.5) takes the form

$$\begin{aligned} E(\mathcal{T}) = \sum_{n=1}^N -m \log \frac{e^{2d}}{(e^{2d} + 1)^2} + \frac{2m+1}{2} \log(1 - \frac{4\rho e^{2d}}{(e^{2d} + 1)^2}) - \log \left( \frac{2(1 - \rho)^m}{D\beta(m)} \right) = \\ \sum_{n=1}^N \frac{2m+1}{2} \log(\cosh^2(d) - \rho) - \frac{1}{2} \log(\cosh^2(d)) - \log \left( \frac{(1 - \rho)^m}{D2^{2m-1}\beta(m)} \right), \end{aligned} \quad (5.31)$$

where  $d_n$  is a scaled difference between the intensities of corresponding image voxels:  $d_n = \frac{J_n(\mathcal{T}) - I_n}{D}$ . By convention, we shall call this similarity function  $MS_2$  similarity measure. The last term is just a constant if  $m$  and  $\rho$  parameters are fixed. As a check, for  $m = 1$  the  $MS_2$  similarity measure simplifies to  $MS$  in Eq. 5.22.

#### 5.4.6 Analysis

The  $CD_2$  similarity measure, we believe, is of general interest for image registration. We can rewrite the Eq. 5.14 in different but equivalent form as  $E(\mathcal{T}) \propto \sum_n -\log \text{sech}^2(d_n)$ , where  $\text{sech}$  is a hyperbolic secant function. Thus, we could have achieved the same similarity measure by simply assuming an additive image noise with logistic pdf. The logistic distribution resembles the Gaussian distribution, but has heavier tails (higher kurtosis) and thus belongs to the class of super-gaussian distributions.

From the optimization perspective, we can analyze the shape of the voxel-wise similarity function and its derivative (influence function), rather than the distribution itself. The shape of  $CD_2$  similarity function resembles the shape of the absolute value function ( $L_1$  norm between the images), but is also differentiable at the origin. The derivative of the  $CD_2$  function is  $\tanh$ , which approximates the step function. Figure 5.1 shows the plot of  $CD_2$  similarity function and its derivative versus  $L1$  norm ones.  $L1$  norm belongs to the family of robust similarity measures with many applications in statistics, machine learning and computer vision areas. This supports the robust and accurate performance of  $CD_2$  measure in image registration.



The  $MS$  similarity measure generalizes the  $CD_2$  measure. Figure 5.2 shows the plot of  $MS$  similarity function and its derivative for different values of correlation coefficient  $\rho$ . For  $\rho = 0$ , the  $MS$  simplifies to  $CD_2$ . For the larger values of correlation coefficient,  $MS$  has a sharper peak in a vicinity of the origin and otherwise asymptotically approaches the  $L1$  norm function. The  $MS_2$  similarity measure further generalizes  $MS$ . Additional parameter  $m$  controls the angle of side-lobes of the similarity function, which adds extra control on the difference between smaller and larger intensity distance between the images.

From the optimization perspective, the higher the correlation value, the higher weight is assigned to the closer intensity differences, which makes the similarity function more robust to outliers (large differences in intensity). The behavior of  $MS$  and  $MS_2$  function becomes similar to the behavior of functions in the robust M-estimator approach. The robustness to outliers are crucial in registration of ultrasound images, which typically includes multiple intensity artifacts.

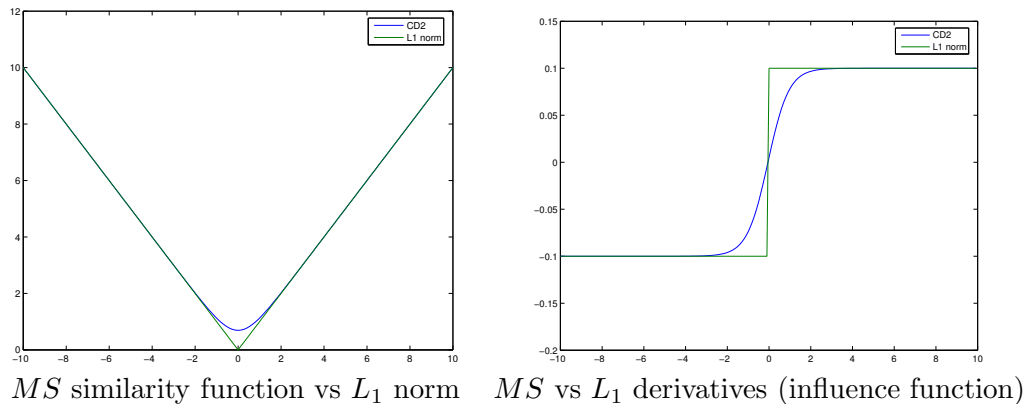


Figure 5.1: The plot of  $CD_2$  similarity function and its derivative versus  $L1$  norm function.

We note that some other distribution models have been proposed to model speckle pdf including K-distribution [153], generalized Nakagami distribution [155] and Rician Inverse Gaussian (RiIG) distribution [51]. Among them, generalized Nakagami and RiIG distributions showed to have better performance [51].

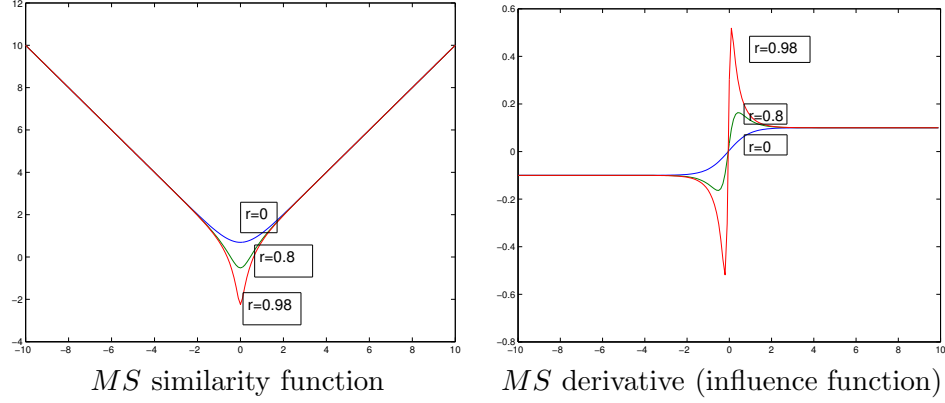


Figure 5.2: The plot of  $MS$  similarity function and its derivative for different values of correlation coefficient  $\rho$ . When  $\rho = 0$ , the  $MS$  similarity measure simplifies to  $CD_2$ .

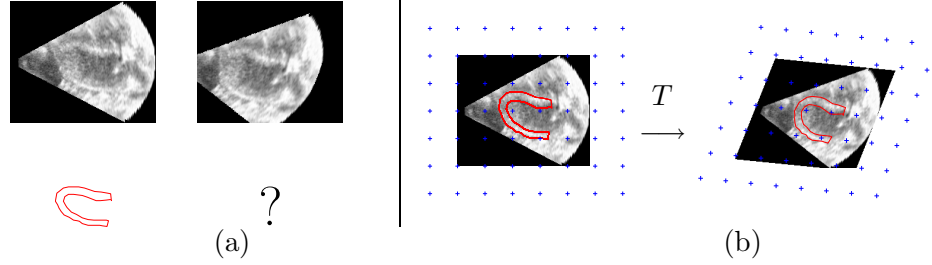


Figure 5.3: Joint non-rigid image registration and segmentation schema. (a) Given two 3D ultrasound images  $I$  and  $J$ , and the initial position of the contour  $C_t$  (b) We use FFD to parametrize the transformation  $T$  by a mesh of control points (+). By manipulating the control points positions, we want to solve two tasks simultaneously: to deform the image  $I$  to align it with  $J$ , and to deform contour  $C_t$  to align it to the high intensity gradient region of the image  $J$ .

## 5.5 Shape Constraints

Using the speckle pattern (intensity-based similarity measure) alone is prone to errors due to multiple 3D ultrasound image artifacts. We use the heart shape boundary constraints to add robustness to the non-rigid image registration tasking [115]. Consider two consecutive 3D ultrasound images,  $I$  and  $J$ , obtained at time  $t$  and  $t+1$  respectively. Assume an initial boundary segmentations of the LV endocardium and epicardium are available at time  $t$ . These two boundaries define the LV contour  $C_t$ .

We seek the transformation  $\mathcal{T}$  that not only optimizes the intensity-based similarity

measure but also aligns the initial contour  $C_t$  to its new position that lies onto the high intensity gradient region. We adopt the active contour segmentation approach [82], with the following objective function

$$E_{shape}(\mathcal{T}) = \left\| \nabla \tilde{J}(\mathcal{T}(C_t)) \right\|^2 \quad (5.32)$$

where  $\tilde{J}$  is the image  $J$  after preprocessing by anisotropic diffusion to remove noise and emphasize the boundaries. Thus our tracking method between two consecutive volumes is a joint non-rigid image registration and active contour segmentation using a shared transformation  $\mathcal{T}$ . We minimize the following objective function

$$E(\mathcal{T}) = E_{sim}(\mathcal{T}) + \lambda E_{shape}(\mathcal{T}) \quad (5.33)$$

where  $E_{sim}$  is the non-rigid image registration similarity measure that utilizes speckle information, and  $E_{shape}$  is an active contour energy function that accounts for shape structures. Parameter  $\lambda$  represents a trade-off between the intensity (speckle) and structural information influences. Figure 5.3 illustrates the schema of the method.

To initialize the contour  $C_t$  in the first volume we use the registration-based image segmentation. We have selected a 3D volume with a clear LV view, and manually segmented its endo- and epi-cardium. We have used 2D cross-section image planes to allocate the boundary contours and then form a 3D finite element model. We use this 3D volume together with the boundary countour as a template image and register it to the first volume of a tested 3D echo sequence to allocate the initial contour  $C_t$ .

## 5.6 Transformation Model

We use the Free Form Deformation (FFD) model [152, 146] to parameterize the transformation of the imaged area  $\mathcal{T}$  between two consecutive volumes. Such transformation represents a continuous mapping from every 3D location of one volume to the corresponding 3D location of the next volume.

FFD is defined a tensor product of cubic B-splines basis functions. The basic idea of FFD is to deform a 3D object by manipulating a mesh of B-spline control points. FFD has been successfully used for non-rigid medical image registration [146], for tracking of

cardiac MRI [8] and in 3D echocardiography [115]. The main advantage of FFD is that the complex non-rigid transformation is defined by a small number of parameters (control points positions). Since FFD is volumetric transformation, we can transform several objects (e.g., 3D image, endocardium and epicardium model contours) simultaneously with a single FFD. More details on the FFD transformation model can be found in Section 2.3.1.

A level of smoothness provided by parametric form of FFD model is not always sufficient, because locally large deformations or unnatural foldings of the B-spline control points can occur. We use an addition regularization term for FFD; we penalize the Euclidean distance between all the neighboring displacements of B-spline control points to prevent unnatural wraps. Such regularization penalizes first order derivative over the displacements of control points, which allows fast computations.

## 5.7 Optimization

For a given pair of consecutive 3D image volumes, we minimize the following objective function

$$E(\mathbf{u}) = E_{sim}(\mathbf{u}) + \lambda E_{shape}(\mathbf{u}) + \frac{\alpha}{2} \|\nabla \mathbf{u}\|^2 \quad (5.34)$$

where  $E_{sim}(\mathbf{u})$  is the  $MS_2$  similarity measure (Eq. 5.31),  $E_{shape}(\mathbf{u})$  is the contour energy function (Eq. 5.32),  $\mathbf{u}$  is the displacement vector of all B-splines control points, which uniquely specify the transformation,  $\nabla$  is the gradient operator and  $\|\nabla \mathbf{u}\|^2$  is the regularization term. We do not use a multiresolution registration approach, because consecutive images are already reasonably close (in terms of the transformation), which alleviates the need for multiresolution. We equate the gradient of the objective function with respect to the transformation parameters (displacements of B-spline control points) to zero

$$\nabla E_{sim}(\mathbf{u}) + \lambda \nabla E_{shape}(\mathbf{u}) + \alpha \Delta \mathbf{u} = 0 \quad (5.35)$$

where  $\Delta$  is the Laplacian operator. We use the implicit Euler time marching method to minimize the objective function [23]:

$$\mathbf{u} = (\mathbf{Id} + \gamma \alpha \Delta)^{-1} (\mathbf{u} - \gamma (\nabla E_{sim}(\mathbf{u}) + \lambda \nabla E_{shape}(\mathbf{u}))) \quad (5.36)$$

where  $\gamma$  is a time step parameter, which is similar to the gradient descent step size. The stopping condition was either when the current change in the objective function drops below the threshold of  $10^{-8}$  or the number of iterations reaches the maximum of 100.

## 5.8 Sequential Registration

After aligning two consecutive 3D volumes, we repeat the process in the similar fashion for the next pair of images, until reaching the last image pair. Thus we track the mesh of B-spline control points throughout the 3D sequence, which gives us a continuous transformation field as a function of time (volume number).

Sequential image registration, in contrast to the registration to a single template, benefits in terms of computational effort, allows to incorporate the dynamic model and supports our assumption of speckle correlation between the consecutive volumes, which we use to derive the  $MS_2$  similarity measure. However, we recognize that such sequential approach could potentially lead to a slow error accumulation from volume to volume.

## 5.9 Dynamic Constrains

Speckle motion is also subject to temporal constraint, characterized by the dynamics of the heart motion. Here, we use a simple dynamic system, and update it with a discrete Kalman filter.

