

# **Three-dimensional Ultrasound Fusion for Transesophageal Echocardiography**

**by Zhehua Mao**

Thesis submitted in fulfilment of the requirements for  
the degree of

**Doctor of Philosophy**

under the supervision of Liang Zhao and Shoudong Huang

University of Technology Sydney  
Faculty of Engineering and Information Technology

August 2023

# CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Zhehua Mao*, declare that this thesis is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *Faculty of Engineering and Information Technology* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature:	Production Note: Signature removed prior to publication.
<hr/>	

Date:	29/08/2023
<hr/>	

# *Abstract*

UNIVERSITY OF TECHNOLOGY SYDNEY

Faculty of Engineering and Information Technology

Robotics Institute

Doctor of Philosophy

by [Zhehua Mao](#)

In cardiology, transesophageal echocardiography (TEE) is a widely used diagnostic test that uses an ultrasound transducer passed into the patient's esophagus or stomach to capture images of the heart. Since the esophagus and stomach are close to the heart, detailed images can be obtained by TEE to facilitate the diagnosis, management, and follow-up of patients with various cardiac diseases. In recent decades, progress in ultrasound technology has promoted the evolution of TEE from two-dimensional (2D) imaging to three-dimensional (3D) imaging modality. 3D TEE has overcome several important limitations of 2D TEE such as the lack of anatomy and orientation information, which makes it increasingly used in the perioperative period for assessing cardiac anatomy and function. However, current 3D TEE still suffers from a narrow field of view (FoV) due to manufacturing limitations on 2D array transducers, which makes it difficult to inspect a complete region of interest in a single volume.

Image registration and fusion is a fundamental task for many medical image analysis problems where valuable information conveyed by two or more images needs to be combined and examined. This two-step technique is able to enlarge the FoV of images through combining images captured from different viewpoints. Although registration and fusion of multiview images is widely used for enlarging the FoV in the computer vision community, there are some challenges when dealing with 3D TEE images. For example, 3D ultrasound images lack distinct features such that commonly used feature-based methods tend

to fail when handling 3D TEE images. And it is difficult for existing non-rigid registration methods to deal with large cardiac motion in the 3D TEE images. Furthermore, 3D TEE images usually contain a large number of voxels, which requires a lot of computational resources and time for the registration and fusion. Therefore, this study is conducted to enlarge the FoV of 3D TEE images in both rigid and non-rigid scenarios and overcome some major challenges in multiview registration and fusion of 3D medical images.

In this thesis, five works are presented to enlarge the FoV of 3D TEE images in rigid and deformable scenarios. To circumvent the problems of feature extraction and matching in 3D TEE images, all the proposed algorithms are based on the raw image data of 3D TEE images directly. In the first work, a multiview registration and fusion framework is proposed to fuse 3D TEE images efficiently in the rigid scenario. The images are obtained using ECG-gating such that cardiac motion is avoided. In the framework, a direct method with efficient optimization and fast implementation is proposed for pairwise registration. A 3D fusion method is proposed to fuse the aligned images seamlessly and efficiently. To fuse multiview images, a sequential fusion strategy is employed, which means we repeatedly fuse one image into the panoramic image until all images are fused together. Our second work is based on dynamic images of 3D TEE in which the cardiac motion of the heart is retained, instead of using ECG-gated images. In this work, we propose a direct dynamic fusion framework to fuse multiview 3D TEE images with deformations sequentially. A direct embedded deformation algorithm is proposed for pairwise registration of 3D TEE images with large cardiac motions. Different from the original embedded deformation method that relied on explicit point correspondences, we propose to use intensity information as constraints directly to estimate the deformation field. To fuse multiview images, a sequential fusion strategy is used as well.

Our first two works are deduced from pairwise registration and fusion. Although they are efficient for multiview registration and fusion, they probably bring in accumulating errors in the results. Therefore, in the following three works, we present three simultaneous registration and fusion algorithms to further improve the accuracy, which are very useful in practice when fast performance is not required. In the third work, we propose a novel direct bundle adjustment (DBA) algorithm for registering and fusing a collection of 3D images simultaneously in the rigid scenario. It optimizes the global poses of local



images and the intensities of a predefined panoramic image using the 3D image data directly. Without any reference image, correspondences, or information loss or reuse, the proposed DBA algorithm is an elegant way to obtain the optimal panoramic image and poses of local images. To our best knowledge, compared to conventional two-step registration and fusion methods, this is the first work that could complete the whole task in one go. It is also the first work of bundle adjustment which is based on the direct method for 3D image registration and fusion. When investigating DBA further, we found a very interesting property and have theoretically proved that when solving DBA with Gauss-Newton iterations, the optimization of poses is independent of the intensities of the panoramic image. This finding has a potential of reducing the computational complexity of simultaneous image registration and fusion. Based on the finding, a direct simultaneous registration (DSR) algorithm is proposed to optimize the image poses only without considering the intensities of the panoramic image but obtains the same optimal poses as the DBA. Furthermore, inspired by the DBA and DSR, in the fifth work, we propose a novel dynamic direct simultaneous registration (D-DSR) method to estimate panoramic images at all times/phases corresponding to local images simultaneously. This is achieved by substituting the rigid transformations in DBA with embedded deformation graphs and maximizing the similarity between a predefined panoramic image and all local images. We also prove a similar property as DBA in this work that the estimation of the embedded graphs is independent of the intensities of the panoramic image during the optimization process.

In the above five works, rigorous experiments have been performed to test our algorithms on simulated and in-vivo datasets. All results show that our algorithms can enlarge the FoV of 3D TEE images effectively and accurately, which is potential for future clinical use.

## *Acknowledgements*

Time is fleeting! It fills me with great happiness to think about all the three and a half years that I spent at the University of Technology Sydney (UTS) and all the friendships, support, guidance, and supervision that I have found here. I really enjoyed my journey of Ph.D. study in Sydney.

I would like to take this opportunity to express my sincere gratitude to many people. My Ph.D. study and the works presented in this thesis could not have been possible without their support and help. First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Liang Zhao and Prof. Shoudong Huang, for providing me with a valuable opportunity to pursue the Ph.D. degree and guiding me on the road of research. They are rigorous in academic research and sincere in treating others. They are not only good supervisors but also helpful friends to me. During my work and studies, I benefited a lot from their profound academic foundation and keen scientific insights. They always gave me support and encouragement when I was most confused in research and helped me overcome difficulties. Their passion towards research has had a profound influence on me and will encourage me to achieve more success in my future life.

In addition, I would like to express my sincere thanks to Prof. Dikai Liu for introducing me to the lab members. I would like to thank Dr. Raphael Guenot-Falque, A/Prof. Jaime Valls Miro, and A/Prof. Teresa Vidal Calleja. Thanks a lot for your valuable suggestions on my research.

Furthermore, I would like to express my gratitude to the UTS Graduate Research School (GRS) for providing me with financial support to complete my Ph.D. studies. And I deeply appreciate GRS research officers like A/Prof. Andrea Trianni and Angel Lawah for their great help and support.

Moreover, I would like to express my heartfelt thanks to my wife Caili Yao, who has been a great support and encouragement to me over the years, sharing my joys and sorrows. It is her love and care that make me more firm in my choice and future direction.

Finally, many thanks to my family, who always give me selfless love. Their deep love is with me every step of the way, and their support and encouragement are powerful sources of motivation for me to make positive progress and continue to climb the mountain of life.

# Contents

<b>CERTIFICATE OF ORIGINAL AUTHORSHIP</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	3
1.2 Research Aims . . . . .	4
1.3 Main Contributions . . . . .	5
1.4 Publications . . . . .	7
1.5 Thesis Outline . . . . .	8
<b>2 Related Works</b>	<b>11</b>
2.1 Feature-based vs. Intensity-based Registration . . . . .	12
2.2 Rigid vs. Deformable Registration . . . . .	14
2.2.1 Rigid and Affine Transformation . . . . .	14
2.2.2 Deformable Transformation . . . . .	15
2.2.2.1 Free-Form Deformation . . . . .	15
2.2.2.2 Demons-based Registration . . . . .	17
2.2.2.3 Learning-based Registration . . . . .	18
2.3 Pairwise vs. Simultaneous Registration . . . . .	22
2.3.1 Accumulated Pairwise Estimates . . . . .	22
2.3.2 Bundle Adjustment . . . . .	23
2.4 Chapter Summary . . . . .	24

<b>I</b>	<b>PAIRWISE REGISTRATION AND SEQUENTIAL FUSION</b>	<b>25</b>
<b>3</b>	<b>Fast Direct Ultrasound Fusion for 3D Transesophageal Echocardiography</b>	<b>26</b>
3.1	Methods . . . . .	27
3.1.1	Direct Registration . . . . .	27
3.1.1.1	Formulation of the Registration Problem . . . . .	27
3.1.1.2	Solving the Optimization Problem . . . . .	29
3.1.1.3	Efficient Implementation Based on Gradient Space and Orthogonal Planes . . . . .	30
3.1.2	3D Fusion . . . . .	31
3.2	Experiments and Results . . . . .	33
3.2.1	Simulated Experiments . . . . .	33
3.2.1.1	Effectiveness of Registration . . . . .	33
3.2.1.2	Three-orthogonal-plane Approximation vs Full Image . . . . .	38
3.2.1.3	Robustness Assessment via Monte-Carlo Simulation — Intensity Noise . . . . .	40
3.2.1.4	Robustness Assessment via Monte-Carlo Simulation — Initial Guess . . . . .	41
3.2.2	In-vivo Experiments . . . . .	43
3.2.2.1	Three-orthogonal-plane Approximation vs Full Image . . . . .	44
3.2.2.2	Robustness Assessment to Initial Guess . . . . .	46
3.2.2.3	Comparison with PCA-based Registration . . . . .	47
3.2.2.4	Image Fusion of In-vivo Data . . . . .	49
3.2.3	Clinical Applications of 3D TEE Fusion and Further Processing . . . . .	50
3.3	Chapter Summary . . . . .	51
<b>4</b>	<b>Direct Dynamic Fusion for 3D Transesophageal Echocardiography</b>	<b>53</b>
4.1	Methods . . . . .	54
4.1.1	Direct Embedded Deformation . . . . .	55
4.1.1.1	Local Deformation . . . . .	55
4.1.1.2	Energy Function . . . . .	56
4.1.1.3	Optimization . . . . .	58
4.2	Experiments and Results . . . . .	59
4.2.1	Experimental Setup and Evaluation . . . . .	59
4.2.2	Pairwise Non-rigid Registration . . . . .	62
4.2.3	Sequential Fusion . . . . .	65
4.3	Chapter Summary . . . . .	68
<b>II</b>	<b>SIMULTANEOUS REGISTRATION AND FUSION</b>	<b>70</b>
<b>5</b>	<b>Direct Bundle Adjustment for Fusing Multiple 3D Images</b>	<b>71</b>
5.1	Methods . . . . .	72
5.1.1	Problem Statement . . . . .	72

5.1.2	Formulation of Direct Bundle Adjustment . . . . .	73
5.1.3	Solving the Optimization Problem . . . . .	74
5.1.4	Efficient Implementation . . . . .	75
5.2	Experiments and Results . . . . .	77
5.2.1	Simulated Experiments . . . . .	78
5.2.1.1	Accuracy Assessment of DBA . . . . .	79
5.2.1.2	Robustness Assessment of DBA . . . . .	80
5.2.2	In-vivo Experiments . . . . .	81
5.3	Chapter Summary . . . . .	84
<b>6</b>	<b>DSR: Direct Simultaneous Registration for Multiple 3D Images</b>	<b>86</b>
6.1	Methods . . . . .	87
6.1.1	Revisit of Direct Bundle Adjustment . . . . .	87
6.1.2	Simultaneous Registration without Intensity Optimization . . . . .	88
6.2	Experiments and Results . . . . .	90
6.2.1	Simulated Experiments . . . . .	90
6.2.2	In-vivo Experiments . . . . .	92
6.3	Chapter Summary . . . . .	95
<b>7</b>	<b>D-DSR: Dynamic Direct Simultaneous Registration for Multiple 3D Images</b>	<b>97</b>
7.1	Methods . . . . .	98
7.1.1	Problem Statement . . . . .	98
7.1.2	Formulation of Dynamic Direct Bundle Adjustment . . . . .	99
7.1.3	Dynamic Direct Simultaneous Registration without Intensity Optimization . . . . .	101
7.2	Experiments and Results . . . . .	104
7.3	Chapter Summary . . . . .	107
<b>8</b>	<b>Conclusion and Future Plans</b>	<b>111</b>
	<b>Bibliography</b>	<b>116</b>

# List of Figures

1.1	An example of 3D TEE image: (a) the valid image is within a pyramidal region; (b) the 3D TEE image is displayed via three orthogonal slices. . . .	2
1.2	(a) Schematic diagram of LAA occlusion; (b) Clinical measurement for device size selection of LAA occluder. . . . .	4
2.1	B-spline based free-form deformation. A three-orthogonal view of a 3D image shown without control point grid (a) and with control point grid (b). . . .	16
2.2	An example of a U-Net-based registration framework for MRI images. . . .	18
2.3	A schematic diagram of STN. $U$ , $V$ , and $\theta$ denote input, output, and learned transformation parameters, respectively. . . . .	19
2.4	An example of a GAN-based registration framework for MRI images. The framework combines U-Net, STN, and GAN. . . . .	20
2.5	Overview of the VoxelMorph. . . . .	21
3.1	Overview of the direct 3D ultrasound fusion framework. . . . .	27
3.2	Schematic diagram of direct registration for two 3D images. . . . .	28
3.3	Full image vs. three-orthogonal-plane image. . . . .	30
3.4	(a) A 3D image is denoted in three different formats along different axes in the image coordinate system; (b) Weights for 2D slices fusion. . . . .	32
3.5	$I_1$ and $I_2$ are images cropped from 3D heart CT scan. After adding Gaussian noise to these two images, corresponding noisy images $I_1^*$ and $I_2^*$ are obtained. . . . .	34
3.6	Comparison of the convergence and accuracy of the proposed method and MATLAB function <i>imregtform</i> (one-level image pyramid, maximum 100 iterations). . . . .	35
3.7	Comparison of the iterative process of <i>imregtform</i> and the proposed method with full image and three-plane image (one-level image pyramid, maximum 100 iterations). . . . .	37
3.8	Comparison of the computational cost of the proposed tree-plane method and MATLAB function <i>imregtform</i> (three-level image pyramid, maximum 300 iterations). . . . .	37
3.9	3D CT image of the heart and the simulated 3D TEE image sequence. . . .	38
3.10	Overlapping rate of the valid image area of each frame w.r.t. Frame 0. . . .	38
3.11	The accuracy of the full-image method and the three-plane method: (a) and (b) are the MAE of the estimated translations and Euler angles, respectively. . . . .	40
3.12	Robustness test of the proposed method to intensity noise via Monte-Carlo simulation: MAE of the poses under five levels of intensity noise. . . . .	42

3.13	Robustness test of the proposed method to initial guess via Monte-Carlo simulation: the percentages of convergence and un-convergence with 10 levels of noise of initial guesses. . . . .	43
3.14	Mean absolute relative errors of the three-plane method w.r.t. the full-image method. . . . .	45
3.15	Four pairs of images with large mean absolute relative errors (Experiment 3, 4, 5 and 9 in Fig. 3.14) are fused. For the same pairs of images, relative poses are estimated by full-image method and three-plane method respectively. . . . .	45
3.16	Robustness test of the proposed method to initial guess via in-vivo data: the percentages of convergence and un-convergence with 10 levels of noise of initial guesses. . . . .	46
3.17	Four pairs of aligned images based on the PCA-based method and three-plane method are shown in the upper and bottom row, respectively. The same stitching areas are indicated by the rectangular boxes on images. . . . .	48
3.18	The fusion process of in-vivo 3D TEE images from patient # 6. The FoV of the final fused image is 2.04 times larger than the first volume. . . . .	48
3.19	Comparison of original single frame of in-vivo 3D TEE image with fused 3D image using the proposed algorithms. . . . .	49
3.20	(a) Fused image is segmented manually; (b) a, b, c are segmented regions of interest and d is a 3D printed model. . . . .	50
4.1	The framework of the direct dynamic fusion. . . . .	54
4.2	Checkerboard volumes obtained by stitching images after registration. The same positions at the stitches are marked by circles and obvious misalignments are indicated by the red dashed circles in the images. . . . .	63
4.3	Fixed image and deformed moving images. The defects in the volume obtained from the original Demons are indicated by the red dashed cycles. And the white arrow indicates one obvious error in the volume obtained from the diffeomorphic Demons. . . . .	64
4.4	Signed distance and MAD between LA walls segmented from the deformed and fixed images. . . . .	64
4.5	Comparisons of the fused volume of the sequence from Patient # 4. Obvious accumulating errors are marked by red boxes in the images. . . . .	66
4.6	Comparisons of the fused volumes of sequences from Patients # 1, # 2, and # 3. Obvious accumulating errors are marked by red boxes in the images. . . . .	66
4.7	Signed distance from the segmented LA walls obtained from the diffeomorphic Demons and the proposed method to that obtained from the static fusion. . . . .	67
4.8	Comparison of the original single volume (a) and the fused volume (b). . . . .	68
5.1	Direct bundle adjustment (DBA): intensity of panoramic image $\{M(\mathbf{p}_j)\}$ and pose parameters of local frames $\{\xi_i\}$ are optimized simultaneously. . . . .	72
5.2	Sparsity of matrix $H$ : $H_{\xi\xi}$ is block diagonal and $H_{MM}$ is diagonal. The size of $H_{MM}$ is usually much larger than $H_{\xi\xi}$ . The structure of $H_{\xi M}$ and $H_{M\xi}$ indicates the observation relationships between voxels in panoramic image and local frames. . . . .	76

5.3	Accuracy of the sequential fusion method and the proposed DBA method using the data from the Sequence 1 to Sequence 5. (Remark: One-pixel length represents 0.25 mm in practice.) . . . . .	80
5.4	Robustness of the sequential fusion method and the proposed DBA method using Sequence 6 with different intensity noise levels. (Remark: One-pixel length represents 0.25 mm in practice.) . . . . .	82
5.5	Comparison of original single frame of in-vivo 3D TEE image with fused 3D panoramic image using the proposed DBA algorithm. In the last two rows, regions with sharp boundaries are marked by circles for qualitative evaluation of the fused panoramic images. . . . .	83
5.6	Comparison of the aligned images using poses calculated by the DBA and sequential methods. . . . .	84
6.1	Accuracy of DSR method compared to pairwise, Lie normalization, sequential, and APE methods using 5 sequences of simulated 3D TEE images. . .	91
6.2	Comparison of iterative process of DSR and DBA, different colors represent experiments on different in-vivo datasets. . . . .	93
6.3	Comparisons of the aligned images using poses from sequential, APE, and DSR. . . . .	94
6.4	Fused 3D TEE images using registration results from DSR for six in-vivo datasets. LA walls which have sharp structures in the images are indicated by white arrows in selected areas. Colored frames are the boundaries of two registered volumes. . . . .	95
7.1	Schematic diagram of dynamic direct simultaneous registration of multiple 3D images with deformations. . . . .	98
7.2	Comparisons of original local images with reconstructed panoramic images at different phases obtained by D-DSR. . . . .	106
7.3	Evaluation of the reconstructed panoramic images obtained by DDF and D-DSR from the second phase to the fifth phase based on <i>Deform-3</i> from the dataset # 3. The same areas on the images are marked by red boxes for comparison. . . . .	108
7.4	Evaluation of the reconstructed panoramic images obtained by DDF and D-DSR at the sixth phase of each sequence with deformations. The same areas on the images are marked by red boxes for comparison. . . . .	109



# List of Tables

3.1	Details of the Nine In-vivo ECG-gated Datasets . . . . .	44
4.1	Details of the Four In-vivo Datasets . . . . .	60
4.2	Comparisons of Statistical Values of MAD for Pairwise Registration . . . . .	65
4.3	Comparisons of Sequential Fusion Accuracy and the Enlarged FoV after Fusion . . . . .	67
5.1	Ground Truth Poses of six sequences of simulated images . . . . .	79
5.2	Comparison of Intensity Errors between two methods . . . . .	81
5.3	Details of the Five In-vivo ECG-gated Datasets . . . . .	82
6.1	Details of the Six In-vivo ECG-gated Datasets . . . . .	92
6.2	Mean Computational Cost . . . . .	94
7.1	Details of the Four In-vivo Datasets . . . . .	105

# Acronyms & Abbreviations

<b>APE</b>	Accumulated pairwise estimates
<b>AS</b>	Atrial septum
<b>BA</b>	Bundle adjustment
<b>CT</b>	Computed tomography
<b>DBA</b>	Direct bundle adjustment
<b>DDF</b>	Dynamic direct fusion
<b>D-DSR</b>	Dynamic direct simultaneous registration
<b>DoF</b>	Degrees of freedom
<b>DSC</b>	Dice similarity coefficient
<b>DSR</b>	Direct simultaneous registration
<b>ECG</b>	Electrocardiogram
<b>FFD</b>	Free form deformation
<b>FoV</b>	Field of view
<b>GN</b>	Gauss-Newton
<b>GPU</b>	Graphics processing unit
<b>IAS</b>	Interatrial septum
<b>LA</b>	Left atrium

---

<b>LAA</b>	Left atrial appendage
<b>LV</b>	Left ventricle
<b>MAD</b>	Mean absolute distance
<b>MAE</b>	Mean absolute errors
<b>MI</b>	Mutual information
<b>MRI</b>	Magnetic Resonance Imaging
<b>MSE</b>	Mean squared error
<b>NCC</b>	Normalized cross correlation
<b>NLLS</b>	Nonlinear least-squares
<b>PCA</b>	Principal component analysis
<b>SSD</b>	Sum of squared differences
<b>TEE</b>	Transesophageal Echocardiography
<b>TTE</b>	Transthoracic Echocardiography
<b>VO</b>	Visual odometry
<b>WGN</b>	White Gaussian noise

# Chapter 1

## Introduction

Ultrasound imaging, X-ray imaging (radiography, fluoroscopy, and computed tomography (CT)) and magnetic resonance imaging (MRI) are the three mainstream imaging modalities in modern medicine. X-ray imaging has the inherent risk of exposing both the patient and the medical staff to ionizing radiation. MRI is expensive, time-consuming, and may be uncomfortable for some people because it can induce claustrophobia. By comparison, ultrasound imaging is a safe, portable, and cheap tool for the diagnosis of diseases and real-time guidance for surgery or intervention. It combines excellent soft tissue penetration capability with very good spatial resolution. Ultrasound imaging has become the most widely used tool in medical fields such as cardiology, obstetrics, and emergency medicine [1, 2].

Transesophageal echocardiography (TEE) [3] and transthoracic echocardiography (TTE) [4] are two types of tests that use ultrasound to capture images of a patient's heart. In contrast to TTE in which an ultrasound transducer is placed on a patient's chest or abdomen, the ultrasound transducer in TEE is placed into the esophagus or stomach of the patient to obtain the views of the heart. Because the ultrasound transducer inside the esophagus or stomach is closer to the heart than TTE, TEE can provide superior image quality of posterior cardiac structures, such as the left atrium (LA), left atrial appendage (LAA), interatrial septum (IAS), and mitral valve [3]. Moreover, the location of the ultrasound transducer away from the operative field of cardiac surgery and transcatheter procedures

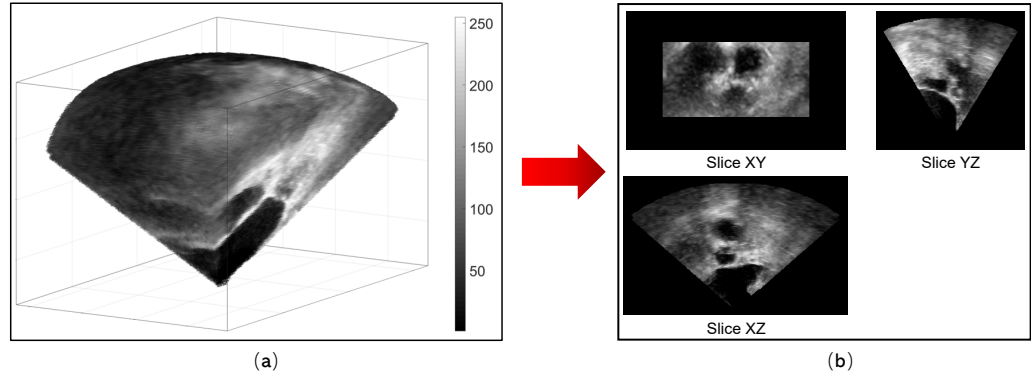


FIGURE 1.1: An example of 3D TEE image: (a) the valid image is within a pyramidal region; (b) the 3D TEE image is displayed via three orthogonal slices<sup>1</sup>.

enables TEE to be used intraoperatively to guide surgery and percutaneous interventions [3].

Thanks to state-of-the-art 2D array transducers, 3D TEE images can be acquired in real-time [5]. An example of 3D TEE image is shown in Fig. 1.1. Although the valid image is within a pyramidal region due to the characteristics of ultrasound imaging [5], an imaging system assigns the intensity value of voxels outside the pyramidal field of view (FoV) to zero and outputs a cuboid 3D TEE image finally. Compared to 2D TEE, the major advantages of 3D TEE imaging include: (1) the improvement of accuracy of echocardiography in evaluating cardiac chamber volumes; (2) the realistic comprehensive views of cardiac structures such as the valves and congenital abnormalities; (3) real-time guidance in intraoperative settings [3, 6]. These advantages have made 3D TEE imaging increasingly used for cardiac imaging in the past decade. However, current 3D TEE still suffers from a narrow FoV due to the manufacturing limitations on 2D array transducers [2], which makes it difficult to inspect complete regions of interest in a single volume. To complete a comprehensive imaging examination, the routine method requires the clinician to move the transducer to obtain multiview images [7] and combine these images through their imagination, which heavily depends on clinicians' knowledge and experience.

Image registration and fusion is an active research topic<sup>2</sup> in the field of medical image

<sup>1</sup>In this thesis, 3D images are usually presented via three orthogonal slices to display details of regions of interest.

<sup>2</sup>Although in many works of literature, registration and fusion are studied separately, we regard them as two steps of one fusion task since our objective is multiview image fusion.

computing, which aims to combine images captured from different viewpoints automatically [8]. Registration that determines the accuracy of anatomical structures in the fused image is a critical step prior to the fusion. The core of medical image registration is to find transformations that can relate corresponding anatomical or functional positions between two or more images [9]. Although registration and fusion of multiview images are widely used for enlarging the FoV in the computer vision community, there are some challenges when dealing with 3D TEE images, such as lack of distinct features, large cardiac motion, and high computational complexity caused by a large number of voxels. Therefore, this study is conducted to enlarge the FoV of 3D TEE images and overcome these challenges.

## 1.1 Motivation

LAA is an small ear-shaped structure attached to the LA body (see Fig. 1.2(a)) [10, 11]. LAA has a highly variable shape and morphology between patients. In patients with nonvalvular atrial fibrillation, studies have shown that over 90% of stroke-causing clots from the heart are formed in the LAA [10, 12]. Anticoagulation therapy is the most common treatment for these patients to reduce atrial fibrillation stroke risk, but with potential bleeding complications and tolerance issues [11]. In these cases, percutaneous LAA occlusion is a clinically recommended alternative treatment for nonvalvular atrial fibrillation patients who present contraindications for oral anticoagulants and high stroke risk [10].

The LAA occlusion is a minimally invasive cardiac intervention in which a foldable device is percutaneously delivered through a transseptal path for obstructing the LAA and reducing the stroke risk in patients with nonvalvular atrial fibrillation [11] (see Fig. 1.2(a)). Due to the highly variable anatomical structure of LAA, there are various occluding devices produced by different manufacturers [11]. Therefore, before the intervention, clinical measurements (see Fig. 1.2(b)) need to be performed to select a suitable device for the patient. Correct device sizing is one of the most challenging tasks in preoperative planning, which has an immediate impact on the treatment efficiency, procedural time, and

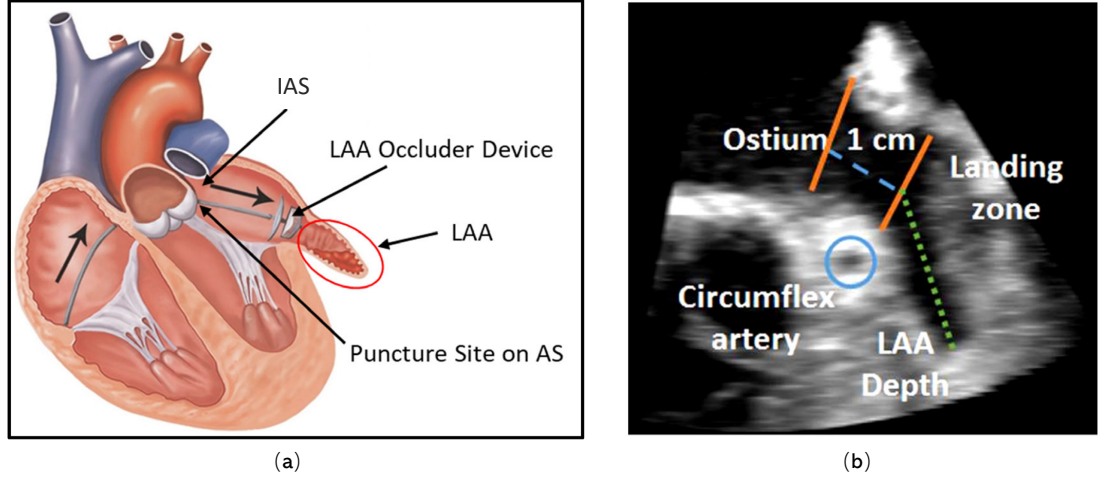


FIGURE 1.2: (a) Schematic diagram of LAA occlusion; (b) Clinical measurement for device size selection of LAA occluder. (a) and (b) are from [13] and [14], respectively.

number of complications. However, due to the limited **FoV** of 3D **TEE**, it is often difficult to inspect a complete **LAA** in a single volume, which makes it hard to complete the measurements for the device size selection.

In addition, in the planning of most cardiac surgery and intervention, it is crucial for cardiac structures of interest to be visualized in a single **FoV** for a complete analysis of the spatial orientation of these structures. For **LAA** occlusion, the angle of approach of the device deployment system is partially dependent on the spatial relationship between the puncture site on the **IAS** and the left heart structures (e.g. **LAA**, mitral valve), and may have an impact upon how a device ‘sits’ in the target structures during and after deployment [15]. Limited by the **FoV** of standard 3D **TEE**, the **IAS**, **LAA**, and the mitral valve can hardly be imaged in their entirety as a single imaging volume.

## 1.2 Research Aims

This research aims at developing innovative multiview registration and fusion algorithms to enlarge the **FoV** of 3D medical images. Although the proposed algorithms could be applied to a wide variety of image types, in this thesis, we mainly focus on 3D **TEE** images with potential applications to assist clinicians’ diagnosis and surgical planning for patients with various cardiac diseases.

We aim at solving some major challenges in multiview registration and fusion for 3D TEE images. Firstly, 3D ultrasound images lack distinct features such that commonly used feature-based methods tend to fail when handling 3D TEE images. Secondly, it is difficult for existing non-rigid registration methods to deal with large cardiac motion in 3D TEE images. Thirdly, 3D TEE images usually contain a large number of voxels which requires a lot of computational resources and time for registration and fusion. Fourthly, there are limited algorithms available for simultaneous registration and they may have problems with optimization, information reuse, and very high computational cost, especially for 3D images [16, 17]. While bundle adjustment (BA) [18] framework has been regarded as a gold standard for simultaneous pose<sup>3</sup> estimation and 3D scene reconstruction when fast performance is not required, almost all of BA methods are feature-based and the goal of them is to refine the positions of scene points and poses of imaging sensors based on 2D images [19, 20] instead of image registration and fusion. To our best knowledge, there is no BA method that can be used for multiple 3D image registration and fusion.

### 1.3 Main Contributions

The main contributions of this thesis are the five frameworks proposed to solve the aforementioned challenges and enlarge the FoV of 3D TEE images by registering and fusing multiview images. The research starts with achieving efficient pairwise registration and fusion, followed by the novel simultaneous registration methods for improving accuracy and reducing computational complexity. Both the rigid and deformable scenarios are considered in pairwise and simultaneous registration and fusion. To circumvent the process of feature extraction and matching, direct (intensity-based) methods are considered in this thesis. In contrast to the classical feature-based BA, we proposed the first intensity-based BA framework named direct bundle adjustment (DBA) that can complete the 3D TEE image registration and fusion simultaneously. Furthermore, the DBA framework is extended to solve dynamic simultaneous registration and fusion problems. Detailed experiments on in-vivo 3D TEE images demonstrate the reconstruction ability and the potential clinical

---

<sup>3</sup>In this thesis, the words ‘transformation’ and ‘pose’ are used interchangeably when we describe the rigid transformation of an image or imaging sensor.



value of our methods to be used for reconstructing the regions of interest of the heart and assist LAA occlusion.

In summary, our main contributions to the 3D image registration and fusion, especially for 3D TEE images include:

- A novel multiview registration and fusion framework to enlarge the FoV of ECG-gated<sup>4</sup> 3D TEE images efficiently.
- A novel dynamic direct fusion (DDF) framework to extend the FoV of 3D TEE images at different cardiac phases sequentially, avoiding the bias toward the specifically chosen phase when ECG-gating is used.
- A novel direct bundle adjustment (DBA) framework to solve multiview 3D image registration and fusion simultaneously in rigid scenarios. It redefines classical BA by optimizing the intensities of a predefined panoramic (global) image (instead of positions of scene points in classical BA) and the poses of the local images simultaneously. Without any reference image, correspondences, or information loss or reuse, the proposed DBA algorithm is an elegant way to obtain the optimal panoramic image and poses of local 3D images in one go. To our best knowledge, this is the first work of BA based on the direct method for 3D image registration and fusion.
- A novel algorithm named direct simultaneous registration (DSR) to solve multiview 3D image registration simultaneously in rigid scenarios. DSR solves the poses only without considering the intensities of the panorama but is proved to be able to obtain the same optimal poses as DBA. The proposed method is potential to reduce the computational complexity of multiview simultaneous registration and fusion for 3D images.
- A novel dynamic direct simultaneous registration (D-DSR) framework used in deformable scenarios to estimate the 3D TEE panoramic images at all times/phases corresponding to the collected local images simultaneously.

---

<sup>4</sup>ECG-gating is a technique that extracts images of the heart at a specified phase in different cardiac cycles by synchronizing 3D TEE images to the electrocardiogram (ECG) signal. In the obtained ECG-gated 3D TEE images, deformations are eliminated.

## 1.4 Publications

The summary of research outcomes in this study is listed as follows. Paper 1, 2, 3, 4, and 6 are corresponding to the proposed five frameworks that are elaborated in Chapter 3 to 7, respectively.

- (1). **Zhehua Mao**, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee, “Direct 3D Ultrasound Fusion for Transesophageal Echocardiography”, *Computers in Biology and Medicine*, vol. 134, no. 104502, p. 104502, 2021. ([Q1] IF = 6.698)
- (2). **Zhehua Mao**, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee, “Complete Region of Interest Reconstruction by Fusing Multiview Deformable Three-Dimensional Transesophageal Echocardiography Images”, *Medical Physics*, 2022. ([Q1] IF = 4.506)
- (3). **Zhehua Mao**, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee, “Direct Bundle Adjustment for 3D Image Fusion with Application to Transesophageal Echocardiography”, in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021. (Accept rate: 45%)
- (4). **Zhehua Mao**, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee, “DSR: Direct Simultaneous Registration for Multiple 3D Images”, in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, Cham, 2022. (Accept rate: 31%)
- (5). Liang Zhao, **Zhehua Mao**, and Shoudong Huang, “Feature-Based SLAM: Why Simultaneous Localisation and Mapping?”, in *Robotics: Science and Systems (RSS)*, 2021. (Accept rate: 27%)
- (6). **Zhehua Mao**, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee, “D-DSR: Dynamic Direct Simultaneous Registration for Multiple 3D Images”, in *preparation(to be submitted to IEEE Robotics and Automation Letters)*
- (7). **Zhehua Mao**, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee, “DSR: Direct Simultaneous Registration for Multiple 3D Images with Application

to 3D Transesophageal Echocardiography ”, *in preparation (to be submitted to IEEE Transactions on Medical Imaging)*

## 1.5 Thesis Outline

The details of the proposed five frameworks are presented in Chapter 3 to 7, which are divided into two parts. The structure of this thesis is as follows.

- Chapter 2: We provide an overview of the state-of-the-art medical image registration methods from three perspectives, namely feature-based and intensity-based registration methods, rigid and deformable registration methods, and pairwise and simultaneous registration methods.

**Part I - Pairwise Registration and Sequential Fusion:** in Chapter 3 and 4, we propose two frameworks that combine pairwise registration and sequential fusion strategy to fuse multiview 3D TEE images in rigid and deformable scenarios, respectively.

- Chapter 3: A novel multiview registration and fusion framework is proposed to enlarge the FoV of 3D TEE images efficiently. Cardiac motion (deformation of the heart) is avoided in the obtained in-vivo 3D TEE images by using the ECG-gating technique. A direct method is proposed to solve the registration problem in the Lie algebra space and its fast implementation is realized by searching voxels on three orthogonal planes between two volumes. Besides, a 3D fusion method is proposed to fuse the aligned images seamlessly and efficiently. For multiple 3D TEE images, a sequential fusion strategy is used to obtain the globally consistent image and poses. Simulated and in-vivo experiments show the proposed registration algorithm outperforms state-of-the-art registration methods in terms of accuracy and efficiency and the framework enlarges the FoV of 3D TEE images effectively.
- Chapter 4: We propose a novel direct dynamic fusion (DDF) framework to extend the FoV of 3D TEE images at different cardiac phases sequentially, avoiding the bias toward the specifically chosen phase when ECG-gating is used. The deformation field

for the registration is defined by an embedded deformation graph and estimated by a direct method. The accuracy of the proposed method is evaluated by comparing it with two B-spline-based methods, two Demons-based methods, and one learning-based method VoxelMorph using in-vivo experiments. Without selecting the static (ECG-gated) images from the same cardiac phase, this work addressed the problem of limited FoV of 3D TEE images in the deformable scenario, obtaining the fused images with high accuracy and good quality.

**Part 2 - Simultaneous Registration and Fusion:** in Chapter 5 to 7, we propose three simultaneous registration and fusion algorithms to further improve the accuracy in the rigid and deformable scenarios.

- Chapter 5: We propose a novel framework named direct bundle adjustment (DBA) that registers a collection of 3D images in a simultaneous fashion without specifying any reference image, feature extraction and matching, or information loss or reuse, used in the rigid scenario. Different from Chapter 3 that deduces the global transformations of local images from the results of pairwise registration, DBA simultaneously optimizes the global transformation parameters of local images and the intensity values of a predefined panoramic image using the 3D image data directly. Compared with conventional two-step methods for image fusion, DBA completes registration and fusion in one go. Simulated and in-vivo experiments on 3D TEE images show that the proposed method is robust to intensity noises and much more accurate than the sequential fusion method in Chapter 3.
- Chapter 6: When investigating DBA further, we found a very interesting property and have theoretically proved that when solving DBA with Gauss-Newton (GN) iterations, the optimization of poses is independent of the intensities of the panoramic image. Based on the finding, a novel algorithm named direct simultaneous registration (DSR) is proposed, which solves the poses only instead of solving full DBA problem but can obtain the same optimal poses as DBA. The proposed method is potential to reduce the computational complexity of simultaneous registration and fusion for 3D images. DSR is evaluated by comparing it with four widely used methods via simulated and in-vivo 3D TEE images. It is shown that the proposed method

outperforms these methods in terms of accuracy and is much more efficient than the state-of-the-art accumulated pairwise estimates (APE).

