UNIVERSITÉ MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ MONTPELLIER II

Discipline : Génie informatique, automatique et traitement du signal

École Doctorale : Information Structures Systèmes

présentée et soutenue publiquement par

Rogério de Almeida Richa

le 23 Avril 2010

Suivi 3D Robuste pour la Chirurgie Cardiaque Robotisée

(Robust 3D Motion Tracking for Robotic-Assisted Beating Heart Surgery)

Rapporteurs:	François Chaumette	Directeur de Recherche, IRISA/INRIA, Rennes
	Jacques Gangloff	Professeur, Université Louis Pasteur, Strasbourg
Examinateurs:	Ezio Malis	Chargé de Recherche, INRIA, Sophia Antipolis
	Guang-Zhong Yang	Professor, Department of Computing, Imperial College, London
	Roland Demaria	PU-PH, Chirurgie Thoracique et Cardio-vasculaire, CHU Montpellier
Directeur:	Philippe Poignet	Professeur, LIRMM CNRS, Université Montpellier II

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
161 rue Ada - 34392 Montpellier Cedex 5 - France

· *

aos Seareiros.

Remerciements

Je souhaiterais tout d'abord remercier toute l'équipe de l'administration du Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) pour m'avoir aidé dès mon arrivé en France. Je tiens remercier tout particulièrement mon directeur de thèse, M. Philippe Poignet de m'avoir accordé l'opportunité de réaliser cette thèse et pour la liberté et la confiance qu'il m'a accordées. J'aimerais aussi remercier Nabil Zemiti, Ahmed Chemori, Jean Triboulet, Olivier Strauss, Etienne Dombre et Chao Liu pour leur aide et leurs précieux conseils.

Je remercie sincièrement mes collegues et amis thésards et ex-thésards pour tous les moments de bonheur partagés ces 3 dernières années, je garderai toujours un très agréable souvenir de votre compagnie: Arturo Gil Pinto 'Numero 9', Ashvin Sobhee, Mourad 'Mohhad' Benoussaad, Michel 'Rage Against the Machine' Dominici, Nicolas Riehl, Sébastien Cotton, Milan Djilas, Walid Zarrad, Frederick Van Meer, Mickael Sauvée, Alfredo 'Compadre' Palacios, 'L'autre fille' Chikh, Jovana 'Jovic-Jovan' Jovic, Alonso 'Bandeiras' Sanchez Secades, Hai 'Hi!' Yang, Peter Muel, Andreea 'Calamity' Ancuta et les brésiliens Pedro Moreira et Guilherme Sartori.

*Finalement, je tiens à remercier ma famille à Montpellier, Brasilia et Goiânia. Merci Antonio, Bruno, Polyana, Alice, Carlinha (*obrigado, minha amiga, voce estara sempre no meu coraçao!*) et Vincent 'inho' Bonnet from St. Laurent d'Aigouze (qui m'a appris le français et la vie in the 'south of France', merci beaucoup beaucoup mec), mon oncle et tante Aldo et Telma (sans vous je ne serais pas arrivé jusqu'ici) et Célia, Ismael, Mariana, Lucas et Vovo (pelos seus incriveis paos de queijo e tantas preces). Merci Sara pour ton soutien et ta compagnie pendant ces deux dernières années.*

Obrigado à todos, eu dedico esta tese à voces!

Contents

General Introduction **11**

I Introduction **13**
I.1 The traditional CABG procedure . . . 13
I.2 Off-pump CABG . . . 14
I.3 Minimally invasive CABG . . . 15
I.4 Robot assistance for cardiac surgery . . . 18
I.4.1 Robotic assistance for minimally invasive surgery . . . 18
I.4.2 Limitations . . . 19
I.4.3 Towards robotic cardiac surgery with motion compensation . . . 20
I.4.3.1 Modeling of the heart . . . 21
I.4.3.2 Robot control for beating heart surgery . . . 22
I.4.3.3 Visual-based motion estimation . . . 22

II Estimating the heart motion **25**
II.1 Introduction . . . 25
II.2 Challenges . . . 25
II.3 Previous works . . . 26
II.3.1 Feature-based methods . . . 27
II.3.2 Region-based methods . . . 28
II.3.3 Discussion . . . 30
II.4 Tracking the beating heart . . . 30
II.4.1 The Thin-Plate Spline Warping . . . 31
II.4.2 Extension to 3D tracking . . . 33
II.4.3 Efficient Warp Estimation . . . 34
II.4.4 Illumination Compensation . . . 35
II.4.4.1 Specular highlights detection . . . 36
II.4.4.2 Adopted strategy . . . 36
II.4.5 New minimization problem . . . 37
II.5 Experimental results . . . 38
II.5.1 *Ex vivo* experiments . . . 39
II.5.1.1 Comparative analysis . . . 39
II.5.1.2 Discussion . . . 40
II.5.2 *In vivo* experiments . . . 41
II.5.2.1 Experimental data . . . 42
II.5.2.2 Tracking results . . . 43
II.5.2.3 Implementation on a Graphics Processor Unit (GPU) . . . 46