We define the FFD control points locations as the underlying state. We assume a simple process model given by

$$\mathbf{s}_t = \mathbf{F}\mathbf{s}_{t-1} + n_{t-1}. \quad (5.37)$$

The state at time  $t - 1$  is defined as the FFD B-spline control points positions and their velocities for image  $I$ :  $\mathbf{s}_{t-1} = [\mathbf{X}; \mathbf{V}]$ . Matrix  $\mathbf{F}$  is defined as

$$\mathbf{F}_{6M \times 6M} = \begin{pmatrix} \mathbf{I}_{3M \times 3M} & \mathbf{I}_{3M \times 3M} \\ \mathbf{0}_{3M \times 3M} & \mathbf{I}_{3M \times 3M} \end{pmatrix}$$

where  $M$  is total number of control points. Process noise  $n_{t-1}$  is assumed to have Gaussian

distribution  $p(n) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ . We define the measurement model by

$$\mathbf{z}_t = \mathbf{H}\mathbf{s}_t + \mu_t \quad (5.38)$$

Here the measurement (observation) is defined as locations of control points obtained at the optimum of joint registration and segmentation objective function. Matrix  $\mathbf{H}_{3M \times 6M}$  is given by  $[\mathbf{I}_{3M \times 3M}; \mathbf{0}_{3M \times 3M}]$ . The measurement noise  $\mu_t$  is assumed to have Gaussian distribution with  $p(\mu) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ . The discrete Kalman filter can be used to update the above system. We empirically determined the process and the measurement noise covariances to be  $\mathbf{Q} = 0.2\mathbf{I}$ ,  $\mathbf{R} = 0.01\mathbf{I}$ . The positions of B-spline control points after the joint registration and segmentation are used as a noisy measurements at time  $t$  to update the predicted state and achieve the final transformation parameters.

## 5.10 Spherical Coordinates

One factor to consider for quantitative analysis of 3D echo images is the format of the ultrasound signal. A typical digital ultrasound processing chain consists of the following four steps: 1) Multiple analog-to-digital converters convert the analog radio frequency (RF) signal from the ultrasound transducer to a digital RF signal; 2) Beamforming is applied to individual RF signals by applying time delays and summations as a function of time and transformed into a single RF signal, which is often called “beamformed RF” data; 3) The RF signal is filtered, is run through an envelope detector and is log compressed into a gray scale format. At this stage the data is represented as a sequence of 3D arrays, each of which is a 3D image in the raw spherical coordinate system; 4) Finally, the envelope-detected signal is scan converted to Cartesian coordinate system and interpolated at the regular grid to form the digital 3D image (also called B-mode data). The data corresponding to the third and fourth steps (spherical vs. Cartesian coordinates of post-converted image) is shown in Figure 5.4.

We propose to use envelope-detected 3D echo image in the raw spherical coordinates (step 3), before the final conversion to the Cartesian coordinates. In other words, we can take one step back in 3D echo image formation. We note that such approach differs

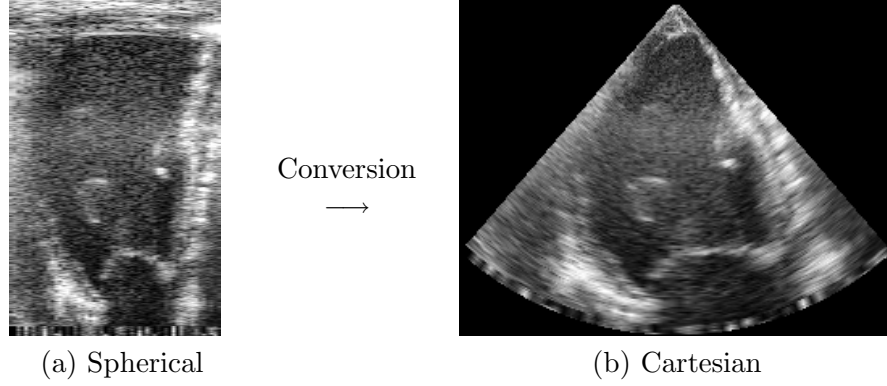


Figure 5.4: (a) An example of a 2D slice out of 3D spherical data displayed as a 2D array with  $416 \times 62 \times 56$  resolution. (b) The spherical data is converted to the Cartesian data with  $208 \times 297 \times 269$  resolution.

significantly from the registration methods [25] that build a spherical or cylindrical transformation model along the heart orientation axis, using Cartesian post-converted images (step 4). We use envelope-detected images in raw spherical coordinates (step 3), which may or may not be in alignment with heart orientation. To implement the non-rigid image registration in spherical coordinates we define the FFD B-spline control points in spherical system with uniform spacing. Image interpolation required during the registration iterations is done also from the image voxels in spherical system. At the end of registration the transformation field can be easily converted to Cartesian system and vice versa.

The main benefits of using envelope-detected image in spherical coordinates, in contrast to the final Cartesian coordinates, is the preservation of speckle statistics and avoidance of information loss associated with scan conversion and interpolation. The envelope-detected data has higher resolution in vicinity of transducer and lower resolution away from the transducer. Also, the axial resolution is much higher than the lateral. After conversion to the Cartesian system some information is lost, including high axial resolution and high resolution closer to the transducer, and additional errors are introduced, for instance, the new voxels that appear after regular grid interpolation in the areas far from the transducer. Even though the envelope-detected data in spherical coordinates still contains less information than RF data, it represents a reasonable compromise between

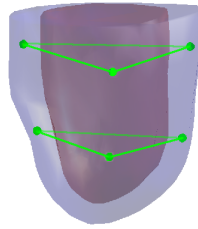


Figure 5.5: Placement of sono crystals on the LV.

level of the signal details and computational complexity.

### 5.11 Validation by Sonomicrometry

To validate our approach, we perform the ground truth strain and torsion estimation with implanted sonomicrometers (sono crystals) on open chest piglets. An array of six sonomicrometers (sono crystals) was implanted as two sets: three sonomicrometers at apex level and three sonomicrometers closer to the base level in the LV myocardium (see Figure 5.5). Using Sonometrics ®software (ON, Canada), we computed off-line the pairwise crystal distances. From these distances, we computed pairwise strains and between-plane torsion, which were used as a ground truth for validation. The strain between a pair of sonomicrometers was computed as the relative distance change with respect to the initial distance, so that the positive values correspond to the dilation. The torsion was computed as the angle of rotation between the two planes defined by the two sets of sono crystals, which are approximately parallel. We use a geometrical center of crystal triplets in LV as a center of rotation, which is approximately true due to the intentional crystal placement around the myocardium. The actual sonomicrometer motion also involves off-plane complex motion; we use projections of sonomicrometer positions onto the initial plane to measure the rotation.

To compare strain and torsion derived from the registration algorithm to those from the sonomicrometry, we visually located, based on anatomical markers, the sono crystals positions on the myocardium in the end diastole volume. These locations were propagated across time by the displacement field obtained from the tracking algorithm. From their



moving positions we computed strain and torsion in the same fashion as those for sono crystals. All algorithm-derived and sonomicrometry-derived strains and torsions were computed relative to the end diastole volume.

We note, that the algorithm and sonomicrometry measurements have different temporal resolutions. We upsample the algorithm derived measurement to the higher sonomicrometer resolution using cubic interpolation.

## 5.12 Experimental Results

We have implemented the algorithm in Matlab, and tested it on a AMD Opteron CPU 2GHz Linux machine with 4GB RAM. We acquired the 3D echo sequences using Philips iE33 X7-2 probe with ECG gating from 7 open-chest pigs. We used maximal spatial resolution option. The data were combined from 7 cycles (7 stitches). The spatial resolution was  $208 \times 297 \times 269$  voxels, with 24 volumes per sequence on average. We use the Cartesian coordinate system. We put the FFD control points uniformly with 20 voxels spacing. We set  $D = 1$ ,  $\rho = 0.8$  and  $m = 0.5$ , which we empirically found to give the best performance. For the boundary shape constraints we use the trade-off parameter  $\lambda = 0.1$ . The weight of the regularization term was set to  $\alpha = 0.01$ . On average, the tracking algorithm between two volumes requires around 5 minutes (120 minutes for the full sequence).

### 5.12.1 Animal Preparation

We studied 7 pigs of both sexes approximately 12 weeks of age (18-22 kg). The pigs were placed in the supine position during the experiment and anesthesia was induced with 25 mg/kg IV thiopental and maintained with an infusion of 25 mg/mL solution (100 mL/h) of fentanyl citrate. A median sternotomy was performed and the pericardium split from apex to base to expose the heart. The pigs were ventilated with a constant-volume ventilator and body temperature was maintained in the physiological range with a heating pad. LV, RV and aortic pressures were measured with 8F micromanometer-tipped catheters. Catheters were inserted via peripheral vessels. Ventricular and aortic pressures were referenced using their respective fluid-filled lumens and atrial pressures, by

comparison to ventricular values during diastole. Aortic and pulmonary arterial flow was measured using ultrasonic flowmeters.

Additional perturbation was undertaken by subjecting the piglets to a left anterior descending (LAD) coronary artery ligation and the inferior vena cava (IVC) occlusion. Atraumatic vascular occluders were placed around the proximal third of the left anterior descending coronary artery to induce ischemia. The inferior vena cava was isolated by blunt dissection in posterior mediastinum and a pneumatic occluder was placed on the IVC to transiently decrease preload (typically down to an LV end diastolic pressure of 3-4 mmHg). At end-expiration, the respirator was switched off for very short periods to allow for successive beats to be recorded without the effects of breathing.

### 5.12.2 Data Acquisition

Three sono crystals were implanted along the LV circumference close to the apex and three close to the base in the short axis plane. Atraumatic surgical and fine suturing techniques were used to secure crystals subepicardially to minimize myocardial damage. To achieve reproducible and parallel planes, the crystals at each level were placed at fixed distances from LV apex (20% of LV length for apical plane and 80% of LV length for basal plane, as shown in Figure 5.5). Signals from crystals were acquired at a rate between 250-300 Hz after optimizing parameters such as inhibit delay and transmit impulse, and after observing a clear signal.

Full volume 4D image loops were acquired at 24 volumes per heart cycle with placed directly on the cardiac apex separated only by a small piece of fresh liver as standoff. Frequency and focus places were optimized to get the best possible myocardial texture throughout the LV full volume. Baseline images were acquired before and after placing sono crystals. After acquiring baseline (normal steady state) data, preload to heart was altered by a graded occlusion of inferior vena cava (IVC) for a short period of time. Sono data and 4D apical long axis full volume images were acquired during periodic occlusion. After releasing the IVC, another set of baseline data was acquired. The study was repeated with ligation of the mid segment of the left anterior descending (LAD) coronary artery for 5-7 minutes to produce myocardial ischemia.

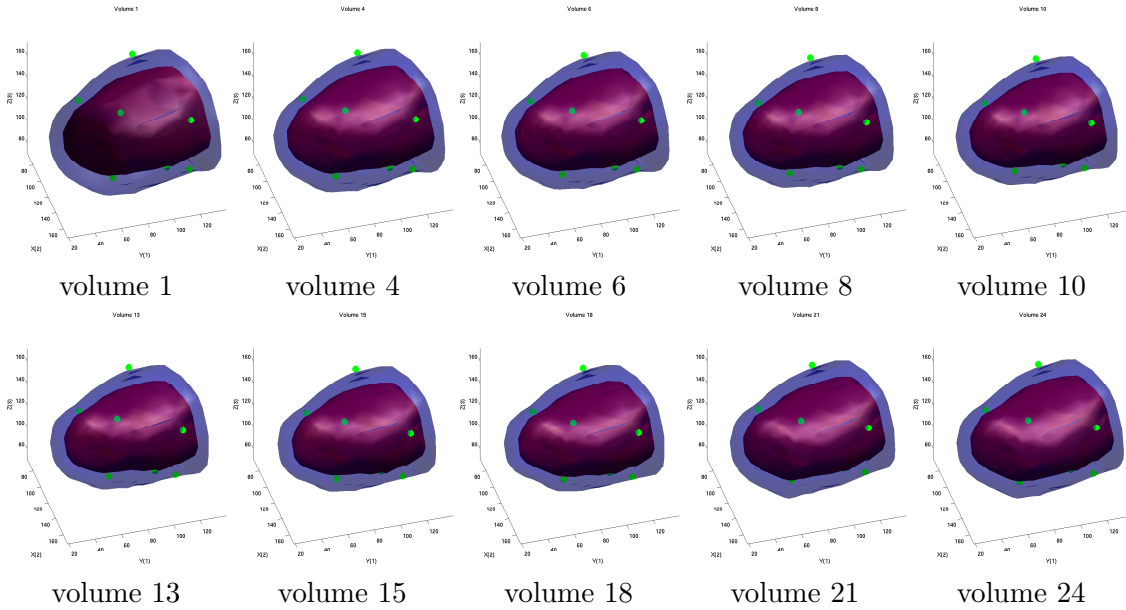


Figure 5.6: The LV contour and sonomicrometer motion found during the cardiac cycle (24 consecutive volumes). LV achieves the maximum contraction at volume 13, then the LV dilates (diastolic phase) up to the volume 22 and starts contracting again (systolic phase).

### 5.12.3 LV Strain Analysis

We show an example of the estimated dense transformation through motion of the LV endo- and epicardium contours with and without original 4D echo overlay in Figure 5.6 and Figure 5.7. The green markers show the estimated motion of sonomicrometers. Figure 5.8 demonstrates the color-coded surface strain on the LV endocardium for an example of baseline scan. Strain between a pair two particular crystals at the apical level is shown in Figure 5.9. The strain values are plotted for the algorithm and the corresponding groundtruth sonomicrometry over time for the baseline, IVC occlusion and LAD ligation experiments. Clearly, the algorithm-derived strains follows the sonomicrometry-derived strains closely both in the amplitude and shape. Both algorithm and sonomicrometry strain measurements were able to differentiate between the ischemic and non-ischemic states caused by LAD ligation. This is also observed in the end systolic (ES) strain plot for LAD ligated and baseline states in Figure 5.10.