- Chapter 7: Inspired by the DBA and DSR, we propose a novel dynamic direct simultaneous registration (D-DSR) framework in the deformable scenario to estimate the 3D TEE panoramic images at all phases corresponding to the collected local images simultaneously. Similar to the DSR framework, we first formulate our problem as a dynamic direct bundle adjustment problem by substituting the rigid transformations in DBA with embedded deformation graphs. Secondly, we prove a similar property as DBA that the estimation of embedded deformation graphs is independent of the intensities of the panorama during the optimization process and propose the D-DSR algorithm. Detailed validation of the proposed D-DSR algorithm is performed on the in-vivo 3D TEE images.
- Chapter 8: We summarize the thesis and present the potential future research directions.

## Chapter 2

# Related Works

Medical image registration and fusion is a fundamental task for many medical image analysis problems in which valuable information conveyed by two or more images needs to be combined and examined [9, 17]. In general, registration and fusion are two sequential steps that fuse multiple local images together to get one global (panoramic) image with enlarged FoV. Medical image registration aims at finding transformations that map any anatomical or functional positions from one image to its correspondences in other images [21]. Once local images are registered w.r.t. each other using corresponding transformations, we can fuse all these aligned images together to get a panoramic image. Most studies focus on registration since it determines the accuracy of anatomical features in the fused images, which is critical for various clinical applications such as computer-assisted diagnosis, therapy, and surgery [21]. In recent decades, mainstream medical imaging techniques, such as CT, MRI, and ultrasound, have evolved from 2D to 3D, which proposes new challenges to medical image registration, such as feature extraction and high computational complexity [22].

In order to avoid confusion and to allow readers to have a sufficient amount of knowledge about the 3D medical image registration and fusion problem, herein we provide an overview of medical image registration methods from three perspectives, i.e. feature-based and intensity-based (direct) registration methods, rigid and deformable registration methods, and pairwise and simultaneous registration methods.

## 2.1 Feature-based vs. Intensity-based Registration

According to whether registration methods extract features from images or not, they can generally be divided into two categories: feature-based methods [22, 23] and intensity-based methods [24–26]. Feature-based methods decouple the transformation estimation problem into two sequential steps: firstly, distinct features are extracted from images and matched to find the correspondences between the images [27–29]; secondly, transformations are estimated by only using these features, discarding all other information of the images [30]. Although feature-based methods simplify the overall problem, the accuracy of the transformation estimation might be also compromised due to the loss of much valuable information. In addition, a 3D image like Fig. 1.1(a) contains millions of valid voxels. Feature extraction and matching from such a large number of voxels take much time and fast implementations of the feature-based methods can only be realized by using accelerated frameworks on graphics processing units (GPUs) [22]. Furthermore, classical algorithms for feature extraction and matching such as SIFT [27], SURF [28], and ORB [29] are designed for 2D images. Developing robust feature extraction algorithms for 3D images is still challenging [31], especially for 3D TEE images that lack distinct features.

On the contrary, intensity-based (direct) methods use the intensity information of images directly without feature extraction and matching [32]. They are well-known for their high accuracy and robustness in environments with few features, which is very suitable for ultrasound image registration. Direct methods estimate the transformation by maximizing the similarity between the images. Commonly used similarity metrics include the sum of squared differences (SSD), normalized cross correlation (NCC), and mutual information (MI) [33–35]. SSD and NCC are usually used for monomodal registration, and MI is commonly used for multimodal registration [21].

Let  $I_F$  and  $I_M$  be two images,  $\mathbf{p}_j$  be a point in image  $I_M$  with intensity  $I_M(\mathbf{p}_j)$ , and  $\tilde{\mathbf{p}}_j$  be the corresponding point in image  $I_F$  with intensity  $I_F(\tilde{\mathbf{p}}_j)$ , we aim at registering the image  $I_M$  to the image  $I_F$ . Suppose the transformation from  $I_M$  to  $I_F$  can be parameterized by a function  $f_T(\cdot)$ , then the spacial relationship between  $\mathbf{p}_j$  and  $\tilde{\mathbf{p}}_j$  can be established by

$\tilde{\mathbf{p}}_j = f_T(\mathbf{p}_j)$ . The SSD similarity [32] can be defined as

$$\text{SSD} = \sum_{\mathbf{p}_j \in \Omega} (I_M(\mathbf{p}_j) - I_F(f_T(\mathbf{p}_j)))^2, \quad (2.1)$$

and the NCC similarity [21] is defined as

$$\text{NCC} = \frac{\sum_{\mathbf{p}_j \in \Omega} (I_M(\mathbf{p}_j) - \bar{I}_M)(I_F(f_T(\mathbf{p}_j)) - \bar{I}_F)}{\sqrt{\sum_{\mathbf{p}_j \in \Omega} (I_M(\mathbf{p}_j) - \bar{I}_M)^2 \sum_{\mathbf{p}_j \in \Omega} (I_F(f_T(\mathbf{p}_j)) - \bar{I}_F)^2}}, \quad (2.2)$$

where  $\Omega$  is the overlapping area between  $I_F$  and  $I_M$ ;  $\bar{I}_F$  and  $\bar{I}_M$  are intensity averages of the images  $I_F$  and  $I_M$ , respectively.

The MI similarity is designed based on the Shannon–Wiener entropies of images [21]. For an image  $I_F$ , its entropy is defined as

$$\mathcal{H}(I_F) = - \sum_{a \in I_F} p(a) \log p(a), \quad (2.3)$$

where  $p(a)$  is the probability that the intensity of a voxel in image  $I_F$  is  $a$ . Similarly, the entropy of an image  $I_M$  is

$$\mathcal{H}(I_M) = - \sum_{b \in I_M} p(b) \log p(b), \quad (2.4)$$

where  $p(b)$  is the probability that the intensity of a voxel in image  $I_M$  is  $b$ . Furthermore, a joint entropy of  $I_F$  and  $I_M$  over the overlapping area can be defined as

$$\mathcal{H}(I_F, I_M) = - \sum_{a \in I_F} \sum_{b \in I_M} p(a, b) \log p(a, b), \quad (2.5)$$

where  $p(a, b)$  is the joint probability that the intensities of a voxel in the overlapping area of  $I_F$  and  $I_M$  are  $a$  and  $b$ . Then the MI similarity can be defined by

$$\text{MI} = \mathcal{H}(I_F) + \mathcal{H}(I_M) - \mathcal{H}(I_F, I_M). \quad (2.6)$$

In addition to these conventional metrics, Grau et al. [36] proposed a phase-based method to estimate the poses between ultrasound images. To solve the problem that the phase-based method is sensitive to the initial value, Housden et al. [37] proposed to involve an



X-ray tracking system in the registration process. But this strategy not only makes the system complex but also leads to the latent risks of ionizing radiation. Recently, Peressutti et al. [38] proposed a novel subspace error metric for registration of multiview 3D+t ultrasound images. The method is based on the principal component analysis (PCA), which estimates the poses of image sequences in a low-dimensional space. It is reported that the method outperforms the conventional intensity-based methods and the phase-based method in terms of accuracy, robustness, and computational efficiency. Since intensity-based methods usually involve a large number of voxels of 3D images, fast implementation usually still needs to use accelerated frameworks on GPUs [39].

## 2.2 Rigid vs. Deformable Registration

We can also categorize registration methods according to the number of degrees of freedom (DoF) that transformation models can describe, ranging from simple transformations such as rigid or affine ones to complex transformations such as deformable ones [21].

### 2.2.1 Rigid and Affine Transformation

In medical image registration, a rigid transformation is commonly used when the anatomical structures keep the same shape and size in multiview images. In a 3D space, a rigid transformation has six DoF, i.e. three translations and three rotations. If we use homogeneous coordinates, it can be expressed in Euclidean space  $SE(3)$  by

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \in SE(3), \quad (2.7)$$

where  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  and  $\mathbf{t} \in \mathbb{R}^3$  denote rotation matrix and translation vector, respectively. A rotation matrix in 3D space is orthogonal and subject to six constraints, i.e. each column should be a unit length and should be orthogonal to one another.

In medical image registration, affine transformation is another widely used transformation model when shears or/and scaling exist between the obtained images. In a 3D space,

an affine transformation as shown below has twelve [DoF](#) which represent translations, rotations, shears, and scaling:

$$\mathbf{T}_A = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \in \mathbb{R}^{4 \times 4}. \quad (2.8)$$

Different from the rigid transformation  $\mathbf{T}$ , an affine transformation only requires  $\mathbf{A} \in \mathbb{R}^{3 \times 3}$  to be an invertible matrix, not necessarily orthogonal. Affine transformation is also called orthogonal projection.

In the clinic, echocardiography examination can be assisted by the [ECG](#)-gating technique to extract images of the heart at a specified phase in different cardiac cycles, typically during diastole when the heart is moving the least [40]. Therefore, most of the deformation of the heart can be eliminated in these [ECG](#)-gated images. In these cases, image registration can be simplified to rigid registration problems.

## 2.2.2 Deformable Transformation

There is another group of registration methods that consider much more [DoF](#) in the transformation than rigid/affine ones to take the deformation in images into account, which is called deformable (non-rigid) registration [41]. Deformable registration has been a popular topic in recent decades due to its great potential for assisting motion correction, longitudinal studies, multimodality information fusion, and population studies [41].

### 2.2.2.1 Free-Form Deformation

**B-Spline-based FFD:** Free form deformation ([FFD](#)) [42] is one of the most widely used transformation models in non-rigid registration. The idea of [FFD](#) is to embed the moving image into a space and then deform it as the space is deformed. The [FFD](#) model has achieved great success in the medical image computing field since it was coupled with B-splines [43]. These registration methods use B-spline (usually cubic B-splines) curves that are controlled by a sparse set of points to define a continuous deformation field. Then, a moving image can be registered to the fixed image by the deformation field [44].

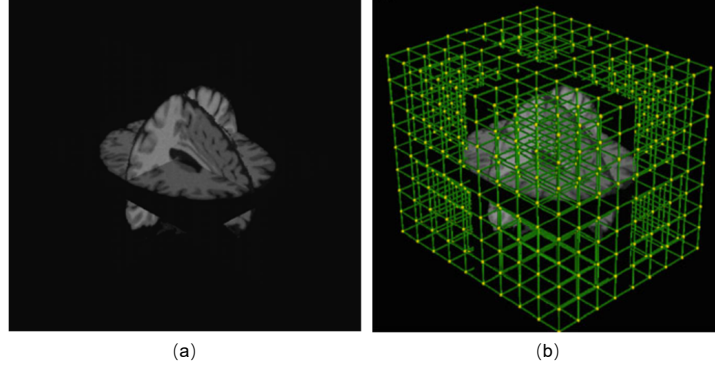


FIGURE 2.1: B-spline based free-form deformation. A three-orthogonal view of a 3D image shown without control point grid (a) and with control point grid (b). Figures are from [21].

In Fig. 2.1(a), there is a 3D image displayed by three orthogonal planes and  $\mathbf{p} = [x, y, z]^\top$  is one point position in the image. Suppose  $\mathcal{N}_c$  represents a grid of control points (see Fig. 2.1(b)) with uniform control point spacing  $\delta$ , and  $\phi_{i,j,k}$  is one control point, B-spline-based FFD can be described as a tensor product of three one-dimensional cubic B-splines:

$$D(\mathbf{p}) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 \beta_l(u) \beta_m(v) \beta_n(w) \phi_{i+l, j+m, k+n} \quad (2.9)$$

where  $i = \lfloor x/\delta \rfloor - 1, j = \lfloor y/\delta \rfloor - 1, k = \lfloor z/\delta \rfloor - 1, u = x/\delta - \lfloor x/\delta \rfloor, v = y/\delta - \lfloor y/\delta \rfloor, w = z/\delta - \lfloor z/\delta \rfloor$  ( $\lfloor \cdot \rfloor$  denotes the floor of a value) and  $\beta_l$  represents the  $l$ -th basis function of the B-spline:

$$\begin{aligned} \beta_0(u) &= (1 - u)^3 / 6 \\ \beta_1(u) &= (3u^3 - 6u^2 + 4) / 6 \\ \beta_2(u) &= (-3u^3 + 3u^2 + 3u + 1) / 6 \\ \beta_3(u) &= u^3 / 6 \end{aligned}$$

B-splines have local supports, which means local deformation can be calculated from only a couple of surrounding control points [41, 45]. This property is beneficial for local deformation modeling and fast implementation of the B-spline based methods [41, 46, 47].

**Embedded Deformation:** Based on the FFD model, Sumner et al. [48] proposed an embedded deformation method in the computer graphics community, which is notable due to its simplicity, versatility, and efficiency. Embedded deformation defines the deformation

field with a collection of affine transformations organized in a graph structure and estimates the deformation graph making use of the specified corresponding landmarks of the embedded object before and after the deformation. Song et al. [49] adopted the method to reconstruct the surface of soft tissues from 2D laparoscopy images captured with different deformations. Extracted and matched feature points between moving and fixed images are used to estimate the deformation in their work. However, such a method relies on the distinct features in images and discards all other information, which tends to fail when handling images with few distinct features such as 3D TEE images [9, 41].

### 2.2.2.2 Demons-based Registration

Demons-based methods are another group of well-known non-rigid registration methods. Different from the transformation models that are controlled by a set of parameters such as rigid, affine, and B-spline based FFD, Demons-based registration models are usually non-parametric, in which a dense displacement field that describes the deformation of each voxel is computed.

Inspired by Maxwell's Demons and optical flow equations, Thirion proposed the original Demons algorithm [50] that considers image registration as a diffusion process and computes the deformation field through iterating between the computation of the Demons forces and a Gaussian smoothing regularization. Again, suppose that  $I_F$  and  $I_M$  are two images,  $\mathbf{p}_j$  is a point in image  $I_M$  with intensity  $I_M(\mathbf{p}_j)$ , and  $\tilde{\mathbf{p}}_j$  is the corresponding point in image  $I_F$  with intensity  $I_F(\tilde{\mathbf{p}}_j)$ , we want to register  $I_M$  to  $I_F$ . In the original Demons algorithm, the estimated displacement  $u(\mathbf{p}_j) \in \mathbb{R}^3$  for  $\mathbf{p}_j$  at the current iteration is:

$$u(\mathbf{p}_j) = (I_M(\mathbf{p}_j) - I_F(\tilde{\mathbf{p}}_j)) \frac{\nabla I_F(\tilde{\mathbf{p}}_j)}{\|\nabla I_F(\tilde{\mathbf{p}}_j)\|^2 + (I_M(\mathbf{p}_j) - I_F(\tilde{\mathbf{p}}_j))^2}, \quad (2.10)$$

where  $\nabla I_F$  is the intensity gradient of the image  $I_F$ . In Demons-based methods,  $\nabla I_F(\tilde{\mathbf{p}}_j)$  is called internal force originating from the image  $I_F$ , which represents the relationship between the neighboring points in the image  $I_F$ .  $(I_M(\mathbf{p}_j) - I_F(\tilde{\mathbf{p}}_j))$  is called external force, which represents the difference between the images  $I_M$  and  $I_F$ . Since (2.10) computes the displacement based on local image information only, the Gaussian smoothing regularization

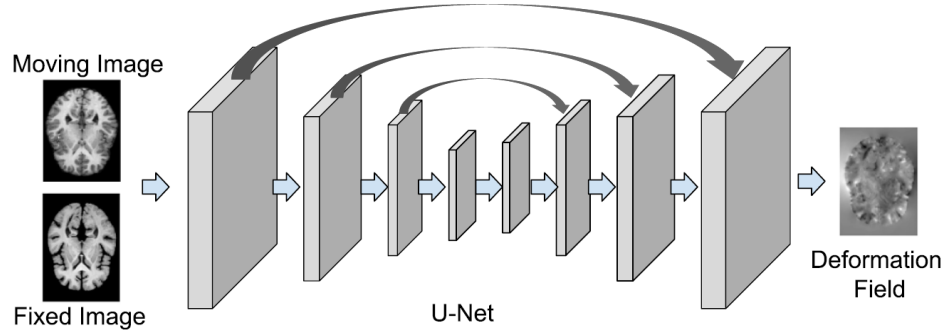


FIGURE 2.2: An example of a U-Net-based registration framework for MRI images. Figure is from [59].

is an essential step for suppressing noise and preserving the geometric continuity of the deformed image [51].

Following Thirion’s Demons method, a range of variants were proposed to improve convergence speed and precision [52]. Among these, Vercauteren et al. [53] proposed to search for the displacement  $u(\mathbf{p}_j)$  in the Lie algebra space and then map it to the space of diffeomorphisms through the exponential map to provide diffeomorphic transformations. The diffeomorphic Demons algorithm [53] that can preserve the topology of the objects achieves remarkable success in various computational anatomy studies, such as breast CT [54], brain MRI [55], and lung CT [56]. Since the optical flow model is less constrained than B-splines and the deformation field is estimated at every voxel which results in an extremely large number of DoF [57], Demons-based methods can deal with larger local deformation in the images than B-spline based FFD methods [52, 58].

### 2.2.2.3 Learning-based Registration

In the past decade, there has been a significant increase in the application of learning-based techniques in medical image computing and analysis [59, 60]. While learning-based medical image registration methods have yet to achieve notable advancements in terms of registration accuracy in comparison to conventional methods, they have attracted a lot of attention because of their high efficiency after training [59, 60], with the support of high-performance computational hardware such as GPUs. In the field of learning-based image registration, encoder-decoder convolutional neural network (CNN) [61], spatial transformer

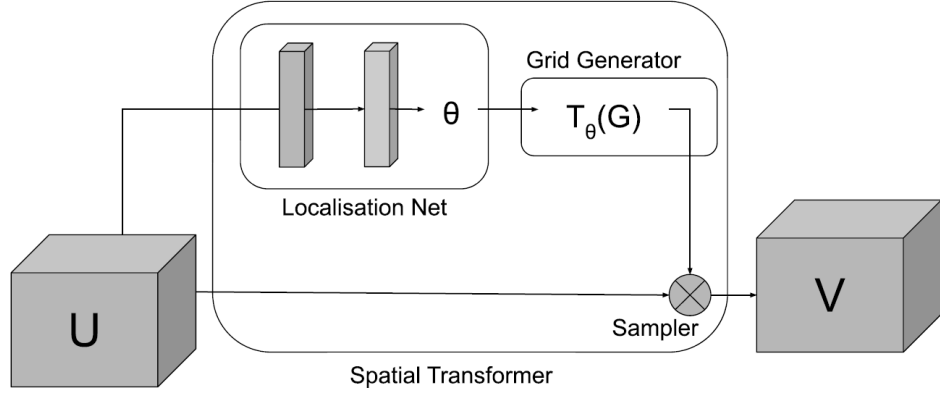


FIGURE 2.3: A schematic diagram of STN.  $U$ ,  $V$ , and  $\theta$  denote input, output, and learned transformation parameters, respectively. Figure is from [59].

network (STN) [62], and generative adversarial network (GAN) [63] are three widely used and fundamental components of medical image registration networks.

The success of machine learning in computer vision is closely linked to the use of CNNs. This type of network utilizes successive convolution layers to automatically learn multi-scale features for specific tasks [61]. For medical image registration, U-Net [64] is one of the most widely used CNN architectures. A classic U-Net network is shown in Fig. 2.2. It is an encoder-decoder network with skip connections between the encoding and decoding layers. The U-Net employs several down-sampling layers of the encoder and un-sampling layers of the decoder to learn features of input images at varying resolutions [64]. Given its flexibility, U-Net is widely used in most state-of-the-art medical image registration frameworks [65–67].

In addition, STN [62] is another key component of most learning-based image registration methods. This type of network learns to spatially transform feature maps in a way that is beneficial to particular tasks. As shown in Fig. 2.3, STN consists of a localization network, a grid generator, and a sampler. The localization network learns the transformation parameters from the input. Then, the grid generator uses the learned transformation parameters to generate a transformation grid. The sampler uses bilinear sampling to sample coordinates from the input to the output. The entire network is differentiable, enabling automatic parameter updates during back-propagation. Since the networks learn features

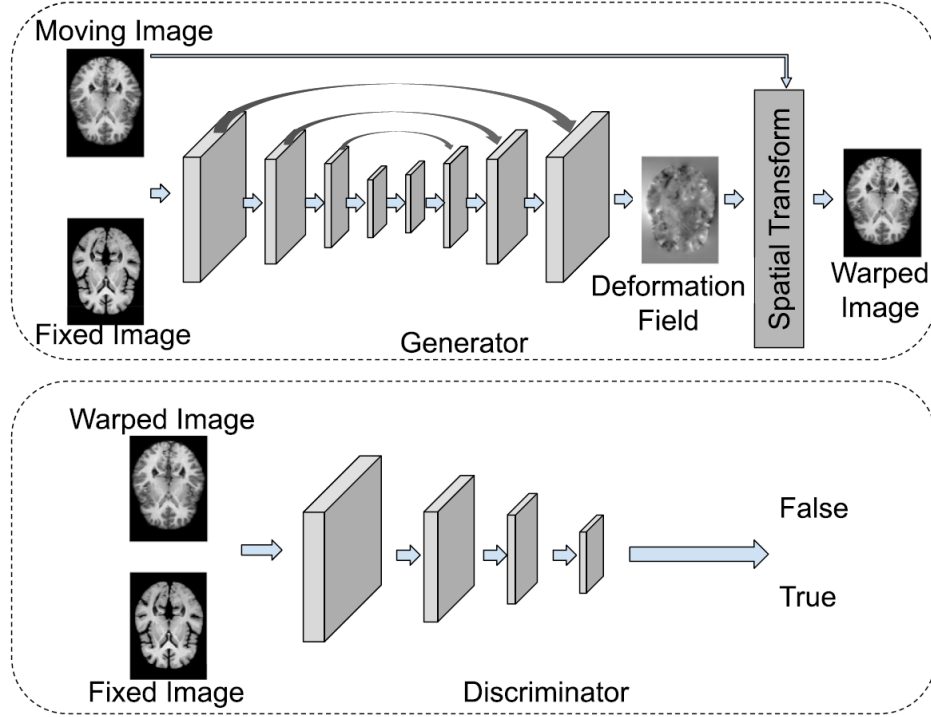


FIGURE 2.4: An example of a GAN-based registration framework for MRI images. The frameworks combines U-Net, STN, and GAN. Figure is from [59].

in a way that is invariant to rigid and non-rigid transformations, they serve as the basis for most unsupervised registration algorithms [68–70].

Furthermore, GAN [63] is also a commonly used component of learning-based registration methods. GANs are the most widely used generative model for data augmentation, image synthesis, and image-to-image translation. A typical GAN is comprised of two CNN sub-networks which are called generator and discriminator [63], respectively. The generator is used to sample from the data distribution and can be used for synthesizing new data. The discriminator is used to discriminate synthesized samples and real samples to compete with the generator [63]. An example of a GAN-based image registration framework is shown in Fig. 2.4. When a pair of fixed and moving images are input to the network, the generator network will predict the warped moving image first and then, the discriminator will evaluate the similarity between the warped image and the fixed image. Since the discriminator of GANs provides a novel learnable method to evaluate the similarity between two images, they have the potential to create learnable and adaptable similarity metrics, which makes it possible for multimodal image registration [71, 72].

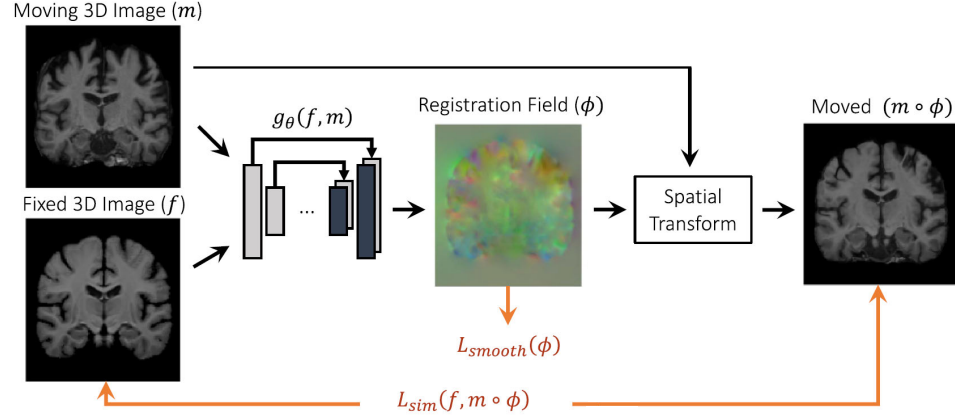


FIGURE 2.5: Overview of the VoxelMorph. Figure is from [74].

Based on the above three fundamental networks, a variety of supervised and unsupervised networks are proposed in recent years. While supervised deep learning methods could allow for fast and robust registration, they either use the results from non-learning-based methods, the labels created by experts, or the simulated datasets with known deformation [59, 60, 65, 73] for training the model because ground-truth deformation is usually not available in real datasets. Therefore, these methods are not only cumbersome for training but also constrained by the accuracy of the non-learning-based methods and experience of the experts, and may result in implausible or over-simplified deformation [59].

In [74], an unsupervised learning framework named VoxelMorph is proposed for deformable medical image registration. Fig. 2.5 is an overview of VoxelMorph which combines U-Net with STN for 3D brain MRI image registration. It optimizes parameters  $\theta$  for a parameterized registration function  $g_{\theta}(\cdot, \cdot)$  by only using the collected 3D images. The model learns a common representation that enables the alignment of a new pair of volumes from the same distribution [74]. The state-of-the-art learning-based method VoxelMorph has shown comparable accuracy to the non-learning-based methods [74].

While the learning-based methods have computational advantages in the registration step, they have not demonstrated superior accuracy to traditional non-learning-based methods [59, 60, 74]. In addition, learning-based methods usually require much time for training. When processing images from different regions of interest, different patients, or different imaging modalities, learning-based frameworks usually need to be trained separately [59, 60].



## 2.3 Pairwise vs. Simultaneous Registration

Most of the registration methods such as B-spline based FFD [46, 47], embedded deformation [48], Demons-based registration [53], and VoxelMorph [74] are developed for pairwise registration. The process involves designating one image as the fixed (or reference) image and registering the other image (moving image) to the fixed one, such as aforementioned registration from  $I_M$  to  $I_F$  in Section 2.1 and 2.2. When more than two images need to be fused, most of the existing methods obtain the panoramic image based on results from pairwise registration such as [36, 75, 76]. The advantage of these methods is that they simplify the whole multiview registration and fusion problem into several pairwise registration and fusion tasks, which can effectively reduce the computational cost and are appropriate for the scenarios where fast performance is preferred [49].

The other group of registration methods is simultaneous registration which estimates the transformations of all local images at the same time. Compared with pairwise registration, simultaneous registration is still an open problem, especially for 3D images. In [77], a framework called congealing is proposed, which uses underlying entropic information of images for alignment. A large number of images are necessary for congealing because the estimation is done with the information at one location at a time [17]. Therefore, it is not suitable for situations where the number of images is limited such as 3D TEE that only captures around six to ten images for each patient. And as [16] pointed out, employing entropy for congealing is problematic due to its poor optimization characteristics. Simultaneous registration methods usually have advantages in accuracy but are commonly used in offline scenarios where fast performance is not required.

### 2.3.1 Accumulated Pairwise Estimates

Recently in [17], an accumulated pairwise estimates (APE) method is proposed for simultaneous registration. Let  $\mathbf{x}_T = \{\mathbf{T}_1, \dots, \mathbf{T}_n\}$  ( $\mathbf{T}_i \in SE(3)$ ) denote the global transformations that map the points from each of the image spaces to the joint image space. APE

solves the simultaneous registration problem by minimizing the following energy function:

$$E(\mathbf{x}_T) = \sum_{i \neq j} F_{i,j}(\mathbf{x}_T) = \sum_{i \neq j} \frac{1}{2} \|f_{i,j}(\mathbf{x}_T)\|^2, \quad (2.11)$$

where  $f_{i,j}$  represents the pairwise similarity measure between local images  $i$  and  $j$ , which can be designated as [SSD](#), [NCC](#), [MI](#), et al.

From (2.11), we can see that [APE](#) may have a problem with information reuse since the overlapping areas are taken into account in the objective function multiple times, which also increases the extra complexity in the optimization. For example, suppose there are  $m$  images, [APE](#) considers all pairwise combinations of images, i.e.  $(m(m-1)/2)$  pairs of images [78], in the optimization.

### 2.3.2 Bundle Adjustment

Bundle adjustment ([BA](#)) [18] is regarded as the gold standard for obtaining optimal global scene points and poses of an imaging sensor at different viewpoints from multiple 2D images in the computer vision community.

Suppose we have an initial estimate of the scene points  $\{\mathbf{p}_j\}_{j=1}^n$  and sensor's poses  $\{\boldsymbol{\xi}_i\}_{i=1}^m$ , and  $\mathbf{p}_{ij}$  is the projection of a scene point  $\mathbf{p}_j$  to the imaging plane of the sensor at the  $i_{th}$  position, the reprojection error is defined as

$$\epsilon_{ij}(\boldsymbol{\xi}_i, \mathbf{p}_j) = \|\mathbf{p}_{ij} - \pi(\boldsymbol{\xi}_i, \mathbf{p}_j)\|, \quad (2.12)$$

where  $\pi(\cdot, \cdot)$  is a function that projects the scene point  $\mathbf{p}_j$  to the imaging plane of the  $i_{th}$  sensor position.

Given initial values of the pose parameters and scene points, we may refine the results of them by minimizing the sum of squared reprojection errors of (2.12), i.e.

$$\{\hat{\boldsymbol{\xi}}_i, \hat{\mathbf{p}}_j\} = \underset{\boldsymbol{\xi}_i, \mathbf{p}_j}{\operatorname{argmin}} \sum_{i=1}^m \sum_{j=1}^n \frac{1}{2} \sigma(\mathbf{p}_{ij}) \epsilon_{ij}^2(\boldsymbol{\xi}_i, \mathbf{p}_j), \quad (2.13)$$

where  $\sigma(\mathbf{p}_{ij}) = 1$  if the  $\mathbf{p}_j$  is observed by the sensor at  $i_{th}$  position, otherwise  $\sigma(\mathbf{p}_{ij}) = 0$ . (2.13) is a classical formulation of the BA problem.

Please note, most of the BA algorithms are feature-based, which require finding feature correspondences between the scene points and their projections prior to the BA optimization. Importantly, the goal of classical BA methods is to refine the positions of scene points and poses of the sensor based on the 2D images [19, 20] instead of image registration and fusion. To our best knowledge, there is no BA method that can be used for multiple 3D image registration and fusion.

## 2.4 Chapter Summary

In this chapter, we reviewed works that are related to medical image registration and fusion from three perspectives, namely feature-based and intensity-based registration methods, rigid and deformable registration methods, and pairwise and simultaneous registration methods. Widely used and state-of-the-art registration methods such as B-spline based FFD [46, 47], embedded deformation [48], Demons-based registration [50, 53], VoxelMorph [74], and bundle adjustment [18] are introduced briefly.

In the remaining thesis, five frameworks for 3D image registration and fusion will be presented to extend the FoV of 3D TEE images in rigid or deformable scenarios. They are divided into two parts: Part I involves two frameworks deduced from pairwise registration, which are used for rigid and deformable scenarios, respectively. Part II involves three simultaneous registration and fusion algorithms in which the first two are used for rigid scenarios and the last one is used for deformable scenarios.

## Part I

# PAIRWISE REGISTRATION AND SEQUENTIAL FUSION

## Chapter 3

# Fast Direct Ultrasound Fusion for 3D Transesophageal Echocardiography

In this chapter, we present our first framework that extends **FoV** of 3D **TEE** images efficiently by combining pairwise registration with sequential fusion to fuse multiview images. Cardiac motion (deformation of the heart) is avoided in the obtained in-vivo 3D **TEE** images by using **ECG**-gating technique. Since 3D **TEE** images lack distinct features, the proposed framework is completely based on the intensity of the obtained images to circumvent the feature extraction and matching processes used in conventional feature-based registration methods. Our registration is solved in the Lie algebra space and its fast implementation is realized by directly searching the corresponding three orthogonal planes in the whole voxel spaces of the fixed and moving images. In addition, an intuitive and efficient weighted-average fusion method is proposed to fuse the aligned 3D images seamlessly. For multiple 3D **TEE** images, a sequential fusion strategy is used to obtain the globally consistent image and poses. Monte-Carlo simulations and in-vivo experiments are performed to validate our algorithms systematically. Based on the fused in-vivo images, segmentation and 3D printing are performed. For the first time, the whole **LA** with both the **IAS** and **LAA** in a single model is segmented successfully. This work shows good

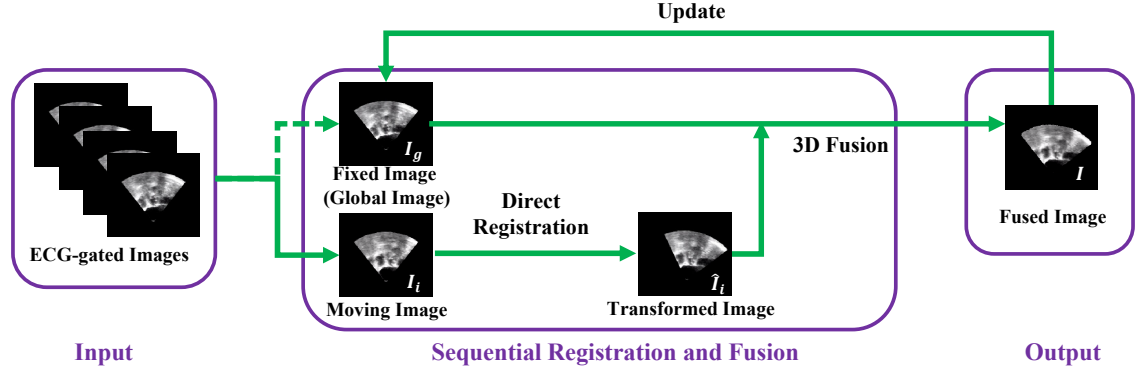


FIGURE 3.1: Overview of the direct 3D ultrasound fusion framework. (Remark: 3D TEE images are shown via 2D slices.)

potential in practice to overcome the drawback of small FoV of 3D TEE, expanding the clinical utility of 3D TEE.

### 3.1 Methods

The overview of the entire framework is illustrated in Fig. 3.1. Multiview 3D TEE images are acquired with the assistance of ECG-gating. A sequential strategy is used to register and fuse the collected images. Suppose there are  $m$  frames of 3D TEE images, the global (panoramic) image  $I_g$  is initialized by the first frame of image and is designated as the fixed image for a new coming image  $I_i$ . Relative pose from  $I_i$  to  $I_g$  is calculated by the proposed direct registration method. After registering  $I_i$  to  $I_g$ , they are fused together by using the proposed weighted-average 3D fusion method. Then, the fused image  $I$  is used to update the global image  $I_g$  and designated as the fixed image for the next coming image. The same processes are performed repeatedly until all images are fused. The system centers on the direct registration method and the 3D fusion method.

#### 3.1.1 Direct Registration

##### 3.1.1.1 Formulation of the Registration Problem

To fuse 3D TEE images captured from different viewpoints, one key step is to align the image data. Since the ECG-gating is used, the process of registration can be simplified to

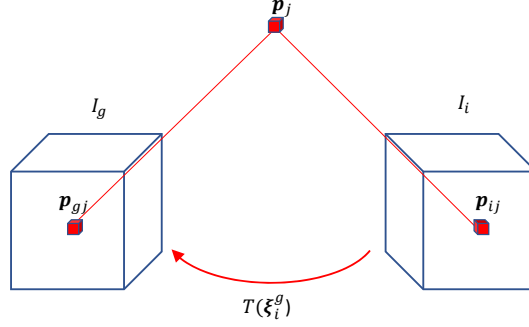


FIGURE 3.2: Schematic diagram of direct registration for two 3D images.

estimate the rigid transformation (pose) between 3D TEE images.

Let  $\mathbf{p}_j$  represent the coordinates of a point in the heart, which is shown in Fig. 3.2.  $\mathbf{p}_{gj}$  and  $\mathbf{p}_{ij}$  are the projections of  $\mathbf{p}_j$  in the global image  $I_g$  and the local image  $I_i$ . Since  $\mathbf{p}_{gj}$  and  $\mathbf{p}_{ij}$  are in different image spaces, they should be transformed to the same coordinate system before fusion. We estimate the pose from image  $I_i$  to  $I_g$  in Lie algebra space. Suppose the pose parameters are denoted by the vector elements of Lie algebra  $\boldsymbol{\xi}_i^g \in \mathbb{R}^6$ . A rigid transformation is formally defined as follows:

$$\mathbf{p}_{gj} = T(\boldsymbol{\xi}_i^g)\mathbf{p}_{ij}^1, \quad (3.1)$$

where  $T(\cdot) \in SE(3)$  maps the pose parameters  $\boldsymbol{\xi}_i^g$  to a rigid transformation matrix in Euclidean space  $SE(3)$  [79]. A rigid transformation matrix is shown in (2.7).

Assuming that the intensity value of an image  $I$  is a function w.r.t. the local coordinates of voxel  $\mathbf{p}$ , the intensity difference of  $\mathbf{p}_j$ 's projections in  $I_i$  and  $I_g$  can be written as:

$$e_j(\boldsymbol{\xi}_i^g) = I_i(\mathbf{p}_{ij}) - I_g(\mathbf{p}_{gj}). \quad (3.2)$$

During the process of optimization,  $\mathbf{p}_{gj}$  is calculated from  $\mathbf{p}_{ij}$  by using (3.1).

---

<sup>1</sup>In (3.1), while  $\mathbf{p}_{gj}$  and  $\mathbf{p}_{ij}$  should be homogeneous coordinates, we implicitly perform the conversion between 3D Euclidean coordinates and homogeneous coordinates in this thesis when the meaning is clear from the context.

Suppose  $n$  spatial points of the heart are observed in both  $I_g$  and  $I_i$ , we estimate the relative pose from  $I_i$  to  $I_g$  by minimizing the sum of squared intensity differences as follows,

$$\hat{\xi}_i^g = \underset{\xi_i^g}{\operatorname{argmin}} \sum_{j=1}^n \|e_j(\xi_i^g)\|^2. \quad (3.3)$$

### 3.1.1.2 Solving the Optimization Problem

In (3.2), only the term  $I_g(\mathbf{p}_{gj})$  is dependent on  $\xi_i^g$ . According to the chain rule, the derivative of intensity difference  $e_j(\xi_i^g)$  w.r.t. pose parameters  $\xi_i^g$  can be written as:

$$J_j(\xi_i^g) = \frac{\partial e_j(\xi_i^g)}{\partial \xi_i^g} = -\frac{\partial I_g}{\partial \mathbf{p}_{gj}} \frac{\partial \mathbf{p}_{gj}}{\partial \xi_i^g}. \quad (3.4)$$

It is found from (3.4) that the derivative consists of two parts:

- 1) The first part  $\partial I_g / \partial \mathbf{p}_{gj}$  is the partial derivative of intensity  $I_g$  w.r.t. the coordinates of point  $\mathbf{p}_{gj}$ .
- 2) The second part  $\partial \mathbf{p}_{gj} / \partial \xi_i^g$  is the derivative of  $\mathbf{p}_{gj}$  w.r.t. Lie algebra  $\xi_i^g$ .

Assuming that the coordinates of  $\mathbf{p}_{gj}$  are  $[x, y, z]^\top$ , then

$$\frac{\partial \mathbf{p}_{gj}}{\partial \xi_i^g} = [\mathbf{I} | -\mathbf{p}_{gj}^\wedge] = \begin{bmatrix} 1 & 0 & 0 & 0 & z & -y \\ 0 & 1 & 0 & -z & 0 & x \\ 0 & 0 & 1 & y & -x & 0 \end{bmatrix}.$$

The overall intensity differences over  $n$  voxels can be built as a collection of  $e_j(\xi_i^g)$ , i.e.  $e(\xi_i^g) = [e_1(\xi_i^g), \dots, e_n(\xi_i^g)]^\top$ . The Jacobian matrix of intensity differences  $e(\xi_i^g)$  w.r.t.  $\xi_i^g$  can be built as a collection of  $J_j(\xi_i^g)$ , i.e.  $J(\xi_i^g) = [J_1^\top(\xi_i^g), \dots, J_n^\top(\xi_i^g)]^\top$ .