II.5.2.4 Discussion and perspectives 48
II.6 Conclusion 50

III Robust visual tracking **59**
III.1 Introduction 59
III.2 Predicting the heart motion 60
III.2.1 Beating heart motion dynamics 60
III.2.2 Previous works 66
III.2.3 Non-stationary dual Fourier series model 67
III.2.4 The Extended Kalman Filter 68
III.2.5 Experiments 71
III.2.5.1 EKF parameter configuration 71
III.2.5.2 Experiments description 72
III.2.5.3 Discussion on the predictive filter performance 73
III.2.5.4 Comparative Study 80
III.3 Robust Visual Tracking 85
III.3.1 A unified heart motion estimation framework 85
III.3.2 Evaluating tracking quality 86
III.3.2.1 The alignment error 86
III.3.2.2 Shape analysis 88
III.3.3 Experimental results 88
III.3.3.1 System initialization 89
III.3.3.2 Specular reflections 89
III.3.3.3 Long occlusions 89
III.3.3.4 Discussion 90
III.4 Conclusion 90

IV General Conclusion **97**
IV.1 Conclusion 97
IV.2 Perspectives 98

Publications **101**

Appendice **103**

A Error Measures **105**
A.1 Quantifying the alignment error 105
A.2 The normalized cross-correlation (NCC) 105
A.3 A metric based on the Bhattacharyya coefficient 106
A.4 Mutual Information 106
A.5 Experiments 108

B The Efficient Second-order Minimization (ESM) **111**
B.1 The template tracking problem 111
B.2 The second-order approximation 112
B.3 Comparative analysis 113

C Lens distortion correction **115**

D Parametric deformable models **117**
D.1 Depth from disparity 117
D.2 Experimental results 117
D.2.1 Setup 117
D.2.2 Discussion 117

E Estimated Fourier series coefficients **121**
E.1 The estimated Fourier series harmonics 121

F The Jacobian Matrix **125**
F.1 Building the Jacobian matrix 125

Acronyms **127**

Bibliography **129**

List of tables

II.1 The approximation errors for the selected deformable models. 40
II.2 Computational time in the GPU implementation 47

III.1 The values of the diagonal elements of the process covariance matrix **Q**. . 72
III.2 Predictive filter performance on *in vivo* data 74
III.3 The parameter values for each experimental dataset. 81
III.4 The Extended Kalman Filter setup for the comparative study. 82
III.5 The prediction errors for each method. 82

C.1 The calibration data for the example in figure C.1 116

List of figures

I.1 Illustration of two grafts for bypassing an occluded artery 14
I.2 Pictures of mechanical heart stabilizers 15
I.3 The Cardiolock active stabilizer 16
I.4 Surgical instruments for minimally invasive surgery 16
I.5 A surgical team performing a minimally invasive surgery 17
I.6 The HeartLander robot 18
I.7 The hand-held robotic device for Mitral valve repairment surgery. 18
I.8 Endoscope holders for minimally invasive surgery 19
I.9 Prototype lightweight robots for minimally invasive surgery 20
I.10 The Zeus surgical platform 20
I.11 The DaVinci surgical platform 21
I.12 Surgical platform prototypes 22
I.13 The 'heartbeat synchronization' concept proposed by Nakamura et al. . . . 23

II.1 Images of a minimally invasive CABG using the DaVinci platform 26
II.2 The heart motion sampled at 500 Hz and 25 Hz 27
II.3 The feature tracking method proposed in Ortmaier *et al.* 28
II.4 The 3D feature tracking method proposed by Stoyanov *et al.* 29
II.5 Intra-operative depth recovery methods using depth from disparity 30
II.6 Silicon heart reconstructed in 3D using shape-from-shading techniques . . 30
II.7 Non-rigid object tracking methods using Thin-Plate Splines 31
II.8 The concept behind the parameterization for tracking in the stereo scenario 34
II.9 The proposed specular detection method 36
II.10 Control point placement of the TPS surface $\hat{\mathbf{I}}$ 38
II.11 Illustration of the laser profilometer used for the *ex vivo* experiments . . 39
II.12 *Ex vivo* experiments - The TPS approximation of the heart surface shape 40
II.13 *Ex vivo* experiments - The heart surface shape approximation using the B-spline and piecewise bilinear models 41
II.14 A schematic overview of the tracking algorithm 42
II.15 The acquisition setup of the LIRMM1 sequence 43
II.16 An example of a stereo endoscope integrated into the DaVinci surgical platform 44
II.17 *In vivo* experiments - Selecting region of interest and control points . . . 44
II.18 Trade-off between accuracy and the number of iterations 49
II.19 Experiments on the DAVINCI sequence - 3D motion plots 51
II.20 Experiments on the DAVINCI sequence - The heart surface deformation in 3D 52