Table 5.1 shows the average peak strain (ES strain) values, derived by the algorithm,

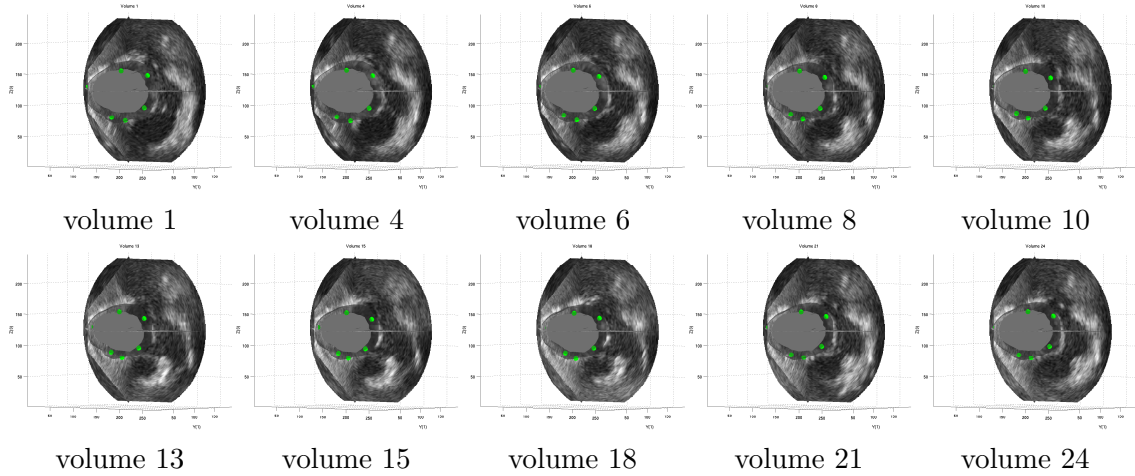


Figure 5.7: Estimated motion of the LV contours and sonomicrometers superimposed on the 3D echo images in Cartesian coordinates.

	Baseline	LAD	IVC
Apical	$-19.14 \pm 3.78$	$-13.57 \pm 1.21$	$-24.3 \pm 1.44$
Basal	$-18.87 \pm 2.35$	$-16.73 \pm 1.16$	$-21.6 \pm 4.37$
Apex-Base	$-18.11 \pm 4.52$	$-17.28 \pm 2.12$	$-18.9 \pm 4.13$

Table 5.1: Algorithmically estimated peak strain (ES strain) for different segments between the sono crystals in the apical, basal and in apical-to-basal levels. The measurements were averaged among the combinations of sono crystal within a particular level and among all scans. The decrease in the peak strain caused by LAD ligation is clear especially in the apical region. Also the reduction of preload due to controlled IVC occlusion led to the slight increase of the overall strain.

between the sono crystals in the apical, basal and in apical-to-basal levels. The apical level strain was derived using the three apical sono crystals that provide three pairwise strain measurements, as shown in Fig. 5.5. Similarly the basal level strain was derived from three intercombinations of basal sono crystals, and the apical-to-basal strains were from the pairs of the corresponding apical and basal sono crystals. The measurements were averaged among the combinations of sono crystal within a particular level and among all scans. The decrease in the peak strain caused by LAD ligation is clear especially in the apical region. Also the reduction of preload due to controlled IVC occlusion led to the slight increase of the overall strain.

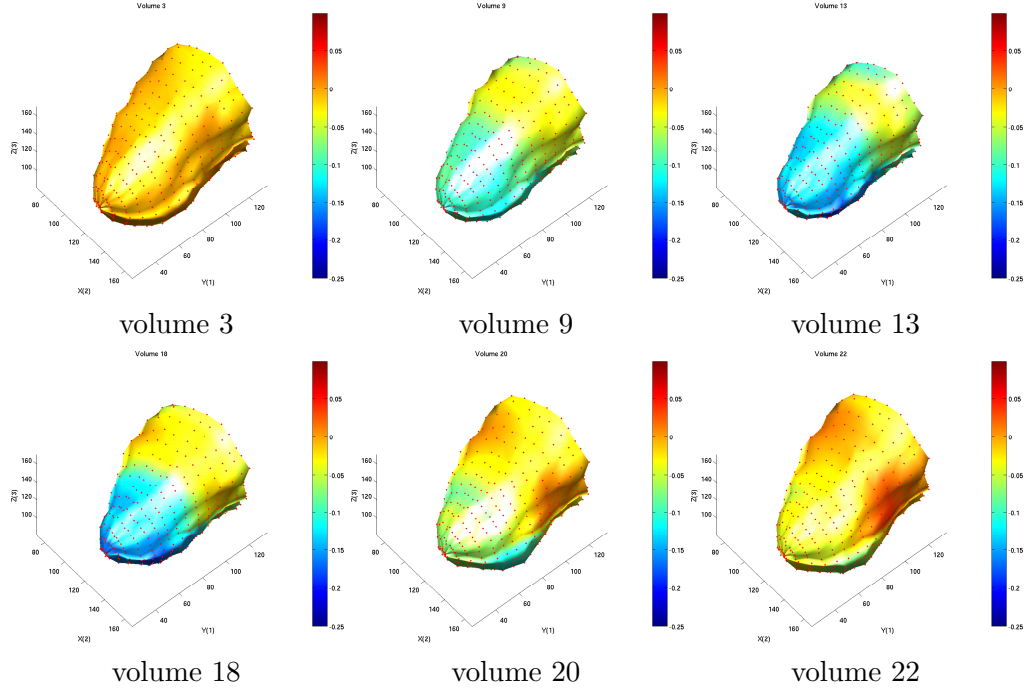


Figure 5.8: Color-coded display of LV endocardium strain is visualized at 6 particular volumes (out of 24 total). LV achieves the maximum contraction at volume 13 (end diastolic phase).

We compared the algorithm strain to the sonomicrometry-derived strain for matching segments. Table 5.2 shows the mean absolute error values between the sonomicrometry and the algorithm strain averaged over time, pairs within the levels (Apical, Basal, Apical to Basal and overall) and among all 4D echo scans. Our method performs well with the average error within a few percent from the groundtruth strain.

	Baseline	LAD	IVC
Apical	$1.83 \pm 0.83$	$1.12 \pm 0.62$	$1.92 \pm 2.03$
Basal	$2.71 \pm 1.39$	$2.47 \pm 1.57$	$3.17 \pm 1.27$
Apical-to-Basal	$1.76 \pm 2.13$	$1.54 \pm 1.83$	$1.63 \pm 1.86$
Overall	$2.22 \pm 1.57$	$1.69 \pm 1.47$	$2.39 \pm 1.78$

Table 5.2: Mean absolute error between algorithm and groundtruth sonomicrometry strain. The error is shown for different levels (apical, basal, apical-to-basal and overall) against different states (baseline, LAD ligation and IVC occlusion).

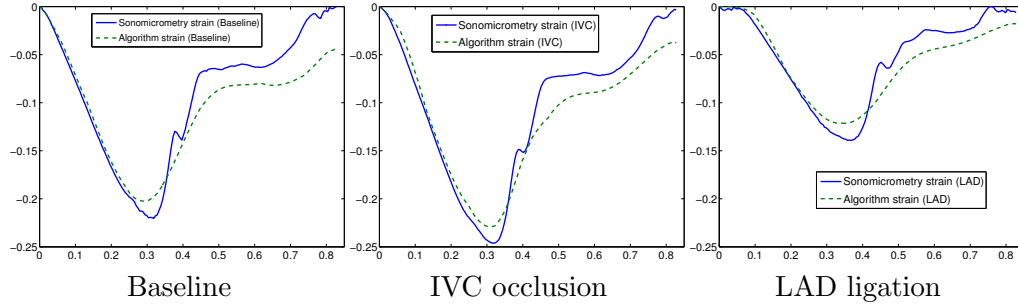


Figure 5.9: Strain values against time (sec) from the apical region (between 2 particular sono crystals), derived from algorithm and sonomicrometry for the Baseline, IVC occlusion and LAD ligation states. The algorithm strain follows the sonomicrometry strains closely both in the amplitude and shape.

	Baseline	LAD	IVC
Algorithm	$3.85^\circ \pm 1.65^\circ$	$5.62^\circ \pm 1.43^\circ$	$2.79^\circ \pm 1.21^\circ$
Sonomicrometry	$5.25^\circ \pm 1.20^\circ$	$7.85^\circ \pm 0.90^\circ$	$3.65^\circ \pm 0.35^\circ$

Table 5.3: The average peak torsion as estimated by the algorithm and by the sonomicrometry.

### 5.12.4 LV Torsion Analysis

We calculated torsion from the instantaneous differences of apical and basal rotations in each volume. We can visually see the rotation patterns from the displacement field between the end diastolic (ED) and end systolic (ES) phases in Figure 5.11.

We illustrate the estimated torsion for different steady states in Figure 5.12. Table 5.3 shows the average peak torsion values derived by the algorithm and by the sonomicrometry. On average, at baseline peak torsion values were  $5.25^\circ \pm 1.20^\circ$  as detected by sonomicrometry and  $3.85^\circ \pm 1.65^\circ$  as detected by the algorithm. With the reduction of preload due to controlled temporary IVC occlusion, peak torsion was increased to  $7.85^\circ \pm 0.90^\circ$  by sonomicrometry and  $5.62^\circ \pm 1.43^\circ$  by the algorithm. With anterior myocardial ischemia produced by atraumatic occlusion of the distal half of the LAD, peak global torsion was reduced to  $3.65^\circ \pm 0.35^\circ$  by sonomicrometry and  $2.79^\circ \pm 1.21^\circ$  by the algorithm. Overall the algorithm consistently underestimates the torsion values. Nonetheless, we achieve a significant correlation,  $r = 0.89$ , between the algorithm and sonomicrometer torsions. We computed the correlation as an average correlation coefficient between the corresponding

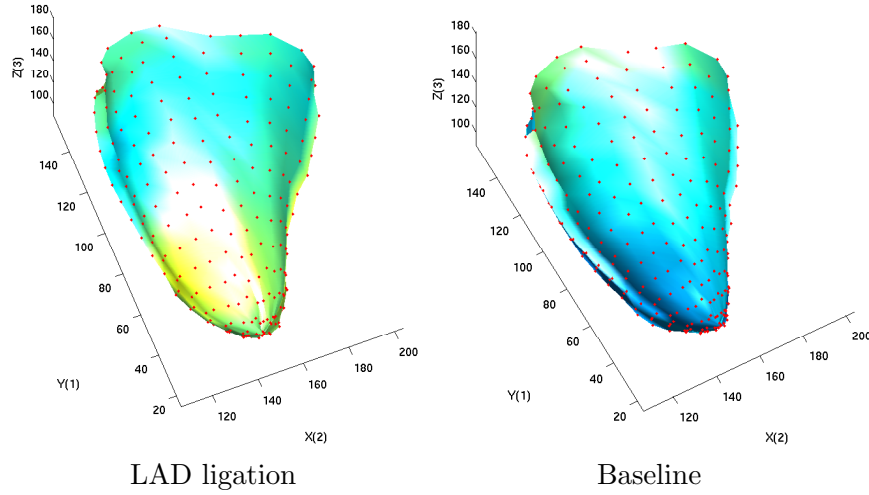


Figure 5.10: Color-coded display of LV endocardium ES-strain for LAD and Baseline states.

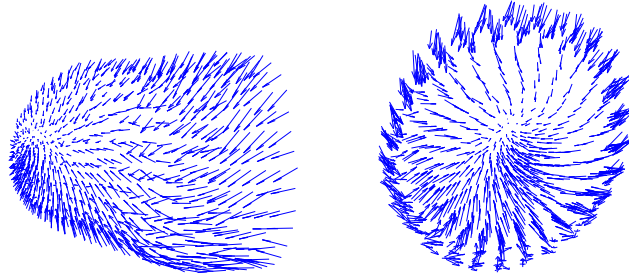


Figure 5.11: The displacement vector fields on LV endocardium between end diastole and end systole volumes at two different viewing angles.

algorithm and sonomicrometer torsion measurements averaged over all scans.

Underestimation of the torsion values is likely caused by a larger speckle decorrelation due to myocardium rotation and by the regularization of the transformation. Nonetheless, a high temporal correlation between the algorithm and sonomicrometry torsions and the ability to differentiate between different steady states can make the algorithm torsion measurements a valuable indicator for the actual torsion assessment.

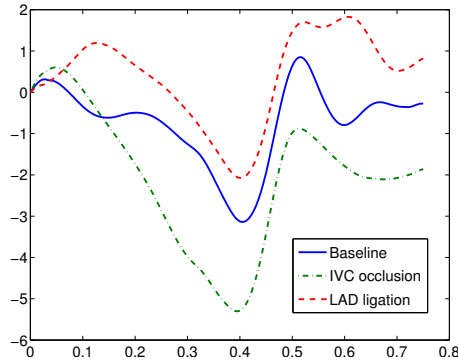


Figure 5.12: The plots of LV torsion (degrees) against time (sec) in a cardiac cycle at different steady states. Torsion was increased with reduction in preload due to controlled IVC occlusion (green) and decreased with myocardial ischemia induced with ligation of LAD (red) as compared to baseline (blue).

### 5.13 Discussion and Conclusion

Accurate dense 3D speckle tracking methods are not well developed yet neither by commercial vendors nor by research community. Algorithms to estimate quantitative characteristics from 4D echo, such as 3D strain and torsion, are currently of great interest. While providing comprehensive 3D information, 3D echo suffers from ultrasound artifacts, low spatial and temporal resolutions. To overcome low image quality all available information must be taken into account, including shape priors, dynamic priors and speckle statistics.

Recently, several researches has proposed methods for motion estimation from 3D echocardiography based on non-rigid image registration [115, 50, 116]. Image registration explicitly constrains smoothness of the transformation by the non-rigid transformation parametrization and takes into account full volume speckle information by minimizing the intensity-based similarity measure.

Here, we proposed a new robust similarity measure ( $MS_2$ ), which takes into account speckle formation, correlation and practical acquisition process. Eventhough,  $MS_2$  accounts for speckle formation specifics, it still assumes the voxel-wise independence. Accounting for spatial dependencies in a similarity measures is an ongoing research in image registration community. Including such information in the ultrasound similarity measure will likely further improve the registration performance.



To assist the non-rigid image registration, our method utilizes shape and dynamic constraints, which add extra information to the registration procedure and improve the overall performance. Our dynamic model simply ensures temporal smoothness, but does not account for pseudo-periodic heart dynamic. Application of more complex dynamic model requires extensive training sets. Such models can benefit the registration accuracy, but also can lead to biased results if an analysed cardiac motion had no analogs in the training set.

We validate our method in vivo on open-chest pigs, which resembles human heart and allows for groundtruth invasive measurements. In contrast to phantom validation [50], animal data provide a realistic evaluation of the method accuracy, without relying on synthetic noise model as in phantom images. We obtained accurate algorithm strain measurements both in amplitude and shape within a few percent from the groundtruth ones. Similar to sonomicrometry, our algorithm is able to differentiate between the baseline, LAD ligation and IVC occlusion states. A decrease in strain was observed with LAD ligation, and a slight increase caused by IVC occlusion as expected.