Gauss-Newton (GN) method [80] is commonly used to solve nonlinear least-squares (NLLS) problems like (3.3). By using the GN method, the optimal solution  $\hat{\xi}_i^g$  of objective function (3.3) can be obtained by starting with an initial guess  $\xi_0$  and iterating with  $T(\xi_i^g) \leftarrow T(\Delta \xi_i^g)T(\xi_i^g)$ . The step change  $\Delta \xi_i^g$  can be calculated from GN equation:

$$J^\top(\xi_i^g)J(\xi_i^g)\Delta \xi_i^g = -J^\top(\xi_i^g)e(\xi_i^g). \quad (3.5)$$



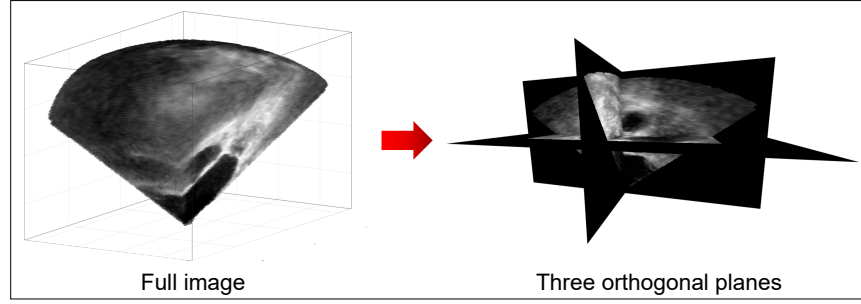


FIGURE 3.3: Full image vs. three-orthogonal-plane image.

After the optimal solution of the relative pose  $\hat{\xi}_i^g$  is obtained,  $I_i$  can be registered to  $I_g$  by transforming all of its voxels using (3.1).

### 3.1.1.3 Efficient Implementation Based on Gradient Space and Orthogonal Planes

Since a 3D TEE image generally contains millions of valid voxels, calculating all these voxels is time-consuming and needs many computation resources. In order to estimate the poses of images efficiently, the following two strategies are used.

**Pre-computation of Gradient Space:** According to (3.4) and (3.5), calculating  $\partial I_g / \partial \mathbf{p}_{gj}$  is an important but computationally costly step in the process of GN iterations. Since this calculation is based on the fixed image, we can pre-compute the gradient space of intensity of the fixed image before the iterations. And then, the gradient value at any voxel point can be read directly from the gradient space, which means there is no need for any calculation of the first part in Jacobian (3.4) during GN iterations. This strategy can greatly improve the effectiveness of computation. In addition, the intensity and gradient value of voxels are interpolated using trilinear interpolation to reduce the error of pose estimation, which are also processed before the iterations.

**Simplification by Using Three Orthogonal Planes:** In order to reduce the computational time of pose estimation, we simplify the method by using voxels from three orthogonal planes instead of a full image, which is shown in Fig. 3.3. For the sake of distinction, pose estimation by using three-orthogonal-plane voxels and full image are called the three-plane method and full-image method respectively in this chapter. To verify the

**Algorithm 1:** Direct Method for Pose Estimation**Input:** Fixed image  $I_g$  and a moving image  $I_i$ .**Output:** Relative pose between two images.

---

```

1 Step 1: Pre-computation:
2 Calculate mask matrices mask1 and mask2 for  $I_g$  and  $I_i$ , respectively;
3 Calculate gradient space of intensity of  $I_g$ ;
4 Interpolate intensity/gradient space of  $I_g$ ;
5 Step 2: Optimization:
6 Initialize the pose  $\xi_i^g \leftarrow \xi_0$ ;
7 while Algorithm not converged do
8   for every voxel  $p_{ij}$  in valid areas of mask2 do
9      $p_{gj} = T(\xi_i^g)p_{ij}$ ;
10    if  $p_{gj}$  is in valid areas of mask1 then
11      Calculate intensity difference  $e_j(\xi_i^g)$  by (3.2)
12      Calculate Jacobian  $J_j(\xi_i^g)$  by (3.4);
13    end
14  end
15  Construct  $J(\xi_i^g) = [J_1^\top(\xi_i^g), \dots, J_n^\top(\xi_i^g)]^\top$  and  $e(\xi_i^g) = [e_1(\xi_i^g), \dots, e_n(\xi_i^g)]^\top$ ;
16  Calculate step change  $\Delta\xi_i^g$  via (3.5);
17  Update the pose  $T(\xi_i^g) \leftarrow T(\Delta\xi_i^g)T(\xi_i^g)$ ;
18 end

```

---

effectiveness of this strategy, pose estimation by using three-orthogonal-plane voxels and full image are compared in Section 3.2.1.1, 3.2.1.2, and 3.2.2.1.

In Fig. 1.1, it is shown that the valid area of a TEE image is fan-shaped (for 2D image) or pyramid-shaped (for 3D image) and the rest area of a cuboid TEE image is black (i.e. intensity value is 0) which is invalid for pose estimation. Therefore, we create mask matrices to block these invalid voxels at the preprocessing stage to speed up the computation. The steps of pose estimation via GN method are shown in Algorithm 1.

### 3.1.2 3D Fusion

Although some sophisticated fusion methods such as wavelet-based methods [81] may get the fused images with higher quality, they involve several time-consuming steps such as multiresolution analysis, subband decomposition, and are not easy to realize fast performance, especially when it comes to 3D volumes. By contrast, image-average methods

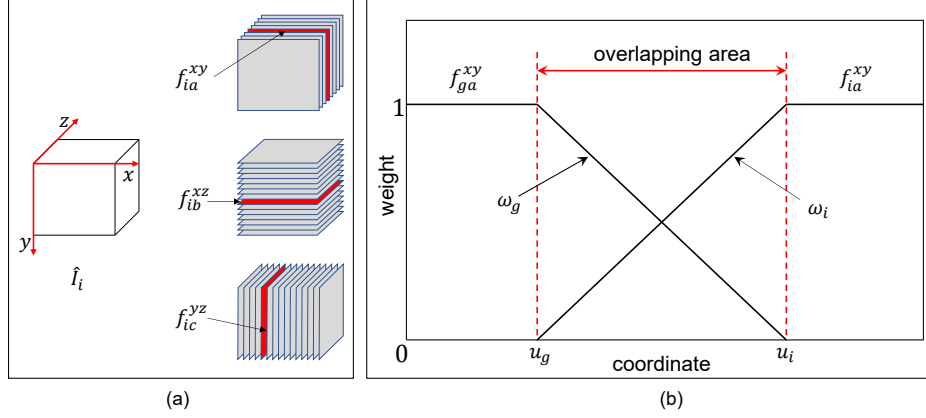


FIGURE 3.4: (a) A 3D image is denoted in three different formats along different axes in the image coordinate system; (b) Weights for 2D slices fusion.

are computationally efficient without any information analysis. Therefore, we propose a weight-average 3D fusion method to fuse the aligned images.

Suppose  $\hat{I}_i$  is the transformed image of  $I_i$  after registration, it can be viewed as a collection of 2D slices stacked along different axes in the image coordinate system, i.e.  $\hat{I}_i = \{f_{ia}^{xy}\} = \{f_{ib}^{xz}\} = \{f_{ic}^{yz}\}$ , which is shown in Fig. 3.4(a). Similarly,  $I_g$  can also be denoted in this way:  $I_g = \{f_{ga}^{xy}\} = \{f_{gb}^{xz}\} = \{f_{gc}^{yz}\}$ .

In order to fuse 3D images, corresponding slices are fused first along  $x$ ,  $y$ , and  $z$  axis respectively. For any two corresponding 2D slices  $f_{ga}^{xy}$  and  $f_{ia}^{xy}$ , the fused image  $f_a^{xy}$  can be calculated by:

$$f_a^{xy}(u, v) = \begin{cases} f_{ga}^{xy}(u, v) & (u, v) \in f_{ga}^{xy} \\ \omega_g f_{ga}^{xy}(u, v) + \omega_i f_{ia}^{xy}(u, v) & (u, v) \in (f_{ga}^{xy} \cap f_{ia}^{xy}) \\ f_{ia}^{xy}(u, v) & (u, v) \in f_{ia}^{xy} \end{cases} \quad (3.6)$$

where  $\omega_g = |u - u_g| / |u_g - u_i|$  and  $\omega_i = 1 - \omega_g$ .  $(u, v)$  are the coordinates of voxels in the corresponding 2D slices.  $u_g$  and  $u_i$  are the upper and lower boundaries of overlapping areas of 2D slices. This method uses the distance from points to the boundaries of the overlapping area as weights, which is illustrated in Fig. 3.4(b).  $f_b^{xz}$  and  $f_c^{yz}$  along  $y$  and  $x$  axes can also be calculated in a similar way as (3.6).

After three fused 3D images  $I_x = \{f_a^{xy}\}$ ,  $I_y = \{f_b^{xz}\}$ , and  $I_z = \{f_c^{yz}\}$  are obtained, the final fused 3D image  $I$  is calculated by:

$$I = (I_x + I_y + I_z)/3. \quad (3.7)$$

## 3.2 Experiments and Results

In this section, simulated and in-vivo experiments are performed to validate our method in terms of accuracy, robustness, and efficiency.

### 3.2.1 Simulated Experiments

Based on the simulated datasets, this section starts with the validation of the proposed solution to the optimization of our direct registration method, followed by detailed assessments of the accuracy and robustness of the proposed three-plane method for registration.

#### 3.2.1.1 Effectiveness of Registration

Many intensity-based registration methods use numerical Jacobian in the optimization process, which makes the solving process sensitive to initial guess and hard to converge. Thus, image pyramid is usually used to guide the optimization process. One example is the built-in function *imregtform* in the MATLAB Image Processing Toolbox. The function can estimate the relative pose between two input images and specify mean squared difference as similarity metric, which is similar to the proposed method. But different from the analytical expression of Jacobian used in our optimization process, *imregtform* uses numerical method combined with image pyramid. The function *imregtform* has been optimized to have excellent performance in MATLAB so the proposed method is compared with it to verify the effectiveness of our method in this section.

Since volumes with the pyramidal FoV like Fig. 1.1 have a large number of invalid voxels whose intensity values are zero, directly applying *imregtform* to such images will make the function try to align these invalid areas and obtain wrong results. Thus, simulated

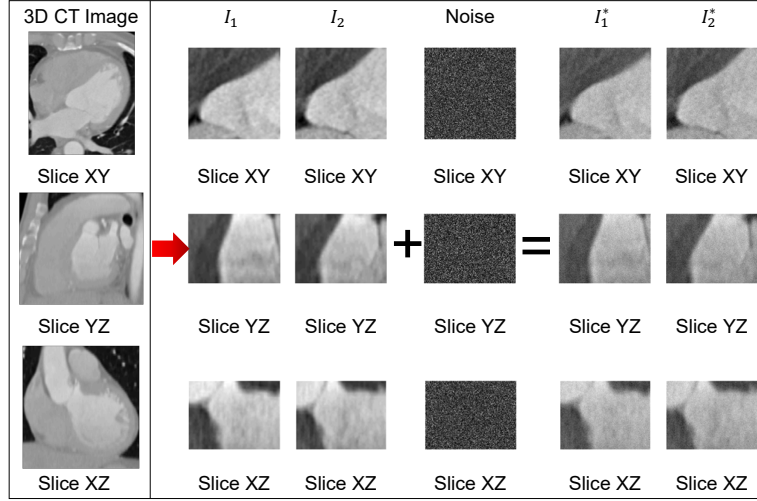


FIGURE 3.5:  $I_1$  and  $I_2$  are images cropped from 3D heart CT scan. After adding Gaussian noise to these two images, corresponding noisy images  $I_1^*$  and  $I_2^*$  are obtained.

images without invalid voxels are used to compare our method with the MATLAB function *imregtform*. Two cuboid 3D volumes are cropped from a 3D heart CT scan and denoted as  $I_1$  and  $I_2$ , as shown in Fig. 3.5. The intensity of CT scan has been adjusted to a range of 0 – 255 before cropping to make it in line with in-vivo ultrasound images. The volume size of two images is set as  $200 \times 200 \times 150$  voxels to make them have roughly the same number of voxels as real 3D TEE images. The transformation from  $I_2$  to  $I_1$  is  $[\pi/36, \pi/36, \pi/36]^\top$  rad for three rotations and  $[5, 5, 5]^\top$  pixels for three translations.

Speckle noise is an inherent property of medical ultrasound imaging [82–84]. It is an interference effect caused by the scattering of the ultrasonic beam from microscopic tissue inhomogeneities. Speckle noise generally results in the change of intensity of images, thereby influencing the robustness of the direct method. In commercial ultrasound imaging systems, logarithmic compression is usually applied to the envelope detected image so that multiplicative speckle noise is converted into a kind of additive noise and is close to white Gaussian distribution [84–86]. An obtained image after compression can be described by the following mathematical model [84]:

$$g(x, y, z) = f(x, y, z) + u(x, y, z), \quad (3.8)$$

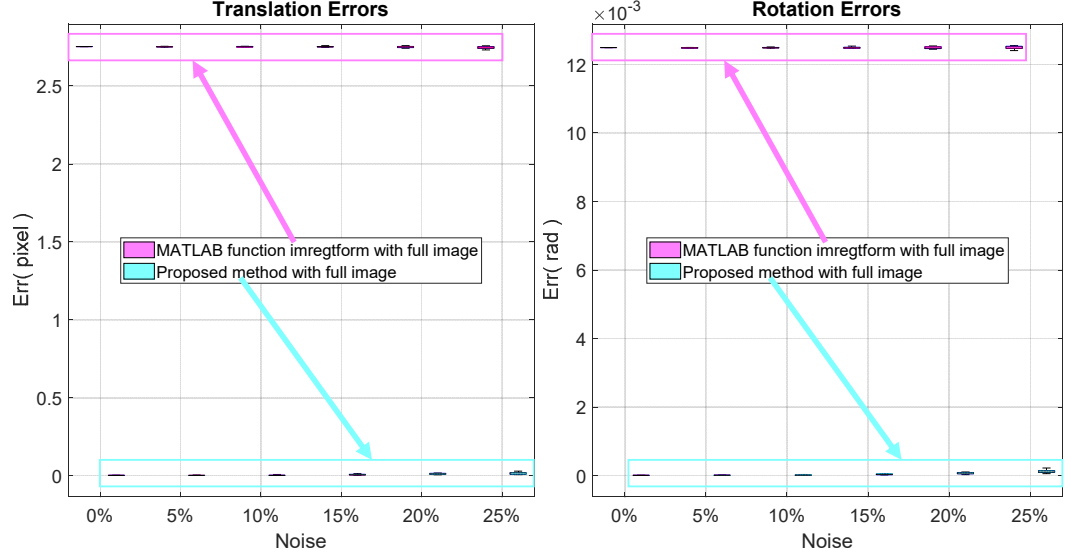


FIGURE 3.6: Comparison of the convergence and accuracy of the proposed method and MATLAB function *imregtform* (one-level image pyramid, maximum 100 iterations).

where  $g(x, y, z)$ ,  $f(x, y, z)$ , and  $u(x, y, z)$  denote the observed image (with noise), the original image (without noise), and speckle noise after logarithmic compression respectively.  $(x, y, z)$  represent coordinates of voxels. We use the ratio of standard deviation of the added intensity noise to standard deviation of the intensities of an image to define the noise level. According to model (3.8), five levels of white Gaussian noise (WGN) are added, with ratios of 5%, 10%, 15%, 20%, and 25%, respectively. For each level of intensity noise, ten pairs of noisy images are generated based on  $I_1$  and  $I_2$ . Finally, we obtain a dataset with a total of 51 pairs of images (including the pair of  $I_1$  and  $I_2$ ).

Firstly, the accuracy and convergence of the proposed solution to the optimization of our method are validated. The default configuration of *imregtform* using regular step gradient descent optimizer is as follows: the maximum number of iterations of each level of image pyramid is 100, the minimum step length of algorithm convergence is  $10^{-5}$ , and the initial value of the algorithm is set as the identity matrix. Since the proposed method does not require the image pyramid, the pyramid level of *imregtform* is specified as one level with the maximum number of iterations 100 for comparison. And accordingly, the same configuration of the maximum number of iterations, minimum step length, and initial guess is set in the proposed method. Full images are used in both *imregtform* and the proposed registration method for evaluating the accuracy and convergence of

the optimization. Mean absolute errors (MAE) of translations and rotations against the ground-truth poses are compared in Fig. 3.6 (comparisons of the accuracy of the proposed method with full image and three-orthogonal-plane image are detailed in Section 3.2.1.2 and 3.2.2.1). It is shown that none of the results obtained by *imregtform* can converge to the accurate result even if the function reaches the maximum number of iterations. But the proposed method can always obtain the accurate results at different noise levels. For further comparison, we configure the function *imregtform* with an image pyramid with three levels (100 iterations for each level) and perform all the experiments again. Compared with the configuration with one level of pyramid which uses original images directly, when *imregtform* is configured with a three-level image pyramid, it employs a coarse-to-fine registration strategy: firstly, the algorithm generates two more pairs of images by halving the resolution of the original images each time and constructs a three-level coarse-to-fine image pyramid. Then, the algorithm starts the registration from the coarsest image level and takes the result as the initial value of the next level of image registration. The process continues until the highest resolution images, i.e. the input original images, are registered. It is found that *imregtform* configured with a three-level of image pyramid can obtain almost the same accurate results as our method that uses original images directly. The MAE of translations are all less than 0.1 pixels and MAE of Euler angles are within a  $10^{-3}$  order of magnitude radians.

Additionally, in terms of convergence, we found that the proposed solution to the optimization of our method can converge to accurate results quickly not only when using the full image but also the three-plane image, but MATLAB function *imregtform* usually needs to reach the maximum number of iterations in the optimization process. For example, when using the images with 20% intensity noise, the optimization process of *imregtform* and the proposed method with full image and three-plane image is shown in Fig. 3.7 (*imregtform* uses an one-level image pyramid with the maximum number of iterations 100).

Secondly, the efficiency of the proposed strategy of using a three-plane image in the optimization is validated. The computational cost of the proposed three-plane method is compared with the MATLAB function *imregtform* for estimating the poses of the above 51 pairs of simulated images. All experiments are tested on the same computer (Intel Core i7 processor, 8th generation @ 4.2 GHz, 16 GB RAM) and the computational cost

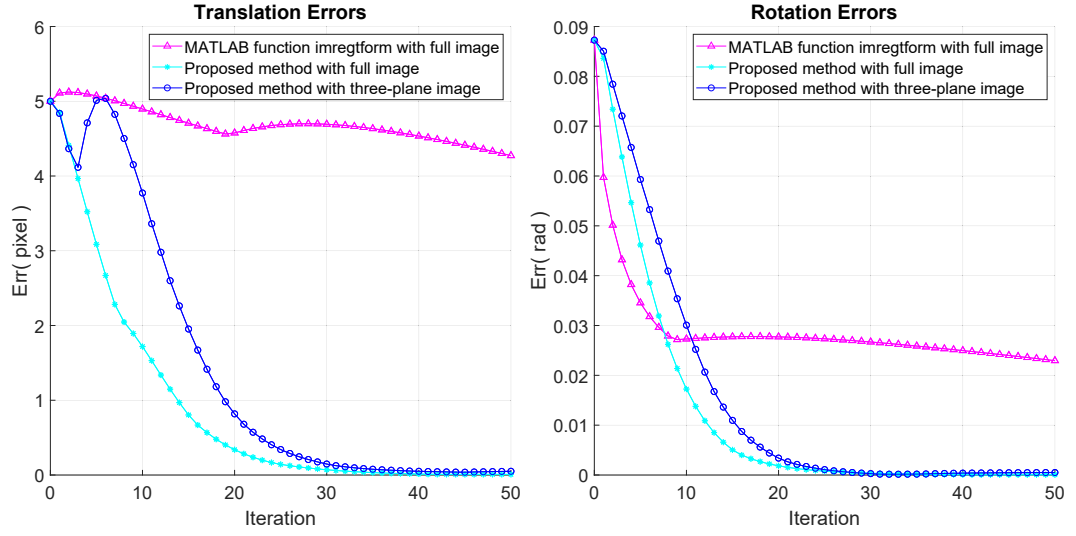


FIGURE 3.7: Comparison of the iterative process of *imregtform* and the proposed method with full image and three-plane image (one-level image pyramid, maximum 100 iterations).

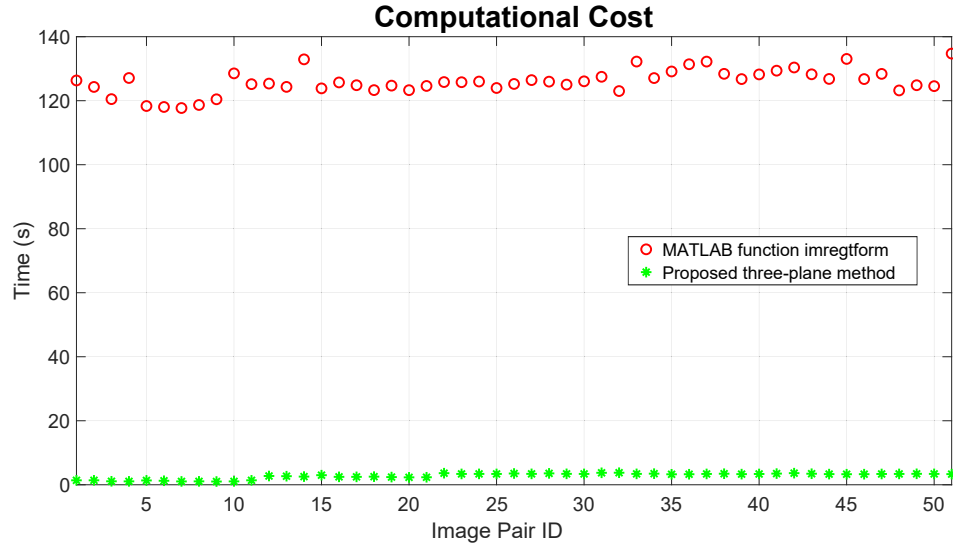


FIGURE 3.8: Comparison of the computational cost of the proposed tree-plane method and MATLAB function *imregtform* (three-level image pyramid, maximum 300 iterations).

is shown in Fig. 3.8. The mean computational cost of the pose estimation is 2.7s for the proposed three-plane method and 126.7s for the MATLAB *imregtform*.

From the above results, it is evident that compared to the conventional pairwise registration method like *imregtform* that uses numerical optimization combined with hierarchical strategy, the proposed registration method can converge to accurate results with fewer iterations. And using voxels on three orthogonal planes is an efficient way to implement our



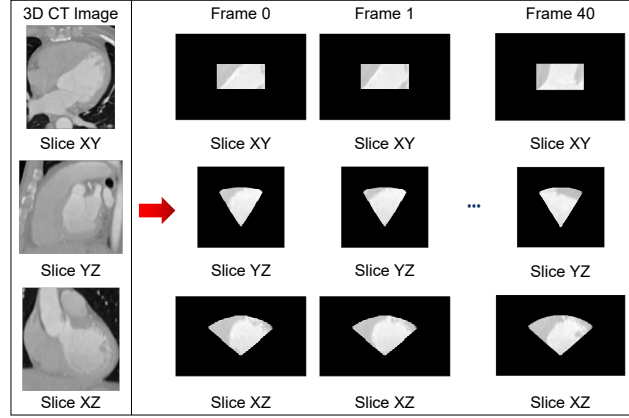


FIGURE 3.9: 3D CT image of the heart and the simulated 3D TEE image sequence.

algorithm. In the following parts of the chapter, the proposed direct registration method will be validated in detail in terms of accuracy and robustness.

### 3.2.1.2 Three-orthogonal-plane Approximation vs Full Image

In Section 3.2.1.1, we have shown that the full-image method has higher accuracy and better convergence than conventional intensity-based methods like *imregtform* without using the image pyramid. In this section, poses calculated by the three-plane method are compared with those calculated by the full-image method to evaluate the accuracy of the three-plane method.

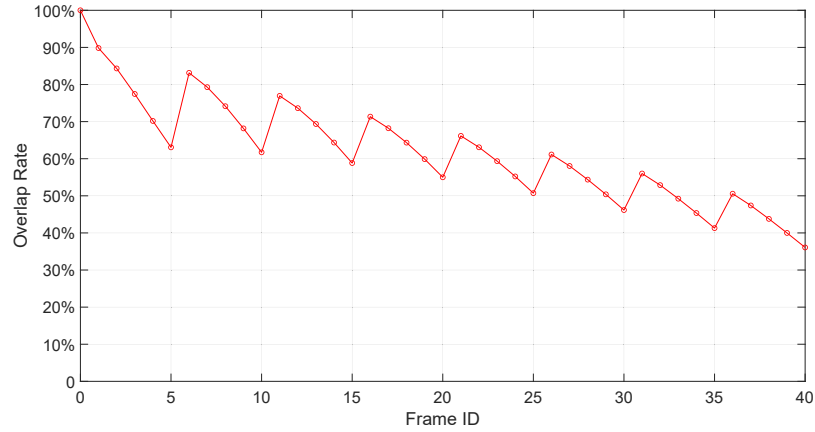


FIGURE 3.10: Overlapping rate of the valid image area of each frame w.r.t. Frame 0.

Similar to Section 3.2.1.1, we generate simulated 3D TEE images by firstly cropping images from the 3D heart CT scan (see Fig. 3.9), and then adding WGN to them. First, the valid image area of a 3D TEE image shown in Fig. 1.1 is deleted to get a frame as our ‘viewfinder’. Then, an image sequence with 41 frames of 3D images is obtained by transforming the ‘viewfinder’ to the same coordinate system as the 3D heart CT scan with various poses and cropping the corresponding image areas, which are shown in Fig. 3.9. The magnitude of these transformations varies between  $[0, 2\pi/15]$  rad for three rotations and  $[0, 25]$  pixels for three translations. Frame 0 is in the same global coordinate system as the 3D heart CT scan. The overlapping rate of each frame w.r.t. Frame 0 changes as the pose changes, which is shown in Fig. 3.10. It decreases with fluctuation from Frame 1 to Frame 40. The overlapping rates of Frame 1-29, Frame 31-32, and Frame 36 w.r.t. Frame 0 are more than 50%, but those of the rest eight frames w.r.t. Frame 0 are less than 50%. Finally, by adding five levels of intensity noise (5%, 10%, 15%, 20%, and 25% intensity noise) to the 41 frames of images, a total of six sequences of simulated 3D TEE images with different levels of noise are obtained (including the image sequence without adding intensity noise).

Algorithm 1 is performed to minimize (3.3) and the initial guesses of poses of the images are given by adding an offset to the ground-truth poses ( $[\pi/60, \pi/60, \pi/60]^\top$  rad to three rotations and  $[8, 8, 8]^\top$  pixels to three translations). For each noise level, relative poses between the last 40 frames and Frame 0 are calculated by using the full-image method and three-plane method, respectively. The MAE of translations and rotations are compared respectively in Fig. 3.11. From the results, it is found that both the full-image method and the three-plane method can obtain results that are the same as the ground-truth poses when there is no intensity noise. The accuracy of the results obtained by both methods decreases with the increase of the noise level, but errors are still within a relatively small range. For the results of the rotation part, both methods can converge to results with high accuracy. For the results of the translation part, MAE of the full-image method are within 0.2 pixels and most of the MAE of the three-plane method are within 1 pixel. According to the distribution of the translation and rotation errors, the accuracy of pose estimation of the two methods is close for most of the cases. Further investigation finds that large differences in results between the three-plane method and the full-image method

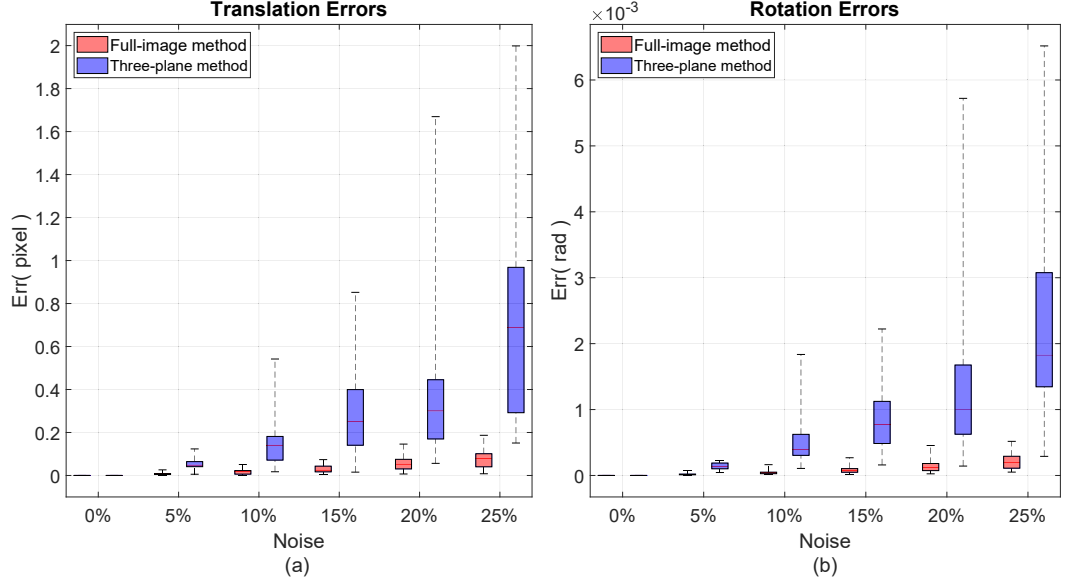


FIGURE 3.11: The accuracy of the full-image method and the three-plane method: (a) and (b) are the MAE of the estimated translations and Euler angles, respectively.

are mainly caused by the reduction of the overlapping rate. The detailed results of errors w.r.t. the overlapping rate will be shown in Section 3.2.1.3.

While the accuracy of the three-plane method is influenced by big noise and small overlap, it is still promising to use the method in actual situations. First, an ultrasound transducer can be controlled in practice by operators to avoid large rotations and translations between two consecutive imaging positions. In our in-vivo datasets, overlapping rates between consecutive images are more than 50%. In addition, for 3D TEE, it is possible to avoid main artifacts that affect the obtained images in the imaging stage [87]. Importantly, the number of voxels used for calculation in the three-plane method is around 1/45 of all voxels, which is very beneficial for fast registration when compared with the full-image method. Therefore, in the following sections of simulated experiments, we will only focus on the three-plane method.

### 3.2.1.3 Robustness Assessment via Monte-Carlo Simulation — Intensity Noise

In Section 3.2.1.2, the results indicate that the intensity noise and overlapping rate can influence the accuracy of pose estimation. In this section, detailed tests are performed to assess the influences on the three-plane method.

For each sequence with added intensity noise mentioned in Section 3.2.1.2, i.e. those with 5%, 10%, 15%, 20%, and 25% intensity noise, we generate a total of 10 sequences of images with the same level of intensity noise and then apply the three-plane method to them for poses estimation. We perform a total of 2000 runs of our algorithm with different intensity noises and the MAE of poses for each noise level are shown in Fig. 3.12. We can see it more clearly from Fig. 3.12 that both intensity noise and overlapping rate can influence the accuracy of the results. With the increase of intensity noise, the accuracy of the results obtained by the three-plane method starts to decrease. For the accuracy of rotations, even if they are influenced by intensity noise and overlapping rate, the errors of Euler angles still maintain a  $10^{-3}$  order of magnitude radians. For the accuracy of translations, the MAE of most of the translations are within 1 pixel. Big errors usually appear when the images have relatively small overlapping rates, such as the last few frames of images in each sequence which only have 35-45% overlap. It can be seen from Fig. 3.12 that translation errors of Frame 34, 35, 39, and 40 whose overlapping rates are below 50% are larger than other frames in the same noise level.

The results presented in Fig. 3.11 and Fig. 3.12 indicate if we continue to increase the intensity noise, the performance of the algorithm will decrease further. However, it is worth noting that the proposed method exhibits robustness to the intensity noise within a wide range. In fact, the method consistently converges to accurate results when intensity noise is less than 25%, especially when the overlapping rate surpasses 50%. This finding is instructive for practical applications because clinicians can easily control the overlapping rate of consecutive frames in 3D TEE examinations to be greater than 50% to obtain high accurate registration results.

#### 3.2.1.4 Robustness Assessment via Monte-Carlo Simulation — Initial Guess

For a direct method, a good initial guess of pose is usually important for the algorithm's convergence and speed. In Section 3.2.1.2 and Section 3.2.1.3, initial guesses are given by adding an offset to the ground-truth poses. Here, the tolerance of the proposed method to the initial guess is tested. To eliminate the influence of intensity noise, in this section, experiments are performed without adding intensity noise. Excluding the eight frames of

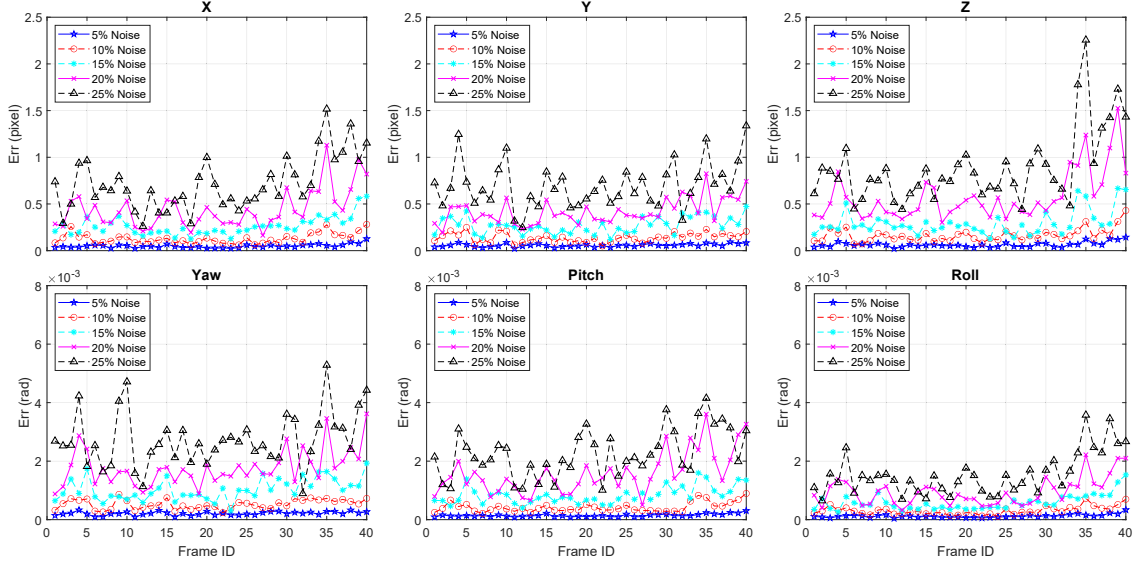


FIGURE 3.12: Robustness test of the proposed method to intensity noise via Monte-Carlo simulation: MAE of the poses under five levels of intensity noise.

images whose overlapping rates with frame 0 are less than 50%, we conduct tests on the remaining 32 frames.

Different levels of zero mean uniformly distributed noises are added to the ground truth of the relative poses. We separately add noise to the Euler angles of rotations and translations to test the influence of initial guess on the convergence of the algorithm. In order to add noise quantitatively, the uniformly distributed vector is normalized to a unit vector first, and then it is added to the ground-truth pose by multiplying different amplification factors:

$$\begin{aligned}\theta^* &= \theta + k_1 \alpha \\ t^* &= t + k_2 \beta\end{aligned}\tag{3.9}$$

where  $\theta$  and  $t$  are  $3 \times 1$  vectors of Euler angles and translations of the ground-truth pose respectively.  $\theta^*$  and  $t^*$  are the simulated Euler angles and translations of the initial pose.  $\alpha$  and  $\beta$  are zero mean uniformly distributed vectors that have been normalized, respectively.  $k_1$  and  $k_2$  denote the amplification factors which represent the noise level.  $k_1$  is set from  $\pi/180$  radians to  $10\pi/180$  radians and  $k_2$  is set from 1 pixel to 10 pixels.

By adding ten levels of noises to the ground truth of 32 relative poses and performing ten runs of the pose estimation algorithm for each noise level, we perform a total of 6400

runs of the algorithm with different initial guesses. The percentages of convergence and un-convergence of the experiments are shown in Fig. 3.13.

From the results, it is found that the proposed method is also robust to the error on initial guess in a wide range. With the same length of noise vector, the algorithm has a higher tolerance to translation error than Euler angle error. When only adding noises to Euler angles, more than 60% of runs can converge to the ground truth even if the noise is up to 0.175 radians (equal to 10 degrees) in three directions. And when only adding noises to translation, more than 80% of runs can converge to the ground truth when translation noise is up to 50 pixels in three directions.

### 3.2.2 In-vivo Experiments

In this section, in-vivo experiments are performed to validate the proposed methods. In-vivo datasets for the experiments are collected from nine patients using an iE33 ultrasound imaging system (Philips Medical Systems) equipped with an X7-2 real-time 3D transducer. A 3D ultrasound transducer can generally record a sequence of 3D TEE images with different deformations at one position. By moving the transducer, multiple sequences of 3D TEE images can be acquired from one patient. We collect a total of 76 sequences

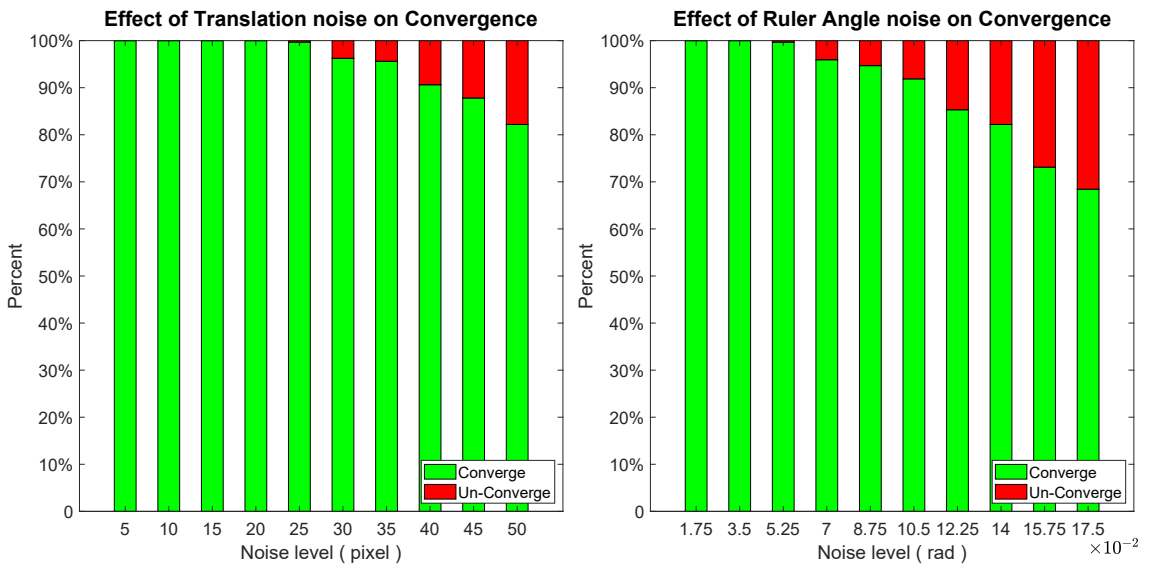


FIGURE 3.13: Robustness test of the proposed method to initial guess via Monte-Carlo simulation: the percentages of convergence and un-convergence with 10 levels of noise of initial guesses.

TABLE 3.1: Details of the Nine In-vivo ECG-gated Datasets

Patient	No. volumes	Volume size (voxel)	Resolution (mm/voxel)	FoV w.r.t original
# 1	9	$240 \times 160 \times 208$	$0.69 \times 0.98 \times 0.73$	2.22
# 2	6	$240 \times 160 \times 208$	$0.77 \times 1.11 \times 0.82$	2.05
# 3	6	$240 \times 160 \times 208$	$0.64 \times 0.92 \times 0.68$	1.82
# 4	8	$272 \times 208 \times 208$	$0.63 \times 0.90 \times 0.68$	2.09
# 5	6	$272 \times 208 \times 208$	$0.58 \times 0.87 \times 0.63$	2.01
# 6	17	$336 \times 224 \times 208$	$0.69 \times 0.72 \times 0.77$	2.04
# 7	5	$336 \times 224 \times 208$	$0.55 \times 0.83 \times 0.63$	1.70
# 8	8	$272 \times 224 \times 208$	$0.72 \times 1.02 \times 0.78$	1.83
# 9	11	$272 \times 224 \times 208$	$0.54 \times 0.77 \times 0.58$	2.13

of TEE volumes from nine patients and accordingly, 76 ECG-gated TEE volumes are extracted from the sequences with the assistance of ECG-gating. The movement of the transducer between consecutive positions is controlled within  $[-50, 50]$  mm and  $[-10, 10]$  degrees range to ensure that the overlapping rate between images collected at consecutive positions is more than 50%. In addition, ultrasound parameters of the imaging system are fixed during each collection. The details of the ECG-gated datasets are listed in the first four columns of Table 3.1.