II.21 Experiments on the DAVINCI sequence – The required number of iterations during tracking . . . 53
II.22 Experiments on the DAVINCI sequence - The evolution of the illumination compensation parameters . . . 54
II.23 Experiments on the LIRMM1 sequence - 3D motion plots . . . 55
II.24 Experiments on the LIRMM1 sequence - The heart surface deformation in 3D . . . 56
II.25 Experiments on the LIRMM1 sequence - The required number of iterations 57
II.26 Experiments on the LIRMM1 sequence - The evolution of the illumination compensation parameters . . . 57

III.1 Recorded motion using artificial markers . . . 62
III.2 The frequency spectrum of the 3D motion recovered from the tracking experiment on the DAVINCI sequence . . . 63
III.3 The frequency spectrum of the 3D motion recovered from the tracking experiment on the LIRMM1 sequence . . . 64
III.4 The frequency spectrum of the 3D motion recovered from the LIRMM2 experimental data . . . 65
III.5 The beating heart motion estimation using a dual Fourier series . . . 68
III.6 Prediction errors using data from the DAVINCI experiment . . . 75
III.7 Prediction errors using data from the LIRMM1 experiment . . . 76
III.8 Prediction errors using data from the LIRMM2 experiment . . . 77
III.9 The predictive filter performance under amplitude changes . . . 78
III.10 The predictive filter performance under arrhythmic heart behavior . . . 79
III.11 The mean and peak prediction error values for each method analyzed in the comparative study for all Cartesian coordinates of the LIRMM1 experimental data. . . . 83
III.12 Prediction error values . . . 84
III.13 The unified tracking framework . . . 86
III.14 Evaluating the tracking quality . . . 87
III.15 Tracking convergence problems . . . 92
III.16 The TPS bending energy for the estimated heart surface shape . . . 93
III.17 Tracking under specular reflections . . . 94
III.18 The evolution in time of the NCC coefficients . . . 95
III.19 Detected occlusions . . . 95
III.20 A simulated 3 second occlusion . . . 96

IV.1 The schematic overview of a surgical robotic platform with a motion compensation system . . . 99

A.1 Examples of similarity measures . . . 106
A.2 Histogram examples . . . 107
A.3 Joint Entropy . . . 108
A.4 The similarity measures for evaluating the image alignment error analyzed in our studies. . . . 109

B.1 Lenna image . . . 113
B.2 Convergence experiments . . . 114

C.1 A lens distortion correction example 116

D.1 Region of interest on the DAVINCI sequence 118
D.2 Comparative Analysis 119
D.3 Number of iterations 119

E.1 The estimated respiratory and cardiac frequencies in the experiments with the LIRMM1 data. Notice that no significant frequency variations can be verified. 122
E.2 The amplitude of the Fourier series coefficients that model the LIRMM1 experimental data. 123

General Introduction

The past decades have witnessed the notable development of minimally invasive surgery (MIS), in which the surgical gesture is performed through small incisions in the patient's body. The benefits of this modality of surgery for patients are numerous, shortening convalescence, reducing trauma and surgery costs. However, difficulties such as decreased mobility, reduced visibility, uncomfortable working posture and the loss of tactile feedback are imposed to the surgeon. In this context, robotic assistance aims to aid surgeons to overcome such difficulties, making the surgical act more intuitive and safer.

Nevertheless, current surgical robotic platforms display certain limitations, notably regarding the loss of tactile sensations and the absence of active physiological motion compensation, which impedes their use in a broader range of surgical scenarios. In recent years, several research teams have been working on the development of active physiological motion compensation systems for increasing the precision of the surgical gesture. The ultimate goal is the creation of a virtual stable operating site, where the surgeon is given the impression to be working on a motionless environment. The potential improvement in the precision and repeatability of the surgeon gestures is significant.