Our torsion estimation showed a high correlation, but includes a consistent underestimation of torsion values. This errors were most likely caused by higher rotational speckle decorrelation. Nevertheless, our algorithm torsion measurement showed a high correlation with the groundtruth sonomicrometry ones, and were able to differentiate between the simulated traumatic states.

We computed only strain and torsion, however our methods provides a dense displacement field throughout the cardiac cycle, which can be used to derive any other regional or global quantitative characteristics. On average, the tracking algorithm between two volumes requires less than 10 minutes. The computational time of our method can be greatly improved by implementing the code in C++, which will reduce the processing time to several minutes for the full scan. Furthermore, parallel implementation and/or implementation of GPU can make the processing nearly realtime. We discuss the future directions in motion estimation from 3D+T echocardiography in Section 6.2.

## Chapter 6

# Conclusions and Future Directions

### 6.1 Summary

In this thesis we have contributed several new methods to the field of non-rigid image registration. We have developed a novel similarity measure, called Residual Complexity (RC), that is robust to spatially-varying intensity distortions. RC is minimized when the residual image can be sparsely coded using only a few known basis functions. We have applied RC for non-rigid motion stabilization of time lapse microscopic videos. We have developed a novel probabilistic method for rigid and non-rigid point set registration. The algorithm is called Coherent Point Drift, and is currently the state-of-the-art in point set registration. We have develop an automated system for motion estimation from 3D+T echocardiography, based on sequential non-rigid image registration.

We have released two Matlab toolboxes: a) We have developed Medical Image Registration Toolbox (MIRT). The toolbox allows 2D and 3D non-rigid image registration. The toolbox implements many standard and state-of-the-art non-rigid image registration methods. MIRT also provides a GUI interface for non-rigid video stabilization. b) We have developed and implemented Point Set Registration Toolbox for MATLAB, based on Coherent Point Drift (CPD) algorithm. The toolbox finds the correspondences between two multidimensional point sets, as well as rigid or non-rigid transformation to align the point sets. The toolbox has been downloaded thousands of times since 2007.

## 6.2 Possible Extensions & Future Work

In this section, we describe some possible extensions of our work. These are mostly speculations, open questions and promising directions we can take to improve the main components of non-rigid image registration that are presented in this thesis.

**Similarity Measure:** Similarity measure is one of the key components in image registration. Its accurate definition is still problematic. How can we measure the similarity between two images mathematically? Is there a universal image similarity measure exist that is optimized at the correct image alignment? Despite many research works, these questions are not answered.

We have proposed the Residual Complexity similarity measure. It is optimized when the residual image can be sparsely coded, that is the difference image at the correct alignment must have a low complexity measured in terms of the basis functions. Abstractly speaking, at the correct alignment images share some information, and the remaining information (residual) should be minimal. From this perspective, we can improve the similarity measure in the following ways:

- Is there a better way to define residual image than a simple difference image? If one can generalize the definition of the residual image so that it represents the remaining information, then this new residual can be used instead.
- Can we generalize the measure of complexity of the residual image? In RC, we measure the sparseness of the residual image in terms of global pre-defined basis functions. One can make such basis functions image dependent or adaptive. Alternatively one can try to use measures of local sparseness in terms of local basis functions.

Another generalization of RC is possible by analyzing the roots of Mutual Information (MI). MI is maximized when the joint entropy is low, whereas individual entropies are high.

$$MI = \mathcal{H}(I) + \mathcal{H}(J) - \mathcal{H}(I, J) \quad (6.1)$$

Entropy is often considered to be a measure of information content. Another way to measure the information content would be the coding complexity, or sparseness in terms of some basis functions. From this perspective, we can define a similarity measure that is optimized when two images have low complexity jointly, whereas each of the image complexities is high. For instance, the coding complexity of two  $2D$  images assembled as a  $3D$  image should be small in terms of some  $3D$  basis functions, whereas the coding complexity of the individual images remains high.

**Point Set Registration:** Coherent Point Drift algorithm have a relatively high computational complexity. We have proposed several ways to improve the computational speed including FGT and low-rank matrix approximations. Another way to speed up the performance is to directly use parametric non-rigid transformation, e.g. B-splines over the point set domain. By using locally supported B-splines one can improve the speed and also allow local non-rigid deformations that do not influence the rest of the points, which can be advantageous in some applications.

Another thing to consider is whether only the point set locations should be used. If point sets prepresent a shape, then other features, such as curvature or normal vector, can be also used to improve the performance.

**Motion Estimation from 3D+T echocardiography:** Regional deformation analysis of myocardium requires automated system for motion estimation from 3D+T echocardiography. Ideally such system should be accurate and nearly real-time. Here we discuss directions of future research to create such a state-of-the-art automated system. We mostly concentrate on the accuracy of the system, which is crucial in clinical settings. As the technology rapidly develops, real-time implementations of such system can be possible in the nearest future. An accurate myocardium tracking system is not a single image processing method, but rather a complex combination of multiple research fields including ultrasound segmentation, speckle and boundary tracking, state estimation and atlas construction.

The key component of such tracking system is the myocardial motion estimation between two consecutive 3D volumes. We believe that the dense speckle tracking, in particular non-rigid image registration, is the most promising approach to estimate such motion. This way we track speckle, which follows the myocardium. In contrast, the boundary based tracking methods provide only sparse motion estimates along the boundaries, which are prone to errors due to the aperture problem. To reliably track speckles one needs to develop a better similarity measure that is robust to speckle decorrelations. Here, we have proposed the  $MS_2$  similarity measure, which takes into account speckle acquisition process. Nevertheless, we have made several simplifying assumptions, such as strictly multiplicative noise and voxel-wise stationarity of conditional probabilities. A better similarity measure should not assume voxel-wise independence and stationarity. On the other hand, a better similarity measure should still have a reasonably low computational complexity.

Traditionally, image registration operates over the whole image domain. To reduce the computation load, one can define the registration only over the region of interest - the myocardium. Thus, one has to segment the myocardium, e.g. endo- and epi-cardium boundaries, in order to defined the myocardial region, which will be used during the registration. Such segmentation should be done only for a single, e.g. the first, image volume. Once the myocardial region is defined one can evaluate the similarity measure only within such region. Another big advantage of segmenting the myocardial boundaries is the ability to apply a heart-specific dynamic model.

Heart has distinct shape and motion cycles. In this thesis we have used a random walk dynamic model on the B-spline control points to simply smooth the deformation through time. Such model does not require any training sets and is fast to implement. As the B-splines control points are on a regular grid in spatial domain without any knowledge of the heart location, we cannot apply the heart-specific dynamic model to them. We can apply the heart-specific dynamic model only to the heart position, which is another reason to pre-segment the heart boundaries.

There are several approaches to build the heart dynamic models [24, 182, 194]. First, one has to decide on what substitutes the so-called state of the system, which may include

heart boundary positions, positions of all myocardial points and/or its intensity levels. With the current computational capabilities it is feasible to use only the boundaries, e.g. endo- and epi-cardium. Such boundaries will also define the region of the myocardium. Second, one has to obtain a training set of the heart boundary positions over time from the population. Such training set can be obtained from the MRIs of the heart, which itself requires manual or automated algorithms for boundary tracking. As far as MRIs have higher spatial resolution and less artifacts, the segmented boundaries will have a higher accuracy compared to the ones obtained from echocardiography.

The training set should be large enough to explain all the possible heart shape variations. The dynamic model learned from an insufficient training set can bias the result, e.g. in some abnormal heart cases. Finally, once the training set is available, one can build the probability distribution of the boundary position during the cardiac cycle [194], or build a transition probability that provides the next time step prediction of the boundaries given one or several previous positions [24, 182]. It is not only the dynamic, but also the shape model that is learned from the training set. We can integrate the learned dynamic and shape model within the recursive Bayesian estimation to predict the boundary position in the next 3D volume. The predicted boundary position can be used to initialize the dense speckle tracking in the next volume and/or to constrain the dense speckle tracking from deviating from the predicted position.

Finally, to reduce the computational complexity one should consider fast GPU and parallel implementations. For instance, motion estimation between each pair of 3D volumes can be calculated on a separate CPU. The computational time will be reduced proportional to the number of CPUs used. Such approach, however, will not allow implementation of dynamic constraints, because all 3D volumes are processed in parallel. Alternatively and can parallelize the motion estimation between a pair of 3D volumes. Specifically, GPU implementation provides highly parallel evaluation of simple and some complex image processing methods, including image interpolation and linear transforms, which will significantly improve the processing speed.

### 6.3 Conclusion

Non-rigid image registration is the key part of medical image analysis and computer vision. As the technology progressively improves, so does the field of image registration. Many successful new algorithms have been proposed recently, but many challenges still remain. This thesis overviews common strategies in non-rigid image registration as well as contributes several new methods and algorithms to the field of non-rigid image and point set registration. The algorithms described in this thesis are of general interest and can be applied to many computer vision problems. This thesis aims to contribute its part to the growing body of knowledge in non-rigid image registration. Some of the methods described in this work, including CPD and RC similarity measure, have already become popular among many researchers in image analysis.

# Bibliography

- [1] AHA. Heart disease and stroke statistics. American Heart Association, 2008.
- [2] ARDEKANI, B. A., GUCKEMUS, S., BACHMAN, A., HOPTMAN, M. J., WOJTASZEK, M., AND NIERENBERG, J. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain mri scans. *Journal of Neuroscience Methods* 142, 1 (2005), 67 – 76.
- [3] ARONSAJN, N. Theory of reproducing kernels. *Transactions of the American Mathematical Society* 68, 3 (1950), 337–404.
- [4] ARSIGNY, V., COMMOWICK, O., PENNEC, X., AND AYACHE, N. A log-euclidean framework for statistics on diffeomorphisms. In *Proc. of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06)* (2006), no. 4190 in LNCS, pp. 924–931.
- [5] ARUN, K., HUANG, T. S., AND BLOSTEIN, S. D. Least-squares fitting of two 3-D point sets. *IEEE PAMI* 9, 5 (1987), 698–700.
- [6] ASHBURNER, J. A fast diffeomorphic image registration algorithm. *NeuroImage* 38, 1 (2007), 95–113.
- [7] ASHBURNER, J., AND FRISTON, K. Unified segmentation. *NeuroImage* 26 (2005), 839–851.
- [8] BARDINET, E., COHEN, L., AND AYACHE, N. Tracking and motion analysis of the left ventricle with deformable superquadrics. *Medical Image Analysis* 1, 2 (1996).
- [9] BECKER, M., ADAMUS, G., MARTIN, T., CRESPO, S., PLANCK, S., OFFNER, H., AND ROSENBAUM, J. Serial imaging of the immune response in vivo in a t cell mediated autoimmune disease model. *The FASEB Journal* 14 (2000), A1118.
- [10] BECKER, M., NOBILING, R., PLANCK, S., AND ROSENBAUM, J. Digital video-imaging of leukocyte migration in the iris: intravital microscopy in a physiological model during the onset of endotoxin-induced uveitis. *J Immunol Meth* 240 (2000), 23–27.



- [11] BEG, F., MILLER, M., TROUV'E, A., AND YOUNES, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61 (2005), 139–157.
- [12] BELONGIE, S., MALIK, J., AND PUZICHA, J. Shape matching and object recognition using shape contexts. *IEEE Transaction on Pattern Analysis Machine Intelligence* 24, 4 (2002), 509–522.
- [13] BESL, P. J., AND MCKAY, N. D. A method for registration of 3-D shapes. *IEEE PAMI* 14, 2 (Feb. 1992), 239–256.
- [14] BISHOP, C. M. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., 1995.
- [15] BISTOQUET, A., OSHINSKI, J., AND SKRINJAR, O. Myocardial deformation recovery from cine mri using a nearly incompressible biventricular model. *Medical Image Analysis* 12, 1 (2008), 69 – 85.
- [16] BLAKE, A., AND ISARD, M. *Active Contours*. Springer, 1998.
- [17] BOOKSTEIN, F. L. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 11, 6 (June 1989), 567–585.
- [18] BOSCH, J., MITCHELL, S., LILIEVELDT, B. P., NIJLAND, F., KAMP, O., SONKA, M., AND REIBER, J. H. Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Transactions on Medical Imaging* 21, 11 (2002), 1374–1383.
- [19] BOUKERROUI, D., BASKURT, A., NOBLE, J. A., AND BASSET, O. Segmentation of ultrasound images: multiresolution 2D and 3D algorithm based on global and local statistics. *Pattern Recogn. Lett.* 24, 4-5 (2003), 779–790.
- [20] BOUKERROUI, D., NOBLE, J. A., AND BRADY, M. Velocity estimation in ultrasound images: A block matching approach. *Lecture Notes in Computer Science* 2732 (Aug. 2003), 586–598.
- [21] BROWN, L. G. A survey of image registration techniques. *ACM Computing Surveys* 24 (1992), 325–376.
- [22] BURCKHARDT, C. B. Speckle in ultrasound b-mode scans. *IEEE Trans. on Sonics and Ultrasonics* 25, 1 (Jan. 1978).