### 3.2.2.1 Three-orthogonal-plane Approximation vs Full Image

Here three-plane method and full-image method are also compared by using the dataset of patient # 6. Since the ground truth of relative pose between frames in in-vivo experiments is unknown, initial guess is given by using a feature-based algorithm SIFT3D which is proposed in [31]. Although there are many mismatches or very few of extracted points when using SIFT3D in our 3D TEE images because of the lack of distinct features, it is proved in practice that the method is useful for getting an effective initial guess for our direct registration method. And in the in-vivo experiments, accuracy and convergence of the proposed method are determined according to the minimum step size and objective function value.

Given the fact that the ground truth of relative pose is unknown in in-vivo experiments and the results from full-image method are arguably the best results one can get for the

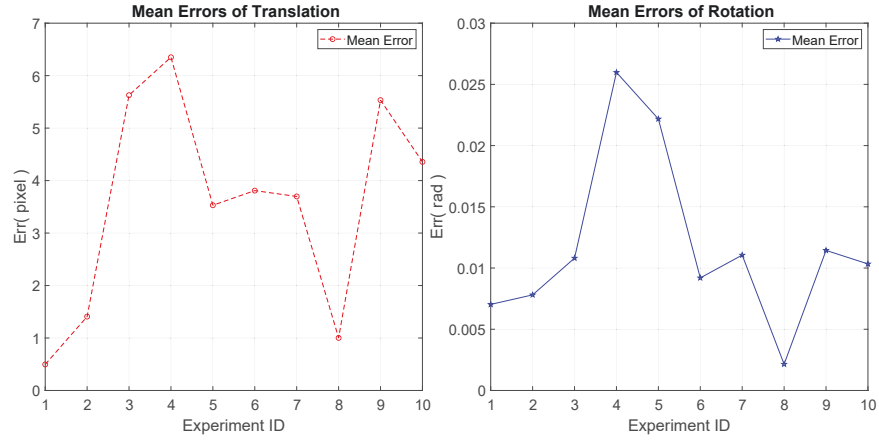


FIGURE 3.14: Mean absolute relative errors of the three-plane method w.r.t. the full-image method.

	Full-image Method	Three-plane Method	Full-image Method	Three-plane Method	Full-image Method	Three-plane Method	Full-image Method	Three-plane Method
Slice XY								
Slice YZ								
Slice XZ								
	Experiment 3		Experiment 4		Experiment 5		Experiment 9	

FIGURE 3.15: Four pairs of images with large mean absolute relative errors (Experiment 3, 4, 5, and 9 in Fig. 3.14) are fused. For the same pairs of images, relative poses are estimated by full-image method and three-plane method respectively.

proposed optimization problem, comparisons are performed by calculating mean absolute relative errors of three-plane method against the results from the full-image method. 10 experiments with different pairs of images are performed and mean absolute relative errors are shown in Fig. 3.14. From the results, it is found that the relative errors are larger than our simulated ones but still relatively small. One important reason is that although in-vivo data are acquired with ECG-gating, they still contain some deformation of the heart which could bring in errors. In addition, since the ground truth of relative poses between frames is unknown, mean absolute relative errors of three-plane method are calculated w.r.t. the poses obtained by full-image method. However, poses estimated by full-image method also contain errors which may enlarge the relative errors. To further assess the results of three-plane method, four pairs of images with large relative errors (Experiment 3, 4, 5,



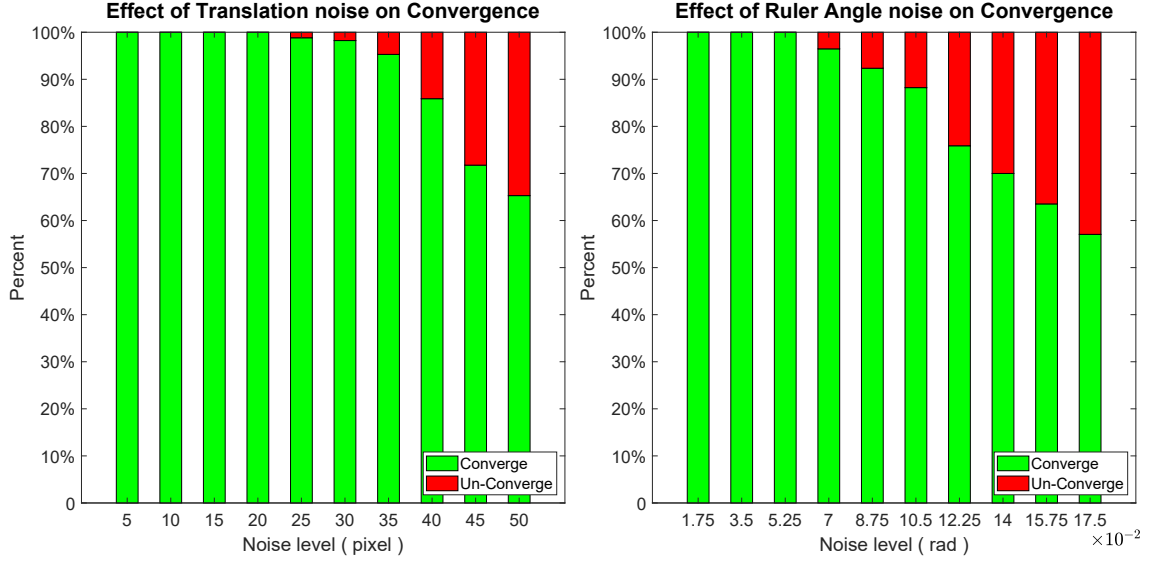


FIGURE 3.16: Robustness test of the proposed method to initial guess via in-vivo data: the percentages of convergence and un-convergence with 10 levels of noise of initial guesses.

and 9 shown in Fig. 3.14) are aligned firstly by using the poses estimated by full-image method and three-plane method respectively. Then, the aligned images are fused by using the proposed 3D fusion method. Comparisons are shown in Fig. 3.15. As we can see from the fused image, the results of full-image method and three-plane method are almost identical.

### 3.2.2.2 Robustness Assessment to Initial Guess

Based on the results of pairwise registration performed on the in-vivo 3D TEE images of patient # 6, robustness of the proposed three-plane method to initial guess is tested again. Ten pairs of 3D TEE images are used. Similar to Section 3.2.1.4, ten levels of noise are added to the Euler angles and translations of the poses, respectively.  $k_1$  is set from  $\pi/180$  radians to  $10\pi/180$  radians and  $k_2$  is set from 1 pixel to 10 pixels. For each level of noise, ten runs are performed. We perform a total of 2000 runs of our algorithm with different initial guesses and the percentages of convergence and un-convergence of the experiments are shown in Fig. 3.16.

Very similar to the results in Section 3.2.1.4, the proposed three-plane method is robust to different initial guesses in a wide range. With the same length of noise vector, the

algorithm has a higher tolerance to errors in translation than errors in Euler angles. When only adding noises to Euler angles, more than 55% of runs can converge to the same results even if the noise is up to 0.175 radians (equal to 10 degrees) in three directions. And when only adding noises to translation, more than 65% of runs can converge to the same results when translation noise is up to 50 pixels (around 35 mm) in three directions.

### 3.2.2.3 Comparison with PCA-based Registration

To verify the accuracy of the proposed three-plane method, the algorithm is compared with a state-of-the-art PCA-based method [38] for pairwise registration. The PCA-based method has shown superior performance in accuracy and computational efficiency to conventional intensity-based methods [38]. In the experiments, twenty pairs of 3D TEE images from the nine patients are registered by using the three-plane method and the PCA-based registration methods, respectively. Same as [38], we use two sequences of in-vivo 3D TEE images observed from different viewpoints as the inputs of the PCA-based registration method to estimate the relative pose of the transducer between these two positions. Each sequence of the images within one cardiac cycle are identified with the assistance of ECG-gating. The inputs of the proposed three-plane method are two ECG-gated images extracted from the corresponding two sequences. In the experiments, two algorithms started from the same initial poses.

Analyzing the accuracy of a registration algorithm based on in-vivo datasets is complex because the ground-truth poses are usually not available. If images are aligned using the estimated poses, visually we can confirm that the stitching areas of the aligned images should be smooth and without misalignment if the poses are accurate. To assess the accuracy of calculated poses, ECG-gated images are aligned with the poses estimated by the PCA-based method and our method, respectively. From the comparisons, it is found that the aligned images based on the poses from the two methods are almost the same for most of the cases with a smooth transition. However, since the PCA-based registration includes the deformation of the heart in the cardiac cycles, in some cases, apparent misalignments are found in the images aligned with the poses from the PCA-based method. Four examples are shown in the upper row of Fig. 3.17. On the contrary,

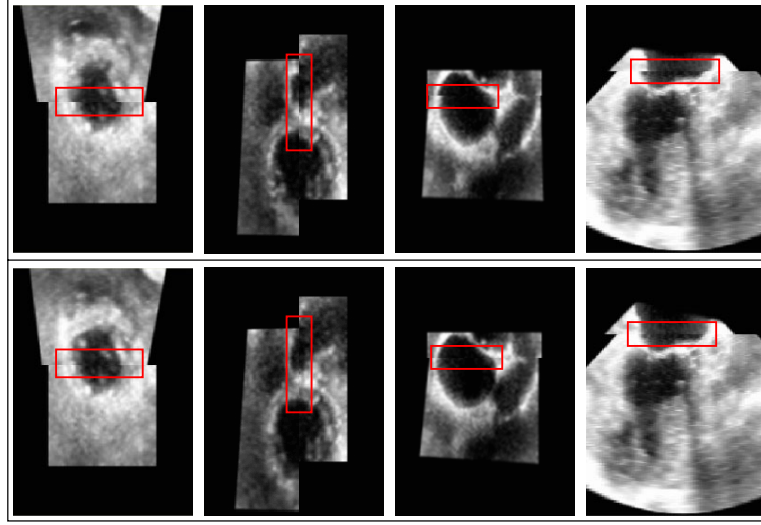


FIGURE 3.17: Four pairs of aligned images based on the PCA-based method and three-plane method are shown in the upper and bottom row, respectively. The same stitching areas are indicated by the rectangular boxes on images.

No. of fused volumes	1 volume	2 volumes	4 volumes	6 volumes	9 volumes	12 volumes	17 volumes
Slice XY							
Slice YZ							
Slice XZ							

FIGURE 3.18: The fusion process of in-vivo 3D TEE images from patient # 6. The FoV of the final fused image is 2.04 times larger than the first volume.

no obvious misalignment is found from the results obtained by the proposed three-plane method. In the bottom row of Fig. 3.17, four results obtained by the proposed three-plane method show accurate alignment and smooth transition. Comparisons indicate the overall accuracy of the proposed method outperforms the PCA-based method. In addition, it is found from the experiments that mean computational cost of one iteration for the PCA-based method is 81.1 s, while the mean computational cost of one iteration for the proposed three-plane method is only 40 ms using the same computer.

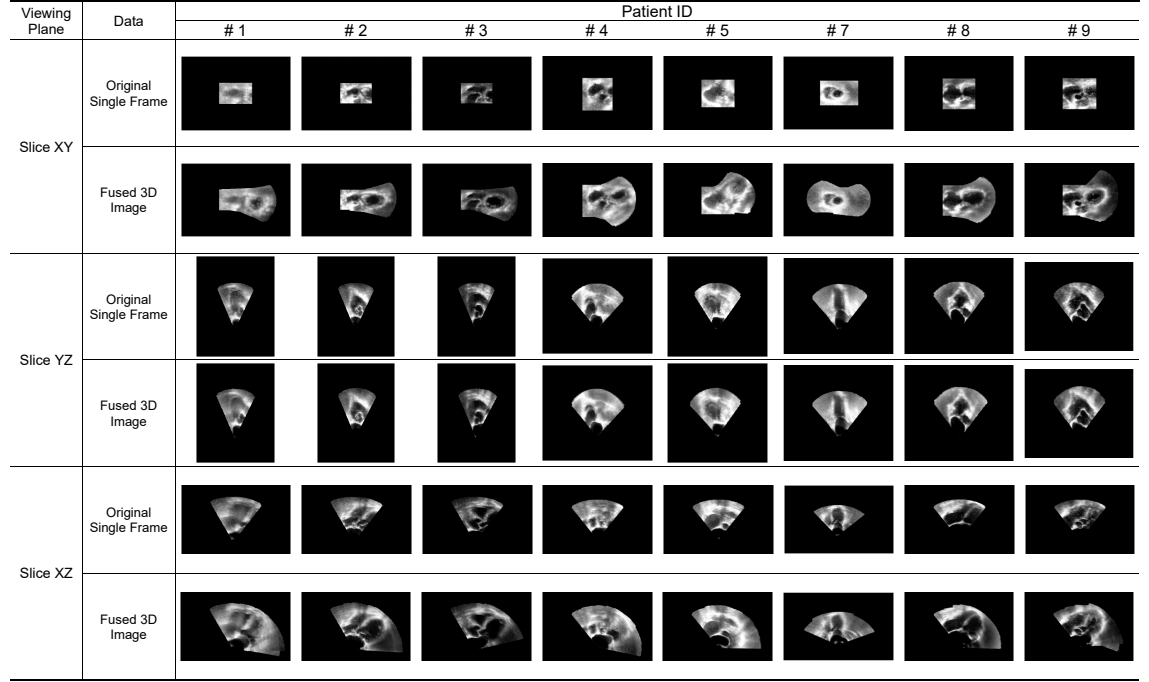


FIGURE 3.19: Comparison of original single frame of in-vivo 3D TEE image with fused 3D image using the proposed algorithms.

### 3.2.2.4 Image Fusion of In-vivo Data

In this section, we apply the three-plane registration method and the 3D fusion method to the 76 ECG-gated 3D TEE images captured from nine patients to conduct the final validation. The sequential fusion strategy is used to fuse the images into the panoramic image one by one for each dataset. The fusion process of the in-vivo 3D TEE images of patient # 6 is shown in Fig. 3.18. With the continuous fusion of the images, the FoV of the global image is enlarged gradually. By counting the number of voxels, it is found that the FoV of the final fused image is 2.04 times larger than the original one's. In Fig. 3.19, 3D TEE images of patient # 1, # 2, # 3, # 4, # 5, # 7, # 8, and # 9 are fused respectively and the final results are compared with the original single frame of images. The comparisons show that the FoV in the fused images is significantly enlarged. In addition, no ghosting or misalignment is found from the fused images, which indicates the images are aligned accurately using the estimated poses and fused with good quality. By counting the number of voxels, it is found that the FoV of the fused images is enlarged to 2.22, 2.05, 1.82, 2.09, 2.01, 1.70, 1.83, and 2.13 times as compared with the FoV of

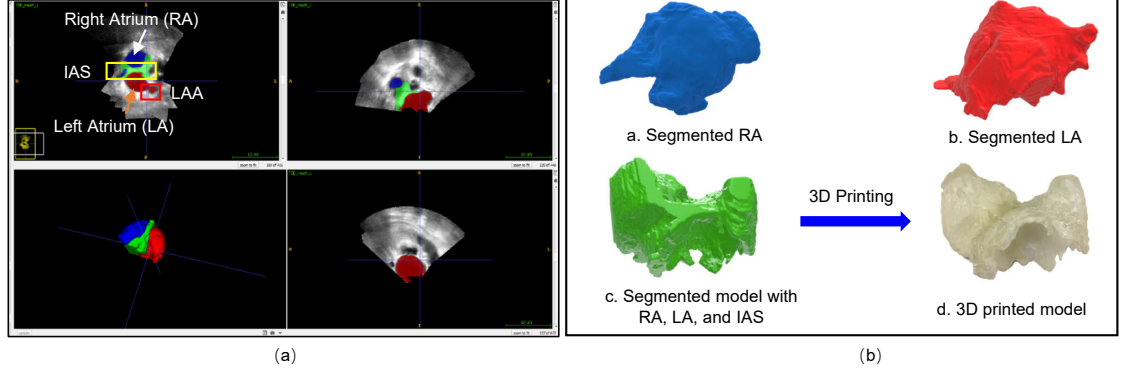


FIGURE 3.20: (a) Fused image is segmented manually; (b) a, b, c are segmented regions of interest and d is a 3D printed model.

the original single frame of TEE image of # 1, # 2, # 3, # 4, # 5, # 7, # 8, and # 9 respectively. The results are listed in the last column of Table 3.1.

In the process of data collection, we fixed the ultrasound parameters in the imaging system to avoid the need for additional processing of the data in our method. From the perspective of the algorithm, artifacts or operations which significantly change the appearances of the TEE images, such as big ultrasound imaging artifacts or adjusting gain in a wide range during the data collection, should be avoided since they may affect accuracy and robustness of the proposed method. But the operations that change the spatial resolution such as adjusting FoV or depth may not affect the algorithm since captured images can be converted to the same resolution with the known scale.

### 3.2.3 Clinical Applications of 3D TEE Fusion and Further Processing

Achieving a wider FoV in 3D TEE imaging has important clinical applications. In the planning of most cardiac surgery and intervention such as the LAA occlusion, it is crucial for cardiac structures of interest to be observed in a single image for a complete analysis of the spatial orientation of these structures. Limited by the FoV of standard 3D TEE, the IAS, LAA, and the mitral valve can hardly be imaged in their entirety as a single volume. Our techniques of direct 3D TEE fusion allow visualization of all these cardiac structures in a single volume, allowing measurement of distances and angles related to these structures, similar to CT and MRI.

Compared to direct visualization and measurement on 3D images, 3D-printed model is a more intuitive method for the planning of cardiac surgery and intervention. For example, in the planning of LAA occlusion, 3D-printed model is increasingly used in recent years to improve the accuracy of evaluating the size of LAA [88, 89]. A complete structure of the regions of interest is a prerequisite for 3D printing. Based on the final fused image of patient # 6, segmentation is performed manually to further evaluate the result [90]. For the first time, regions of interest are well observed and the whole LA with both the IAS and LAA in a single model is segmented successfully, as shown in Fig. 3.20. Segmentation of the fused ultrasound volumes allows modeling and 3D printing of the LA, IAS, and LAA as a single entity. In Fig. 3.20(b), a segmented model containing right atrium, LA, and IAS was 3D printed in a 1 : 1 scale on a high-resolution ( $32\ \mu\text{m}$ ) 3D printer Objet350 Connex3 (Stratasys, Eden Prairie, MN) using a translucent photopolymer material Agilus30 (Stratasys, Eden Prairie, MN) [91] in the hospital. It is very convenient to make measurements on the complete structure of LAA in the printed 3D model. In addition, simulation of surgical and transcatheter procedures is also performed on the model by clinicians as part of the preoperative planning for LAA occlusion [91, 92]. Importantly, the location of transseptal puncture and the angle for approaching the catheter inside the left heart can be precisely planned to maintain catheter coaxiality with the target structures in the in vitro setting resembling clinical implantation. This has an important advantage over performing procedural simulation on a limited anatomic model.

### 3.3 Chapter Summary

In this chapter, a novel multiview registration and fusion framework is proposed to enlarge the FoV of 3D TEE images efficiently. A direct method with efficient optimization is proposed for the pairwise registration of 3D TEE images. Fast implementation of the method is realized by using a three-orthogonal-plane approximation strategy. Compared with many intensity-based registration methods that use numerical Jacobian in the optimization process such as *imregtform* in MATLAB, an analytical expression of Jacobian is used in the proposed registration method, which shows higher accuracy and better convergence. Besides, a 3D fusion method is proposed to fuse images seamlessly and efficiently.

Monte-Carlo simulations and in-vivo experiments are performed to verify the effectiveness of the proposed methods. The results of the in-vivo experiments indicate that the proposed registration algorithm outperforms the state-of-the-art PCA-based registration method in terms of accuracy and efficiency. Image registration and fusion are performed on 76 in-vivo 3D TEE images from nine patients, and the results show apparent enlargement of FoV (enlarged around two times) in the final fused images.

The current framework uses ECG-gating to avoid the deformation of the heart in 3D TEE images, which requires synchronizing 3D TEE images to the ECG signal and builds the panoramic image at a selected phase of cardiac cycles. The next chapter will introduce a direct dynamic fusion framework that is able to construct the 3D TEE images with enlarged FoV at different phases by fusing multiview dynamic images sequentially, without using ECG-gating.

## Chapter 4

# Direct Dynamic Fusion for 3D Transesophageal Echocardiography

In Chapter 3, we proposed a framework that can enlarge the FoV of 3D TEE images effectively by fusing multiview ECG-gated images. ECG-gated images are obtained by synchronizing 3D TEE images to ECG signal at a specific phase of cardiac cycles such that deformations are avoided in these images. While ECG-gating simplifies the registration problem to a rigid one, the panoramic image is only constructed at the selected time point and motion information of the anatomy of interest [93] is not available in the fused images.

In this chapter, we propose a direct dynamic fusion framework (DDF) based on the embedded deformation graph to address the problem of limited FoV of 3D TEE images by sequentially fusing images captured at different phases of cardiac cycles and from different viewpoints. Different from the embedded deformation method in [48] and [49] that relies on explicit point constraints, the proposed non-rigid registration method in the framework uses intensity information as constraints directly to estimate the deformation, thus making use of all information from the images. Since registration and fusion are performed directly on the dynamic 3D TEE images, a 3D image with extended FoV is likely to be reconstructed at different phases of cardiac cycles and therefore, avoid the bias to the specifically chosen time. The main contributions of the work include: (1) we propose an direct embedded deformation method for deformable registration based on the intensity of



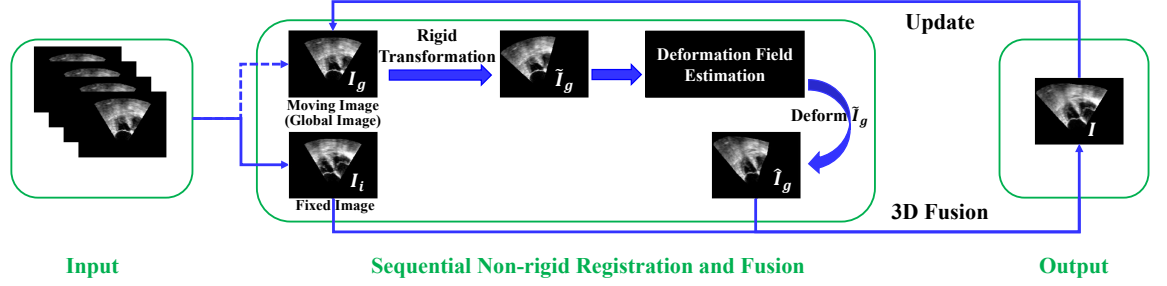


FIGURE 4.1: The framework of the direct dynamic fusion. (Remark: 3D TEE images are shown via 2D slices)

the medical images directly instead of specifying point constraints in [48] and [49]; (2) we propose a **DDF** framework that can fuse multiview dynamic 3D images to enlarge the **FoV** of 3D **TEE** images such that bias towards the specifically chosen phase in rigid registration and fusion can be avoided; (3) we evaluate the accuracy of the proposed method compared with the widely used B-spline based methods implemented in Plastimatch [46] and Elastix [47] respectively, the original Demons [50], the diffeomorphic Demons [53], and the state-of-the-art learning-based method VoxelMorph [74], qualitatively and quantitatively using in-vivo 3D **TEE** datasets.

## 4.1 Methods

Our **DDF** framework for enlarging the **FoV** of 3D **TEE** images with deformation is presented in Fig. 4.1, which mainly consists of non-rigid registration and 3D image fusion. A sequential fusion strategy is used to perform the registration and fusion repeatedly so that multiview 3D **TEE** images captured at different phases of cardiac cycles and different viewpoints can be fused one by one. Suppose there are  $m$  frames of 3D **TEE** images with various deformations to be fused. First, the global (panoramic) image  $I_g$  is initialized by the first image. And then, in each loop, we designate the global image as a moving image and every new coming image  $I_i$  as a fixed image and perform non-rigid registration and fusion. To deal with global motion between the moving and the fixed images, the global image  $I_g$  is rigidly transformed to an image space that is close to image  $I_i$  and denoted as  $\tilde{I}_g$  first. The rough relative pose used for the transformation is provided by SIFT3D [31]. Although the feature-based algorithm SIFT3D cannot precisely extract and

match key features from the TEE images, it is still useful for getting an initial pose to guide our registration. After pre-registration, the deformation field  $\Phi_g^i$  is estimated by using the proposed direct embedded deformation method. Then, the transformed image  $\tilde{I}_g$  is deformed using the deformation field and fused with image  $I_i$ . The fused image  $I$  is used as the new global image  $I_g$  for the next new coming image. The same processes are performed repeatedly until all the images are fused.

At the end of each loop, we reconstruct the global 3D TEE image at the same phase as the fixed image. Besides the area which can be observed in the fixed image, deformation in unobserved regions can also be predicted according to the principle of as-rigid-as-possible [94]. Therefore, through continuously fusing the global image with the remaining images, the FoV of 3D TEE images is enlarged.

#### 4.1.1 Direct Embedded Deformation

Our non-rigid registration method employs the embedded deformation graph to describe the deformation field, i.e. non-rigid transformation. But due to the relatively low signal-to-noise ratio of ultrasound imaging, it is difficult to extract and match feature points from the 3D TEE images for deformation estimation as in [49]. Therefore, we propose to use a direct method to compute the deformation by maximizing the similarity between the moving and the fixed images.

##### 4.1.1.1 Local Deformation

The embedded deformation graph is represented by a set of nodes associated with affine matrices. Assuming that the graph nodes are selected uniformly from  $\tilde{I}_g$  by downsampling it ( $\tilde{I}_g$  is obtained by pre-registering the moving image), let  $\mathbf{g}_r \in \mathbb{R}^3$  denote the position of node  $r$  and an associated affine transformation consists of a matrix  $\mathbf{A}_r \in \mathbb{R}^{3 \times 3}$  and a translation vector  $\mathbf{t}_r \in \mathbb{R}^3$ . The affine transformation applies a deformation to its nearby space. Conversely, for any voxel  $\mathbf{p} \in \mathbb{R}^3$  in image  $\tilde{I}_g$ , its deformation is influenced by its nearby nodes. To maintain the consistency of the deformation graph, nodes that are close to one another should have similar transformations. Thus, we limit the influence

of the deformation on voxel  $\mathbf{p}$  to its  $k$  nearest graph nodes to keep the consistency and efficiency of the deformation estimation. And the final deformation of point  $\mathbf{p}$  is defined as a weighted average of the effects of its nearby  $k$  nodes by:

$$\hat{\mathbf{p}} = \sum_{r=1}^k \hat{w}_r(\mathbf{p}) [\mathbf{A}_r(\mathbf{p} - \mathbf{g}_r) + \mathbf{g}_r + \mathbf{t}_r] \quad (4.1)$$

where weight  $\hat{w}_r(\mathbf{p})$  denotes the influence of a nearby node on the voxel,  $\sum_{r=1}^k \hat{w}_r = 1$ .  $\hat{w}_r$  can be pre-calculated by the following formula and then normalized.

$$w_r(\mathbf{p}) = (1 - \|\mathbf{p} - \mathbf{g}_r\|/d_{max}) \quad (4.2)$$

where  $d_{max}$  represents the distance from the voxel to the  $(k+1)_{th}$  nearest node. We use  $k = 6$  for all experiments presented in this chapter.

#### 4.1.1.2 Energy Function

To estimate the parameters  $\Phi_g^i = \{\mathbf{A}_1, \mathbf{t}_1 \dots \mathbf{A}_{m_g}, \mathbf{t}_{m_g}\}$  for a deformation graph with  $m_g$  nodes, an energy function as the following is minimized:

$$E(\Phi_g^i) = w_{rot}E_{rot}(\Phi_g^i) + w_{reg}E_{reg}(\Phi_g^i) + w_{con}E_{con}(\Phi_g^i), \quad (4.3)$$

where  $E_{rot}(\Phi_g^i)$ ,  $E_{reg}(\Phi_g^i)$ , and  $E_{con}(\Phi_g^i)$  represent rotation term, regularization term, and constraint term respectively and  $w_{rot}$ ,  $w_{reg}$ , and  $w_{con}$  are the corresponding weights. In this energy function, rotation and regularization terms are used to guide the optimization to a reasonable and smooth deformation [48, 49]. And the constraint term constrain the deformation by maximizing the similarity between the fixed and moving images. Additionally, the number of graph nodes  $m_g$  determines the expressibility of the deformation graph. The more nodes are selected from one image, the more details that the graph could express. According to our tests, we select around 1000 nodes from a moving image that contains around five million of voxels for pairwise registration, which could balance the accuracy and efficiency of the algorithm.

**Rotation Term:** To preserve the topological structure of the deformed images, the objective function (4.3) directly specifies a matrix  $\mathbf{A}_r$  at each node should be close to a rotation matrix [95] and local voxels are deformed as rigidly as possible. A rotation matrix in 3D space is subject to six constraints, i.e. each column should be a unit length and should be orthogonal to one another. Thus, the rotation term sums the rotation error over all affine transformations of the deformation graph [48]:

$$E_{\text{rot}}(\Phi_g^i) = \sum_{r=1}^{m_g} \text{Rot}(\mathbf{A}_r) \quad (\text{where } \mathbf{A}_r = [\mathbf{a}_{r1}, \mathbf{a}_{r2}, \mathbf{a}_{r3}]) \quad (4.4)$$

$$\begin{aligned} \text{Rot}(\mathbf{A}_r) = & \left( \mathbf{a}_{r1}^\top \cdot \mathbf{a}_{r2} \right)^2 + \left( \mathbf{a}_{r1}^\top \cdot \mathbf{a}_{r3} \right)^2 + \left( \mathbf{a}_{r2}^\top \cdot \mathbf{a}_{r3} \right)^2 \\ & + \left( \mathbf{a}_{r1}^\top \cdot \mathbf{a}_{r1} - 1 \right)^2 + \left( \mathbf{a}_{r2}^\top \cdot \mathbf{a}_{r2} - 1 \right)^2 + \left( \mathbf{a}_{r3}^\top \cdot \mathbf{a}_{r3} - 1 \right)^2. \end{aligned} \quad (4.5)$$

**Regularization Term:** To keep the consistency and smoothness of the deformation graph, the position of node  $r$  deformed by nearby  $k$  nodes should be conformed with the actual position calculated by node  $r$  itself. Thus, the regularization term sums the squared error between predicted positions of nodes and their actual positions [48]:

$$E_{\text{reg}}(\Phi_g^i) = \sum_{r=1}^{m_g} \sum_{s \in \mathbb{N}(r)} \alpha_{rs} \|\mathbf{A}_r(\mathbf{g}_s - \mathbf{g}_r) + \mathbf{g}_r + \mathbf{t}_r - (\mathbf{g}_s + \mathbf{t}_s)\|^2, \quad (4.6)$$

where  $\mathbb{N}(r)$  denotes the index of nearby  $k$  nodes which influences node  $r$ . And we also follow [48] to set weight  $\alpha_{rs}$  as 1. In the energy function (4.3), increasing the weight of  $E_{\text{reg}}(\Phi_g^i)$  will make the non-rigid transformation become closer to the rigid transformation.

**Constraint Term:** Our method estimates the deformation by directly minimizing intensity differences between the two images instead of specifying point constraints [48, 49]. Suppose the intensity value of a voxel  $\mathbf{p}$  in image  $I$  is a function w.r.t. its coordinates and denoted as  $I(\mathbf{p})$ , for a spatial point  $\mathbf{p}_j$  in the heart and its projections  $\mathbf{p}_{gj}$  and  $\hat{\mathbf{p}}_{ij}$  in image  $\tilde{I}_g$  and  $I_i$ , the intensity difference between  $\mathbf{p}_{gj}$  and  $\hat{\mathbf{p}}_{ij}$  can be written as:

$$e_j = \tilde{I}_g(\mathbf{p}_{gj}) - I_i(\hat{\mathbf{p}}_{ij}). \quad (4.7)$$

Then, we can sum the squared intensity differences over  $n$  voxels in overlapping areas between the two images as the constraint term:

$$E_{\text{con}}(\Phi_g^i) = \sum_{j=1}^n \|e_j\|^2. \quad (4.8)$$

In this work, we use iterative optimization to obtain the optimal solution of (4.3). In every iteration,  $\hat{\mathbf{p}}_{ij}$  is computed from  $\mathbf{p}_{gj}$  via (4.1). Since  $\hat{\mathbf{p}}_{ij}$  may not be integers and thus not on the grid of the fixed image  $I_i$ , the intensity of  $I_i(\hat{\mathbf{p}}_{ij})$  is obtained using trilinear interpolation to reduce the error for difference computation in (4.7).

#### 4.1.1.3 Optimization

Minimizing (4.3) is a routine non-linear least-squares (NLLS) problem and the solution can be obtained by using Gauss-Newton (GN) method [80]. Firstly, we rewrite the energy function (4.3) as a standard NLLS format as follows

$$E(\Phi_g^i) = f(\Phi_g^i)^\top \Sigma f(\Phi_g^i), \quad (4.9)$$

where  $\Sigma$  is the weight matrix of (4.3) and  $f(\Phi_g^i)$  is the cost vector that concatenates the rotation, regularization, and constraint terms together.

Let  $J(\Phi_g^i)$  denote the Jacobian matrix of  $f(\Phi_g^i)$  w.r.t. deformation parameter  $\Phi_g^i$ . Then, the optimal solution  $\hat{\Phi}_g^i$  can be obtained by initializing with  $\Phi_0$  and iterating with  $\Phi_g^i \leftarrow \Phi_g^i + \Delta\Phi_g^i$ . In our algorithm, the initial value of  $\Phi_g^i$  is set as  $\Phi_0 = \{\mathbf{I}_1, \mathbf{0}_1, \dots, \mathbf{I}_{m_g}, \mathbf{0}_{m_g}\}$ , where  $\mathbf{I} \in \mathbb{R}^{3 \times 3}$  and  $\mathbf{0} \in \mathbb{R}^3$  are identity matrix and column vector of zeros, respectively. The step change  $\Delta\Phi_g^i$  in each iteration can be obtained from the GN equation:

$$J^\top(\Phi_g^i) \Sigma J(\Phi_g^i) \Delta\Phi_g^i = -J^\top(\Phi_g^i) \Sigma f(\Phi_g^i). \quad (4.10)$$

The steps of deformation estimation is summarized in Algorithm 2.

After the optimal deformation parameters  $\hat{\Phi}_g^i$  is obtained, the deformed image  $\hat{I}_g$  can be obtained by deforming  $\tilde{I}_g$  using (4.1). And then, the deformed image  $\hat{I}_g$  is fused with the fixed image  $I_i$  by using a 3D fusion method proposed in Chapter 3.

**Algorithm 2:** Direct Embedded Deformation**Input:** A fixed image  $I_i$  and a moving image  $I_g$  (global image).**Output:** Non-rigid transformation  $\Phi_g^i$  from  $I_g$  to  $I_i$  and deformed image  $\hat{I}_g$ .

- 1 **Step 1:** Pre-computation:
- 2 Pre-register  $I_g$  to  $I_i$  rigidly using the pose from SIFT3D [31], denoted as  $\tilde{I}_g$ ;
- 3 Downsample  $\tilde{I}_g$  evenly to get a graph with  $m_g$  nodes  $\{\mathbf{g}_r\}_{r=1}^{m_g}$ ;
- 4 Calculate mask matrices  $mask_g$  and  $mask_i$  for  $\tilde{I}_g$  and  $I_i$ , respectively;
- 5 Calculate gradient space of intensity of  $I_i$ ;
- 6 Interpolate intensity and gradient spaces of  $I_i$ ;
- 7 **Step 2:** Optimization:
- 8 Initialize the transformation  $\Phi_g^i \leftarrow \Phi_0$ ;
- 9 **while** *Algorithm not converged* **do**
- 10     Calculate cost vector  $f(\Phi_g^i)$  over rotation, regularization, and constraint terms;
- 11     Calculate Jacobian matrix  $J(\Phi_g^i)$ ;
- 12     Construct weight matrix  $\Sigma$  of the energy function (4.3);
- 13     Calculate step change  $\Delta\Phi_g^i$  via (4.10) ;
- 14     Update the pose  $\Phi_g^i \leftarrow \Phi_g^i + \Delta\Phi_g^i$ ;
- 15 **end**
- Calculate deformed image  $\hat{I}_g$  by (4.1);

## 4.2 Experiments and Results

Since we aim to solve the cardiac motion in 3D TEE images, in-vivo datasets are used to validate our method. Same as the data collection process introduced in Section 3.2.2, in-vivo datasets for the experiments are collected from patients using a real-time 3D transducer. Based on the finding in Chapter 3 about the impact of image overlapping rate on the algorithm performance, consecutive 3D TEE images collected at different viewpoints are controlled to have more than 50% overlap. In total, twenty-nine sequences of 3D TEE images with various deformations are collected from four patients. The details of the in-vivo datasets from the four patients # 1, # 2, # 3, and # 4 are listed in Table 4.1.

### 4.2.1 Experimental Setup and Evaluation

The proposed method is validated by comparing it with two B-spline based methods, two Demons-based methods, and the learning-based method VoxelMorph [74]. For the B-spline based and Demons-based methods, we use the implementations of them in 3D Slicer [96] (version 4.10) for comparisons. Since the two B-spline based methods are

TABLE 4.1: Details of the Four In-vivo Datasets

Patient	No. Sequences	Volume Size (voxel)	Resolution (mm/voxel)
# 1	9	$240 \times 160 \times 208$	$0.69 \times 0.98 \times 0.73$
# 2	6	$240 \times 160 \times 208$	$0.77 \times 1.11 \times 0.82$
# 3	6	$240 \times 160 \times 208$	$0.64 \times 0.92 \times 0.68$
# 4	8	$272 \times 208 \times 208$	$0.63 \times 0.90 \times 0.68$

implemented based on the software package Plastimatch [46] and ElastiX [47] respectively, the packages' names are used to indicate the methods. Additionally, the original Demons and the diffeomorphic Demons methods in 3D Slicer are implemented based on Insight Segmentation and Registration Toolkit (ITK) [97]. The codes of VoxelMorph published by the author are used in the experiments and the smoothness regularization parameter  $\lambda$  is set as 0.02 following the suggestion of the paper [74]. For the proposed method, coefficient for each term in the energy function (4.3) is specified as:  $w_{\text{rot}} = 1$ ,  $w_{\text{reg}} = 10000$ ,  $w_{\text{con}} = 0.1$ . The minimum step length and the maximum number of iterations are used as stop criteria for convergence, which are set as  $10^{-5}$  and 100, respectively. The following two groups of in-vivo experiments are performed to validate the proposed DDF framework.

- (i) Pairwise registration experiments: in each experiment, two 3D TEE images collected at different cardiac phases and from different viewpoints in one patient are used for pairwise registration. Five pairs of 3D TEE images from the four patients are registered by using the proposed method and the other five non-rigid registration methods. To mitigate the influences of global motion on the registration, rigid transformation obtained from SIFT3D [31] is used to do the pre-registration for the six methods in each experiment (all methods use the same rigid transformation to perform the pre-registration in the same experiment). For VoxelMorph, we train each patient's dataset individually with 3D TEE images of more than one cardiac cycle until the algorithm converges. Then, the trained models are used for registration. Since VoxelMorph cannot deal with non-overlapping areas between images [74], only overlapping areas between the moving and fixed images are used for non-rigid registration. The detailed results of pairwise registration are reported in Section 4.2.2.

- (ii) Sequential fusion experiments: firstly, twenty-nine ECG-gated TEE volumes are extracted from the twenty-nine sequences by synchronizing acquired 3D TEE images to ECG signal at late diastole. These ECG-gated 3D TEE images extracted from the same dataset are static (i.e., deformations in images are eliminated). The static images are fused together and the results are used as the references for evaluating the results obtained by proposed and competing methods. Secondly, we perform dynamic fusion on the selected volumes sequentially by using the DDF and competing methods, respectively. The final volume is reconstructed at the same phase as the ECG-gated volume so that the fused volumes from both static fusion and dynamic fusion should be at the same phase and have the same images. Finally, the fused images obtained from the dynamic fusion are compared with the images from the static fusion to evaluate the proposed method. In the second step of dynamic fusion, we select one volume from each sequence in each dataset (from one patient) with different deformations and therefore, four sequences containing eight, six, six, and nine 3D TEE images respectively are obtained. In each sequence, the final volume is selected as an ECG-gated one to make sure the final fused images are at the same phase as the referenced ones. The detailed results of sequential fusion are reported in Section 4.2.3

Verifying accuracy is difficult in non-rigid registration since ground-truth deformation is usually not available in practical data [98]. In addition to checking the quality of the aligned images visually, there are some quantitative metrics used in many works of literature [99, 100] to assess the accuracy of non-rigid registration, such as computing the distance between the segmented anatomical structures or overlap of the segmented anatomical structures after registration. Since different metrics have different sensitivity to the true deformation [99], the following four metrics are used jointly in this chapter to evaluate the accuracy of registration and fusion:

- (i) The alignment of stitches in the checkerboard volumes [101]. A checkerboard volume is created by taking a part from each of the two aligned images and stitching them together. By checking alignment at the stitches, we can assess the accuracy of registration qualitatively.



- (ii) Signed distance between segmented LA walls. LA walls are sharp boundaries that can be easily recognized in the 3D TEE images. We segment the LA walls from both the deformed/fused images and their reference images manually and then, calculate the signed distance of each vertex on the LA walls segmented from deformed/fused images to the closest vertex on the LA walls segmented from the reference images. The results are displayed via heatmaps in Section 4.2.2 and 4.2.3.
- (iii) Mean absolute distance (MAD) between segmented LA walls. MAD is a statistical value that is calculated by averaging the absolute value of the metric (ii).
- (iv) Dice similarity coefficient (DSC) [100] between segmented anatomical structures from the fused images and those from the reference images. Generally, DSC is calculated by

$$\text{DSC}(M, N) = \frac{2|M \cap N|}{(|M| + |N|)}, \quad (4.11)$$

where  $M$  and  $N$  represent the target regions of the two segmentations. DSC computes the overlap of segmented anatomical structures. The maximum value of DSC is 1 and a higher DSC indicates a better accuracy of registration and fusion.

In addition,  $P$ -value approach is used to perform paired  $t$ -tests to examine the statistical significance of the results with the significance level of 0.05.

#### 4.2.2 Pairwise Non-rigid Registration

We jointly use the qualitative and quantitative metrics to evaluate the accuracy of the proposed algorithm as comprehensively as possible. The qualitative evaluation of pairwise registration accuracy is performed by checking the alignment of stitches in the checkerboard volumes. The checkerboard volumes based on the results from different methods are shown in Fig. 4.2. Through checking the stitches in sharp areas such as the walls of LA in the checkerboard volumes, obvious misalignments which are indicated by red dashed cycles can be found in images obtained from the two B-spline based methods and VoxelMorph method, especially in the images obtained from the Plastimatch method. In contrast, checkerboard volumes from the two Demons-based and the proposed method have smooth transitions in the stitching areas. No obvious misalignment can be found in

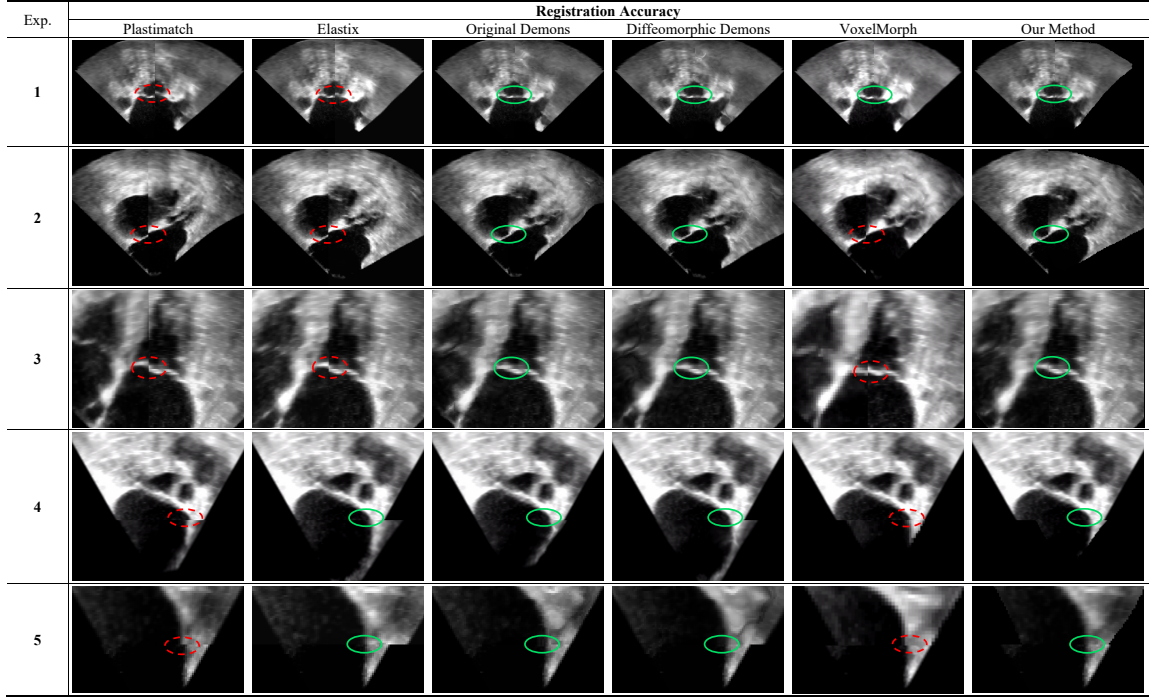


FIGURE 4.2: Checkerboard volumes obtained by stitching images after registration. The same positions at the stitches are marked by circles and obvious misalignments are indicated by the red dashed circles in the images.

the results from these three methods, which indicates that the Demons-based methods and the proposed method are more accurate than the compared B-spline based methods and VoxelMorph method. Since the original Demons method cannot guarantee the topology in the deformed images, some defects are found from the images deformed by the method. One example is shown in Fig. 4.3 (the defects are indicated by the red dashed cycles). Please note, one clear error in anatomical structure is also found from the result obtained by the diffeomorphic Demons method in Fig. 4.3, which is indicated by the white arrow. Such conditions are not found from the results obtained from the proposed method in our experiments.

Furthermore, to assess the registration accuracy quantitatively, we first manually segment the LA walls from both deformed and fixed images in each pair and then calculate the distance between them [102]. Un-overlapping areas and areas prone to segmentation errors such as mitral valves are omitted to reduce computational and human errors. Both the signed distance and MAD are calculated and shown in Fig. 4.4. The corresponding statistical values of MAD (provided by mean  $\pm$  standard deviation) are shown in Table 4.2.

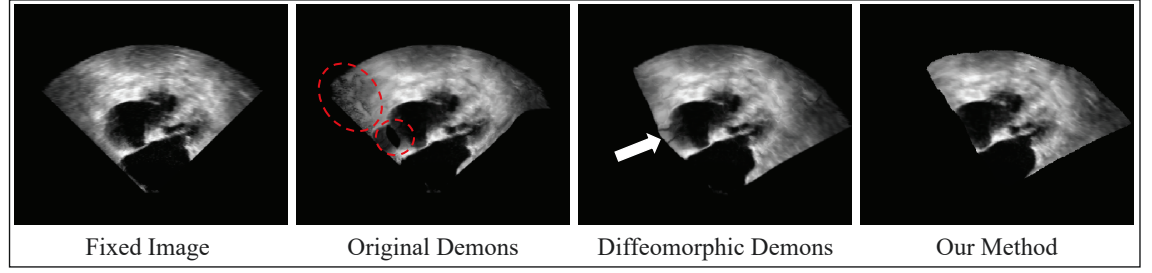


FIGURE 4.3: Fixed image and deformed moving images. The defects in the volume obtained from the original Demons are indicated by the red dashed cycles. And the white arrow indicates one obvious error in the volume obtained from the diffeomorphic Demons.

Exp.	Registration Accuracy					
	Plastimatch	Elastix	Original Demons	Diffeomorphic Demons	VoxelMorph	Our Method
1						
MAD	1.15 mm	0.95 mm	0.46 mm	0.29 mm	0.83 mm	0.10 mm
2						
MAD	1.61 mm	0.70 mm	0.42 mm	0.17 mm	0.48 mm	0.07 mm
3						
MAD	0.90 mm	0.40 mm	0.17 mm	0.17 mm	0.31 mm	0.03 mm
4						
MAD	1.61 mm	0.38 mm	0.12 mm	0.08 mm	0.38 mm	0.08 mm
5						
MAD	1.35 mm	0.67 mm	0.40 mm	0.29 mm	0.71 mm	0.07 mm

FIGURE 4.4: Signed distance and MAD between LA walls segmented from the deformed and fixed images.

Similar to our qualitative evaluation, it is found from Fig. 4.4 that the distribution of signed distance from the two Demons-based methods and our method is concentrated near zero and within a smaller range than that from the two B-spline based methods and VoxelMorph, which indicates the higher accuracy of the two Demons-based methods and our method than the other three. Similarly, it is also shown from Fig. 4.4 that Elastix method and VoxelMorph have higher accuracy than Plastimatch method. In addition, from the statistical values shown in Table 4.2, the MAD of the proposed method, the original Demons method, the diffeomorphic Demons method, and VoxelMorph are around

TABLE 4.2: Comparisons of Statistical Values of MAD for Pairwise Registration

	Mean $\pm$ Std (mm)	$P$ -value
Plastimatch	$1.32 \pm 0.31$	$3.41 \times 10^{-4}$
Elastix	$0.62 \pm 0.24$	$2.50 \times 10^{-3}$
Original Demons	$0.31 \pm 0.16$	$9.90 \times 10^{-3}$
Diffeomorphic Demons	$0.20 \pm 0.09$	$1.39 \times 10^{-2}$
VoxelMorph	$0.54 \pm 0.22$	$3.30 \times 10^{-3}$
Our Method	<b><math>0.07 \pm 0.03</math></b>	-

$0.07 \pm 0.03$  mm,  $0.31 \pm 0.16$  mm,  $0.20 \pm 0.09$  mm, and  $0.54 \pm 0.22$  mm respectively, while the results from the two B-spline based methods are greater than 0.50 mm and the results from Plastimatch method are even greater than 1.00 mm.  $P$ -values of paired  $t$ -tests are less than 0.05, which indicates that the proposed method outperforms the other five methods in terms of accuracy with statistical significance.

In summary, it is found from the comparisons that: 1) the proposed method has the best accuracy (less than 0.10 mm) in pairwise registration and can preserve the topology in the deformed images; 2) the diffeomorphic Demons is more accurate than the original Demons method and can preserve topology in the images; 3) both the proposed method and Demons-based methods are more accurate than the B-spline based methods and VoxelMorph method. Since the proposed method and the diffeomorphic Demons outperformed the other methods in terms of accuracy and VoxelMorph cannot deal with non-overlapping areas, in the next section, the proposed method will be evaluated by comparing it with the diffeomorphic Demons method only.

### 4.2.3 Sequential Fusion

The four sequences of 3D TEE images with deformations are registered and fused respectively by using the proposed DDF method and the diffeomorphic Demons method. The results of the static fusion are used as references for evaluation. Fig. 4.5 shows the final fused 3D volume of the sequence from Patient # 4. Through the comparisons, it is found that the fused image obtained from the proposed method has almost identical anatomical structures as the image from the static fusion. However, obvious accumulating errors are found in the volume fused by the diffeomorphic Demons method, which are indicated by

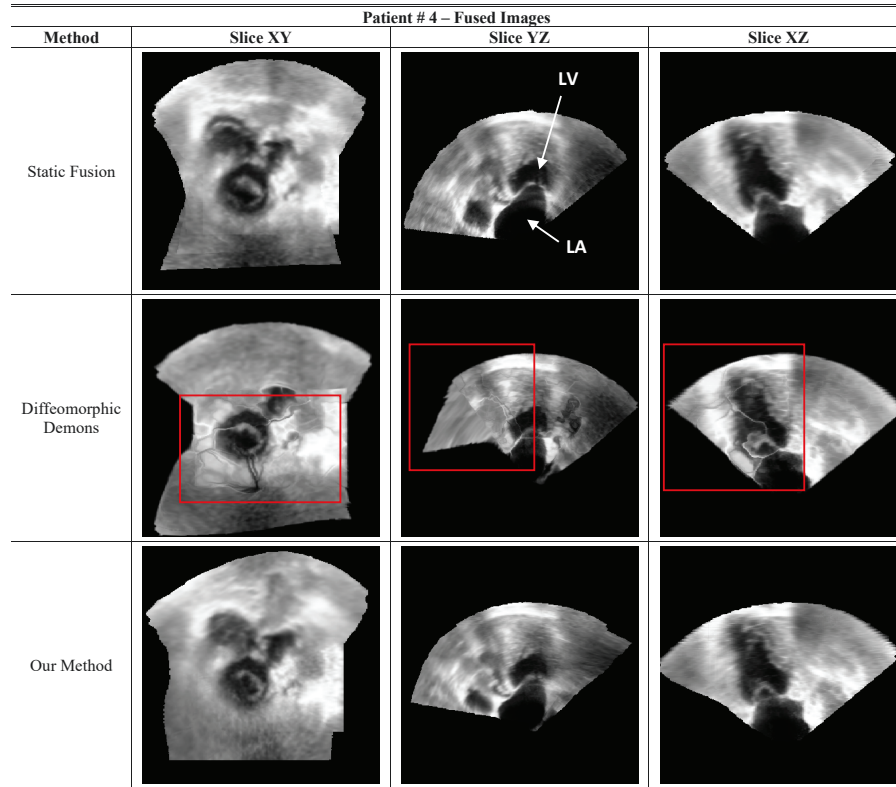


FIGURE 4.5: Comparisons of the fused volume of the sequence from Patient # 4. Obvious accumulating errors are marked by red boxes in the images.

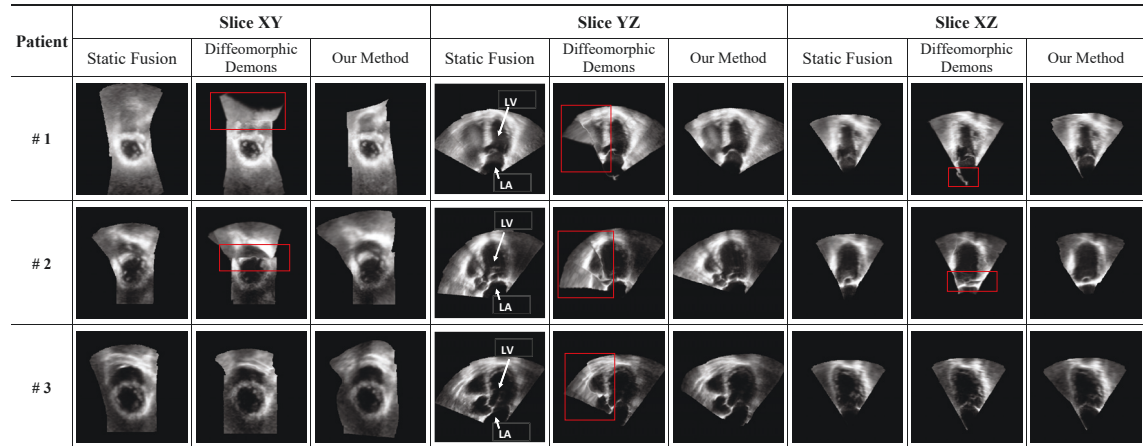


FIGURE 4.6: Comparisons of the fused volumes of sequences from Patients # 1, # 2, and # 3. Obvious accumulating errors are marked by red boxes in the images.

red boxes on images in the figure. Similar results are found from the 3D volumes fused by the other three sequences, which are shown in Fig. 4.6.

In addition, similar to Section 4.2.2, distances from the segmented LA walls obtained from

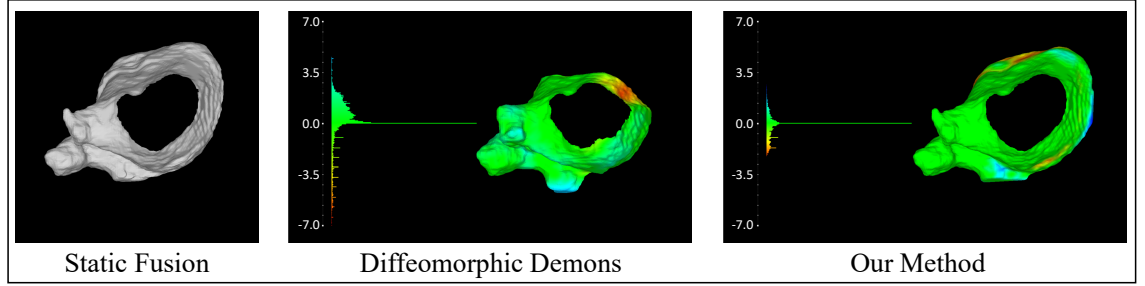


FIGURE 4.7: Signed distance from the segmented LA walls obtained from the diffeomorphic Demons and the proposed method to that obtained from the static fusion.

TABLE 4.3: Comparisons of Sequential Fusion Accuracy and the Enlarged FoV after Fusion

	MAD (mm)		DSC of LV		FoV w.r.t. Original
	Diffeomorphic Demons	Our Method	Diffeomorphic Demons	Our Method	
# 1	0.81	0.17	0.91	0.94	1.97
# 2	1.42	0.19	0.86	0.88	2.06
# 3	0.30	0.21	0.74	0.79	1.84
# 4	1.04	0.18	0.78	0.89	1.83
Mean $\pm$ Std	$0.89 \pm 0.47$	<b><math>0.19 \pm 0.02</math></b>	$0.82 \pm 0.08$	<b><math>0.88 \pm 0.06</math></b>	<b><math>1.93 \pm 0.11</math></b>
<i>P</i> -value	<b>0.03</b>	-	<b>0.04</b>	-	-

the two dynamic fusion methods to those obtained from the static fusion are calculated respectively. An example of the signed distance is presented in Fig. 4.7, which shows that the LA wall segmented from the image fused by the proposed method is more accurate than that segmented from the image fused by the diffeomorphic Demons method. The MAD for each sequence and the corresponding statistical results are summarized in Table 4.3. Furthermore, since the left ventricle (LV) can be observed completely in the reconstructed images (see Fig. 4.5 and Fig. 4.6), DSC value between the segmented LV from the dynamic fusion and that from the static fusion is also calculated for each experiment and summarized in Table 4.3. The results show that the MAD of the results obtained from the proposed method are around  $0.19 \pm 0.02$  mm, while the results from the diffeomorphic Demons are around  $0.89 \pm 0.47$  mm. Moreover, DSC values calculated on the results from the proposed method ( $0.88 \pm 0.06$ ) are greater than those calculated on the results from the diffeomorphic Demons method ( $0.82 \pm 0.08$ ). *P*-values of paired *t*-tests indicate that the proposed method outperforms diffeomorphic Demons in terms of accuracy with statistical significance.

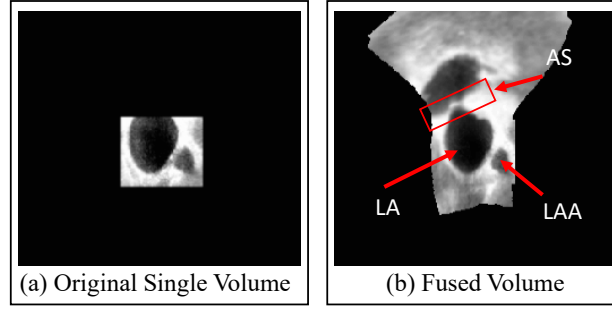


FIGURE 4.8: Comparison of the original single volume (a) and the fused volume (b).

It is found from the experiments that the proposed method can enlarge the **FoV** of the single volume effectively. One example of comparison of the original single volume and the final fused volume is shown in Fig. 4.8. By comparing the number of voxels in the original single volume and the fused volume, it is found that the **FoV** of the fused images is enlarged to 1.97, 2.06, 1.84, and 1.83 times as compared with the original single volume of **TEE** image of # 1, # 2, # 3, and # 4 respectively, which are listed in the last column of Table 4.3. Please note, the fused results from diffeomorphic Demons have obvious accumulating errors which indicate the method fails to fuse such a number of images. Therefore, the **FoV** of the fused images from diffeomorphic Demons is not comparable with those obtained by our method.

From the example shown in Fig. 4.8, one can see that the complete structures of both the atrial septum (**AS**) and **LAA** are clearly visible in the fused image, which is beneficial to the preoperative planning of transcatheter **LAA** occlusion. Furthermore, since the proposed method is capable of reconstructing 3D **TEE** images at various phases of cardiac cycles, it makes motion estimation [93, 103] in the fused images possible. In clinical practices, it is often desirable to estimate the motion of the anatomical structures such as to build motion and deformation models for surgical or therapeutic planning and guidance and to analyze motion-related markers of diseases quantitatively.

### 4.3 Chapter Summary

In this chapter, a novel direct dynamic fusion (**DDF**) framework is proposed to enlarge the **FoV** of 3D **TEE** images by fusing multiview images captured at different phases of cardiac



cycles sequentially. The in-vivo experiments are performed to evaluate our method compared with the other five widely used non-rigid registration methods, namely two B-spline based methods implemented in Plastimatch [46] and ElastiX [47], the original Demons [50], the diffeomorphic Demons [53], and the state-of-the-art learning-based method Vox-elMorph [74]. In the performed pairwise and sequential fusion experiments, qualitative and quantitative results show that the proposed method outperforms these five methods in terms of accuracy. The final fused volumes of the four patients show that the proposed method can reconstruct the 3D TEE images with good quality and enlarge the FoV of images in the fused volumes around two times compared with that in the original single volume.

In Chapter 3 and 4, we have presented two frameworks to enlarge the FoV of 3D TEE images in rigid and deformable scenarios, respectively. While the proposed frameworks have shown superior accuracy than the state-of-the-art methods, these frameworks are deduced from the pairwise registration which inevitably brings in accumulating errors in the obtained results. In the next part of the thesis, we will introduce three simultaneous registration and fusion algorithms to further improve the accuracy of registration and fusion in rigid and dynamic scenarios.



## Part II

# SIMULTANEOUS REGISTRATION AND FUSION

## Chapter 5

# Direct Bundle Adjustment for Fusing Multiple 3D Images

In Chapter 3, we enlarge the FoV of 3D TEE images through combining pairwise registration with the sequential fusion strategy. While we have shown the effectiveness and efficiency of the method over the competing methods, it inevitably involves accumulating errors as the fusion process proceeds. In this chapter, we propose our third algorithm named direct bundle adjustment (DBA) to further improve the accuracy of registration and fusion of multiview ECG-gated 3D TEE images. This algorithm simultaneously optimizes the global pose parameters of images and the intensity values of the fused panoramic image using the 3D image data directly. This one-step 3D image fusion approach is achieved by formulating the problem as an optimization problem to minimize the intensity differences between the predefined panoramic image and local images. We validate the proposed method via simulated and in-vivo 3D TEE images. It is shown that the proposed method is robust to intensity noises and much more accurate than the sequential fusion method presented in Chapter 3. It should be noted that although the method is called DBA, it is quite different from most of the bundle adjustment (BA) algorithms which belong to the feature-based methods, require feature correspondences between images, and are mostly for using 2D images to estimate the 3D feature positions (or shapes) in the reconstruction of the environments [18–20]. The novel DBA algorithm presented in this chapter has filled an important gap in 3D image registration and fusion. Without any reference

image, correspondences, or information loss or reuse, the proposed DBA algorithm is an elegant way to obtain the optimal panoramic image and poses of local images. To our best knowledge, this is the first work of BA based on direct method for 3D image registration and fusion.

## 5.1 Methods

### 5.1.1 Problem Statement

Fig. 5.1 illustrates a scenario in which there are  $m$  frames of 3D images and each image captures only a portion of the actual scene. We use the vector elements  $\xi_i \in \mathbb{R}^6$  of Lie algebra [104] to represent the pose parameters of a 3D image  $I_i$  at position  $i$ . Suppose  $M$  is the panoramic image which fuses all the information of local images, it consists of  $n$  voxels and the intensity of one voxel  $\mathbf{p}_j$  in  $M$  is denoted as  $M(\mathbf{p}_j)$  ( $\mathbf{p}_j = [u, v, w]^\top$ , where  $u, v, w$  are integers which present a voxel location on the 3D grid). Assuming that a part of intensity information of voxel  $\mathbf{p}_j$  in the panoramic image is from a point  $\mathbf{p}_{ij}$  in the local image  $I_i$ , in the rigid scenario, transformation in Euclidean space from  $\mathbf{p}_j$  to  $\mathbf{p}_{ij}$  can be

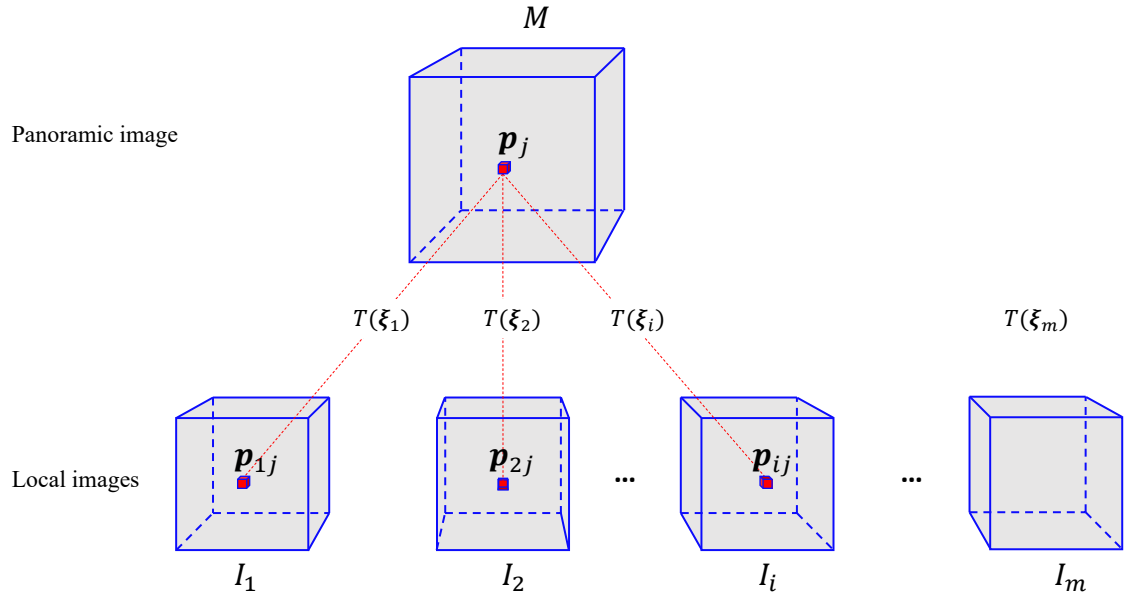


FIGURE 5.1: Direct bundle adjustment (DBA): intensity of panoramic image  $\{M(\mathbf{p}_j)\}$  and pose parameters of local frames  $\{\xi_i\}$  are optimized simultaneously.

formally written as:

$$\mathbf{p}_{ij} = f(\boldsymbol{\xi}_i, \mathbf{p}_j) = T(\boldsymbol{\xi}_i)\mathbf{p}_j, \quad (5.1)$$

where  $T(\cdot) \in SE(3)$  maps the pose parameters  $\boldsymbol{\xi}_i$  to a rigid transformation matrix in Euclidean space.

In this study, we seek to optimize the pose parameters of local images and the intensities of voxels in the panoramic image simultaneously. For a  $n$ -voxel panoramic image  $M$  and  $m$  frames of 3D images, the overall state parameters considered in the proposed algorithm are:

$$\mathbf{x} = [\mathbf{x}_\xi^\top, \mathbf{x}_M^\top]^\top, \quad (5.2)$$

where  $\mathbf{x}_\xi = [\boldsymbol{\xi}_1^\top, \dots, \boldsymbol{\xi}_i^\top, \dots, \boldsymbol{\xi}_m^\top]^\top$  are pose parameters of the local images, and  $\mathbf{x}_M = [M(\mathbf{p}_1), \dots, M(\mathbf{p}_j), \dots, M(\mathbf{p}_n)]^\top$  represent intensities of voxels in the panoramic image  $M$ .

### 5.1.2 Formulation of Direct Bundle Adjustment

In Fig. 5.1, the intensity difference between  $M(\mathbf{p}_j)$  and  $I_i(\mathbf{p}_{ij})$  is:

$$\begin{aligned} e_{ij}(\boldsymbol{\xi}_i, M(\mathbf{p}_j)) &= M(\mathbf{p}_j) - I_i(f(\boldsymbol{\xi}_i, \mathbf{p}_j)) \\ &= M(\mathbf{p}_j) - I_i(\mathbf{p}_{ij}), \end{aligned} \quad (5.3)$$

where  $\mathbf{p}_j$  is constant and  $\mathbf{p}_{ij}$  can be calculated from  $\mathbf{p}_j$  using (5.1) during the optimization process.  $\mathbf{p}_{ij}$  may not be integers thus not on the voxel of the grid of the local image. Therefore, the intensity of  $I_i(\mathbf{p}_{ij})$  is obtained using interpolation to reduce the error of the intensity difference computation.

Based on the direct method, we propose the DBA method to obtain the optimal estimation of  $\mathbf{x}$  by minimizing the sum of the squared intensity differences between panoramic image and local images:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}_\xi, \mathbf{x}_M}{\operatorname{argmin}} \sum_{j=1}^n \sum_{i=1}^m \sigma(\mathbf{p}_{ij}) (e_{ij}(\boldsymbol{\xi}_i, M(\mathbf{p}_j)))^2, \quad (5.4)$$

where  $\sigma(\mathbf{p}_{ij}) = 1$  if the transformed point  $\mathbf{p}_{ij}$  locates within the image region of  $I_i$ , otherwise  $\sigma(\mathbf{p}_{ij}) = 0$ . Such a formulation of DBA circumvents the process of feature

extraction and matching in most BA problems. Additionally, since we take the intensities of the panoramic image into account, DBA can also obtain the optimal panoramic image besides the global poses of local images.

### 5.1.3 Solving the Optimization Problem

From (5.3), it is found that intensity difference  $e_{ij}$  is only dependent on  $\xi_i$  and intensity of voxel  $\mathbf{p}_j$  in panoramic image, i.e.  $M(\mathbf{p}_j)$ . Therefore, according to the chain rule, we can derive the partial derivatives of intensity difference  $e_{ij}$  w.r.t.  $\xi_i$  and  $M(\mathbf{p}_j)$ :

$$\frac{\partial e_{ij}(\xi_i, M(\mathbf{p}_j))}{\partial \xi_i} = -\frac{\partial I_i}{\partial f(\xi_i, \mathbf{p}_j)} \frac{\partial f(\xi_i, \mathbf{p}_j)}{\partial \xi_i}, \quad (5.5)$$

$$\frac{\partial e_{ij}(\xi_i, M(\mathbf{p}_j))}{\partial (M(\mathbf{p}_j))} = 1. \quad (5.6)$$

It is shown in (5.5) that the partial derivative of  $e_{ij}$  w.r.t. pose parameters  $\xi_i$  consists of two parts:

- 1) The first term  $\partial I_i / \partial f$  in (5.5) is the partial derivative of intensity  $I_i$  w.r.t.  $f$ , which is a  $1 \times 3$  vector.
- 2) The second term  $\partial f / \partial \xi_i$  in (5.5) is the partial derivative of  $f$  w.r.t. pose parameters  $\xi_i$ , which is a  $3 \times 6$  matrix.

Thus, the result of (5.5) will be a  $1 \times 6$  vector. In addition, it is found in (5.6) that the partial derivative of intensity difference w.r.t. intensity at point  $\mathbf{p}_j$  in the panoramic image is a constant 1.

In our DBA algorithm, the state vector  $\mathbf{x}$  to be optimized includes pose parameters of all local images and intensities of all voxels in the panoramic image. Thus, we can derive the Jacobian of one intensity difference  $e_{ij}$  w.r.t. state vector  $\mathbf{x}$  as the following format:

$$J_{ij}(\mathbf{x}) = [\underbrace{\mathbf{0}_{1 \times 6}, \dots, \frac{\partial e_{ij}}{\partial \xi_i}, \dots, \mathbf{0}_{1 \times 6}}_{e_{ij} \text{ w.r.t. } \mathbf{x}_\xi}, \underbrace{0, \dots, \frac{\partial e_{ij}}{\partial (M(\mathbf{p}_j))}, \dots, 0}_{e_{ij} \text{ w.r.t. } \mathbf{x}_M}], \quad (5.7)$$

where  $\mathbf{0}_{1 \times 6}$  denotes a  $1 \times 6$  zero vector.

If we write the overall intensity differences as a concatenation vector  $e(\mathbf{x}) = [\dots, e_{ij}, \dots]^\top$ , the objective function of (5.4) to be optimized can be written as:

$$f(\mathbf{x}) = e(\mathbf{x})^\top e(\mathbf{x}). \quad (5.8)$$

According to (5.7), for all intensity differences  $e(\mathbf{x})$ , the overall Jacobian matrix can be built as a collection of  $J_{ij}$  as  $J(\mathbf{x}) = [\dots, J_{ij}(\mathbf{x})^\top, \dots]^\top$ .

In this study, Gauss-Newton (GN) method is adopted to solve the NLLS problem in (5.4). During GN iterations, step change  $\Delta \mathbf{x}$  in each iteration can be calculated from the GN equation:

$$H \Delta \mathbf{x} = b, \quad (5.9)$$

where

$$H = J(\mathbf{x})^\top J(\mathbf{x}), \quad b = -J(\mathbf{x})^\top e(\mathbf{x}). \quad (5.10)$$

#### 5.1.4 Efficient Implementation

**Sparsity and marginalization:** GN method linearizes the nonlinear optimization problem (5.4). During the GN iterations, we need to solve the linear system (5.9). But this step is cumbersome because the size of matrix  $H$  in our problem is huge. For a case with  $m$  frames of 3D images and  $n$  voxels in the panoramic image, the size of matrix  $H$  is  $(6m + n) \times (6m + n)$ , as shown in Fig. 5.2. And in matrix  $H$ ,  $n$  is usually much larger than  $m$ . Take a 3D TEE image as an example, it usually contains millions of voxels in one single volume. Reconstructing a panoramic image with several 3D TEE images, the size of matrix  $H$  is in the tens of millions multiply tens of millions. Calculating such a large matrix directly is time-consuming and impossible in most cases.

When inspecting (5.7), we find that  $J_{ij}(\mathbf{x})$  is sparse and this sparse property will cause the sparsity in the matrix  $H$  finally. In Fig. 5.2, we use an intuitive structure to illustrate the sparse property of  $H$ . The upper sub-figure of  $J_{ij}(\mathbf{x})$  is corresponding to the row vector in (5.7), the blocks with color in the figure means values in these blocks are nonzero. Since  $J_{ij}(\mathbf{x})$  only has nonzero blocks at block  $i$  and  $j$  (block  $j$  is a scalar 1), it makes nonzero contribution to only four blocks of matrix  $H$ , namely block  $(i, i)$ ,  $(i, j)$ ,  $(j, i)$  and  $(j, j)$ .

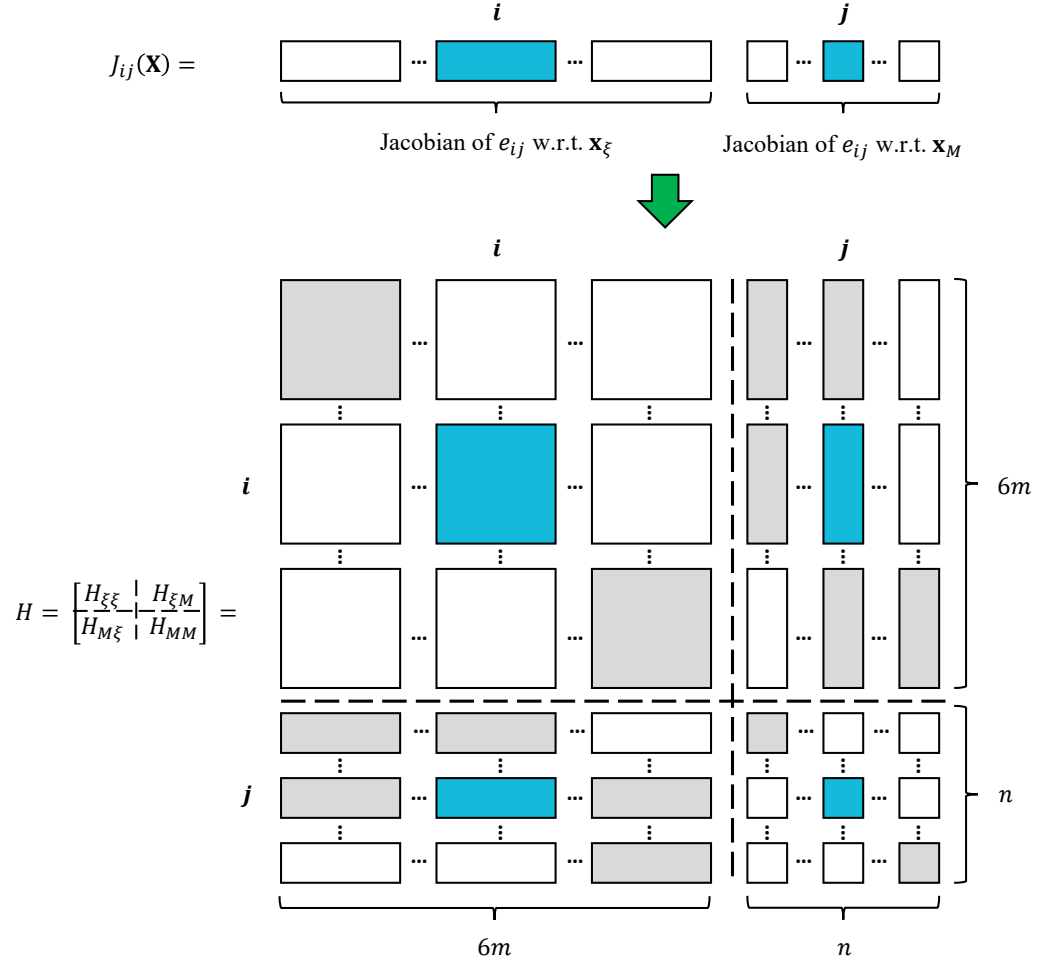


FIGURE 5.2: Sparsity of matrix  $H$ :  $H_{\xi\xi}$  is block diagonal and  $H_{MM}$  is diagonal. The size of  $H_{MM}$  is usually much larger than  $H_{\xi\xi}$ . The structure of  $H_{\xi M}$  and  $H_{M\xi}$  indicates the observation relationships between voxels in panoramic image and local frames.