- [23] BUTCHER, J. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2003.
- [24] CARNEIRO, G., AND NASCIMENTO, J. C. Multiple dynamic models for tracking the left ventricle of the heart from ultrasound data using particle filters and deep learning architectures. In *CVPR* (2010).
- [25] CHANDRASHEKARA, R., MOHIADDIN, R. H., AND RUECKERT, D. Daniel rueckert. In *Surgery Simulation and Soft Tissue Modeling* (2003), vol. 2673 of *LNCS*, pp. 88–99.
- [26] CHEN, X., XIE, H., ERKAMP, R., KIM, K., JIA, C., RUBIN, J., AND O'DONNELL, M. 3-D correlation-based speckle tracking. *Ultrasonic Imaging* 27 (2005), 21–36.
- [27] CHEN, Z., AND HAYKIN, S. On different facets of regularization theory. *Neural Computation* 14, 12 (2002), 2791–2846.
- [28] CHRISTENSEN, G. Consistent linear-elastic transformations for image matching. In *Information Processing in Medical Imaging* (1999), vol. 1613, Springer, pp. 224–237.
- [29] CHRISTENSEN, G., RABBITT, R. D., AND MILLER, M. I. Deformable templates using large deformation kinematics. *IEEE Trans. Image Processing* 5, 10 (1996).
- [30] CHUANG, B., MYRONENKO, A., ENGLISH, R., AND NOBLE, A. Interpreting ultrasound elastography: Image registration of breast cancer ultrasound elastography to histopathology images. In *IEEE International Symposium on Biomedical Imaging (ISBI)* (Apr. 2010).
- [31] CHUI, H., AND RANGARAJAN, A. A feature registration framework using mixture models. In *IEEE Workshop on MMBIA* (June 2000), pp. 190–197.
- [32] CHUI, H., AND RANGARAJAN, A. A new point matching algorithm for non-rigid registration. *CVIU* 89, 2-3 (Feb. 2003), 114–141.
- [33] CHUI, H., WIN, L., SCHULTZ, R., DUNCAN, J., AND RANGARAJAN, A. A unified non-rigid feature registration method for brain mapping. *Medical Image Analysis* 7 (2003), 113–130.
- [34] COBBOLD, R. S. C. *Foundations of Biomedical Ultrasound*. Biomedical Engineering Series. Oxford University Press, 2006.
- [35] COCOSCO, C., V.KOLLOKIAN, KWAN, R.-S., AND EVANS, A. Brainweb: Online interface to a 3D mri simulated brain database. In *ICFMHB* (1997), p. S425.

- [36] COHEN, B., AND DINSTEIN, I. Motion estimation in noisy ultrasound images by maximum likelihood. *ICPR 03* (2000), 182–185.
- [37] COHEN, B., AND DINSTEIN, I. New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition* 35, 2 (feb 2002), 455–463.
- [38] COPPINI, G., POLI, R., AND VALLI, G. Recovery of the 3-d shape of the left ventricle from echocardiographic images. *IEEE Transactions on Medical Imaging* 14, 2 (1995).
- [39] CRAENE, M. D., CAMARA, O., BIJNENS, B. H., AND FRANGI, A. F. Large diffeomorphic FFD registration for motion and strain quantification from 3D-US sequences. In *Functional Imaging and Modeling of the Heart*, vol. 5528, pp. 437–446.
- [40] CROSBY, J., AMUNDSEN, B. H., HERGUM, T., REMME, E. W., LANGELAND, S., AND TORP, H. 3-D speckle tracking for assessment of regional left ventricular function. *Ultrasound in Med. & Biol.* 35, 3 (2008), 458–471.
- [41] CROSS, A. D., AND HANCOCK, E. R. Graph matching with dual step em algorithm. *IEEE PAMI* 20, 1 (1998), 1236–1253.
- [42] CRUM, W. R., HARTKENS, T., AND HILL, D. L. G. Non-rigid image registration: Theory and practice. *The British Journal of Radiology* 77, special issue (Dec. 2004), S140–S153.
- [43] CULLUM, J. K., AND WILLOUGHBY, R. A. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*, vol. 1. Cambridge, 2002.
- [44] DEMPSTER, A., LAIRD, N., AND RUBIN, D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society. Series B (Methodological)* 39, 1 (1977), 1–38.
- [45] DUAN, Q., ANGELINI, E. D., HERZ, S. L., INGRASSIA, C. M., COSTA, K. D., HOLMES, J. W., HOMMA, S., AND LAINE, A. F. Region-based endocardium tracking on real-time three-dimensional ultrasound. *Ultrasound in Med. & Biol.* 35, 2 (2009).
- [46] DUAN, Q., PARKER, K. M., LORSAKUL, A., ANGELINI, E. D., HYODO, E., HOMMA, S., HOLMES, J. W., AND LAINE, A. F. Quantitative validation of optical flow based myocardial strain measures using sonomicrometry. In *ISBI* (2009), no. 454-457.

- [47] DUCHON, J. Spline minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive Theory of Functions of Several Variables* (1977), vol. 571, pp. 85–100.
- [48] EASTMAN, R., LE MOIGNE, J., AND NETANYAHU, N. Research issues in image registration for remote sensing. In *Fusion* (2007), pp. 1–8.
- [49] EL-BAZ, A., FARAG, A., GIMEL’FARB, G., AND ABDEL-HAKIM, A. Image alignment using learning prior appearance model. In *ICIP* (Oct 2006), pp. 341–344.
- [50] ELEN, A., CHOI, H. F., LOECKX, D., GAO, H., CLAUS, P., SUETENS, P., MAES, F., AND D’HOOGHE, J. Three-dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: A feasibility study. *TMI* 27, 11 (Nov. 2008), 1580–1591.
- [51] ELTOFT, T. Modeling the amplitude statistics of ultrasonic images. *IEEE Transactions on Medical Imaging* 25, 2 (Feb. 2006), 229–240.
- [52] FISCHER, B., AND MODERSITZKI, J. Curvature based image registration. *Journal of Mathematical Imaging and Vision* 18, 1 (2003), 81–85.
- [53] FITZGIBBON, A. W. Robust registration of 2D and 3D point sets. *Image and Vision Computing* 21 (2003), 1145–1153.
- [54] FRISTON, K., ASHBURNER, J., FRITH, C., POLINE, J., HEATHER, J. D., AND FRACKOWIAK, R. Spatial registration and normalization of images. *Human Brain Mapping* 2 (1995), 165–189.
- [55] GIROSI, F., JONES, M., AND POGGIO, T. Regularization theory and neural networks architectures. *Neural Computation* 7, 2 (1995), 219–269.
- [56] GOLD, S., LU, C. P., RANGARAJAN, A., PAPPU, S., AND MJOLSNESS, E. New algorithms for 2D and 3D point matching: Pose estimation and corresp. In *NIPS* (1995), vol. 7, The MIT Press, pp. 957–964.
- [57] GOLUB, G. H., AND LOAN, C. F. V. *Matrix Computations*, second ed. Johns Hopkins University Press, Baltimore, MD, 1989.
- [58] GOODMAN, J. W. Some effects of target-induced scintillation on optical radar performance. *Proceedings of the IEEE* 53, 11 (Nov. 1965), 1688–1700.
- [59] GOODMAN, J. W. *Speckle Phenomena in Optics. Theory and Applications*. 2006.
- [60] GOSHTASBY, A. *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Wiley, 2005.

- [61] GOTTE, M., GERMANS, T., RUSSEL, I., ZWANENBURG, J., MARCUS, J., VAN ROSSUM, A., AND VAN VELDHUISEN, D. Myocardial strain and torsion quantified by cardiovascular magnetic resonance tissue tagging. *J Am Coll Cardiol* 48 (2006), 2002–2011.
- [62] GREENGARD, L., AND STRAIN, J. The fast gauss transform. *SIAM JSSC* 12, 1 (1991), 79–94.
- [63] GUCCIONE, J. M., AND MCCULLOCH, A. D. Finite element modeling of ventricular mechanics. *Theory of Heart* (1991), 122–144.
- [64] HABER, E., AND MODERSITZKI, J. Intensity gradient based registration and fusion of multi-modal images. In *MICCAI* (2006).
- [65] HARRIS, C., AND STEPHENS, M. A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference* (1988), pp. 147–151.
- [66] HELLIER, P., BARILLOT, C., MEMIN, E., AND PEREZ, P. Hierarchical estimation of a dense deformation field for 3-D robust registration. *IEEE Transactions on Medical Imaging* 20, 5 (2001).
- [67] HERMOSILLO, G., AND FAUGERAS, O. Dense image matching with global and local statistical criteria: A variational approach. In *Comp. Vision and Patt. Recognition (CVPR)* (2001), pp. 73–78.
- [68] HILL, D. L. G., BATCHELOR, P. G., HOLDEN, M., AND HAWKES, D. J. Medical image registration. *Physics in medicine and biology* 46, 3 (Mar. 2001), R1–R45.
- [69] HINTON, G. E., WILLIAMS, C. K. I., AND REVOW, M. D. Adaptive elastic models for hand-printed character recognition. In *NIPS* (1992), vol. 4, pp. 512–519.
- [70] HO, J., YANG, M.-H., RANGARAJAN, A., AND VEMURI, B. A new affine registration algorithm for matching 2D point sets. In *IEEE Workshop on ACV* (2007), p. 25.
- [71] HOLDEN, M. A review of geometric transformations for nonrigid body registration. *IEEE Transactions on medical imaging* 27, 1 (2008).
- [72] HORN, B., AND SCHUNK, B. Determining optical flow. *Artif. Intell.* 17 (1981), 185–203.
- [73] HUANG, X., PARAGIOS, N., AND METAXAS, D. N. Shape registration in implicit spaces using information theory and free form deformations. *IEEE PAMI* 28, 8 (2006), 1303–1318.

- [74] HUNG, J., LAND, R., FLACHSKAMPF, F., SHERNAN, S. K., MCCULLOCH, M. L., ADAMS, D. B., THOMAS, J., VANNAN, M., AND RYAN, T. 3D echocardiography: A review of the current status and future directions. *Jour. of the Am. Society of Echocardiography* 20 (2007), 213–233.
- [75] JACOB, G., NOBLE, A., BEHRENBRUCH, C., KELION, A. D., AND BANNING, A. P. A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. *IEEE Transactions on Medical Imaging* 21, 3 (2002), 226–238.
- [76] JIAN, B., AND VEMURI, B. C. A robust algorithm for point set registration using mixture of gaussians. In *IEEE ICCV* (Oct. 2005), vol. 2, pp. 1246–1251.
- [77] JOSHI, A., AND LEE, C.-H. On the problem of correspondence in range data and some inelastic uses for elastic nets. *IEEE Trans. on Neural Networks* 6, 3 (May 1995), 716–723.
- [78] JOSHI, S., DAVIS, B., JOMIER, M., AND GERIG, G. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 23 Suppl. 1 (2004), S151–S160.
- [79] KABUS, S., NETSCH, T., FISCHER, B., AND MODERSITZKI, J. B-spline registration of 3D images with levenberg-marquardt optimization. In *Image Processing* (may 2004), J. Fitzpatrick and M. Sonka, Eds., vol. 5370 of *SPIE Medical Imaging*, pp. 304–313.
- [80] KALLEL, F., BERTRAND, M., AND MEUNIER, J. Speckle motion artifact under tissue rotation. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control* 41, 1 (Jan. 1994), 105–122.
- [81] KARVANEN, J., AND CICHOCKI, A. Measuring sparseness of noisy signals. In *4th Int. Symp. on Independent Component Analysis and Blind Signal Separation* (2003), pp. 125–130.
- [82] KASS, M., WITKIN, A., AND TERZOPOULOS, D. Snakes: Active contour models. *Inter. J. Computer Vision* 1, 4 (Jan. 1988), 321–331.
- [83] KIMELDORF, G., AND WAHBA, G. Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications* 33, 1 (1971), 82–95.
- [84] KLEIN, S., STARING, M., MURPHY, K., VIERGEVER, M. A., AND PLUIM, J. elastix: a toolbox for intensity based medical image registration. *IEEE Trans. Med. Imaging* (2009).

- [85] KLEIN, S., STARING, M., AND PLUIM, J. P. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Trans. Image Processing* 16, 12 (Dec. 2007).
- [86] KLEIN, S., VAN DER HEIDE, U. A., LIPS, I. M., VAN VULPEN, M., STARING, M., AND PLUIM, J. P. Automatic segmentation of the prostate in 3D MR image by atlas matching using localized mutual information. *Medical Physics* 35, 4 (2008), 1407 – 1417.
- [87] LANG, M., MOR-AVI, V., SUGENG, L., NIEMAN, P., AND SAHN, D. J. Three-dimensional echocardiography. *Jour. of the Am. Coll. of Cardiology* 48, 10 (July 2006), 2053–2069.
- [88] LEDESMA-CARBAYO, M., KYBIC, J., DESCO, M., SANTOS, A., SÜHLING, M., HUNZIKER, P., AND UNSER, M. Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *IEEE Transactions on Medical Imaging* 24, 9 (Sept. 2005), 1113–1126.
- [89] LI, M., CHEN, X., LI, X., MA, B., AND VITANYI, P. M. B. The similarity metric. *IEEE Transactions on information theory* 50, 12 (Dec. 2004).
- [90] LIN, N., YU, W., AND DUNCAN, J. S. Combinative multi-scale level set framework for echocardiographic image segmentation. *Medical Image Analysis* 7, 4, 529–537.
- [91] LINGURARU, M. G., VASILYEV, N. V., MARX, G. R., TWORETZKY, W., NIDO, P. J. D., AND HOWE, R. D. Fast block flow tracking of atrial septal defects in 4D echo. *MIA* (2008).
- [92] LIU, X., ABD-ELMONIEM, K. Z., AND PRINCE, J. L. Incompressible cardiac motion estimation of the left ventricle using tagged mr images. In *MICCAI* (2009), vol. 5762, pp. 331–338.
- [93] LOECKX, D., SLAGMOLEN, P., MAES, F., VANDERMEULEN, D., AND SUETENS, P. Nonrigid image registration using conditional mutual information. *IEEE Transactions on Medical Imaging* (2009).
- [94] LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 2 (November 2004), 91–110.
- [95] LUCAS, B., AND KANADE, T. An iterative image restoration technique with an application to stereo vision. In *DARPA IU Workshop* (1981), pp. 121–130.

- [96] LUO, B., AND HANCOCK, E. R. Structural graph matching using the em algorithm and singular value decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 10 (2001), 1120–1136.
- [97] LUO, B., AND HANCOCK, E. R. Iterative procrustes alignment with the em algorithm. *Image and Vision Computing* 20, 5-6 (Apr. 2002), 377–396.
- [98] LUO, B., AND HANCOCK, E. R. A unified framework for alignment and correspondence. *CVIU* 92, 1 (2003), 26–55.
- [99] MAILLOUX, G., LANGLOIS, F., SIMARD, P. Y., AND BERTRAND, M. Restoration of the velocity field of the heart from two-dimensional echocardiograms. *IEEE Transactions on Medical Imaging* 8, 6 (1989), 143–153.
- [100] MAINTZ, J. B. A., AND VIERGEVER, M. A. A survey of medical image registration. *Medical Image Analysis* 2, 1 (Mar. 1998), 1–37.
- [101] MAKELA, T., CLARYSSE, P., SIPILA, O., PAUNA, N., PHAM, Q. C., KATILA, T., AND MAGNIN, I. A review of cardiac image registration methods. *IEEE Transactions on Medical Imaging* 21, 9 (2002), 1011–1021.
- [102] MARSLAND, S., TWINING, C., AND TAYLOR, C. A minimum description length objective function for groupwise non-rigid image registration. *Image and Vision Computing* (2006). In press.
- [103] MCNEILL, G., AND VIJAYAKUMAR, S. A probabilistic approach to robust shape matching. In *IEEE ICIP* (2006), pp. 937–940.
- [104] MEUNIER, J. Tissue motion assessment from 3D echographic speckle tracking. *Physics in medicine and biology*. 43, 5 (1998), 1241–1254.
- [105] MILLER, K., BERNSTEIN, R., AND BLUEMENSON, L. Generalized rayleigh processes. *Quarterly of Applied Mathematics* 16 (1958), 137–145.
- [106] MINKA, T. Divergence measures and message passing. Tech. rep., Microsoft Research, 2005.
- [107] MITCHELL, S., BOSCH, J., LILIEVELDT, B. P., VAN DER GEEST, R. J., REIBER, J. H., AND SONKA, M. 3-D active appearance models: Segmentation of cardiac mr and ultrasound images. *IEEE Transactions on Medical Imaging* 21, 9 (2002).
- [108] MODERSITZKI, J. *Numerical Methods for Image Registration*. Oxford university press, 2004.



- [109] MODERSITZKI, J., AND WIRTZ, S. Combining homogenization and registr. In *WBIR06* (2006), pp. 257–263.
- [110] MYRONENKO, A., AND SONG, X. Image registration by minimization of mapping complexity. In *MMBIA* (2009), vol. 0, IEEE Computer Society, pp. 17–24.
- [111] MYRONENKO, A., AND SONG, X. Image registration by minimization of residual complexity. In *IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 49–56.
- [112] MYRONENKO, A., AND SONG, X. On the closed-form solution of the rotation matrix arising in computer vision problems. Tech. Rep. arXiv:0904.1613v1, Oregon Health and Science Univ., 2009.
- [113] MYRONENKO, A., AND SONG, X. Point set registration: Coherent point drift. *Patter Analysis and Machine Intelligence* (2010).
- [114] MYRONENKO, A., SONG, X., AND CARREIRA-PERPIÑÁN, M. Á. Non-rigid point set registration: Coherent Point Drift. In *Advances in Neural Information Processing Systems 19 (NIPS'06)* (Cambridge, MA, 2007), B. Schölkopf, J. Platt, and T. Hoffman, Eds., MIT Press, pp. 1009–1016.
- [115] MYRONENKO, A., SONG, X., AND SAHN, D. J. LV motion tracking from 3D echocardiography using textural and structural information. In *MICCAI* (Oct. 2007), vol. 4792, pp. 428–435.
- [116] MYRONENKO, A., SONG, X., AND SAHN, D. J. Maximum likelihood motion estimation in 3D echocardiography through non-rigid registration in spherical coordinates. In *Functional Imaging and Modeling of the Heart (FIMH)* (2009), vol. 5528, pp. 427–436.
- [117] NAKAGAMI, M. The m-distribution - a general formula of intensity distribution of rapid fading. *Statistical Methods in Radio Wave Propagation* (1960).
- [118] NASCIMENTO, J., AND MARQUES, J. Robust shape tracking with multiple models in ultrasound images. *IEEE Trans on Image Processing* 17, 3 (March 2008), 392–406.
- [119] NEAL, R. M., AND HINTON, G. E. A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models* (1998), M. I. Jordan, Ed., Kluwer.
- [120] NOBLE, A. J., AND BOUKERROUI, D. Ultrasound image segmentation: a survey. *IEEE Transactions on medical imaging* 25, 8 (Aug. 2006), 987–1010.

- [121] NOCEDAL, J., AND WRIGHT, S. J. *Numerical Optimization*. Springer, 1999.
- [122] NOTOMI, Y., SETSER, R., SHIOTA, T., MARTIN-MIKOLOVIK, M., WEAVER, J., ZORAN, B., POPOVIC, M., YAMADA, H., GREENBERG, N., WHITE, R., AND THOMAS, J. Assessment of left ventricular torsional deformation by doppler tissue imaging. *Circulation* 111 (2005), 1141–1147.
- [123] NUNES, J., BOUAOUNE, Y., DELECHELLE, E., AND BUNEL, P. A multiscale elastic registration scheme for retinal angiograms. *CVIU* 95, 2 (June 2004), 129–149.
- [124] O'DONNELL, M., AND SILVERSTEIN, S. Optimum displacement for compound image generation in medical ultrasound. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control* 35, 4 (July 1988), 470–476.
- [125] OLSHAUSEN, B. A., AND FIELD, D. J. Sparse coding with an overcomplete basis set: a strategy employed by v1. *Vision Research* 37 (1997), 3311–3325.
- [126] ORDERUD, F., AND RABBEN, S. Real-time 3d segmentation of the left ventricle using deformable subdivision surfaces. In *Comp. Vision and Patt. Recognition (CVPR)* (2008), pp. 1–8.
- [127] PAPADEMETRIS, X., AND DUNCAN, J. S. Cardiac image analysis: Motion and deformation. In *SPIE Handbook on Medical Imaging - Volume III: Medical Image Processing and Analysis*, J. M. Fitzpatrick and M. Sonka, Eds. SPIE, 2000.
- [128] PAPADEMETRIS, X., SINUSAS, A. J., DIONE, D. P., CONSTABLE, R. T., AND DUNCAN, J. S. Estimation of 3-d left ventricular deformation from medical images using biomechanical models. *IEEE Transactions on Medical Imaging* 21, 7 (2002), 786–800.
- [129] PAPADEMETRIS, X., SINUSAS, A. J., DIONE, D. P., AND DUNCAN, J. S. Left ventricular deformation from 3D echocardiography. *Medical Image Analysis* 5, 1 (2001), 17–28.
- [130] PAPOULIS, A. *Probability, Random Variables, and Stochastic Processes*. 1991.
- [131] PARIS, S. Research web page.
- [132] PENNEY, G. P., WEESE, J., LITTLE, J. A., DESMEDT, P., HILL, D. L., AND HAWKES, D. J. A comparison of similarity measures for use in 2D-3D medical image registration. In *MICCAI* (1998), vol. 1496, pp. 1153–1161.

- [133] PLUIM, J. P. W., MAINTZ, J. B. A., AND VIERGEVER, M. A. Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on* 22, 8 (2003), 986–1004.
- [134] POHL, K., FISHER, J., LEVITT, J., SHENTON, M., KIKINIS, R., GRIMSON, W., AND WELLS, W. A unifying approach to registration, segmentation, and intensity correction. In *MICCAI* (2005), vol. 3749, pp. 310–318.
- [135] POHL, K. M., FISHER, J., GRIMSON, W. E. L., KIKINIS, R., AND WELLS, W. M. A bayesian model for joint segmentation and registration. *NeuroImage* 32 (2006), 228–239.
- [136] RANGARAJAN, A., CHUI, H., MJOLSNESS, E., DAVACHI, L., GOLDMAN-RAKIC, P. S., AND DUNCAN, J. S. A robust point matching algorithm for autoradiograph alignment. *MIA* 1, 4 (1997), 379–398.
- [137] RESMERITA, E., AND ANDERSSEN, R. S. Joint additive kullback-leibler residual minimization and regularization for linear inverse problems. *Mathematical Methods in the Applied Sciences*, 30, 13 (2007), 1527–1544.
- [138] REVELL, J., MIRMEHDI, M., AND MCNALLY, D. Combined ultrasound speckle pattern similarity measures. In *MIUA* (Sept. 2004), BMVA Press, pp. 149–153.
- [139] REVOW, M., WILLIAMS, C. K. I., AND HINTON, G. E. Using generative models for handwritten digit recognition. *IEEE PAMI* 18, 6 (1996), 592–606.
- [140] RICE, S. Mathematical analysis of random noise. *Bell System Technical Journal* 24 (1945), 46–156.
- [141] ROCHE, A., MALANDAIN, G., AND AYACHE, N. Unifying maximum likelihood approaches in medical image registration. *IJIST* 11 (2000), 71–80.
- [142] ROCHE, A., MALANDAIN, G., PENNEC, X., AND AYACHE, N. The correlation ratio as a new similarity measure for multimodal image registration. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)* (Cambridge, USA, Oct. 1998), vol. 1496 of *LNCS*, Springer Verlag, pp. 1115–1124.
- [143] ROHR, K. *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer, 2001.
- [144] ROSENBAUM, J.T.AND PLANCK, S., MARTIN, T., CRANE, I., XU, H., AND FORRESTER, J. Imaging ocular immune responses by intravital microscopy. *Int Rev Immunol* 21 (2002), 255–272.

- [145] RUECKERT, D., SONODA, L., HAYES, C., HILL, D., LEACH, M., AND HAWKES, D. Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on* 18, 8 (Aug. 1999), 712–721.
- [146] RUECKERT, D., SONODA, L. I., HAYES, C., HILL, D. L. G., LEACH, M. O., AND HAWKES, D. J. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Image Processing* 18, 8 (Aug. 1999), 712–721.
- [147] RUSINKIEWICZ, S., AND LEVOY, M. Efficient variants of the ICP algorithm. In *International Conference on 3D Digital Imaging and Modeling (3DIM)* (2001), pp. 145–152.
- [148] RYAN, N., HENEGHAN, C., AND DE CHAZAL, P. Registration of digital retinal images using landmark correspondence by expectation maximization. *Image and Vision Computing* 22, 11 (Sept. 2004).
- [149] SANCHEZ-ORTIZ, G. I., DECLERCK, J., MULET-PARADA, M., AND NOBLE, J. A. Automating 3D echocardiographic image analysis. In *MICCAI* (2000), vol. 1935, pp. 687–696.
- [150] SCHÖLKOPF, B., AND SMOLA, A. J. *Learning with Kernels*. The MIT Press, Cambridge, MA, 2002.
- [151] SCOTT, G. L., AND LONGUET-HIGGINS, C. An algorithm for associating the features of two images. *Proceedings of the Royal Society London: Biological Sciences* 244, 1309 (Apr. 1991), 21–26.
- [152] SEDERBERG, T. W., AND PARRY, S. R. Free-form deformation of solid geometric models. In *SIGGRAPH* (1986), pp. 151–160.
- [153] SHANKAR, P. M. A model for ultrasonic scattering from tissues based on k distribution. *Physics in medicine and biology* 40 (1995), 1633–1649.
- [154] SHANKAR, P. M. A general statistical model for ultrasonic backscattering from tissues. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control* 47, 3 (May 2000), 727–736.
- [155] SHANKAR, P. M. Ultrasonic tissue characterization using a generalized nakagami model. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control* 48, 06 (Nov. 2001), 1716–1720.
- [156] SIBSON, R., AND STONE, G. Computation of thin-plate splines. *SIAM JSSC* 12, 6 (Nov. 1991), 1304–1313.

- [157] SLOANE, N. The on-line encyclopedia of integer sequences A002457., 2008.
- [158] SLOANE, N. Online enc. of integer sequences. A002457,A002802,A020918., 2008.
- [159] SONG, X., MYRONENKO, A., PLANK, S. R., AND ROSENBAUM, J. T. Registration of microscopic iris image sequences using probabilistic mesh. In *MICCAI* (Oct. 2006), vol. 4191 of *LNCS*, Springer, pp. 553–560.
- [160] SONG, X., MYRONENKO, A., AND SAHN, D. J. Speckle tracking in 3D echocardiography with motion coherence. In *Comp. Vision and Patt. Recognition (CVPR)* (June 2007).
- [161] STRANG, G. The discrete cosine transform. *SIAM Review* 41, 1 (1999), 135–147.
- [162] STUDHOLME, C., DRAPACA, C., IORDANOVA, B., AND CARDENAS, V. Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change. *Medical Imaging, IEEE Transactions on* 25, 5 (2006), 626–639.
- [163] STUDHOLME, C., HILL, D., AND HAWKES, D. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* 32, 1 (Jan. 1999), 71–86.
- [164] SUHLING, M., ARIGOVINDAN, M., JANSEN, C., HUNZIKER, P., AND UNSER, M. Myocardial motion analysis from b-mode echocardiograms. *IEEE Trans. Image Processing* 14, 4 (2005).
- [165] SUN, W., CETIN, M., CHAN, R., REDDY, V., AND HOLMVANG, G. Segmenting and tracking the left ventricle by learning the dynamics in cardiac images. *IPMI* (2005), 553–565.
- [166] SZELISKI, R. Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.* 2, 1 (2006), 1–104.
- [167] TAN, C. C., AND BEAULIEU, N. C. Infinite series representations of the bivariate rayleigh and nakagami-m distributions. *IEEE Trans. Communications* 45, 10 (Oct. 1997).
- [168] THEVENAZ, P., AND UNSER, M. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Image Processing* 9, 12 (2000), 2083–2099.
- [169] TIKHONOV, A. N., AND ARSENIN, V. I. *Solutions of Ill-Posed Problems*. Winston and Sons, Washington, D.C, 1977.