Matrix  $H$  is produced by the sum of these sparse matrices:

$$H = \sum_{i,j} \mathbf{J}_{ij}^\top(\mathbf{x}) \mathbf{J}_{ij}(\mathbf{x}). \quad (5.11)$$

According to the sparsity of matrix  $H$ , it can be divided into four blocks:

$$H = \begin{bmatrix} H_{\xi\xi} & H_{\xi M} \\ H_{M\xi} & H_{MM} \end{bmatrix}, \quad \text{where } H_{M\xi} = H_{\xi M}^\top. \quad (5.12)$$

$H_{\xi\xi}$  is block diagonal and  $H_{MM}$  is diagonal. Since the linear system (5.9) is sparse, Schur complement [105] can be used to efficiently solve linear system (5.9). If we write step

changes of pose parameters and intensities of voxels separately, (5.9) can be rewritten as follows:

$$\begin{bmatrix} H_{\xi\xi} & H_{\xi M} \\ H_{M\xi} & H_{MM} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_\xi \\ \Delta \mathbf{x}_M \end{bmatrix} = \begin{bmatrix} b_\xi \\ b_M \end{bmatrix}. \quad (5.13)$$

Using the Schur complement, we can compute step changes of poses and intensities of voxels sequentially from:

$$\left( H_{\xi\xi} - H_{\xi M} H_{MM}^{-1} H_{\xi M}^\top \right) \Delta \mathbf{x}_\xi = (b_\xi - H_{\xi M} H_{MM}^{-1} b_M), \quad (5.14)$$

$$H_{MM} \Delta \mathbf{x}_M = b_M - H_{\xi M}^\top \Delta \mathbf{x}_\xi. \quad (5.15)$$

Since matrix  $H_{MM}$  is diagonal, its inverse can be easily computed by the inverse of each element on the diagonal. Solving (5.14) then (5.15) is much more efficient than solving (5.9) directly.

**Pre-computation of Gradient and Intensity Space:** According to (5.5), calculation of partial derivative  $\partial I_i / \partial f$  is a computationally costly step in the process of GN iterations. Since  $\partial I_i / \partial f = \partial I_i / \partial \mathbf{p}_{ij}$  where  $\partial I_i / \partial \mathbf{p}_{ij}$  is the partial derivative of  $I_i$  w.r.t the corresponding coordinates of point  $\mathbf{p}_{ij}$  on the local 3D image, we can pre-compute the gradient space of intensity of all the local 3D images before the GN iterations. Then, the gradient value at any point in local images can be read directly from the gradient space during GN iterations. Additionally, since  $\mathbf{p}_{ij}$  are usually not integers, the intensities and gradient values of point  $\mathbf{p}_{ij}$  in the local image are obtained by trilinear interpolation to improve the accuracy of pose estimation, which are also processed before the iterations.

The DBA algorithm is summarized in Algorithm 3. Poses are initialized using visual odometry (VO) estimated by direct registration method [106].

## 5.2 Experiments and Results

Since the DBA involves a large-scale optimization problem, it is challenging to realize fast performance. However, given its potential to achieve highly accurate registration and fusion, it can be very useful in practice when real-time performance is not required, such



**Algorithm 3:** Direct Bundle Adjustment (DBA)**Input:** A collection of 3D images  $\{I_i\}_{i=1}^m$ .**Output:** Global poses  $\mathbf{x}_\xi$  and intensities of panoramic image  $\mathbf{x}_M$ .

```

1 Step 1: Pre-computation:
2 Define a volume  $M$  with zero intensities for all voxels;
3 Calculate gradient space of intensity of 3D images;
4 Interpolate intensity/gradient spaces of 3D images;
5 Step 2: Optimization:
6 Initialize the pose parameters  $\mathbf{x}_\xi \leftarrow \mathbf{x}_{\xi_0}$ ;
7 Initialize  $M$  with initial pose  $\mathbf{x}_{\xi_0}$  from local images;
8 while Algorithm not converged do
9   for every voxel  $\mathbf{p}_j$  in volume  $M$  do
10      $\mathbf{p}_{ij} = T(\xi_i)\mathbf{p}_j$ ;
11     if  $\sigma(\mathbf{p}_{ij}) = 1$  then
12       Calculate intensity difference  $e_{ij}$  by (5.3)
13       Calculate Jacobian  $J_{ij}$  by (5.7);
14     end
15   end
16    $J(\mathbf{x}) = [..., J_{ij}(\mathbf{x})^\top, ...]^\top$ ,  $e(\mathbf{x}) = [..., e_{ij}, ...]^\top$ ;
17   Calculate  $\Delta\mathbf{x}_\xi$  by (5.14), calculate  $\Delta\mathbf{x}_M$  by (5.15) ;
18   Update poses,  $T(\mathbf{x}_\xi) \leftarrow T(\Delta\mathbf{x}_\xi)T(\mathbf{x}_\xi)$ ;
19   Update intensities of  $M$ ,  $\mathbf{x}_M \leftarrow \mathbf{x}_M + \Delta\mathbf{x}_M$ ;
20 end

```

as enlarging the FoV of 3D TEE for assisting in the surgical measurement and planning of LAA occlusion. In this case, image registration and fusion is usually performed offline. Therefore, in this section, we use 3D TEE image data as an example and perform simulated and in-vivo experiments to validate the proposed DBA algorithm.

### 5.2.1 Simulated Experiments

Since currently, there is no algorithm available for completing 3D image registration and fusion simultaneously, sequential method registering two images at each time is commonly used if there are more than two images. In this section, we access the robustness and accuracy of the proposed DBA method by comparing it with sequential fusion method presented in Chapter 3. Since this work focuses more on the accuracy of registration and fusion rather than the efficiency, in the compared sequential fusion method, full images are used instead of three-orthogonal-plane images.

TABLE 5.1: Ground Truth Poses of six sequences of simulated images

Sequence	Pose Parameter	Image ID in Each Group										
		$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$I_9$	$I_{10}$	$I_{11}$
1	E.A. (degree)	0	6	6	9	9	15	15	18	18	21	21
	Trans. (pixel)	0	5	25	5	25	5	25	5	25	5	25
2	E.A. (degree)	0	3	3	9	9	12	12	15	15	18	18
	Trans. (pixel)	0	15	25	15	25	15	25	15	25	15	25
3	E.A. (degree)	0	9	9	12	12	15	15	18	18	21	21
	Trans. (pixel)	0	15	20	15	20	15	20	15	20	15	20
4	E.A. (degree)	0	9	9	12	12	15	15	21	21	24	24
	Trans. (pixel)	0	10	25	10	25	10	25	10	25	10	25
5	E.A. (degree)	0	12	12	15	15	18	18	21	21	24	24
	Trans. (pixel)	0	5	15	5	15	5	15	5	15	5	15
6	E.A. (degree)	0	6	6	9	9	12	12	15	15	18	18
	Trans. (pixel)	0	10	20	10	20	10	20	10	20	10	20

E.A.: Euler angles, rotation order is X-Y-Z.

Trans.: Translations in X-Y-Z direction.

For example: for Sequence 1, image  $I_2$ , Euler angles are  $[6, 6, 6]^\top$  degrees, translations are  $[5, 5, 5]^\top$  pixels.

One-pixel length represents 0.25 mm in practice.

### 5.2.1.1 Accuracy Assessment of DBA

Similar to Section 3.2.1.2, we generate the simulated 3D TEE image data (in grayscale ranged from 0-255) by firstly cropping images from a 3D heart CT scan, and then adding Gaussian intensity noises on them. Five sequences of 3D TEE images are generated from different poses of ‘viewfinder’ and then white gaussian noise (WGN) with a standard deviation of 8 is added to these images. Each sequence contains 11 simulated images and the ground truth pose of each image is listed in Table 5.1 (from Sequence 1 to Sequence 5).

We perform the experiments on these five sequences of image data by using the proposed DBA method and the sequential fusion method, respectively. In each experiment, we treat the first frame as the global frame. Therefore, 10 poses need to be estimated in each sequence.

**Assessment of Pose Accuracy:** In Fig. 5.3, mean absolute errors (MAE) of translation and Euler angles are compared respectively. From the results, it is shown clearly that the proposed DBA method has better accuracy than the sequential method. For the translation, results from the proposed DBA method have an accuracy of 0.1 pixels (0.025

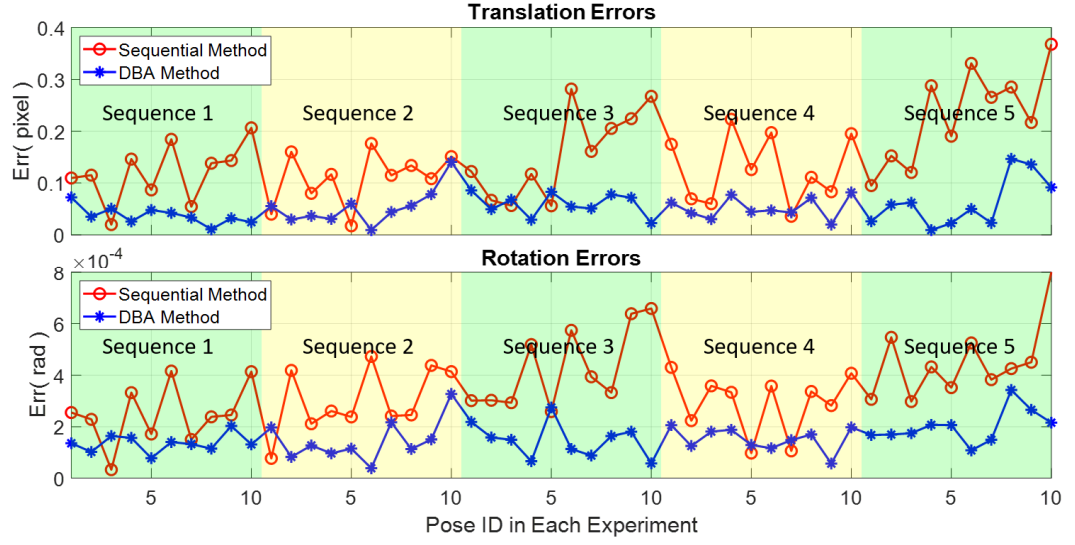


FIGURE 5.3: Accuracy of the sequential fusion method and the proposed DBA method using the data from the Sequence 1 to Sequence 5. (Remark: One-pixel length represents 0.25 mm in practice.)

mm in practice) for most of the cases. Only 3 out of 50 estimated translation errors are larger than 0.1 pixels. On the contrary, 37 out of 50 results from the sequential fusion method have errors larger than 0.1 pixels. For the accuracy of rotation, although both the sequential method and the proposed DBA method can obtain the results with high accuracy of  $10^{-4}$  order of magnitude radians. The results from the proposed DBA method are still more accurate than the sequential fusion method for most of the cases.

**Assessment of Intensity Accuracy:** We also calculate the MAE of intensities of the fuse panoramic images w.r.t. the ground truth image and the results are shown in the upper sub-table in Table 5.2. In the five simulated experiments, The MAE of intensities of the fused image obtained from the proposed DBA method is always smaller than that from the sequential fusion method, which indicates the intensities of the fused image optimized from the proposed DBA method has better accuracy than those from the sequential method.

### 5.2.1.2 Robustness Assessment of DBA

Intensity noise is usually challenging for most of the direct methods. In this section, we assess the robustness of the proposed algorithm with different levels of intensity noises. The comparative experiments are performed using the sequential fusion method as well.

TABLE 5.2: Comparison of Intensity Errors between two methods

Sequence ID	1	2	3	4	5
Sequential	2.4833	2.7201	2.8225	2.6584	2.7031
DBA	2.1732	2.3689	2.4577	2.3613	2.3692
Sequence 6 (5 noise levels)	std 6 (WGN)	std 7 (WGN)	std 8 (WGN)	std 9 (WGN)	std 10 (WGN)
Sequential	1.9603	2.2835	2.6096	2.9332	3.2622
DBA	1.6822	1.9588	2.2364	2.5141	2.7923

The intensity noises are added based on Sequence 6 in Table 5.1. We design five noise levels (Level 1 to Level 5) with the corresponding standard deviations of WGN as 6, 7, 8, 9, and 10 respectively.

**Robustness of Pose Estimation:** The distribution of the MAE of translation and rotation is shown in Fig. 5.4. With the increase of the intensity noise from Level 1 to Level 5, the results of translation from the proposed DBA method can keep an accuracy of 0.1 pixels (0.025 mm in practice) for most of the cases, the largest error is around 0.15 pixel (0.03 mm in practice). By contrast, the error of translation from the sequential fusion method increases with fluctuation from noise Level 1 to Level 5 and the largest translation error is larger than 0.35 pixels. The MAE of rotation shows similar trends as that of translation. From the comparison of pose errors, it is shown that the proposed DBA algorithm has much better robustness to the intensity noise than the sequential fusion method although both methods can obtain a high accuracy.

**Robustness to Intensity Estimation:** The intensity MAE of the reconstructed panoramic image using two methods are compared and results are shown in the bottom sub-table in Table 5.2. It is shown that although the MAE of intensities obtained by both methods increases with the increase of noise level, the results of the proposed DBA method is always better than that of the sequential fusion method.

### 5.2.2 In-vivo Experiments

For the in-vivo experiments, datasets of 3D TEE images from five different patients are collected using a 2D array transducer with the assistance of the ECG-gating. Therefore, we can consider the registration problem as rigid. The details of the datasets are listed

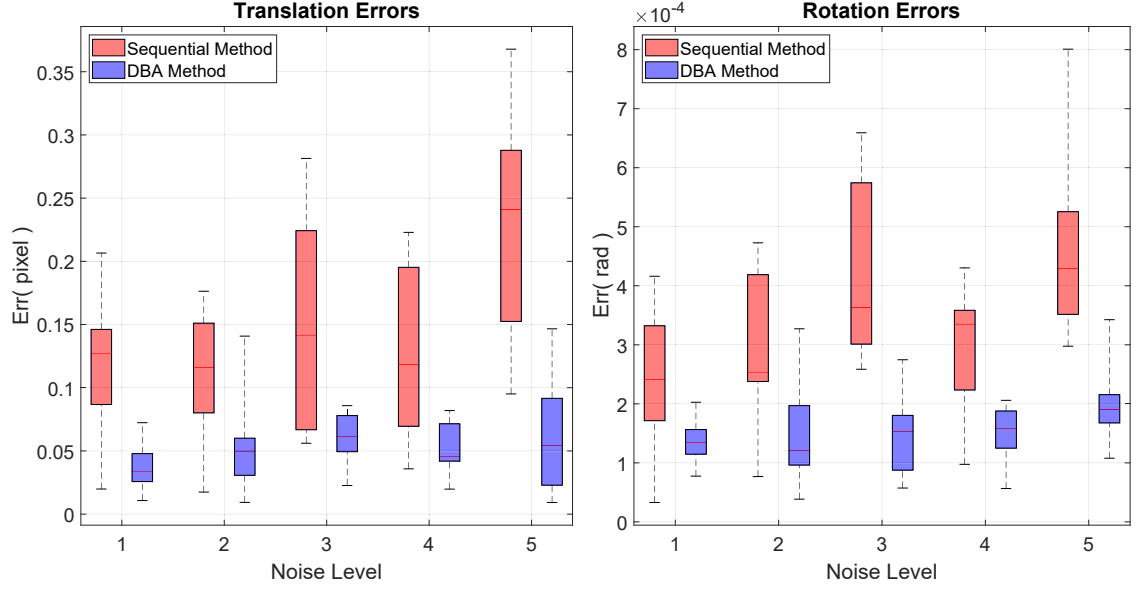


FIGURE 5.4: Robustness of the sequential fusion method and the proposed DBA method using Sequence 6 with different intensity noise levels. (Remark: One-pixel length represents 0.25 mm in practice.)

TABLE 5.3: Details of the Five In-vivo ECG-gated Datasets

Patient ID	No. frames	Volume size (voxel)	Resolution (mm/voxel)	FoV w.r.t original
# 1	9	240×160×208	0.69×0.98×0.73	2.10
# 2	6	240×160×208	0.77×1.11×0.82	2.02
# 3	6	240×160×208	0.64×0.92×0.68	1.80
# 4	8	272×208×208	0.63×0.90×0.68	2.06
# 5	6	272×208×208	0.58×0.87×0.63	2.01

in the first four columns of Table 5.3. In the in-vivo experiments, pose parameters are initialized using the VO estimated by direct method with similarity metric SSD [106] and the panoramic image is initialized by the local 3D images using the initial poses.

Evaluation of the results are performed based on the quality of the panoramic images and aligned images since the ground truth of in-vivo datasets are usually not available in practice. By comparing sharp boundaries in the single frame of images and fused images, such as left atrium (LA) walls indicated by circles in Fig. 5.5, it is found that the panoramic images obtained from the proposed DBA have smooth transition and no misalignment/ghosting is found, which suggests good quality of the estimated panoramic images. From the comparison in Fig. 5.5, it is evident that the FoV of the fused images

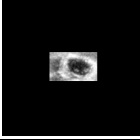
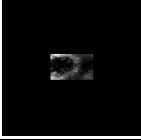
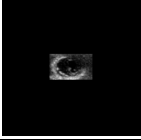
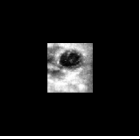
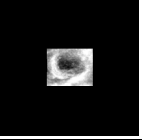
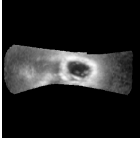
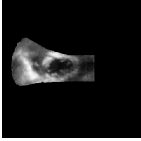
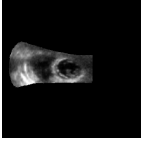
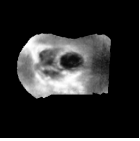
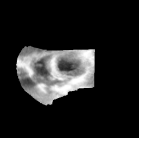
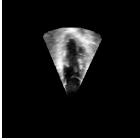
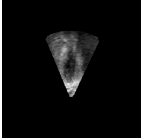
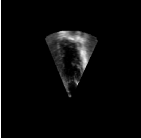
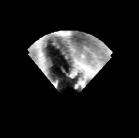
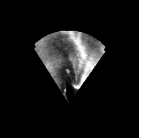
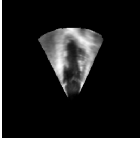
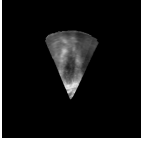
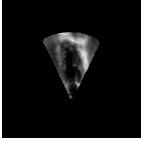
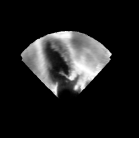
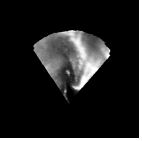
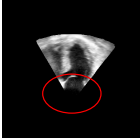
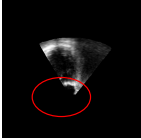
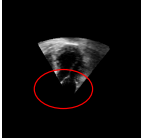
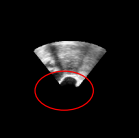
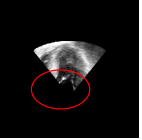
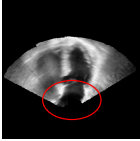
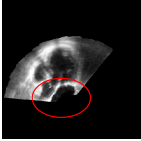
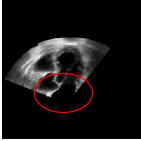
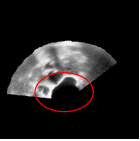
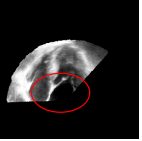
Viewing Plane	Data	Patient ID				
		# 1	# 2	# 3	# 4	# 5
Slice XY	Original Single Frame					
	Fused 3D Image					
Slice YZ	Original Single Frame					
	Fused 3D Image					
Slice XZ	Original Single Frame					
	Fused 3D Image					

FIGURE 5.5: Comparison of original single frame of in-vivo 3D TEE image with fused 3D panoramic image using the proposed DBA algorithm. In the last two rows, regions with sharp boundaries are marked by circles for qualitative evaluation of the fused panoramic images.

are enlarged apparently. By counting the number of voxels, it is found that the FoV of the fused panoramic images is enlarged to 2.10, 2.02, 1.80, 2.06, and 2.01 times as compared with the original single frame of TEE image of # 1, # 2, # 3, # 4, and # 5 respectively, which are listed in the last column of Table 5.3. The results shows good potential of the proposed DBA for overcoming the drawback of limited FoV of 3D TEE images.

Furthermore, to assess the accuracy of the pose estimation, images are aligned with the poses calculated by DBA method and sequential method, respectively. Although in Chapter 3, we have shown that the proposed direct registration method is more accurate than the state-of-the-art PCA-based method for pairwise registration, it is found from the experiments that accumulating errors occur when the sequential fusion strategy is used for

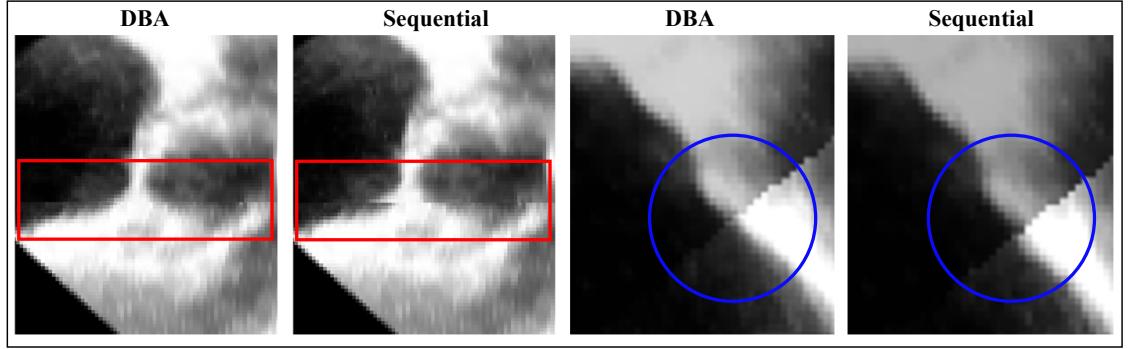


FIGURE 5.6: Comparison of the aligned images using poses calculated by the DBA and sequential methods.

fusing multiple images. Two examples are shown in Fig. 5.6. We can find that the results of DBA have smooth transition, while the results of the sequential method have some apparent misalignments. In addition, mean squared errors (MSE) between panoramic images and registered images (transforming local images to the panoramic image space by using the estimated poses) is calculated for the all five in-vivo datasets. The MSE are (235, 293, 244, 214, 321) and (254, 311, 246, 233, 341), for DBA and sequential method, respectively. Smaller MSE of DBA than sequential method suggest that the results of DBA method have better alignment than those of sequential method. Thus, all these results indicate that the proposed DBA outperforms the sequential method in terms of the pose accuracy.

### 5.3 Chapter Summary

In this chapter, we present a novel direct bundle adjustment (DBA) algorithm for 3D image fusion, which optimizes the intensities of the predefined panoramic image and the poses of the local images simultaneously. Simulated and in-vivo experiments are performed to demonstrate that the proposed algorithm is robust to intensity noises and can obtain more accurate results than the sequential fusion method. Experimental results using in-vivo 3D TEE datasets from five different patients show that the fused 3D TEE images have around twice the FoV than the original image, indicating a significant potential clinical value of the proposed algorithm.

This new algorithm provides an elegant way to obtain the optimal panoramic image and optimal local frame poses in one go, without any information loss or information reuse. Therefore, it can serve as a good framework for 3D image fusion. The current algorithm is still computationally expensive due to the large-scale optimization problem involved. In the next chapter, we will further prove that the pose optimization in [DBA](#) problem is actually independent of the panoramic image in the [GN](#) iterations, which means we do not need to solve the intensities of the panoramic image that have a huge dimension, but can obtain exactly the same optimal poses as solving the complete [DBA](#).



## Chapter 6

# DSR: Direct Simultaneous Registration for Multiple 3D Images

In Chapter 5, we proposed a direct bundle adjustment (DBA) method which redefines the classical BA by jointly optimizing the poses of local images and the intensities of the predefined panorama. DBA uses intensity information directly instead of the extracted and matched feature points of the local 2D images in classical BA. Therefore, the method can deal with images lacking distinct features such as 3D TEE images.

When investigating DBA further, we find a very interesting property and have theoretically proved that “When solving DBA problem with GN iterations, the optimization of poses is independent of the intensities of the panoramic image”. And based on the property, we propose a novel direct simultaneous registration (DSR) algorithm that *only* solves the poses *without* solving the huge dimension of intensities of the panorama but obtains the same poses as DBA. In this chapter, detailed proof of the property is presented first. Then, simulated and in-vivo 3D TEE images are used to evaluate the proposed DSR method compared with pairwise [75], Lie normalization [107], sequential [106], and APE [17] methods. It is shown that the proposed method outperforms these four methods in

terms of accuracy and requires much fewer computational resources than the state-of-the-art accumulated pairwise estimates (APE).

## 6.1 Methods

### 6.1.1 Revisit of Direct Bundle Adjustment

In the DBA problem mentioned in Section 5.1.1, there are  $m$  frames of 3D images taken from different viewpoints, denoted as  $\mathbf{I} = \{I_1, \dots, I_i, \dots, I_m\}$ . Correspondingly, the global transformation for each image is parameterized in Lie algebra space with the pose parameters  $\mathbf{x}_\xi = [\xi_1^\top, \dots, \xi_i^\top, \dots, \xi_m^\top]^\top \in \mathbb{R}^{6m}$ . A predefined 3D panoramic image  $M$  consists of  $n$  voxels  $\{\mathbf{p}_1, \dots, \mathbf{p}_j, \dots, \mathbf{p}_n\}$ , whose intensities are  $\mathbf{x}_M = [M(\mathbf{p}_1), \dots, M(\mathbf{p}_j), \dots, M(\mathbf{p}_n)]^\top$ . DBA aims at optimizing the pose parameters of local images and the intensities of voxels in the panoramic image simultaneously.

The intensity of voxel  $\mathbf{p}_j$  in  $M$  is obtained by fusing different points' intensities in local images. Let  $\mathbf{A}_M = [\dots, M(\mathbf{p}_j), \dots]^\top$  and  $\mathbf{B}(\mathbf{x}_\xi) = [\dots, I_i(\mathbf{p}_{ij}), \dots]^\top = [\dots, I_i(T(\xi_i)\mathbf{p}_j), \dots]^\top$  represent all observed intensities of the panoramic image and their corresponding intensities in local images, respectively. Then, DBA problem (5.4) seeks to obtain the optimal solution  $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_\xi^\top, \hat{\mathbf{x}}_M^\top]^\top$  that minimizes:

$$f(\mathbf{x}) = \|e(\mathbf{x})\|^2 = e(\mathbf{x})^\top e(\mathbf{x}), \quad (6.1)$$

where

$$e(\mathbf{x}) = \mathbf{A}_M - \mathbf{B}(\mathbf{x}_\xi). \quad (6.2)$$

The optimal solution to (6.1) can be obtained by using the GN iterations, which starts with parameters initialization and then updates the parameters using the step changes calculated from GN equation in each iteration until the algorithm converges.

Although the optimal poses and panoramic image can be obtained simultaneously, DBA seems more difficult to solve than traditional multi-image registration problems since a much higher order state vector is involved. However, we can further prove that the pose

optimization is actually independent of the panoramic image in the GN iterations (see Section 6.1.2), which means we do not need to solve the intensities of the panoramic image but can obtain exactly the same optimal poses as solving the complete DBA.

### 6.1.2 Simultaneous Registration without Intensity Optimization

**Theorem:** When solving (6.1) with GN iterations, the optimization of poses is independent of the intensities of the panoramic image.

**Proof:** If we write Jacobian matrix of  $e(\mathbf{x})$  w.r.t.  $\mathbf{x}_\xi$  and  $\mathbf{x}_M$  separately as  $J(\mathbf{x}) = [J_\xi, J_M]$ , then GN equation (5.13) can be written as:

$$\begin{bmatrix} J_\xi^\top J_\xi & J_\xi^\top J_M \\ J_M^\top J_\xi & J_M^\top J_M \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_\xi \\ \Delta \mathbf{x}_M \end{bmatrix} = \begin{bmatrix} -J_\xi^\top e(\mathbf{x}) \\ -J_M^\top e(\mathbf{x}) \end{bmatrix}. \quad (6.3)$$

In Chapter 5, we use Schur complement to solve the step changes of poses  $\Delta \mathbf{x}_\xi$  and intensities of voxels  $\Delta \mathbf{x}_M$  sequentially by (5.14) and (5.15). Now, according to (6.2) and (6.3), we can write the right side of (5.14) as:

$$\begin{aligned} & b_\xi - H_{\xi M} H_{MM}^{-1} b_M \\ &= -J_\xi^\top e(\mathbf{x}) + J_\xi^\top J_M (J_M^\top J_M)^{-1} J_M^\top e(\mathbf{x}) \\ &= -J_\xi^\top (\mathbf{A}_M - \mathbf{B}(\mathbf{x}_\xi)) + J_\xi^\top J_M (J_M^\top J_M)^{-1} J_M^\top (\mathbf{A}_M - \mathbf{B}(\mathbf{x}_\xi)) \\ &= -J_\xi^\top \mathbf{A}_M + J_\xi^\top \mathbf{B}(\mathbf{x}_\xi) + J_\xi^\top J_M (J_M^\top J_M)^{-1} J_M^\top \mathbf{A}_M - J_\xi^\top J_M (J_M^\top J_M)^{-1} J_M^\top \mathbf{B}(\mathbf{x}_\xi) \\ &= -J_\xi^\top (\mathbf{A}_M - J_M (J_M^\top J_M)^{-1} J_M^\top \mathbf{A}_M) - J_\xi^\top (J_M (J_M^\top J_M)^{-1} J_M^\top \mathbf{B}(\mathbf{x}_\xi) - \mathbf{B}(\mathbf{x}_\xi)). \end{aligned} \quad (6.4)$$

It is shown from  $J_{ij}(\mathbf{x})$  (see (5.7)) that there is one and only one nonzero element 1 in each row of  $J_M$ . The nonzero element means the voxel  $\mathbf{p}_j$  is observed in the local frame  $i$  and corresponds to the intensity difference  $e_{ij} = M(\mathbf{p}_j) - I_i(\mathbf{p}_{ij})$ . Therefore, according to the observed status of the panoramic image in the local frames which is indicated by the structure of  $J_M$ , it can be easily deduced that  $\mathbf{A}_M = J_M \mathbf{x}_M$ . Substituting  $\mathbf{A}_M$  to the

first term on the right side of (6.4), we have:

$$-J_\xi^\top (J_M \mathbf{x}_M - J_M (J_M^\top J_M)^{-1} J_M^\top J_M \mathbf{x}_M) = \mathbf{0}. \quad (6.5)$$

Then, (5.14) becomes:

$$\left( H_{\xi\xi} - H_{\xi M} H_{MM}^{-1} H_{\xi M}^\top \right) \Delta \mathbf{x}_\xi = -J_\xi^\top (J_M H_{MM}^{-1} J_M^\top \mathbf{B}(\mathbf{x}_\xi) - \mathbf{B}(\mathbf{x}_\xi)), \quad (6.6)$$

which indicates that the step change  $\Delta \mathbf{x}_\xi$  is independent of intensities  $\mathbf{x}_M$  in every GN iteration. Therefore, obtaining the optimal poses is independent of the intensities of panoramic image  $M$  during the optimization process. **Q.E.D.**

In addition, after  $\hat{\mathbf{x}}_\xi$  is obtained, the NLLS problem in (6.1) becomes a linear least-squares problem to minimize:

$$\|e(\mathbf{x}_M)\|^2 = (\mathbf{A}_M - \mathbf{B}(\hat{\mathbf{x}}_\xi))^\top (\mathbf{A}_M - \mathbf{B}(\hat{\mathbf{x}}_\xi)). \quad (6.7)$$

Therefore, if required, we can calculate the optimal panoramic image easily in only one step from the following closed-form formula:

$$\hat{\mathbf{x}}_M = -H_{MM}^{-1} J_M^\top \mathbf{B}(\hat{\mathbf{x}}_\xi). \quad (6.8)$$

The independent property of DBA is very attractive since it allows us to optimize the poses only using (6.6), which is equivalent to solving the complete DBA problem using Schur complement in two sequential steps as (5.14) and (5.15). A 3D image typically contains millions of voxels but we only need six parameters to represent its pose. Therefore, the independence of optimizing poses to intensities has the potential of helping us reduce the dimension of the solution space. Based on the theorem, we proposed the DSR algorithm which solves the poses only without considering the huge dimension of the intensities. The implementation process of DSR is summarized in Algorithm 4.

**Algorithm 4:** Direct Simultaneous Registration (DSR)**Input:** A collection of 3D images  $\{I_i\}_{i=1}^m$ .**Output:** Global poses  $\mathbf{x}_\xi$  for all local 3D images.

---

```

1 Step 1: Pre-computation:
2 Define a volume  $M$  (grid only) with  $n$  voxels and a vector  $\mathbf{v} \in \mathbb{R}^n$ ;
3 Calculate gradient space of intensity of 3D images  $\{I_i\}_{i=1}^m$ ;
4 Interpolate intensity/gradient spaces of 3D images;
5 Step 2: Optimization:
6 Initialize the pose parameters  $\mathbf{x}_\xi \leftarrow \mathbf{x}_{\xi_0}$ ;
7 while Algorithm not converged do
8   Assign zero to each element of  $\mathbf{v}$ ;
9   for every voxel location  $\mathbf{p}_j$  in volume  $M$  do
10     for every image  $I_i$  do
11        $\mathbf{p}_{ij} = T(\xi_i)\mathbf{p}_j$ ;
12       if  $\sigma(\mathbf{p}_{ij}) = 1$  then
13         Calculate Jacobian  $J_{\mathbf{x}_\xi}^{ij} = [\mathbf{0}^\top, -\frac{\partial I_i}{\partial \omega(\xi_i, \mathbf{p}_j)} \frac{\partial \omega(\xi_i, \mathbf{p}_j)}{\partial \xi_i}, \mathbf{0}^\top]$ ;
14         Calculate Jacobian  $J_{\mathbf{x}_M}^{ij} = [\mathbf{0}^\top, 1, \mathbf{0}^\top]$ ;
15         Update vector  $\mathbf{v}(\mathbf{p}_j) = \mathbf{v}(\mathbf{p}_j) + 1$ ;
16       end
17     end
18   end
19    $J_\xi = [\dots, J_{\mathbf{x}_\xi}^{ij^\top}, \dots]^\top$ ,  $J_M = [\dots, J_{\mathbf{x}_M}^{ij^\top}, \dots]^\top$ ;
20   Calculate  $H_{\xi\xi} = J_\xi^\top J_\xi$ ,  $H_{\xi M} = H_{M\xi}^\top = J_\xi^\top J_M$ , and  $H_{MM}^{-1} = \text{diag}(1./\mathbf{v})$ ;
21   Calculate  $\Delta \mathbf{x}_\xi = [\Delta \xi_1^\top, \dots, \Delta \xi_i^\top, \dots, \Delta \xi_m^\top]^\top$  by (9);
22   Update the pose of each image,  $T(\xi_i) \leftarrow T(\Delta \xi_i)T(\xi_i)$ ;
23 end

```

---

## 6.2 Experiments and Results

In this section, 3D TEE images are used as examples to evaluate the proposed DSR algorithm compared with the pairwise [75], Lie normalization [107], sequential [106], and APE [17] methods. Both simulated and in-vivo experiments are performed.

### 6.2.1 Simulated Experiments

Similar to the data generation process introduced in Section 3.2.1.2, five sequences of 3D images (in grayscale ranging from 0 to 255) are simulated by transforming a 3D TEE volume to a real 3D CT scan of the heart with different poses and cropping the corresponding image area. Each sequence contains 11 frames of 3D images. The magnitude of

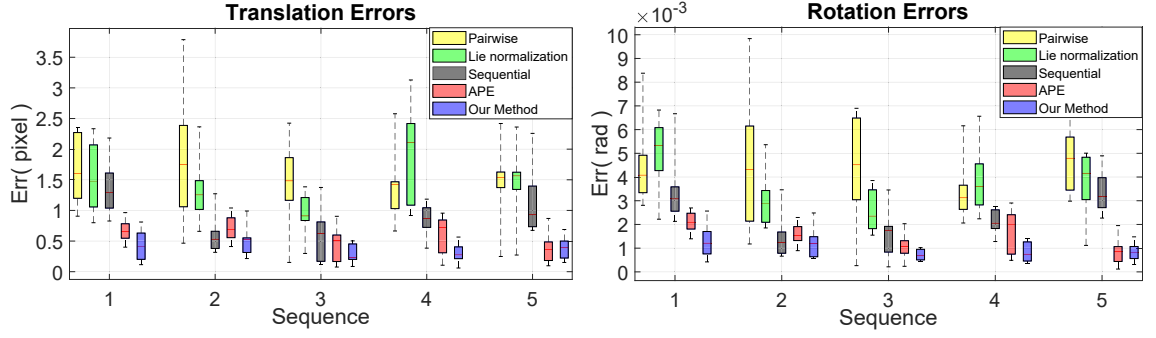


FIGURE 6.1: Accuracy of DSR method compared to pairwise [75], Lie normalization [107], sequential [106], and APE [17] methods using 5 sequences of simulated 3D TEE images.

these transformations varies between  $\pm 12$  degrees for rotations and  $\pm 15$  pixels for translations, which are typical ranges of poses in our obtained in-vivo 3D TEE images. Gaussian noise [84, 108] with a standard deviation of 25 is generated randomly and added to the intensities of these five sequences of images.

The accuracy of the proposed DSR method is evaluated by comparing it with pairwise, Lie normalization, sequential, and APE methods via simulated datasets. For fair comparisons, pairwise, sequential, and APE methods use the SSD as the similarity metric, GN method as optimization solver, and the initial pose parameters in each method are the same. Lie normalization optimizes poses obtained from pairwise methods and does not directly involve images [107]. Thus, we use the results from the pairwise registration as the input to Lie normalization. The mean absolute errors (MAE) of translation and Euler angles obtained from the proposed DSR and other four methods are compared in Fig. 6.1.

It is shown from Fig. 6.1 that MAE of the results obtained from DSR, sequential, and APE methods are much smaller than the pairwise and Lie normalization methods in most of the cases, which indicates the better accuracy of these three methods. Additionally, among DSR, sequential, and APE methods, the accuracy of DSR is within 0.5 pixels for translations in most of the cases and is within  $1 \times 10^{-3}$  rad for rotations in more than half cases. Although the accuracy of APE is the closest to DSR among four competing methods, it still has larger errors than DSR. The errors of APE are greater than 0.5 pixels for translations and greater than  $1 \times 10^{-3}$  rad for rotations in more than half of experiments. And the results show that the accuracy of both translations and rotations

TABLE 6.1: Details of the Six In-vivo ECG-gated Datasets

Patient ID	No. frames	Volume size (voxel)	Resolution (mm/voxel)	FoV w.r.t original
# 1	9	240×160×208	0.69×0.98×0.73	2.10
# 2	6	240×160×208	0.77×1.11×0.82	2.02
# 3	6	240×160×208	0.64×0.92×0.68	1.80
# 4	8	272×208×208	0.63×0.90×0.68	2.06
# 5	6	272×208×208	0.58×0.87×0.63	2.01
# 6	11	236×224×208	0.69×0.72×0.77	2.18

from the sequential method is lower than [DSR](#) and [APE](#). Furthermore, it is seen from Fig. 6.1 that the distribution of errors from [DSR](#) is more concentrated than the others, which indicates it also has better robustness than the other four methods.

The above comparisons indicate that the proposed method has the highest accuracy, followed by [APE](#). Both these two simultaneous registration methods are more accurate than the other three which are deduced from pairwise registration. In addition, compared with the proposed [DSR](#) method, one apparent drawback of [APE](#) is its much higher computational complexity. Since both [DSR](#) and [APE](#) use the sum-of-squared intensity differences as the objective function, the computational complexities for both methods are closely related to the number of intensity differences. Suppose there are  $m$  images, each image has  $h$  pixels, and every two images have around  $\alpha\%$  overlapping area, the computational complexity of [APE](#) is around  $\mathcal{O}(m(m-1)/2 \times \alpha\% \times h)$  [78] since [APE](#) considers all the combinations of images, while that of [DSR](#) is only  $\mathcal{O}(m \times h)$ . In the simulated experiments for each sequence, it is found that [APE](#) calculates around four times as many intensity differences as [DSR](#) and needs around 4-5 times longer time than [DSR](#) for each iteration. Theoretically, the more images involved, the higher the computational complexity of [APE](#) is, and the more time it takes than [DSR](#).