- [170] TSIN, Y., AND KANADE, T. A correlation-based approach to robust point set registration. In *ECCV* (2004), vol. 3, pp. 558–569.
- [171] TURK, G., AND LEVOY, M. Zippered polygon meshes from range images. In *SIGGRAPH* (1994), pp. 311–318.
- [172] TWINING, C., COOTES, T., MARSLAND, S., PETROVIC, V., SCHESTOWITZ, R., AND TAYLOR, C. Information-theoretic unification of groupwise non-rigid registration and model building. In *Proceedings of Medical Image Understanding and Analysis (MIUA)* (2006), vol. 2, pp. 226–230.
- [173] UMEYAMA, S. Least-squares estimation of transformation parameters between two point patterns. *IEEE PAMI* 13, 4 (Apr. 1991), 376–380.
- [174] VEDALDI, A., AND SOATTO, S. A complexity-distortion approach to joint pattern alignment. In *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. MIT Press, Cambridge, MA, 2007, pp. 1425–1432.
- [175] VIOLA, P. A., AND WELLS, W. M. Alignment by maximization of mutual information. In *ICCV* (1995), IEEE Computer Society, pp. 16–23.
- [176] WAHBA, G. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.
- [177] WAN, E., AND VAN DER MERWE, R. *Kalman Filtering and Neural Networks, chap.7: The Unsecented Kalman Filter*. Wiley Publishing, 2001.
- [178] WELLING, M., HINTON, G. E., AND OSINDERO, S. Learning sparse topographic representations with products of student-t distr. In *NIPS* (2003), pp. 1359–1366.
- [179] WELLS, W. M. Statistical approaches to feature-based object recognition. *IJCV* 22, 1-2 (Jan. 1997), 63–98.
- [180] WYATT, P. P., AND NOBLE, J. A. MAP MRF joint segmentation and registration. In *MICCAI* (2002), vol. 2488, Springer-Verlag, pp. 580–587.
- [181] YANG, C., DURAISWAMI, R., GUMEROV, N. A., AND DAVIS, L. Improved fgt and efficient kernel density estimation. In *ICCV* (2003), p. 464.
- [182] YANG, L., GEORGESCU, B., ZHENG, Y., MEER, P., AND COMANICIU, D. 3d ultrasound tracking of the left ventricles using one-step forward prediction and data fusion of collaborative trackers. In *in Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2008).

- [183] YANOVSKY, I., THOMPSON, P. M., OSHER, S., AND LEOW, A. D. Topology preserving log-unbiased nonlinear image registration: Theory and implementation. In *IEEE Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–8.
- [184] YI, Z., AND SOATTO, S. Nonrigid registration combining global and local statistics. In *Comp. Vision and Patt. Recognition (CVPR)* (June 2009).
- [185] YU, W., LIN, N., YAN, P., PURUSHOTHAMAN, K., SINUSAS, A., THIELE, K., AND DUNCAN, J. Motion analysis of 3D ultrasound texture patterns. In *Functional Imaging and Modeling of the Heart (FIMH)* (2003), I. Magnin, Ed., vol. 2674, pp. 252–261.
- [186] YU, W., YAN, P., SINUSAS, A. J., THIELE, K., AND DUNCAN, J. S. Towards pointwise motion tracking in echocardiographic image sequences – comparing the reliability of different features for speckle tracking. *Medical Image Analysis 10* (2006), 496–508.
- [187] YUILLE, A. L., AND GRZYWACZ, N. M. The motion coherence theory. In *Proceedings of the 2nd International Conference on Computer Vision (ICCV’88)* (Dec. 1988), R. Bajcsy and S. Ullman, Eds., vol. 3, IEEE Computer Society press, pp. 344–353.
- [188] YUILLE, A. L., AND GRZYWACZ, N. M. A mathematical analysis of the motion coherence theory. *IJCV 3*, 2 (June 1989), 155–175.
- [189] ZANA, F., MEUNIER, I., AND KLEIN, J. C. A region merging algorithm using mathematical morphology: application to macula detection. In *International symposium on Mathematical morphology and its applications to image and signal processing (ISMM)* (1998), Kluwer Academic Publishers, pp. 423 – 430.
- [190] ZHANG, Z. Iterative point matching for registration of free-form curves and surfaces. *IJCV 13*, 2 (Oct. 1994), 119–152.
- [191] ZHENG, G., AND ZHANG, X. A unifying map-mrf framework for deriving new point similarity measures for intensity-based 2D-3D registration. In *International Conference on Pattern Recognition (ICPR)* (2006), vol. 2, pp. 1181–1185.
- [192] ZHU, H., AND ROHWER, R. Information geometric measurements of generalisation. Tech. rep., Neural Computing Research Group, 1995.

- [193] ZHU, Y., PAPADEMETRIS, X., SINUSAS, A. J., AND DUNCAN, J. S. Segmentation of myocardial volumes from real-time 3D echocardiography using an incompressibility constraint. In *MICCAI* (Oct. 2007), N. Ayache, S. Ourselin, and A. J. Maeder, Eds., vol. 4791 of *Lecture Notes in Computer Science*, Springer, pp. 44–51.
- [194] ZHU, Y., PAPADEMETRIS, X., SINUSAS, A. J., AND DUNCAN, J. S. A dynamical shape prior for lv segmenation from RT3D echocardiography. In *MICCAI* (2009), pp. 206–213.
- [195] ZHU, Y., PAPADEMETRIS, X., SINUSAS, A. J., AND DUNCAN, J. S. A coupled deformable model for tracking myocardial borders from real-time echocardiography using an incompressibility constraint. *Medical Image Analysis* 14, 3 (2010), 429–448.
- [196] ZINI, G., SARTI, A., AND LAMBERTI, C. Application of continuun theory and multi-grid methods to motion evaluation from 3D echocardiography. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control* 44, 3 (1997), 297–308.
- [197] ZITOVA, B., AND FLUSSER, J. Image registration methods: a survey. *Image and Vision Computing* 21 (2003), 977–1000.
- [198] ZOLLEI, L., LEARNED-MILLER, E., GRIMSON, E., AND WELLS, W. Efficient population registration of 3D data. In *ICCV* (2005), pp. 291–301.



# Appendix A

## Upper Bound of the GMM Log-likelihood Function

Here, we briefly show the derivation of the upper bound of the negative log-likelihood function of the Gaussian mixture model [14], which is used on the M-step of the EM algorithm. We can write the change of the negative log-likelihood function for the new and old parameter values as

$$E^{new} - E^{old} = - \sum_{n=1}^N \log \frac{p^{new}(\mathbf{x}_n)}{p^{old}(\mathbf{x}_n)} = - \sum_{n=1}^N \log \frac{\sum_{m=1}^M P^{new}(m) p^{new}(\mathbf{x}_n|m)}{p^{old}(\mathbf{x}_n)} \frac{P^{old}(m|\mathbf{x}_n)}{P^{old}(m|\mathbf{x}_n)} \quad (\text{A.1})$$

where  $p(\mathbf{x}_n)$  is a pdf of GMM in Eq. 4.1,  $P(m|\mathbf{x}_n)$  denotes the posterior probabilities of the mixture components:  $P(m|\mathbf{x}_n) = P(m)p(\mathbf{x}_n|m)/p(\mathbf{x}_n)$  and  $P(m)$  is a membership probability of the mixture component. The last factor inside the brackets is simply the identity. We now use the Jensen's inequality, which says that, given a set of numbers  $\lambda_j \geq 0$  such that  $\sum_j \lambda_j = 1$ ,

$$\log\left(\sum_j \lambda_j x_j\right) \geq \sum_j \lambda_j \log(x_j) \quad (\text{A.2})$$

The probabilities  $P^{old}(m|\mathbf{x}_n)$  in the numerator of Equation A.1 play the role of the  $\lambda_j$ , which gives

$$E^{new} - E^{old} \leq - \sum_{n=1}^N \sum_{m=1}^M P^{old}(m|\mathbf{x}_n) \log \frac{P^{new}(m) p^{new}(\mathbf{x}_n|m)}{p^{old}(\mathbf{x}_n) P^{old}(m|\mathbf{x}_n)} \quad (\text{A.3})$$

We want to minimize  $E^{new}$  with respect to the “new” parameters. If we let  $Q$  be the right-hand side in Equation A.3, then we have  $E^{new} \leq E^{old} + Q$  and so  $E^{old} + Q$  represents an

upper bound on the value of  $E^{new}$ . We can therefore seek to minimize this bound with respect to the “new” values of the parameters, which necessarily leads to a decrease in the value of the  $E^{new}$  unless it is already a local minimum. Ignoring the terms that depends on the “old” parameters, we can rewrite the right-hand side of Equation A.3 as

$$Q = - \sum_{n=1}^N \sum_{m=1}^M P^{old}(m|\mathbf{x}_n) \log(P^{new}(m)p^{new}(\mathbf{x}_n|m)). \quad (\text{A.4})$$

# Appendix B

## Optimal Rotation Matrix

**Lemma 4.** *Let  $\mathbf{R}_{D \times D}$  be an unknown rotation matrix and  $\mathbf{A}_{D \times D}$  be a known real square matrix. Let  $\mathbf{USV}^T$  be a Singular Value Decomposition (SVD) of  $\mathbf{A}$ , where  $\mathbf{UU}^T = \mathbf{VV}^T = \mathbf{I}$ ,  $\mathbf{S} = \text{d}(s_i)$ ,  $s_1 \geq s_2 \geq \dots \geq s_D \geq 0$ . Then the optimal rotation matrix  $\mathbf{R}$  that maximizes  $\text{tr}(\mathbf{A}^T \mathbf{R})$  is*

$$\mathbf{R} = \mathbf{UCV}^T, \quad \text{where } \mathbf{C} = \text{d}(1, 1, \dots, 1, \det(\mathbf{UV}^T)). \quad (\text{B.1})$$

We convert the constrained optimization problem into unconstrained using Lagrange multipliers. Define an objective function  $f$  to be minimized as

$$f(\mathbf{R}) = -\text{tr}(\mathbf{A}^T \mathbf{R}) + \text{tr}((\mathbf{R}^T \mathbf{R} - \mathbf{I})\Lambda) + \lambda(\det(\mathbf{R}) - 1), \quad (\text{B.2})$$

where  $\Lambda$  is a symmetric matrix of unknown Lagrange multipliers and  $\lambda$  is another unknown Lagrange multiplier. Equating to zero the partial derivatives of  $f$  with respect to  $\mathbf{R}$ , we obtain the following system of equations:

$$\frac{\partial f}{\partial \mathbf{R}} = -\mathbf{A} + \mathbf{R}\Lambda + \lambda\mathbf{R} = \mathbf{RB} - \mathbf{A} = 0. \quad (\text{B.3})$$

where  $\mathbf{B}$  is symmetric by construction:  $\mathbf{B} = \Lambda + \lambda\mathbf{I}$ . Thus we need to solve a linear system of equations:

$$\mathbf{A} = \mathbf{RB}, \quad \text{s.t. } \mathbf{R}^T \mathbf{R} = \mathbf{I}, \quad \det(\mathbf{R}) = 1. \quad (\text{B.4})$$

Transposing Eq. B.4 and multiplying from both sides we obtain  $\mathbf{A}^T \mathbf{A} = \mathbf{B}^2$ , where  $\mathbf{A}^T \mathbf{A}$  is guaranteed to be symmetric and positive definite (or semi-definite if  $\mathbf{A}$  is singular), and we can decompose it using the spectral decomposition:

$$\mathbf{B}^2 = \mathbf{A}^T \mathbf{A} = \mathbf{VS}^2 \mathbf{V}^T, \quad (\text{B.5})$$

where  $\mathbf{S}^2$  is real non-negative diagonal matrix of eigenvalues of  $\mathbf{A}^T \mathbf{A}$  as well as  $\mathbf{B}^2$ , so that  $s_1^2 \geq s_2^2 \geq \dots \geq s_D^2 \geq 0$ . Also, note that matrix  $\mathbf{S}$  is real non-negative diagonal matrix of singular values of  $\mathbf{A}$ . Clearly, matrices  $\mathbf{B}$  and  $\mathbf{B}^2$  are both symmetric with commutative property:  $\mathbf{B}\mathbf{B}^2 = \mathbf{B}^2\mathbf{B}$ , hence both share the same eigenvectors, only when  $\mathbf{B}^2$  is not degenerative<sup>1</sup>. Thus matrix  $\mathbf{B}$  is in the form:

$$\mathbf{B} = \mathbf{V}\mathbf{M}\mathbf{V}^T \quad (\text{B.6})$$

where  $\mathbf{M}$  is real diagonal matrix with eigenvalues of  $\mathbf{B}$ , which must be in the form:  $\mathbf{M} = \text{d}(\pm s_1, \pm s_2, \dots, \pm s_D)$ . Substituting  $\mathbf{M}$  into equation Eq. B.4 and then into the objective function, we obtain:

$$\text{tr}(\mathbf{A}^T \mathbf{R}) = \text{tr}(\mathbf{B}^T \mathbf{R}^T \mathbf{R}) = \text{tr}(\mathbf{B}) = \text{tr}(\mathbf{V}\mathbf{M}\mathbf{V}^T) = \text{tr}(\mathbf{M}) \quad (\text{B.7})$$

Taking into account that  $\det(\mathbf{R}) = 1$ , from Eq. B.4 we see that

$$\det(\mathbf{A}) = \det(\mathbf{R}) \det(\mathbf{B}) = \det(\mathbf{B}) = \det(\mathbf{V}) \det(\mathbf{M}) \det(\mathbf{V}^T) = \det(\mathbf{M}), \quad (\text{B.8})$$

hence  $\det(\mathbf{M})$  must have at least the same sign as  $\det(\mathbf{A})$ . Clearly, matrix  $\mathbf{M}$  that maximizes its trace is

$$\mathbf{M} = \text{d}(s_1, s_2, \dots, s_D), \quad \text{if } \det(\mathbf{A}) > 0, \quad (\text{B.9})$$

$$\mathbf{M} = \text{d}(s_1, s_2, \dots, -s_D), \quad \text{if } \det(\mathbf{A}) < 0. \quad (\text{B.10})$$

and the objective function value at the optimum is

$$\text{tr}(\mathbf{A}^T \mathbf{R}) = \text{tr}(\mathbf{M}) = s_1 + s_2 + \dots + s_{D-1} \pm s_D \quad (\text{B.11})$$

where the last sign depends on the determinant of  $\mathbf{A}$ . Now, we can find the optimal rotation matrix  $\mathbf{R}$ , from the Eq. B.4:

$$\mathbf{A} = \mathbf{R}\mathbf{B}, \quad (\text{B.12})$$

$$\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{R}\mathbf{V}\mathbf{M}\mathbf{V}^T, \quad (\text{B.13})$$

$$\mathbf{U}\mathbf{S} = \mathbf{R}\mathbf{V}\mathbf{M}. \quad (\text{B.14})$$

---

<sup>1</sup>Here, by degenerative matrix we mean a matrix with not distinct (repeated) singular values. Note, that a matrix can be non-singular, but still degenerative.

If  $\mathbf{A}$  is non-singular ( $\text{rank}(\mathbf{A}) = D$ ), then  $\mathbf{M}$  is invertable, and the optimal  $\mathbf{R}$  is

$$\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{M}^{-1}\mathbf{V}^T = \mathbf{U}\mathbf{C}\mathbf{V}^T, \quad \text{where } \mathbf{C} = \text{d}(1, 1, \dots, 1, \det(\mathbf{U}\mathbf{V}^T)). \quad (\text{B.15})$$

where  $\det(\mathbf{U}\mathbf{V}^T) = \det(\mathbf{U})\det(\mathbf{V}^T) = \text{sign}(\det(\mathbf{A})) = \pm 1$  depending on a sign of  $\det(\mathbf{A})$ .

For some singular and degenerative cases of  $\mathbf{A}$  the solution is still optimal, but not unique.

We refer to [112] for the detailed discussion.

## Appendix C

### Green's function corresponding to $\phi_{RT}(v)$

Here, we show that the Green's function corresponding to the regularization term

$$\phi_{RT}(v) = \int \frac{|\tilde{v}(\mathbf{s})|^2}{\tilde{G}(\mathbf{s})} d\mathbf{s} \quad (\text{C.1})$$

has a form of the low-pass filter function  $G$ . Here  $\tilde{v}$  indicates the Fourier transform of the velocity function  $v$  and  $\tilde{G}$  is a strictly positive definite function that approaches zero as  $\|\mathbf{s}\| \rightarrow \infty$ . Here  $\tilde{G}$  represents a symmetric low-pass filter, so that its inverse Fourier transform  $G$  is real and symmetric. Such regularization term has been used in the Regularization Theory (RT) [55] to penalize high frequency function content.

Such regularization represents a norm of the function  $v$  after applying an operator  $P$ :

$$\phi(v) = \int \|Pv\|^2 d\mathbf{x} \quad (\text{C.2})$$

We shall find the functional derivative of  $\phi_{RT}(v)$ . First, assume the function  $\hat{v}$  to be the minimizer of the functional  $\phi_{RT}(v)$ . Then, we can define  $v(\mathbf{x}) = \hat{v}(\mathbf{x}) + \epsilon g(\mathbf{x})$ , where  $g$  is an arbitrary infinitely differentiable function and  $\epsilon$  is a small number. Second, we take the derivative of  $\phi_{RT}(v)$  with respect to  $\epsilon$  at zero:

$$\begin{aligned} \left. \frac{d\phi_{RT}(v(\epsilon))}{d\epsilon} \right|_{\epsilon=0} &= \left. \frac{d}{d\epsilon} \int \frac{\frac{1}{(2\pi)^{D/2}} (\int (\hat{v}(\mathbf{x}) + \epsilon g(\mathbf{x})) e^{-i\mathbf{s}^T \mathbf{x}} d\mathbf{x})^2}{\tilde{G}(\mathbf{s})} d\mathbf{s} \right|_{\epsilon=0} = \\ &= \frac{2}{(2\pi)^{D/2}} \int \int \frac{\tilde{v}(\mathbf{s}) g(\mathbf{x}) e^{-i\mathbf{s}^T \mathbf{x}}}{\tilde{G}(\mathbf{s})} d\mathbf{x} d\mathbf{s} = \int \left( \frac{2}{(2\pi)^{D/2}} \int \frac{\tilde{v}(\mathbf{s}) e^{-i\mathbf{s}^T \mathbf{x}}}{\tilde{G}(\mathbf{s})} d\mathbf{s} \right) g(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (\text{C.3})$$

Thus, the functional derivative of  $\phi_{RT}(v)$  is

$$\frac{\delta \phi_{RT}(v)}{\delta v} = \frac{2}{(2\pi)^{D/2}} \int \frac{\tilde{v}(\mathbf{s}) e^{-i\mathbf{s}^T \mathbf{x}}}{\tilde{G}(\mathbf{s})} d\mathbf{s} = 2\hat{P}Pv(\mathbf{x}) \quad (\text{C.4})$$

where  $\hat{P}$  is the adjoint differential operator to  $P$ . The *Green's function* of the self-adjoint operator  $\hat{P}P$  is a function  $K(\mathbf{x}, \mathbf{x}')$  that satisfies

$$\hat{P}PK(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}') \quad (\text{C.5})$$

where  $\delta$  is the Dirac delta function. Substituting  $K(\mathbf{x}, \mathbf{x}')$  in Equation C.4, we achieve

$$\frac{1}{(2\pi)^{D/2}} \int \frac{\tilde{K}(\mathbf{s})e^{-i\mathbf{s}^T \mathbf{x}}}{\tilde{G}(\mathbf{s})} d\mathbf{s} = \delta(\mathbf{x}) \quad (\text{C.6})$$

The equality holds for  $\tilde{K}(\mathbf{s}) = \tilde{G}(\mathbf{s})$ , thus the low-pass filter function  $G$  is itself the Green's function of the self-adjoint operator  $\hat{P}P$ . In general the solution in Eq. 4.19 would also include the null space terms of  $\phi_{RT}(v)$ . However for the positive definite function  $G$  (e.g. Gaussian) the regularization term  $\phi_{RT}(v)$  defines a norm [3], and thus the null space of  $\phi_{RT}(v)$  contains only zero element (see [55] for details).

## Appendix D

### Green's function corresponding to $\phi_{MCT}(v)$

Here, we show that the Green's function corresponding to the regularization term

$$\phi_{MCT}(v) = \int_{\Omega} \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} \left\| D^l v(\mathbf{x}) \right\|^2 d\mathbf{x}, \quad (\text{D.1})$$

has a Gaussian form. Here  $D$  is a derivative operator so that  $D^{2l}v = \nabla^{2l}v$  and  $D^{2l+1}v = \nabla(\nabla^{2l}v)$ , where  $\nabla$  is the gradient operator,  $\nabla^2$  is the Laplacian operator and  $\Omega$  is an open boundary subset on  $\mathbb{R}^N$  with a piecewise smooth boundary  $\partial\Omega$ . Such regularization term has been proposed in the Motion Coherence Theory (MCT) [188]. It penalizes all order derivatives of the function.

Equation D.1 is a norm of the function  $v$  after applying an operator  $P$  (Equation C.2). We shall find the functional derivative of  $\phi_{MCT}(v)$ . First, we denote by  $\hat{v}$  a minimizer of the functional  $\phi_{MCT}(v)$ . We can write that  $v(\mathbf{x}) = \hat{v}(\mathbf{x}) + \epsilon g(\mathbf{x})$ , where  $g$  is an arbitrary infinitely differentiable function and  $\epsilon$  is a small number. The function  $g$  and all its derivative have a zero boundary condition. Second, we take the derivative of  $\phi_{MCT}(v)$  with respect to  $\epsilon$  at zero:

$$\begin{aligned} \left. \frac{d\phi_{MCT}(v(\epsilon))}{d\epsilon} \right|_{\epsilon=0} &= \\ \left. \frac{d}{d\epsilon} \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} \int_{\Omega} \left\| D^l(\hat{v}(\mathbf{x}) + \epsilon g(\mathbf{x})) \right\|^2 d\mathbf{x} \right|_{\epsilon=0} &= 2 \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} \int_{\Omega} (D^l \hat{v}(\mathbf{x})) (D^l (g(\mathbf{x}))) d\mathbf{x} = \\ 2 \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} \int_{\Omega} (-1)^l (D^{2l} \hat{v}(\mathbf{x})) g(\mathbf{x}) d\mathbf{x} &= \int_{\Omega} 2 \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} (-1)^l (D^{2l} \hat{v}(\mathbf{x})) g(\mathbf{x}) d\mathbf{x} \quad (\text{D.2}) \end{aligned}$$



To see that the last simplification is true, consider a few differentiation for  $l = 0, 1, 2$  and use the integration by parts. Thus, the functional derivative of  $\phi_{MCT}(v)$  is

$$\frac{\delta\phi(v)}{\delta v} = 2 \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} (-1)^l (D^{2l}v) = 2\hat{P}Pv \quad (\text{D.3})$$

We can find the Green's function,  $G(\mathbf{x}, \mathbf{x}')$ , of the selfadjoint operator  $\hat{P}P$  using the property of the Green's function (see Equation C.5):

$$\sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} (-1)^l D^{2l}G(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}') \quad (\text{D.4})$$

We use the multidimensional Fourier transforms of  $G(\mathbf{x}, \mathbf{x}')$  and  $\delta(\mathbf{x} - \mathbf{x}')$

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{(2\pi)^{D/2}} \int \tilde{G}(\mathbf{s}) e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s} \quad (\text{D.5})$$

$$\delta(\mathbf{x} - \mathbf{x}') = \frac{1}{(2\pi)^{D/2}} \int e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s} \quad (\text{D.6})$$

and substitute them into Equation D.4:

$$\begin{aligned} \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} (-1)^l D^{2l} \frac{1}{(2\pi)^{D/2}} \int \tilde{G}(\mathbf{s}) e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s} &= \frac{1}{(2\pi)^{D/2}} \int e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s}, \\ \int \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} (-1)^l D^{2l} (\tilde{G}(\mathbf{s}) e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} - e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} ) d\mathbf{s} &= 0, \\ \int \sum_{l=0}^{\infty} \frac{\beta^{2l}}{l!2^l} (-1)^l (i)^{2l} \|\mathbf{s}\|^{2l} \tilde{G}(\mathbf{s}) e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} - e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s} &= 0, \\ \int (\tilde{G}(\mathbf{s}) \sum_{l=0}^{\infty} \frac{\beta^{2l} \|\mathbf{s}\|^{2l}}{l!2^l} - 1) e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s} &= 0. \\ \int (\tilde{G}(\mathbf{s}) e^{\frac{\beta^2 \|\mathbf{s}\|^2}{2}} - 1) e^{-i\mathbf{s}^T(\mathbf{x}-\mathbf{x}')} d\mathbf{s} &= 0. \end{aligned}$$

where we used the Taylor series expansion of the exponential function ( $e^{\mathbf{z}} = \sum_{l=0}^{\infty} \frac{\mathbf{z}^l}{l!}$  with  $\mathbf{z} = \frac{\beta^2 \|\mathbf{s}\|^2}{2}$ ). Thus  $\tilde{G}(\mathbf{s})$  that satisfies the Green's function property is  $\tilde{G}(\mathbf{s}) = e^{-\frac{\beta^2 \|\mathbf{s}\|^2}{2}}$ . Substituting  $\tilde{G}(\mathbf{s})$  back into the Eq. D.5, we find that the Green's function,  $G(\mathbf{x}, \mathbf{x}')$ , is a Gaussian:

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{(2\pi\beta^2)^{D/2}} e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\beta^2}} \quad (\text{D.7})$$

For more details and we refer the reader to [187, 188, 14].

## Biographical Note

Andriy Myronenko was born in Krivoy Rog, Ukraine on February 19, 1984. He earned the BS and MS degrees with honors from Dnepropetrovsk National University, Ukraine in 2004 and 2005 respectively. He joined the Ph.D program in Electrical Engineering at Oregon Health & Science University in 2005 and graduated in 2010. During his Ph.D pursuit he has been working in multidisciplinary teams of researchers, engineers and doctors, from whom he has learned a lot. His research interests broadly include computer vision and machine learning areas. Andriy Myronenko is an author of the following publications:

- Andriy Myronenko and Xubo Song: “Intensity-based Image Registration by Minimizing Residual Complexity”, *Accepted to IEEE Trans. on Medical Imaging*.
- Andriy Myronenko and Xubo Song: “Point-Set Registration: Coherent Point Drift”, *Accepted to IEEE Trans. on Pattern Analysis and Machine Intelligence*.
- Andriy Myronenko and Xubo Song: “Image Registration by Minimization of Residual Complexity”, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 49-56, 2009.
- Andriy Myronenko and Xubo Song: “Global Active Contour-based Image Segmentation via Probability Alignment”, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2798-2804, 2009.
- Andriy Myronenko, Xubo Song and David J. Sahn: “Maximum Likelihood Motion Estimation in 3D Echocardiography through Non-rigid Registration in Spherical Coordinates”, *Functional Imaging and Modeling of the Heart (FIMH)*, pp. 427-436, LNCS vol. 5528, 2009.

- Andriy Myronenko and Xubo Song: "Image Registration by Minimization of Mapping Complexity", *Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pp. 17-24, 2009.
- Andriy Myronenko, Xubo Song and David J. Sahn: "LV Motion Tracking from 3D Echocardiography Using Textural and Structural Information", *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS vol. 4792, Springer, pp. 428-435, 2007.
- Xubo Song, Andriy Myronenko and David J. Sahn: "Speckle Tracking in 3D Echocardiography with Motion Coherence", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-7, 2007.
- Andriy Myronenko, Xubo Song and Miguel A. Carreira-Perpinan: "Free-Form Non-rigid Image Registration Using Generalized Elastic Nets", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8, 2007.
- Houwu Bai, Eric Wan, Xubo Song, Andriy Myronenko and Alexander Bogdanov: "Vision-only Navigation and Control of Unmanned Aerial Vehicles Using the Sigma-Point Kalman Filter", *Proceedings of the Institute of Navigation (ION)*, National Technical Meeting, 2007.
- Andriy Myronenko, Xubo Song and Miguel A. Carreira-Perpinan: "Non-rigid point set registration: Coherent Point Drift", *Advances in Neural Information Processing Systems 19 (NIPS)*, pp. 1009-1016, 2006.
- Xubo Song, Andriy Myronenko, Stephen R. Plank and Stephen R. Plank: "Registration of Microscopic Iris Image Sequences Using Probabilistic Mesh", *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS vol. 4191, Springer, pp. 553-560, 2006.
- Muhammad Ashraf, Andriy Myronenko, Thuan Nguyen, Akio Inage, Wayne Smith, Robert I. Lowe, Karl Thiele, Carol A. Gibbons Kroeker, John V. Tyberg, Jeffrey F. Smallhorn, David J. Sahn and Xubo Song: "Defining Left Ventricular Apex

to Base Twist Mechanics Computed from High-Resolution 3D Echocardiography: Validation against Sonomicrometry”, *Journal of American College of Cardiology (JACC) Imaging*, vol. 3, pp. 227-234, 2010.