### 6.2.2 In-vivo Experiments

In the in-vivo experiments, forty-six 3D [TEE](#) images from six patients (Patient #1 to #6) are collected using a real-time 3D transducer. The details of the [ECG](#)-gated datasets are listed in the first four columns in Table 6.1. Since the proposed [DSR](#), sequential, and [APE](#)

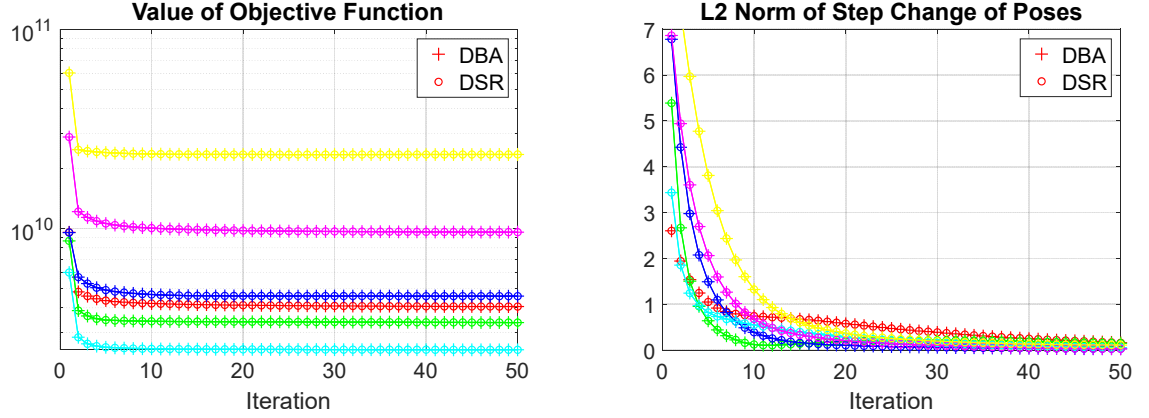


FIGURE 6.2: Comparison of iterative process of DSR and DBA, different colors represent experiments on different in-vivo datasets.

methods outperformed the other two methods in terms of accuracy, in this section, the proposed method is compared with the sequential and [APE](#) methods only. Pose parameters of three methods are initialized using the results from the pairwise method. It is found that such an initialization method is enough for [DSR](#) to converge to the correct results.

To verify the equivalence of proposed [DSR](#) and [DBA](#) for pose estimation, the two methods are performed respectively from the same initial guess of poses. The objective functions and the step changes of the poses in each iteration are shown in Fig. 6.2. Note that in the proposed [DSR](#) method, the intensities of panoramic image are not needed. Panoramic image is computed in [DSR](#) method deliberately only for comparing the objective function. It is found that the values of the objective functions and the calculated step changes of the poses from the two methods are always the same at every iteration, which indicates the estimated poses from the two algorithms at every step are always the same. In addition, the mean computational costs of DBA and DSR for one iteration are compared in Table 6.2. The results indicate that DSR is 14-22% more efficient than DBA when they are run on MATLAB on the same computer. It is important to note that although we have proved DSR has lower computational complexity than DBA, our current focus of DSR is more on accuracy rather than efficiency. Therefore, we directly implement both DBA and DSR in MATLAB on CPU, which may not showcase the full advantages of DSR. We leave the work of fast implementation of DSR for our future work which will be discussed in Chapter 8.



TABLE 6.2: Mean Computational Cost

Patient ID	No. frames	Execution time (seconds/iteration)		
		APE	DBA	DSR
# 1	9	295.55	91.38	74.56
# 2	6	77.41	58.95	49.43
# 3	6	78.97	58.32	50.16
# 4	8	256.53	148.15	124.04
# 5	6	190.37	92.38	78.16
# 6	11	1100.36	240.49	188.74

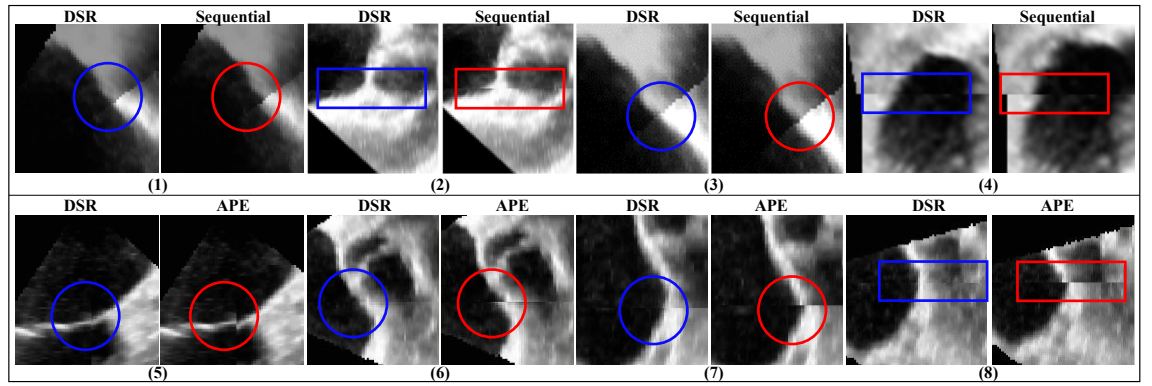


FIGURE 6.3: Comparisons of the aligned images using poses from sequential, APE, and DSR.

To evaluate the accuracy of the proposed [DSR](#) method, pairwise images are aligned using the poses obtained by the sequential, [APE](#), and [DSR](#) methods, respectively. Similar to the results in our simulated experiments, aligned images obtained from [DSR](#) have the best quality, followed by [APE](#) and then the sequential method. In all the aligned images, there is no misalignment found from the proposed [DSR](#) in the experiments, while some apparent misalignment is found from the results of [APE](#) and sequential methods. Several examples are displayed in Fig. 6.3. Additionally, although [APE](#) can obtain results that are closer to those from the proposed method than the sequential method, it requires a much longer time for each iteration. In the experiments, [APE](#) usually needs around 2-6 times longer time than the proposed method for each iteration, which is shown in Table 6.2.

To further evaluate the accuracy of the proposed method, in-vivo 3D [TEE](#) images in each dataset are fused using (6.8) and the estimated poses from [DSR](#). We manually select areas which contain sharp boundaries like the [LA](#) walls in the fused images for evaluation

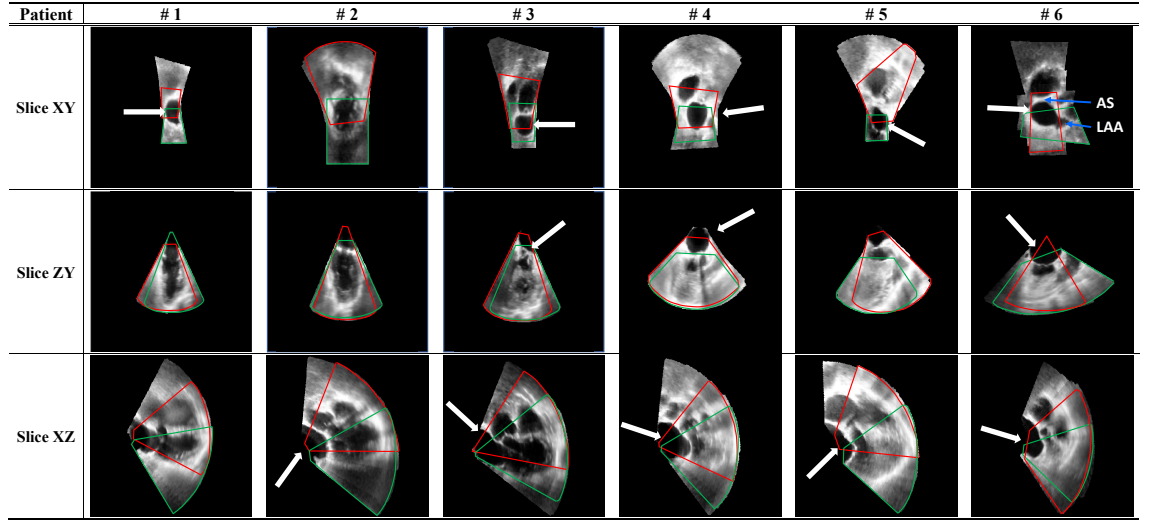


FIGURE 6.4: Fused 3D TEE images using registration results from DSR for six in-vivo datasets. LA walls which have sharp structures in the images are indicated by white arrows in selected areas. Colored frames are the boundaries of two registered volumes.

since generally, misalignment caused by poses with low accuracy can be easily found in these areas. The selected regions are shown in Fig. 6.4 with three orthogonal slices and two of the registered images in the fused images are highlighted in color boundaries. By observing the LA walls which are indicated by white arrows in Fig. 6.4, it is shown that the stitching areas have smooth transition and no misalignment is found in the images, which suggests good quality of alignments have been obtained by DSR.

By counting the number of voxels, it is found that the FoV of the fused image is enlarged to 2.10, 2.02, 1.80, 2.06, 2.01, and 2.18 times as compared with the original single TEE volume of Patient # 1 to # 6, respectively. From the reconstructed panoramic images shown in Fig. 6.4, complete cardiac structures of interest such as LAA and AS are observed again, which could facilitate device size selection for LAA occlusion [14] and allow measuring the relative position and orientation of LAA w.r.t. AS to assist the planning of LAA occlusion [91].

### 6.3 Chapter Summary

Starting from the framework of direct bundle adjustment (DBA), a novel direct simultaneous registration (DSR) algorithm for 3D images is proposed in this chapter. The method

can optimize the poses of a collection of local images simultaneously without any information loss or reuse. Results from the simulated and in-vivo experiments demonstrate that the proposed method outperforms the state-of-the-art [APE](#) method in terms of accuracy and efficiency. From the results of simulated experiments, it is evident that our method improved the accuracy of registration by more than 50% compared to the competing methods for most cases. In-vivo experiments also show accurate structures and extended [FoV](#) of the fused images.

Since both [DBA](#) and [DSR](#) are proposed based on the 3D images without deformation, [ECG](#)-gated 3D [TEE](#) images are used in Chapter 5 and 6. Inspired by [DBA](#) and [DSR](#), in the next chapter, we will propose a novel dynamic direct simultaneous registration framework that can estimate the deformations of the panoramic image at all phases corresponding to the local 3D [TEE](#) images simultaneously.

## Chapter 7

# D-DSR: Dynamic Direct Simultaneous Registration for Multiple 3D Images

In Chapter 4, we proposed a direct dynamic fusion (DDF) framework that can reconstruct 3D TEE images with enlarged FoV from multiview images with deformations. The method combines the pairwise registration method with a sequential fusion strategy to enlarge the FoV of 3D TEE images incrementally. Compared with the fusion methods that are based on ECG-gated images, DDF is capable of building the final panoramic images at different phases of cardiac cycles if we change the order of fusion. Therefore, motion-related information [93, 103] of the heart can be retained and analyzed in the fused panoramic images. We have shown that the DDF framework outperforms the state-of-the-art methods in terms of accuracy and can enlarge the FoV of the 3D TEE images effectively.

Inspired by the DBA and DSR presented in Chapter 5 and 6, we propose a novel dynamic direct simultaneous registration (D-DSR) framework in this chapter to estimate the 3D TEE panoramic images at all phases corresponding to the collected local images simultaneously. Similar to the DBA framework used for rigid scenarios, we firstly formulate our problem as a dynamic direct bundle adjustment problem by substituting the rigid transformations in DBA with embedded deformation graphs. Secondly, we prove a similar

property as DBA that the estimation of embedded deformation graphs in the dynamic direct bundle adjustment problem is independent of the intensities of the panorama during the optimization process and propose the D-DSR algorithm. Detailed validation of the proposed D-DSR algorithm is performed on the in-vivo 3D TEE images. Compared to our second framework DDF presented in Chapter 4, the main contributions of this chapter include: (1) in contrast to DDF that builds the panoramic images incrementally and needs to change the order of the fusion process to construct the final panoramic images at different phases, D-DSR can simultaneously build the panoramic images at all phases corresponding to the local images; (2) in contrast to DDF that estimates the deformation using two images every time in pairwise registration, D-DSR improves the accuracy of the deformation estimation by using simultaneous registration so that information of all collected images is used together.

## 7.1 Methods

### 7.1.1 Problem Statement

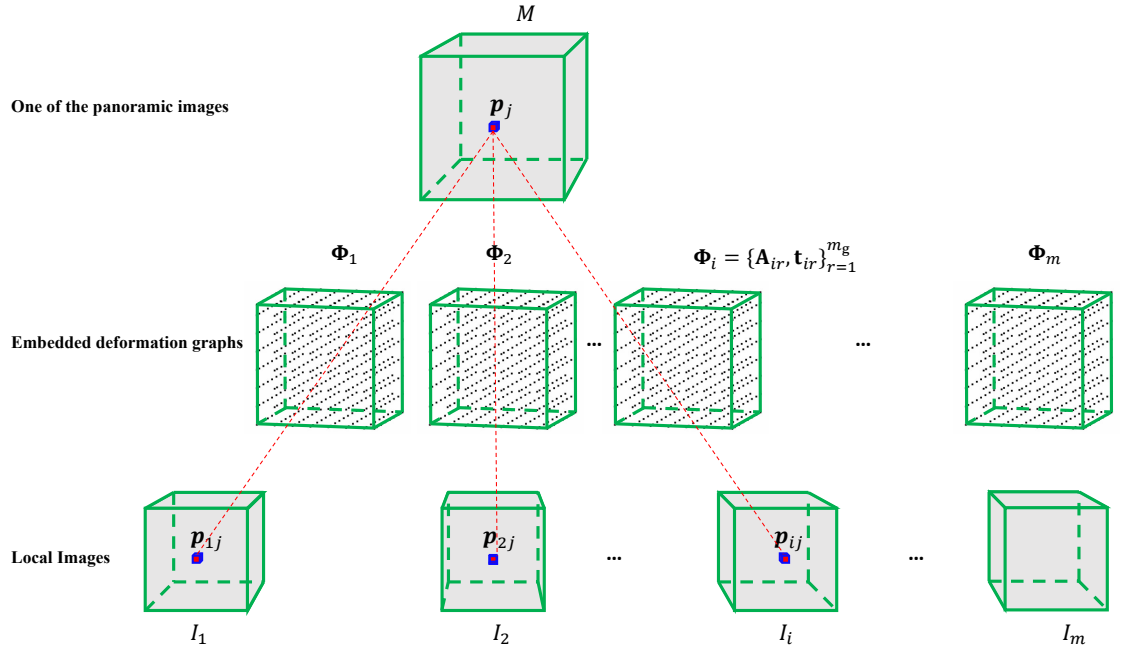


FIGURE 7.1: Schematic diagram of dynamic direct simultaneous registration of multiple 3D images with deformations.

In Fig. 7.1, there are  $m$  frames of local 3D images of a dynamic scene captured at different times/phases and viewpoints. Since we care more about the estimation of deformations of the panoramic image, we assume that all collected local images have been pre-registered before calculation to eliminate the global motion in this chapter. Our objective is to reconstruct  $m$  panoramic images at all phases corresponding to the local images in a simultaneous fashion.

Let  $I_i$  denote a local 3D image captured at the  $i_{th}$  phase and position. Suppose  $M$  is one of the panoramic images that is predefined at  $i_{th}$  phase corresponding to the image  $I_i$ .  $M$  consists of  $n$  voxels and the intensity of one voxel  $\mathbf{p}_j$  in it is denoted as  $M(\mathbf{p}_j)$ .  $\mathbf{p}_j = [u, v, w]^\top$ , where  $u, v, w$  are integers, denote the coordinates of a voxel on the 3D grid of  $M$ . Similar to Chapter 4, we use an embedded deformation graph to represent the transformation between two images. Let  $\Phi_i$  denote the transformation from the panoramic image  $M$  to the local image  $I_i$ . If the panorama  $M$  and  $m$  embedded deformation graphs corresponding to the local images are obtained simultaneously, we can easily reconstruct the deformation of the panoramic image at each phase by deforming the  $M$  with the corresponding embedded deformation graph. Therefore, for the  $m$  frames of 3D images, the overall state parameters we need to solve in the proposed problem are:

$$\mathbf{x} = [\mathbf{x}_\Phi^\top, \mathbf{x}_M^\top]^\top, \quad (7.1)$$

where  $\mathbf{x}_\Phi = [\Phi_1^\top, \dots, \Phi_i^\top, \dots, \Phi_m^\top]^\top$  are transformations from the panoramic image  $M$  to each local 3D image, and  $\mathbf{x}_M = [M(\mathbf{p}_1), \dots, M(\mathbf{p}_j), \dots, M(\mathbf{p}_n)]^\top$  represent the intensities of voxels in the panoramic image  $M$ .

### 7.1.2 Formulation of Dynamic Direct Bundle Adjustment

As illustrated in Fig. 7.1, the embedded deformation graph  $\Phi_i$  is represented by a set of affine matrices associated with nodes. Suppose that  $m_g$  graph nodes of  $\Phi_i$  are selected uniformly from  $M$  by downsampling it. Let  $\mathbf{g}_{ir} \in \mathbb{R}^3$  denote the position of a node  $r$  in graph  $\Phi_i$ , the affine transformation associated with  $r$  consists of a matrix  $\mathbf{A}_{ir} \in \mathbb{R}^{3 \times 3}$  and a translation vector  $\mathbf{t}_{ir} \in \mathbb{R}^3$ . Then, the graph  $\Phi_i$  can be denoted as  $\Phi_i = \{\mathbf{A}_{ir}, \mathbf{t}_{ir}\}_{r=1}^{m_g}$ .

If we set the same structure of  $\Phi_i$  to all the  $m$  embedded deformation graphs, there are  $m \times m_g$  nodes that need to be solved for the transformation part  $\mathbf{x}_\Phi$  in (7.1).

Similar to Chapter 4, for any voxel  $\mathbf{p}_j \in \mathbb{R}^3$  in image  $M$ , we limit the influences of deformation on voxel  $\mathbf{p}_j$  to its  $k$  nearest graph nodes. Suppose that a part of intensity information of voxel  $\mathbf{p}_j$  in the panoramic image is from a point  $\mathbf{p}_{ij}$  in local image  $I_i$ , the transformation from  $\mathbf{p}_j$  to  $\mathbf{p}_{ij}$  can be expressed by using the corresponding embedded graph  $\Phi_i$  by:

$$\mathbf{p}_{ij} = f(\Phi_i, \mathbf{p}_j) = \sum_{r=1}^k \hat{w}_{ir}(\mathbf{p}_j) [\mathbf{A}_{ir}(\mathbf{p}_j - \mathbf{g}_{ir}) + \mathbf{g}_{ir} + \mathbf{t}_{ir}], \quad (7.2)$$

where weight  $\hat{w}_{ir}(\mathbf{p}_j)$  denotes the influence of a nearby node on the deformation of the voxel  $\mathbf{p}_j$  and  $\sum_{r=1}^k \hat{w}_{ir} = 1$ .  $\hat{w}_{ir}$  is pre-calculated by the following formula and then normalized.

$$w_{ir}(\mathbf{p}_j) = (1 - \|\mathbf{p}_j - \mathbf{g}_{ir}\| / d_{max}), \quad (7.3)$$

where  $d_{max}$  denotes the distance from the voxel to its  $(k+1)_{th}$  nearest node. We set  $k = 6$  for all experiments presented in this chapter.

We follow the form of the energy function (4.3) in Chapter 4 to formulate the dynamic direct bundle adjustment over  $m \times m_g$  graph nodes and  $n$  voxels as follows:

$$E(\mathbf{x}_\Phi, \mathbf{x}_M) = w_{\text{rot}} E_{\text{rot}}(\mathbf{x}_\Phi) + w_{\text{reg}} E_{\text{reg}}(\mathbf{x}_\Phi) + w_{\text{con}} E_{\text{con}}(\mathbf{x}_\Phi, \mathbf{x}_M), \quad (7.4)$$

where  $E_{\text{rot}}(\mathbf{x}_\Phi)$ ,  $E_{\text{reg}}(\mathbf{x}_\Phi)$ , and  $E_{\text{con}}(\mathbf{x}_\Phi, \mathbf{x}_M)$  represent rotation term, regularization term, and constraint term respectively and  $w_{\text{rot}}$ ,  $w_{\text{reg}}$ , and  $w_{\text{con}}$  are the corresponding weights. Different from (4.3) that only considers the transformation parameters of nodes in one embedded deformation graph, here we consider the nodes in all embedded deformation graphs and the intensities of the panoramic image  $M$  simultaneously in (7.4).

Rotation term in (7.4) sums rotation errors over affine transformations of all deformation graphs' nodes to preserve the topological structure of the deformed panoramic images.

$$E_{\text{rot}}(\mathbf{x}_\Phi) = \mathbf{f}_{\text{rot}}^\top(\mathbf{x}_\Phi) \mathbf{f}_{\text{rot}}(\mathbf{x}_\Phi) = \sum_{i=1}^m \sum_{r=1}^{m_g} \text{Rot}(\mathbf{A}_{ir}), \quad (7.5)$$

where  $\mathbf{A}_{ir} = [\mathbf{a}_{ir1}, \mathbf{a}_{ir2}, \mathbf{a}_{ir3}]$  and

$$\begin{aligned} \text{Rot}(\mathbf{A}_{ir}) = & \left( \mathbf{a}_{ir1}^\top \cdot \mathbf{a}_{ir2} \right)^2 + \left( \mathbf{a}_{ir1}^\top \cdot \mathbf{a}_{ir3} \right)^2 + \left( \mathbf{a}_{ir2}^\top \cdot \mathbf{a}_{ir3} \right)^2 \\ & + \left( \mathbf{a}_{ir1}^\top \cdot \mathbf{a}_{ir1} - 1 \right)^2 + \left( \mathbf{a}_{ir2}^\top \cdot \mathbf{a}_{ir2} - 1 \right)^2 + \left( \mathbf{a}_{ir3}^\top \cdot \mathbf{a}_{ir3} - 1 \right)^2. \end{aligned} \quad (7.6)$$

Regularization term in (7.4) sums the squared error between the predicted positions of nodes and their actual positions [48] to keep the consistency and smoothness of the deformation graphs.

$$E_{\text{reg}}(\mathbf{x}_\Phi) = f_{\text{reg}}^\top(\mathbf{x}_\Phi) f_{\text{reg}}(\mathbf{x}_\Phi) = \sum_{i=1}^m \sum_{r=1}^{m_g} \sum_{s \in \mathbb{N}(r)} \alpha_{rs} \|\mathbf{A}_{ir}(\mathbf{g}_{is} - \mathbf{g}_{ir}) + \mathbf{g}_{ir} + \mathbf{t}_{ir} - (\mathbf{g}_{is} + \mathbf{t}_{is})\|^2, \quad (7.7)$$

where  $\mathbb{N}(r)$  is the index of nearby  $k$  nodes which influences node  $r$ , and the weight  $\alpha_{rs}$  is set as 1 [48].

Constraint term in (7.4) sums the squared intensity differences between the panoramic image  $M$  and local images:

$$E_{\text{con}}(\mathbf{x}_\Phi, \mathbf{x}_M) = f_{\text{con}}^\top(\mathbf{x}_\Phi, \mathbf{x}_M) f_{\text{con}}(\mathbf{x}_\Phi, \mathbf{x}_M) = \sum_{i=1}^m \sum_{j=1}^n \sigma(\mathbf{p}_{ij}) (e_{ij}(\Phi_i, M(\mathbf{p}_j)))^2, \quad (7.8)$$

where  $e_{ij}(\Phi_i, M(\mathbf{p}_j)) = M(\mathbf{p}_j) - I_i(f(\Phi_i, \mathbf{p}_j)) = M(\mathbf{p}_j) - I_i(\mathbf{p}_{ij})$ , and  $\sigma(\mathbf{p}_{ij}) = 1$  if the transformed point  $\mathbf{p}_{ij}$  is within the image region of  $I_i$ , otherwise  $\sigma(\mathbf{p}_{ij}) = 0$ . Since  $\mathbf{p}_{ij}$  may not be on the integer coordinates of the local image, the intensity of  $I_i(\mathbf{p}_{ij})$  is obtained using trilinear interpolation.

### 7.1.3 Dynamic Direct Simultaneous Registration without Intensity Optimization

Similar to Chapter 6, we can solve energy function (7.4) using GN iterations and the following theorem in the dynamic direct bundle adjustment problem can be proved.

**Theorem:** *When solving (7.4) with GN iterations, the optimization of the embedded deformation parameters is independent of the intensities of the predefined panoramic image.*



**Proof:** We rewrite the energy function (7.4) as a NLLS format:

$$E(\mathbf{x}) = F(\mathbf{x})^\top \Sigma F(\mathbf{x}), \quad (7.9)$$

where  $\Sigma$  is the weight matrix of (7.4) and  $F(\mathbf{x}) = [f_{\text{rot}}^\top(\mathbf{x}_\Phi), f_{\text{reg}}^\top(\mathbf{x}_\Phi), f_{\text{con}}^\top(\mathbf{x}_\Phi, \mathbf{x}_M)]^\top$  is the cost function over the rotation, regularization, and constraint terms. If we write Jacobian matrix of  $F(\mathbf{x})$  w.r.t.  $\mathbf{x}_\Phi$  and  $\mathbf{x}_M$  separately as  $J(\mathbf{x}) = [J_\Phi, J_M]$ , then GN equation can be written as the following format:

$$\begin{bmatrix} H_{\Phi\Phi} & H_{\Phi M} \\ H_{M\Phi} & H_{MM} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_\Phi \\ \Delta \mathbf{x}_M \end{bmatrix} = \begin{bmatrix} b_\Phi \\ b_M \end{bmatrix}, \quad (7.10)$$

where  $H_{\Phi\Phi} = J_\Phi^\top \Sigma J_\Phi$ ,  $H_{\Phi M} = H_{M\Phi}^\top = J_\Phi^\top \Sigma J_M$ ,  $H_{MM} = J_M^\top \Sigma J_M$ ,  $b_\Phi = -J_\Phi^\top \Sigma F(\mathbf{x})$ , and  $b_M = -J_M^\top \Sigma F(\mathbf{x})$ . Then, the step changes of poses and intensities of voxels can be computed sequentially using Schur complement [105]:

$$\left( H_{\Phi\Phi} - H_{\Phi M} H_{MM}^{-1} H_{\Phi M}^\top \right) \Delta \mathbf{x}_\Phi = (b_\Phi - H_{\Phi M} H_{MM}^{-1} b_M), \quad (7.11)$$

$$H_{MM} \Delta \mathbf{x}_M = b_M - H_{\Phi M}^\top \Delta \mathbf{x}_\Phi. \quad (7.12)$$

Let  $\mathbf{A}_M = [\mathbf{0}^\top, M(\mathbf{p}_j), \dots]^\top$  and  $\mathbf{B}(\mathbf{x}_\Phi) = [-F_{RR}^\top, I_i(\mathbf{p}_{ij}), \dots]^\top$ , where  $\mathbf{0}$  represents the zero vector with a compatible dimension,  $F_{RR} = [f_{\text{rot}}^\top(\mathbf{x}_\Phi), f_{\text{reg}}^\top(\mathbf{x}_\Phi)]^\top$  represents the cost vector of rotation and regularization terms, and  $M(\mathbf{p}_j)$  represents an observed intensity of the panoramic image and  $I_i(\mathbf{p}_{ij})$  represents the corresponding intensity in the local image  $I_i$ , respectively. Then, the right side of (7.11) becomes:

$$\begin{aligned} & b_\Phi - H_{\Phi M} H_{MM}^{-1} b_M \\ &= -J_\Phi^\top \Sigma (\mathbf{A}_M - \mathbf{B}(\mathbf{x}_\Phi)) + J_\Phi^\top \Sigma J_M (J_M^\top \Sigma J_M)^{-1} J_M^\top \Sigma (\mathbf{A}_M - \mathbf{B}(\mathbf{x}_\Phi)) \\ &= -J_\Phi^\top \Sigma (\mathbf{A}_M - J_M (J_M^\top \Sigma J_M)^{-1} J_M^\top \Sigma \mathbf{A}_M) - J_\Phi^\top \Sigma (J_M (J_M^\top \Sigma J_M)^{-1} J_M^\top \Sigma \mathbf{B}(\mathbf{x}_\Phi) - \mathbf{B}(\mathbf{x}_\Phi)). \end{aligned} \quad (7.13)$$

Similar to Section 6.1.2 in Chapter 6, according to the observed status of the panoramic image in the local image spaces which is indicated by the structure of  $J_M$ , we can easily deduce that  $\mathbf{A}_M = J_M \mathbf{x}_M$ . Substituting  $\mathbf{A}_M$  to the first term on the right side of (7.13),

we have:

$$-J_{\Phi}^{\top} \Sigma (\mathbf{A}_M - J_M (J_M^{\top} \Sigma J_M)^{-1} J_M^{\top} \Sigma \mathbf{A}_M) = \mathbf{0}. \quad (7.14)$$

Then, (7.13) becomes:

$$\left( H_{\Phi\Phi} - H_{\Phi M} H_{MM}^{-1} H_{\Phi M}^{\top} \right) \Delta \mathbf{x}_{\Phi} = -J_{\Phi}^{\top} \Sigma (J_M (J_M^{\top} \Sigma J_M)^{-1} J_M^{\top} \Sigma \mathbf{B}(\mathbf{x}_{\Phi}) - \mathbf{B}(\mathbf{x}_{\Phi})). \quad (7.15)$$

(7.15) indicates that the step changes of the parameters of the embedded deformation graphs  $\Delta \mathbf{x}_{\Phi}$  are independent of intensities  $\mathbf{x}_M$  in every GN iteration. Therefore, obtaining the optimal parameters of the embedded deformation graphs is independent of the intensities of panoramic image  $M$  throughout the optimization process. **Q.E.D.**

After the optimal solution to deformation graph  $\hat{\mathbf{x}}_{\Phi}$  is obtained, (7.9) becomes a linear least-squares problem and the optimal panoramic image can be calculated easily from the following closed form formula:

$$\hat{\mathbf{x}}_M = -H_{MM}^{-1} J_M^{\top} \Sigma \mathbf{B}(\hat{\mathbf{x}}_{\Phi}). \quad (7.16)$$

Based on the theorem, we propose the dynamic direct simultaneous registration (D-DSR) algorithm which solves the deformation graphs only without considering the intensities during the GN iterations. It calculates the optimal panoramic image  $M$  by (7.16) after the optimal deformation graphs are obtained and estimates the deformations of the  $M$  at the other  $m - 1$  phases by deforming it using the corresponding deformation graphs.

The implementation process of D-DSR is summarized in Algorithm 5. When implementing D-DSR, transformation parameters are initialized by setting each affine matrix associated with a node as an identity matrix. In addition, we fix one of the embedded deformation graphs to avoid singularity during the optimization. If the graph  $\Phi_i$  that represents the transformation from the panoramic image  $M$  to a local image  $I_i$  is fixed, the reconstructed panoramic image  $M$  will be at the same phase as the image  $I_i$ . This is realized by keeping all affine matrices in the graph  $\Phi_i$  as identity matrices during the optimization process.

**Algorithm 5:** Dynamic Direct Simultaneous Registration (D-DSR)**Input:** A collection of local 3D images  $\{I_i\}_{i=1}^m$  with various deformations.**Output:**  $m$  panoramas corresponding local 3D images respectively.

---

```

1 Step 1: Pre-computation:
2 Define a volume  $M$  (grid only) with  $n$  voxels;
3 Downsample  $M$  evenly to get a graph with  $m_g$  nodes, then define a graph set for all
  the  $m$  graphs  $\{\{\mathbf{g}_{ij}\}_{j=1}^{m_g}\}_{i=1}^m$ ;
4 Calculate gradient space of intensity of all the local 3D images;
5 Interpolate intensity and gradient spaces of all the 3D images;
6 Step 2: Optimization:
7 Initialize the transformation parameters  $\mathbf{x}_\Phi \leftarrow \mathbf{x}_{\Phi_0}$  by using identity matrices;
8 Define an empty vector  $\mathbf{c}$ ;
9 while Algorithm not converged do
10   Calculate cost vectors  $F_{RR} = [f_{\text{rot}}^\top(\mathbf{x}_\Phi), f_{\text{reg}}^\top(\mathbf{x}_\Phi)]^\top$  over rotation and regularization
    terms;
11   Calculate Jacobian matrices of  $F_{RR}$  w.r.t.  $\mathbf{x}_\Phi$  and  $\mathbf{x}_M$ , denoted as  $J_\Phi^{RR}$  and  $J_M^{RR}$ ;
12   for every voxel location  $\mathbf{p}_j$  in volume  $M$  do
13     for every image  $I_i$  do
14       Calculate  $\mathbf{p}_{ij}$  by (7.2);
15       if  $\sigma(\mathbf{p}_{ij}) = 1$  then
16         Calculate Jacobian  $J_\Phi^{ij} = [\mathbf{0}^\top, -\frac{\partial I_i}{\partial f(\Phi_i, \mathbf{p}_j)} \frac{\partial f(\Phi_i, \mathbf{p}_j)}{\partial \Phi_i}, \mathbf{0}^\top]$ ;
17         Calculate Jacobian  $J_M^{ij} = [\mathbf{0}^\top, 1, \mathbf{0}^\top]$ ;
18         Save the intensity  $\mathbf{c} = [\mathbf{c}, I_i(\mathbf{p}_{ij})]$ ;
19       end
20     end
21   end
22    $J_\Phi^{\text{con}} = [..., J_\Phi^{ij^\top}, ...]^\top$ ,  $J_M^{\text{con}} = [..., J_M^{ij^\top}, ...]^\top$ ;
23    $J_\Phi = [(J_\Phi^{RR})^\top, (J_\Phi^{\text{con}})^\top]^\top$ ,  $J_M = [(J_M^{RR})^\top, (J_M^{\text{con}})^\top]^\top$ ;
24   Construct weight matrix  $\Sigma$  of the energy function (7.4);
25   Construct  $\mathbf{B}(\mathbf{x}_\Phi) = [-F_{RR}^\top, \mathbf{c}^\top]^\top$ ;
26   Calculate  $H_{\Phi\Phi} = J_\Phi^\top \Sigma J_\Phi$ ,  $H_{\Phi M} = H_{M\Phi}^\top = J_\Phi^\top \Sigma J_M$ , and  $H_{MM} = J_M^\top \Sigma J_M$ ;
27   Calculate  $\Delta \mathbf{x}_\Phi = [\Delta \Phi_1^\top, ..., \Delta \Phi_i^\top, ..., \Delta \Phi_m^\top]^\top$  by (7.15);
28   Update the transformation parameters  $\mathbf{x}_\Phi \leftarrow \mathbf{x}_\Phi + \Delta \mathbf{x}_\Phi$ ;
29 end
30 Calculate the panoramic image  $M$  by (7.16);
31 Deform  $M$  to the other  $m - 1$  phases by (7.2) with corresponding graphs;

```

---

## 7.2 Experiments and Results

We use in-vivo experiments to validate the proposed D-DSR algorithm. A total of twenty-four sequences of in-vivo 3D TEE images with various deformations are collected from four

TABLE 7.1: Details of the Four In-vivo Datasets

Patient	No. Sequences	Volume Size (voxel)	Resolution (mm/voxel)
# 1	6	$240 \times 160 \times 208$	$0.69 \times 0.98 \times 0.73$
# 2	6	$240 \times 160 \times 208$	$0.77 \times 1.11 \times 0.82$
# 3	6	$240 \times 160 \times 208$	$0.64 \times 0.92 \times 0.68$
# 4	6	$272 \times 208 \times 208$	$0.63 \times 0.90 \times 0.68$

patients using a real-time 3D ultrasound transducer. The details of the in-vivo datasets # 1, # 2, # 3, and # 4 used for the experiments are listed in Table 7.1.

In Chapter 4, we have shown superior accuracy of the DDF framework to Plastimatch [46], ElastiX [47], the original Demons [50], the diffeomorphic Demons [53], and the state-of-the-art learning-based method VoxelMorph [74]. The final fused volumes of the four patients show that the DDF can reconstruct the 3D TEE images with good quality and enlarge the FoV of 3D TEE images effectively. In this section, the proposed D-DSR method is compared with the DDF to evaluate the accuracy of D-DSR using the collected in-vivo datasets. In our experiments, the coefficients in (7.4) are specified as  $w_{\text{rot}} = 1$ ,  $w_{\text{reg}} = 60000$ ,  $w_{\text{con}} = 0.1$  for the D-DSR method, while the coefficients for DDF are set as  $w_{\text{rot}} = 1$ ,  $w_{\text{reg}} = 10000$ ,  $w_{\text{con}} = 0.1$ , which are the same as those in Chapter 4.

In the experiments, firstly, six ECG-gated TEE volumes are extracted from six sequences in each dataset by synchronizing acquired 3D TEE images to ECG signals at the same phase. Therefore, four sequences that contain six ECG-gated 3D TEE images in each are obtained and denoted as *ECG-1*, *ECG-2*, *ECG-3*, and *ECG-4*, respectively. Secondly, six TEE volumes with different deformations are selected from six sequences in each dataset at different phases and then, four sequences that contain six 3D TEE images with different deformations in each are obtained and denoted as *Deform-1*, *Deform-2*, *Deform-3*, and *Deform-4*, respectively. In each sequence of the images with deformations, the sixth volume is selected as an ECG-gated one.

The DDF and D-DSR methods are applied to fuse each sequence of the images with deformations, respectively. DDF employs a sequential fusion strategy for fusing multiview images in which we repeatedly fuse one image into the panoramic image until all images

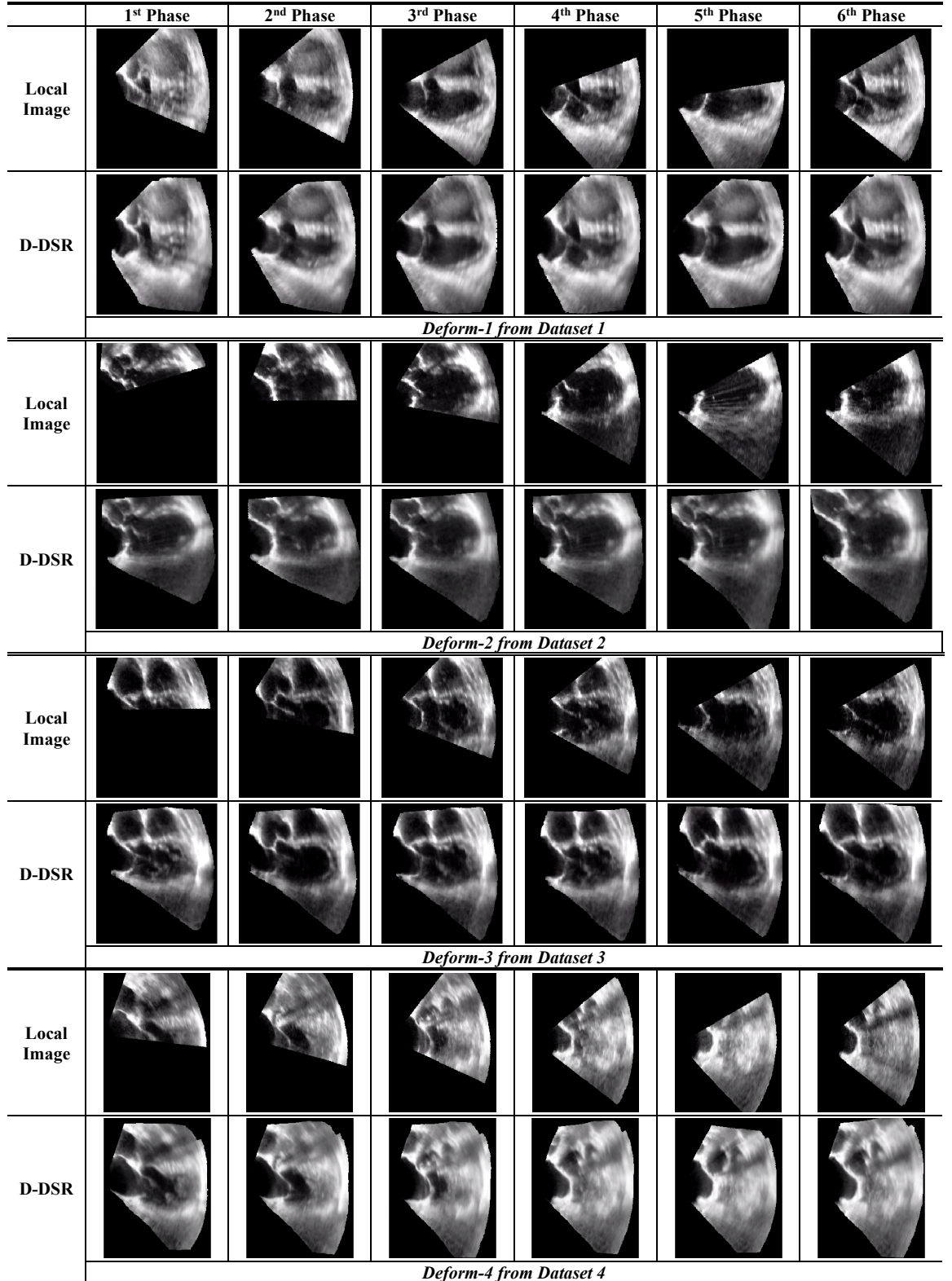


FIGURE 7.2: Comparisons of original local images with reconstructed panoramic images at different phases obtained by D-DSR.

are fused together. For each loop of registration and fusion, it can only estimate the deformation from one moving image (panoramic image) to one fixed image (local image). On the contrary, D-DSR simultaneously estimates the deformations of the panoramic image at all phases corresponding to the local images. Fig. 7.2 shows the estimated deformations of the panoramic image at six phases for each sequence with deformations. It is found that the reconstructed panoramic images at different phases are almost the same as corresponding local images for the overlapping area, which indicates that accurate deformation at each phase is obtained by D-DSR. In addition to obtaining accurate anatomical structures for the overlapping areas, D-DSR is effective to predict the deformation in un-overlapping areas so that extended FoV can be obtained from the reconstructed panoramic images compared with the original local images, which is also shown in Fig. 7.2.

In Fig. 7.3, the results of the deformation estimation of the sequence *Deform-3* at different phases obtained by DDF and D-DSR methods are compared. Local images are used as the references for accuracy evaluation from the second phase to the fifth phase. Since the sixth volume in each sequence with deformations is an ECG-gated one, which means that the estimated deformation at the sixth phase should be the same as the ECG-gated image. Therefore, the estimated deformation at the sixth phase uses the panoramic image fused by ECG-gated images as the reference in each sequence to estimate the overall accuracy, shown in Fig. 7.4.

In Fig. 7.3, through comparing the reference images with the deformed images estimated by DDF and D-DSR, it is found that D-DSR can obtain almost identical results as the references, while some visible differences can be found from the images estimated by DDF. The obvious differences in the images obtained by DDF and the same areas in the reference images and the images obtained by D-DSR are marked by red boxes to facilitate evaluation in Fig. 7.3. In addition, as more images are fused into the panoramic image by DDF, the errors in the deformed images tend to be accumulated, which can be found easily at the fourth and fifth phases in Fig. 7.3. Furthermore, in Fig. 7.4, the overall accuracy of the results obtained by DDF and D-DSR are evaluated. Through the comparisons, it is found that the deformations estimated by the D-DSR method have almost the same anatomical structures as the images from the panoramic images fused by ECG-gated images.





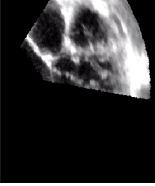





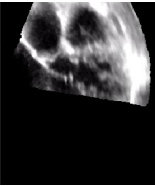
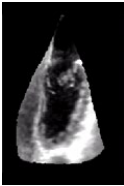






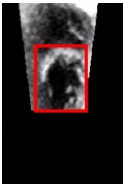
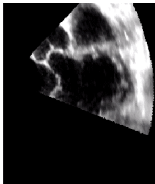
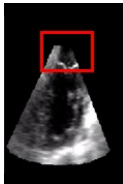
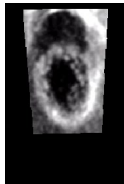
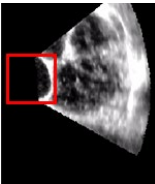

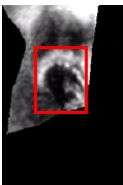

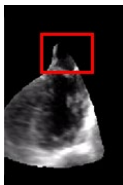
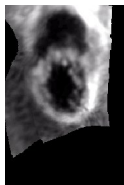
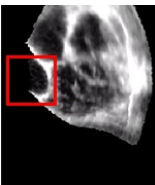
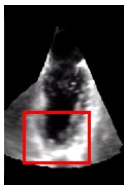
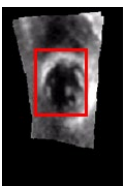
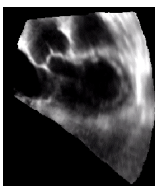

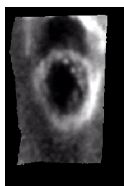
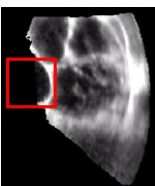
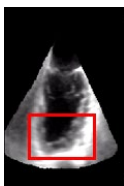
	2 <sup>nd</sup> Phase			3 <sup>rd</sup> Phase		
	Slice XY	Slice YZ	Slice XZ	Slice XY	Slice YZ	Slice XZ
Reference (local image)						
DDF						
D-DSR						
	4 <sup>th</sup> Phase			5 <sup>th</sup> Phase		
	Slice XY	Slice YZ	Slice XZ	Slice XY	Slice YZ	Slice XZ
Reference (local image)						
DDF						
D-DSR						

FIGURE 7.3: Evaluation of the reconstructed panoramic images obtained by DDF and D-DSR from the second phase to the fifth phase based on *Deform-3* from the dataset # 3. The same areas on the images are marked by red boxes for comparison.



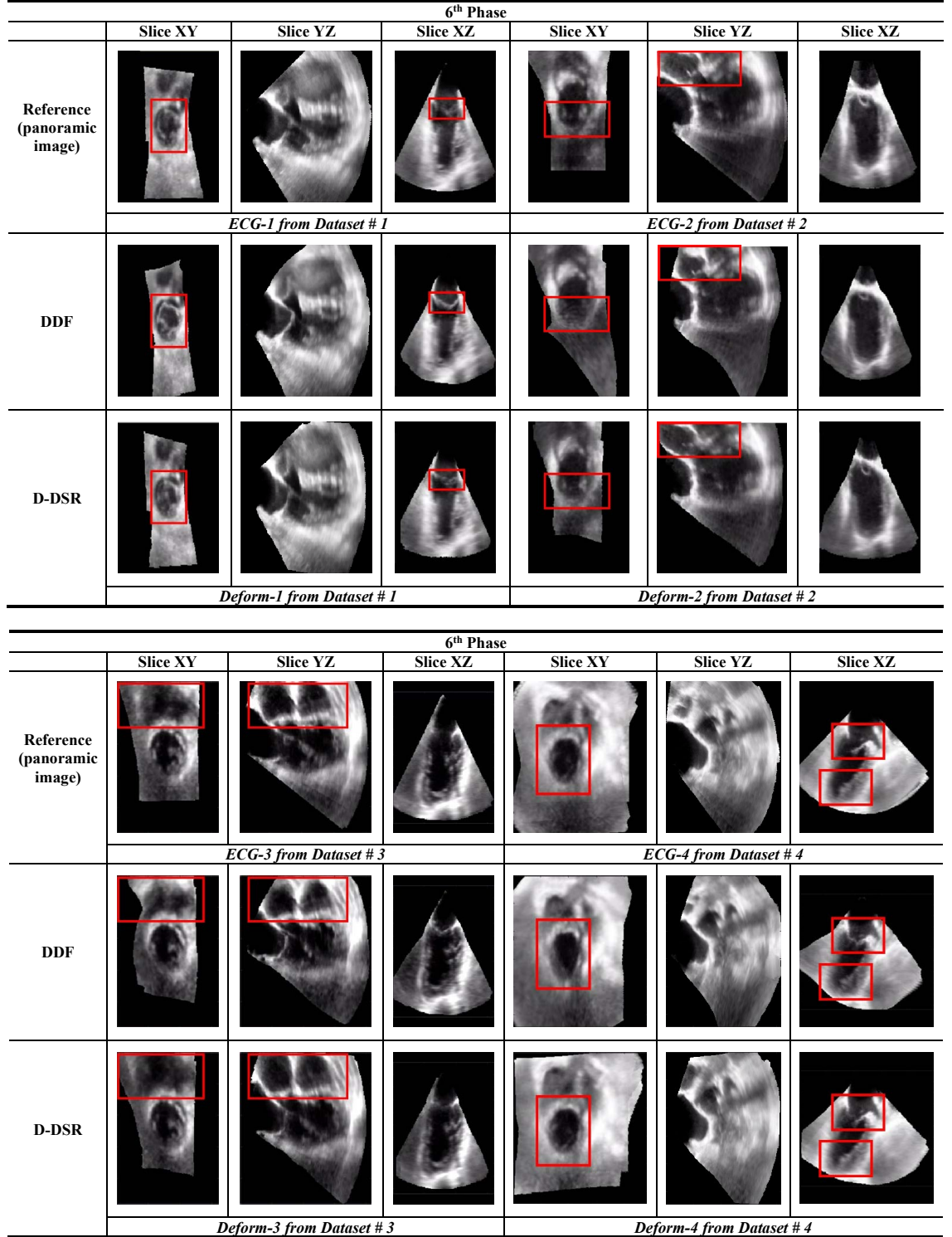


FIGURE 7.4: Evaluation of the reconstructed panoramic images obtained by DDF and D-DSR at the sixth phase of each sequence with deformations. The same areas on the images are marked by red boxes for comparison.



However, some accumulating errors are found in the deformed images by **DDF**. Therefore, both Fig. 7.3 and Fig. 7.4 indicate that the proposed **D-DSR** method outperforms **DDF** in terms of accuracy. This is because **DDF** estimates deformation using two images every time, while **D-DSR** estimates deformations at every phase using all the obtained images together, which provides more image information than **DDF** for the deformation estimation. Moreover, in contrast to **DDF** which enlarges the **FoV** of 3D **TEE** images incrementally, **D-DSR** can use all the collected local images to build the global images at all phases simultaneously, which could better assist clinicians in practice to build motion and deformation models or analyze motion-related markers of diseases on the panoramic images [93, 103].

### 7.3 Chapter Summary

This chapter presents a dynamic direct simultaneous registration (**D-DSR**) framework for enlarging the **FoV** of 3D **TEE** images in deformable scenarios. Different from the previously proposed **DDF** method in Chapter 4 that uses a sequential fusion strategy to fuse multiview images, which accumulates errors as the fusion process proceeds, the proposed **D-DSR** is capable of reconstructing the panoramic images at all phases corresponding to local images simultaneously. The results of comparative experiments based on four in-vivo datasets show that the **D-DSR** method can not only estimate the panoramic images at all phases simultaneously but also obtain the results with higher accuracy than the **DDF** method.

## Chapter 8

# Conclusion and Future Plans

In this thesis, we aim at developing innovative multiview registration and fusion algorithms to enlarge the [FoV](#) of 3D medical images. While the proposed algorithms could be applied to a wide variety of image types, we mainly focus on the registration and fusion of 3D [TEE](#) images with potential applications to assist clinicians with diagnosis, therapy, and surgery for patients with various cardiac diseases, such as facilitating device size selection and path planning for surgical procedures of [LAA](#) occlusion. From the pairwise-based registration and fusion methods to the simultaneous registration and fusion methods, five works have been accomplished to enlarge the [FoV](#) of 3D [TEE](#) images in both rigid and deformable scenarios, solving some major challenges in multiview registration and fusion of 3D [TEE](#) images. In Part [I](#) which consists of Chapter [3](#) and [4](#), two frameworks are presented, which combine pairwise registration with sequential fusion strategies to fuse multiview 3D [TEE](#) images in rigid and deformable scenarios, respectively. In Part [II](#) which consists of Chapter [5](#), [6](#), and [7](#), three simultaneous registration and fusion algorithms are presented to further improve the accuracy of registration and fusion of 3D [TEE](#) images in rigid and dynamic scenarios.

In the first work presented in Chapter [3](#), we propose an efficient multiview registration and fusion framework to extend the [FoV](#) of [ECG](#)-gated 3D [TEE](#) images. A direct method that solves the pairwise registration problem in Lie algebra space is proposed for the registration of 3D [TEE](#) images. Fast implementation of the method is realized by searching for the

corresponding three orthogonal planes in the voxel spaces of moving and fixed images. We employ the analytical expression of Jacobian in the optimization process. Compared to many intensity-based registration methods that use numerical Jacobian in the optimization process such as *imregtform* in MATLAB, our solution to the optimization shows higher accuracy and better convergence in the simulated experiments. In this work, a 3D fusion method is also proposed to fuse aligned images seamlessly and efficiently. A sequential fusion strategy is used for fusing multiple images. Detailed Monte-Carlo simulations and in-vivo experiments are conducted to verify the effectiveness of the proposed methods. In-vivo experiments show that the proposed registration method outperforms the state-of-the-art PCA-based method in terms of accuracy and efficiency. Furthermore, the results of registration and fusion of 76 in-vivo 3D TEE volumes from nine patients demonstrate that the proposed framework can extend the FoV of 3D TEE images effectively. The FoV of the fused image for each dataset is enlarged to around two times as compared with the FoV of an original single 3D TEE image.

In the second work presented in Chapter 4, we propose a novel direct dynamic fusion (DDF) framework to enlarge the FoV of 3D TEE images by sequentially fusing multiview images containing various deformations. A direct embedded deformation method is proposed in the framework to deal with the large cardiac deformation in 3D TEE images. In-vivo experiments are conducted to evaluate our method compared with the state-of-the-art non-learning-based and learning-based registration methods. Qualitative and quantitative results of the experiments show that the proposed method outperforms the competing methods in terms of accuracy. The final fused panoramic images based on the in-vivo datasets show that the proposed DDF framework can reconstruct the 3D TEE images with good quality and enlarge the FoV of images around two times compared with that of the original single images. Without selecting the static (ECG-gated) images from the same cardiac phase, DDF could provide an alternative to the conventional fusion methods that are biased toward the specifically chosen phase.

In the third work presented in Chapter 5, we propose a novel direct bundle adjustment (DBA) framework for completing multiview 3D image registration and fusion simultaneously in the rigid scenario. The method redefines classical bundle adjustment (BA) by optimizing the intensities of the panoramic image (instead of positions of scene points in

classical BA) and the poses of the local images simultaneously. It is achieved by maximizing the similarity between the predefined panoramic image and local images. Detailed experiments are conducted in this chapter and show that the DBA method is robust to intensity noises and can obtain more accurate results than the sequential fusion method presented in Chapter 3. The results of in-vivo experiments show apparent enlargement of the FoV in the final fused images compared with that of the original single 3D TEE images. Without any reference image, correspondences, or information loss or reuse, the DBA method is an elegant way to obtain the optimal global image and poses of local images in one go. This work has filled an important gap in 3D image registration and fusion.

In the fourth work presented in Chapter 6, we propose a novel direct simultaneous registration (DSR) algorithm for 3D images. This algorithm is proposed based on our finding in DBA that “When solving DBA problem with GN iterations, the optimization of poses is independent of the intensities of the panoramic image”. Compared with DBA, the proposed DSR algorithm solves the poses only without solving the huge dimension of intensities of the panoramic image but obtains the same optimal poses as DBA. The method can also optimize the poses of a collection of 3D images simultaneously without any information loss or reuse. Results from the simulated and in-vivo experiments demonstrate that the proposed method outperforms the other competing methods in terms of accuracy and is much more efficient than the state-of-the-art APE method. It is shown from the simulated experiments that our method improves the accuracy of registration by more than 50% compared to the other competing methods for most cases. In-vivo experiments also show accurate anatomical structures and extended FoV in the fused images.

In the fifth work presented in Chapter 7, we propose a novel dynamic direct simultaneous registration (D-DSR) framework that could estimate the panoramic images at all phases corresponding to local images simultaneously. This work is inspired by DBA and DSR which are used in rigid scenarios. Firstly, we formulate our problem as a dynamic direct bundle adjustment problem by substituting the rigid transformations in DBA with embedded deformation graphs. Secondly, a similar theorem as DBA is proved that the estimation of embedded deformation graphs in the dynamic direct bundle adjustment problem is independent of the intensities of the predefined panoramic image during the optimization

process. Then, based on the theorem, the **D-DSR** algorithm is proposed. We validate **D-DSR** by comparing it with the **DDF** method presented in Chapter 4 using the in-vivo 3D **TEE** images. The results show that **D-DSR** can reconstruct the panoramic images with enlarged **FoV** at all phases simultaneously. And the results show a higher accuracy of registration and fusion than the **DDF** method that is deduced from pairwise registration and fusion.

Our methods of enlarging the **FoV** in 3D **TEE** images have significant potential clinical values of assisting clinicians' diagnosis and surgical planning for patients with various cardiac diseases. In the planning of most cardiac surgery and intervention, it is crucial for cardiac structures of interest to be observed in a single image for a complete analysis of the spatial orientation of these structures. The proposed registration and fusion methods allow visualizing complete cardiac structures of interest in a single volume and therefore, measurement of distances and angles related to these structures can be realized easily. For instance, in the pre-operative planning of **LAA** occlusion, firstly, the enlarged **FoV** of 3D **TEE** allows the **LAA** to be observed completely to facilitate device size selection for the optimal occluder. Secondly, a complete structure of **LA** in the enlarged 3D **TEE** image allows measuring the relative position and orientation of **LAA** w.r.t. **AS** to facilitate the path planning for **LAA** occlusion.

In the future, we plan to further improve our current works from three aspects:

- Complete the efficient implementation of **DSR**. Currently, our focus in **DSR** is more on accuracy than efficiency, thus we implemented it in MATLAB on CPU. Since the linear system of the **GN** equation has a special sparse structure that is similar to other **BA** problems, it is very promising for us to use techniques in [109] to achieve parallel implementation of **DSR** on GPU. Additionally, optimization techniques used in g2o [110] and parallax BA [111] could also help us achieve the fast implementation. Our future work will focus on fast implementation of **DSR**.
- Extend our **DSR** work to multimodal registration such as 3D ultrasound to **CT**. Our current **DSR** employs **SSD** as the similarity metric which is usually used to measure the similarity among images obtained from monomodality. We may extend **DSR**

method for multimodal registration through two strategies: (1) change a similarity metric such as [MI](#) that can measure the similarity among images from multimodality; (2) use methods such as structural image representation [\[78\]](#) to convert image from one modality to the other one, and then the current [DSR](#) method can be applied to these images directly.

- Improve the evaluation of the proposed [D-DSR](#) method. Currently, we validated [D-DSR](#) qualitatively using the in-vivo 3D [TEE](#) datasets, compared with the [DDF](#) method. In the future, we are planning to improve the qualitative evaluation of the method by using [CT](#) images as the ground truth.

# Bibliography

- [1] Alexander L. Klibanov and John A. Hossack. Ultrasound in radiology: From anatomic, functional, molecular imaging to drug delivery and image-guided therapy. *Investigative Radiology*, 50(9):657–670, 2015.
- [2] Aaron Fenster, Dónal B Downey, and H Neale Cardinal. Three-dimensional ultrasound imaging. *Physics in Medicine and Biology*, 46(5):R67–R99, apr 2001.
- [3] Alex Pui-Wai Lee, Yat-Yin Lam, Gabriel Wai-Kwok Yip, Roberto M Lang, Qing Zhang, and Cheuk-Man Yu. Role of real time three-dimensional transesophageal echocardiography in guidance of interventional procedures in cardiology. *Heart*, 96(18):1485–1493, 2010.
- [4] Paaladinesh Thavendiranathan, Shizhen Liu, David Verhaert, Anna Calleja, Adrien Nitinunu, Thomas Van Houten, Nathalie De Michelis, Orlando Simonetti, Sanjay Rajagopalan, Thomas Ryan, et al. Feasibility, accuracy, and reproducibility of real-time full-volume 3d transthoracic echocardiography to measure lv volumes and systolic function: a fully automated endocardial contouring algorithm in sinus rhythm and atrial fibrillation. *JACC: Cardiovascular Imaging*, 5(3):239–251, 2012.
- [5] Qinghua Huang and Zhaozheng Zeng. A review on real-time 3d ultrasound imaging technology. *BioMed research international*, 2017:1–20, 2017.
- [6] Mohammad Hamed Mozaffari and Won-Sook Lee. Freehand 3-d ultrasound imaging: a systematic review. *Ultrasound in medicine & biology*, 43(10):2099–2124, 2017.
- [7] Rebecca T Hahn, Theodore Abraham, Mark S Adams, Charles J Bruce, Kathryn E Glas, Roberto M Lang, Scott T Reeves, Jack S Shanewise, Samuel C Siu, William

- Stewart, et al. Guidelines for performing a comprehensive transesophageal echocardiographic examination: recommendations from the american society of echocardiography and the society of cardiovascular anesthesiologists. *Journal of the American Society of Echocardiography*, 26(9):921–964, 2013.
- [8] Fatma El-Zahraa Ahmed El-Gamal, Mohammed Elmogy, and Ahmed Atwan. Current trends in medical image registration and fusion. *Egyptian Informatics Journal*, 17(1):99–124, 2016.
- [9] Francisco PM Oliveira and Joao Manuel RS Tavares. Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, 17(2):73–93, 2014.
- [10] Nina C Wunderlich, Roy Beigel, Martin J Swaans, Siew Yen Ho, and Robert J Siegel. Percutaneous interventions for left atrial appendage exclusion: options, assessment, and imaging using 2d and 3d echocardiography. *JACC: Cardiovascular Imaging*, 8(4):472–488, 2015.
- [11] Jai-Wun Park, Armando Bethencourt, Horst Sievert, Gennaro Santoro, Bernhard Meier, Kevin Walsh, Jose Ramon Lopez-Minquez, David Meerkin, Mariano Valdés, Oliver Ormerod, et al. Left atrial appendage closure with amplatzer cardiac plug in atrial fibrillation: initial european experience. *Catheterization and Cardiovascular Interventions*, 77(5):700–706, 2011.
- [12] Joseph L Blackshear and John A Odell. Appendage obliteration to reduce stroke in cardiac surgical patients with atrial fibrillation. *The Annals of thoracic surgery*, 61(2):755–759, 1996.
- [13] Achille Gaspardone, Maria Benedetta Giannico, Roberto Cancellieri, Laura Velardi, Antonio Giuseppe Posteraro, Alessandra Cinque, and Maria Iamele. Percutaneous retrieval of a left atrial appendage closure device: The device waltz. *Case Reports*, 3(5):766–771, 2021.
- [14] Pedro Morais, João L Vilça, Sandro Queirós, Pieter De Meester, Werner Budts, Joao Manuel RS Tavares, and Jan D’hooge. Semiautomatic estimation of device



- size for left atrial appendage occlusion in 3-d tee images. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 66(5):922–929, 2019.
- [15] Hyemoon Chung, Byunghwan Jeon, Hyuk-Jae Chang, Dongjin Han, Hackjoon Shim, In Jeong Cho, Chi Young Shim, Geu-Ru Hong, Jung-Sun Kim, Yangsoo Jang, et al. Predicting peri-device leakage of left atrial appendage device closure using novel three-dimensional geometric ct analysis. *Journal of cardiovascular ultrasound*, 23(4):211–218, 2015.
- [16] Mark Cox, Sridha Sridharan, Simon Lucey, and Jeffrey Cohn. Least-squares congealing for large numbers of images. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1949–1956. IEEE, 2009.
- [17] Christian Wachinger and Nassir Navab. Simultaneous registration of multiple images: similarity metrics and efficient optimization. *IEEE transactions on pattern analysis and machine intelligence*, 35(5):1221–1233, 2012.
- [18] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999.
- [19] Amaël Delaunoy and Marc Pollefeys. Photometric bundle adjustment for dense multi-view 3d modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1486–1493, 2014.
- [20] Hatem Alismail, Brett Browning, and Simon Lucey. Photometric bundle adjustment for vision-based slam. In *Asian Conference on Computer Vision*, pages 324–341. Springer, 2016.
- [21] Daniel Rueckert and Julia A. Schnabel. *Medical Image Registration*, pages 131–154. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-15816-2.
- [22] Robert J Schneider, Douglas P Perrin, Nikolay V Vasilyev, Gerald R Marx, J Pedro, and Robert D Howe. Real-time image-based rigid registration of three-dimensional ultrasound. *Medical image analysis*, 16(2):402–414, 2012.

- [23] Dong Ni, Yim Pan Chui, Yingge Qu, Xuan Yang, Jing Qin, Tien-Tsin Wong, Simon SH Ho, and Pheng Ann Heng. Reconstruction of volumetric ultrasound panorama based on improved 3d sift. *Computerized medical imaging and graphics*, 33(7):559–566, 2009.
- [24] Michal Irani and P Anandan. About direct methods. In *International Workshop on Vision Algorithms*, pages 267–277. Springer, 1999.
- [25] Chengqian Che, Tejas Sudharshan Mathai, and John Galeotti. Ultrasound registration: A review. *Methods*, 115:128–143, 2017.
- [26] Andrew H Gee, Graham M Treece, Richard W Prager, Charlotte JC Cash, and Laurence Berman. Rapid registration for wide field of view freehand three-dimensional ultrasound. *IEEE transactions on medical imaging*, 22(11):1344–1357, 2003.
- [27] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004.
- [28] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Lecture notes in computer science*, 3951:404–417, 2006.
- [29] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011.
- [30] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part II 13*, pages 834–849. Springer, 2014.
- [31] Blaine Rister, Mark A Horowitz, and Daniel L Rubin. Volumetric image registration from invariant keypoints. *IEEE Transactions on Image Processing*, 26(10):4900–4910, 2017.
- [32] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006.

- [33] Alexis Roche, Gregoire Malandain, and Nicholas Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology*, 11(1):71–80, 2000.
- [34] Derek L G Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):R1–R45, feb 2001.
- [35] Andriy Myronenko and Xubo Song. Intensity-based image registration by minimizing residual complexity. *IEEE transactions on medical imaging*, 29(11):1882–1891, 2010.
- [36] V. Grau, H. Becher, and J. A. Noble. Registration of multiview real-time 3-d echocardiographic sequences. *IEEE Transactions on Medical Imaging*, 26(9):1154–1165, Sep. 2007. ISSN 1558-254X.
- [37] R James Housden, YingLiang Ma, Aruna Arujuna, Niels Nijhof, Pascal Cathier, Geert Gijssbers, Roland Bullens, Jaswinder Gill, C Aldo Rinaldi, Victoria Parish, et al. Extended-field-of-view three-dimensional transesophageal echocardiography using image-based x-ray probe tracking. *Ultrasound in medicine & biology*, 39(6):993–1005, 2013.
- [38] D. Peressutti, A. Gomez, G. P. Penney, and A. P. King. Registration of multiview echocardiography sequences using a subspace error metric. *IEEE Transactions on Biomedical Engineering*, 64(2):352–361, 2017.
- [39] Oliver Kutter, Wolfgang Wein, and Nassir Navab. Multi-modal registration based ultrasound mosaicing. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 763–770. Springer, 2009.
- [40] Benoit Desjardins and Ella A Kazerooni. Ecg-gated cardiac ct. *American Journal of Roentgenology*, 182(4):993–1010, 2004.
- [41] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.

- [42] Thomas W Sederberg and Scott R Parry. Free-form deformation of solid geometric models. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 151–160, 1986.
- [43] Michaël Sdika. A fast nonrigid image registration with constraints on the jacobian using large scale constrained optimization. *IEEE transactions on medical imaging*, 27(2):271–281, 2008.
- [44] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.
- [45] Huizhong Ji, Yusen Li, Enqing Dong, Peng Xue, Wenshuo Xiong, Wenyan Sun, Zhenchao Tang, Dejing Zhang, and Wei Fang. A non-rigid image registration method based on multi-level b-spline and l2-regularization. *Signal, Image and Video Processing*, 12(6):1217–1225, 2018.
- [46] Paolo Zaffino, Patrik Raudaschl, Karl Fritscher, Gregory C Sharp, and Maria Francesca Spadea. Plastimatch mabs, an open source tool for automatic image segmentation. *Medical physics*, 43(9):5155–5160, 2016.
- [47] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009.
- [48] Robert W Sumner, Johannes Schmid, and Mark Pauly. Embedded deformation for shape manipulation. In *ACM SIGGRAPH 2007 papers*, pages 80–es. 2007.
- [49] Jingwei Song, Jun Wang, Liang Zhao, Shoudong Huang, and Gamini Dissanayake. Dynamic reconstruction of deformable soft-tissue with stereo scope in minimal invasive surgery. *IEEE Robotics and Automation Letters*, 3(1):155–162, 2017.
- [50] J-P Thirion. Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical image analysis*, 2(3):243–260, 1998.
- [51] He Wang, Lei Dong, Jennifer O’Daniel, Radhe Mohan, Adam S Garden, K Kian Ang, Deborah A Kuban, Mark Bonnen, Joe Y Chang, and Rex Cheung. Validation

- of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy. *Physics in Medicine & Biology*, 50(12):2887, 2005.
- [52] Monan Wang and Pengcheng Li. A review of deformation models in medical image registration. *Journal of Medical and Biological Engineering*, 39(1):1–17, 2019.
- [53] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.
- [54] Jonathan Santos, Abhijit J Chaudhari, Anand A Joshi, Andrea Ferrero, Kai Yang, John M Boone, and Ramsey D Badawi. Non-rigid registration of serial dedicated breast ct, longitudinal dedicated breast ct and pet/ct images using the diffeomorphic demons method. *Physica Medica*, 30(6):713–717, 2014.
- [55] Siqu Liu, Sidong Liu, Fan Zhang, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng. Longitudinal brain mr retrieval with diffeomorphic demons registration: What happened to those patients with similar changes? In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 588–591. IEEE, 2015.
- [56] Daniel Markel, Ives Levesque, Joe Larkin, Pierre Léger, and Issam El Naqa. A 4d biomechanical lung phantom for joint segmentation/registration evaluation. *Physics in Medicine & Biology*, 61(19):7012, 2016.
- [57] Shouhei Hanaoka, Yoshitaka Masutani, Mitsutaka Nemoto, Yukihiro Nomura, Soichiro Miki, Takeharu Yoshikawa, Naoto Hayashi, Kuni Ohtomo, and Akinobu Shimizu. Landmark-guided diffeomorphic demons algorithm and its application to automatic segmentation of the whole spine and pelvis in ct images. *International journal of computer assisted radiology and surgery*, 12(3):413–430, 2017.
- [58] Arno Klein, Jesper Andersson, Babak A Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E Christensen, D Louis Collins, James Gee, Pierre Hellier, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802, 2009.

- [59] Xiang Chen, Andres Diaz-Pinto, Nishant Ravikumar, and Alejandro F Frangi. Deep learning in medical image registration. *Progress in Biomedical Engineering*, 3(1): 012003, 2021.
- [60] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1):1–18, 2020.
- [61] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 2021.
- [62] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.
- [63] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [64] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [65] Marc-Michel Rohé, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: learning deformable image registration using shape matching. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 266–274. Springer, 2017.
- [66] Koen AJ Eppenhof and Josien PW Pluim. Pulmonary ct registration through supervised learning with convolutional neural networks. *IEEE transactions on medical imaging*, 38(5):1097–1105, 2018.
- [67] Koen AJ Eppenhof, Maxime W Lafarge, Mitko Veta, and Josien PW Pluim. Progressively trained convolutional neural networks for deformable image registration. *IEEE transactions on medical imaging*, 39(5):1594–1604, 2019.

- [68] Jun Lv, Ming Yang, Jue Zhang, and Xiaoying Wang. Respiratory motion correction for free-breathing 3d abdominal mri using cnn-based image registration: a feasibility study. *The British journal of radiology*, 91(xxxx):20170788, 2018.
- [69] Christodoulidis Stergios, Sahasrabudhe Mihir, Vakalopoulou Maria, Chassagnon Guillaume, Revel Marie-Pierre, Mougialakou Stavroula, and Paragios Nikos. Linear and deformable image registration with 3d convolutional neural networks. In *Image Analysis for Moving Organ, Breast, and Thoracic Images: Third International Workshop, RAMBO 2018, Fourth International Workshop, BIA 2018, and First International Workshop, TIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 3*, pages 13–22. Springer, 2018.
- [70] Tobias Fechter and Dimos Baltas. One-shot learning for deformable medical image registration and periodic motion tracking. *IEEE transactions on medical imaging*, 39(7):2506–2517, 2020.
- [71] Jingfan Fan, Xiaohuan Cao, Zhong Xue, Pew-Thian Yap, and Dinggang Shen. Adversarial similarity network for evaluating image alignment in deep learning based registration. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pages 739–746. Springer, 2018.
- [72] Jingfan Fan, Xiaohuan Cao, Qian Wang, Pew-Thian Yap, and Dinggang Shen. Adversarial learning for mono-or multi-modal registration. *Medical image analysis*, 58: 101545, 2019.
- [73] Xiaohuan Cao, Jianhua Yang, Jun Zhang, Dong Nie, Minjeong Kim, Qian Wang, and Dinggang Shen. Deformable image registration based on similarity-steered cnn regression. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 300–308. Springer, 2017.

- [74] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.
- [75] Maria Chiara Carminati, Concetta Piazzese, Lynn Weinert, Wendy Tsang, Gloria Tamborini, Mauro Pepi, Roberto Miguel Lang, and Enrico Gianluca Caiani. Reconstruction of the descending thoracic aorta by multiview compounding of 3-d transesophageal echocardiographic aortic data sets for improved examination and quantification of atheroma burden. *Ultrasound in medicine & biology*, 41(5):1263–1276, 2015.
- [76] Harriët W Mulder, Marijn van Stralen, Ben Ren, Alexander Haak, Max A Viergever, Johan G Bosch, and Josien PW Pluim. Atlas-based mosaicing of left atrial 3-d transesophageal echocardiography images. *Ultrasound in medicine & biology*, 43(4):765–774, 2017.
- [77] Erik G Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):236–250, 2005.
- [78] Christian Wachinger and Nassir Navab. Structural image representation for image registration. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 23–30. IEEE, 2010.
- [79] Timothy D Barfoot. *State estimation for robotics*. Cambridge University Press, 2017.
- [80] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
- [81] Kashif Rajpoot, J Alison Noble, Vicente Grau, Cezary Szmigielski, and Harald Becher. Multiview rt3d echocardiography image fusion. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 134–143. Springer, 2009.



- [82] Oleg V Michailovich and Allen Tannenbaum. Despeckling of medical ultrasound images. *ieee transactions on ultrasonics, ferroelectrics, and frequency control*, 53(1): 64–78, 2006.
- [83] José Seabra and Joao Sanches. Modeling log-compressed ultrasound images for radio frequency signal recovery. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008.
- [84] Christos P Loizou, Constantinos S Pattichis, Christodoulos I Christodoulou, Robert SH Istepanian, Marios Pantziaris, and Andrew Nicolaides. Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 52(10):1653–1669, 2005.
- [85] Xuli Zong, Andrew F Laine, and Edward A Geiser. Speckle reduction and contrast enhancement of echocardiograms via multiscale nonlinear processing. *IEEE transactions on medical imaging*, 17(4):532–540, 1998.
- [86] Ju Zhang and Yun Cheng. *Despeckle Filters for Medical Ultrasound Images*, pages 19–45. Springer Singapore, Singapore, 2020. ISBN 978-981-15-0516-4.
- [87] Huong T Le, Nicholas Hangiandreou, Robert Timmerman, Mark J Rice, W Brit Smith, Lori Deitte, and Gregory M Janelle. Imaging artifacts in echocardiography. *Anesthesia & Analgesia*, 122(3):633–646, 2016.
- [88] Pier Luigi Pellegrino, Gaetano Fassini, Matteo Di Biase, and Claudio Tondo. Left atrial appendage closure guided by 3d printed cardiac reconstruction: Emerging directions and future trends. *Journal of Cardiovascular Electrophysiology*, 27(6): 768–771, 2016.
- [89] Peng Liu, Rijjing Liu, Yan Zhang, Yingfeng Liu, Xiaoming Tang, and Yanzhen Cheng. The value of 3d printing models of left atrial appendage using real-time 3d trans-esophageal echocardiographic data in left atrial appendage occlusion: applications toward an era of truly personalized medicine. *Cardiology*, 135(4):255–261, 2016.

- [90] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.
- [91] Yiting Fan, Fan Yang, Gary Shing-Him Cheung, Anna Kin-Yin Chan, Dee Dee Wang, Yat-Yin Lam, Marco Chun-Kit Chow, Martin Chun-Wing Leong, Kevin Ka-Ho Kam, Kent Chak-Yu So, et al. Device sizing guided by echocardiography-based three-dimensional printing is associated with superior outcome after percutaneous left atrial appendage occlusion. *Journal of the American Society of Echocardiography*, 32(6):708–719, 2019.
- [92] Yiting Fan, Randolph HL Wong, and Alex Pui-Wai Lee. Three-dimensional printing in structural heart disease and intervention. *Annals of Translational Medicine*, 7(20), 2019.
- [93] David J Hawkes, D Barratt, Jane M Blackall, C Chan, Philip J Edwards, K Rhode, Graeme P Penney, Jamie McClelland, and Derek LG Hill. Tissue deformation and shape models in image-guided interventions: a discussion paper. *Medical Image Analysis*, 9(2):163–175, 2005.
- [94] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, volume 4, pages 109–116, 2007.
- [95] Kevin M Lynch and Frank C Park. *Modern robotics*. Cambridge University Press, 2017.
- [96] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.
- [97] Matthew Michael McCormick, Xiaoxiao Liu, Luis Ibanez, Julien Jomier, and Charles Marion. Itk: enabling reproducible research and open science. *Frontiers in neuroinformatics*, 8:13, 2014.

- [98] Keelin Murphy, Bram van Ginneken, Stefan Klein, Marius Staring, Bartjan J de Hoop, Max A Viergever, and Josien PW Pluim. Semi-automatic construction of reference standards for evaluation of image registration. *Medical image analysis*, 15(1):71–84, 2011.
- [99] Andre Santos Ribeiro, David J Nutt, and John McGonigle. Which metrics should be used in non-linear registration evaluation? In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 388–395. Springer, 2015.
- [100] Zhoubing Xu, Christopher P Lee, Mattias P Heinrich, Marc Modat, Daniel Rueckert, Sebastien Ourselin, Richard G Abramson, and Bennett A Landman. Evaluation of six registration methods for the human abdomen on clinically acquired ct. *IEEE Transactions on Biomedical Engineering*, 63(8):1563–1572, 2016.
- [101] Hedyeh Rafii-Tari, Victoria A Lessoway, Allaudin A Kamani, Purang Abolmaesumi, and Robert Rohling. Panorama ultrasound for navigation and guidance of epidural anesthesia. *Ultrasound in medicine & biology*, 41(8):2220–2231, 2015.
- [102] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pages 129–136, 2008.
- [103] Tobias Ortmaier, Martin Groger, Dieter H Boehm, Volkmar Falk, and Gerd Hirzinger. Motion estimation in beating heart surgery. *IEEE Transactions on Biomedical Engineering*, 52(10):1729–1740, 2005.
- [104] Brian Hall. *Lie groups, Lie algebras, and representations: an elementary introduction*, volume 222. Springer, 2015.
- [105] Fuzhen Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.
- [106] Zhehua Mao, Liang Zhao, Shoudong Huang, Yiting Fan, and Alex Pui-Wai Lee. Direct 3d ultrasound fusion for transesophageal echocardiography. *Computers in Biology and Medicine*, 134:104502, 2021.

- 
- [107] Christian Wachinger, Wolfgang Wein, and Nassir Navab. Three-dimensional ultrasound mosaicing. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 327–335. Springer, 2007.
  - [108] Ju Zhang and Yun Cheng. *Despeckling Methods for Medical Ultrasound Images*. Springer, 2020.
  - [109] Changchang Wu, Sameer Agarwal, Brian Curless, and Steven M Seitz. Multicore bundle adjustment. In *CVPR 2011*, pages 3057–3064. IEEE, 2011.
  - [110] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. g2o: A general framework for graph optimization. In *2011 IEEE International Conference on Robotics and Automation*, pages 3607–3613. IEEE, 2011.
  - [111] Liang Zhao, Shoudong Huang, Yanbiao Sun, Lei Yan, and Gamini Dissanayake. Parallaxba: bundle adjustment using parallax angle feature parametrization. *The International Journal of Robotics Research*, 34(4-5):493–516, 2015.