In the domain of cardiac MIS, heartbeat and respiration represent two important sources of disturbances. Even though miniaturized versions of heart stabilizers have been conceived for the MIS scenario, residual motion is still considerable and has to be manually canceled by the surgeon, complicating the realization of delicate surgical tasks. In order to construct a virtually stable operating field for the surgeon, the heart motion must be estimated. In this thesis, the focus is put on computer vision techniques for estimating the 3D motion of the heart relying solely on natural structures on the heart surface. From a practical point of view, vision-based techniques are preferred since only the visual feedback provided by the endoscope is used, therefore avoiding contact with the heart surface and the introduction of additional sensors in the limited workspace.

In this thesis, two main contributions on the subject of motion compensation for robotized cardiac MIS are proposed:

1. A visual tracking method for estimating the 3D deformation of a region of interest on the heart surface. A thin-plate spline model is used for representing the heart surface deformations. The possibility of arbitrarily placing control points offered by the TPS mapping allows us to easily circumvent problems with the estimation of the heart surface deformation due to lack of texture. A novel efficient parameterization for 3D tracking using stereo endoscopic images is proposed. For evaluating its performance, *ex vivo* and *in vivo* experiments have been performed. The method is robust to illumination variations and large tissue deformations.

2. A robust visual tracking method using motion prediction. Two major novelties are introduced: first, a time-varying dual Fourier series for modeling the quasi-periodic

beating heart motion is proposed. For estimating the parameters of the Fourier series, a probabilistic framework based on the Extended Kalman filter (EKF) is used. Second, the visual tracking method is integrated in the heart motion prediction framework, creating an unified framework for estimating the temporal motion and spatial deformation of the heart surface. The performance of the heart motion prediction is assessed through several experiments using *in vivo* data. The resulting improvements in the tracking robustness facing specular reflections and occlusions are also presented.

This thesis is organized as follows: In the following chapter, the context of existing works in the medical robotics domain is presented, highlighting the potential benefits of robotic assistance in cardiac minimally invasive surgery. Clinical scenarios that would significantly benefit from a motion compensation system are studied. In addition, current robotic solutions adopted by the medical community are analyzed and their strong and weak points are discussed.

In the second chapter, the main challenges involved in tracking the heart motion using natural landmarks are exposed. The different approaches proposed in the literature are presented and discussed. The tracking method developed during this thesis is described in detail, together with *ex vivo* and *in vivo* experiments for assessing its performance.

In the third chapter, the current state of the art on heart motion prediction is presented and discussed. For better understanding the heart motion dynamics, a spectral analysis of the heart motion is performed. Next, a motion prediction scheme based on the dual Fourier series model and an EKF estimation framework is introduced. The tracking method proposed in the second chapter is integrated in the prediction scheme, creating a unified heart motion estimation framework. In order to evaluate the performance of the motion prediction scheme and the improvements in the tracking robustness, two sets of experiments on *in vivo* data are used.

Finally, a synthesis of the different contributions of this thesis and the future work perspectives on visual based motion compensation are given in chapter IV.

Chapter I

Introduction

Cardiovascular diseases are the leading cause of mortality in the world, causing around 12 million deaths worldwide per year [WHO, 2009]. In France, 164 900 deaths were registered in 1999, nearly 31% of all registered deaths [MSS, 2009]. The increasing economic impact of this phenomenon has inspired the governments of many countries to take actions to reduce risk factors such as obesity, diabetes and tobacco. Among cardiovascular diseases, coronary heart diseases are the most critical and the improvement of their treatment is a priority.

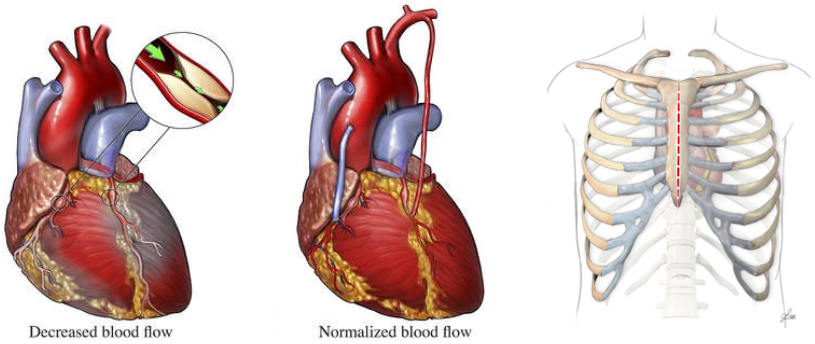
In this chapter a brief overview of the most common surgical intervention on the heart: the coronary artery bypass grafting (CABG) is given. The traditional surgical intervention and recent developments are briefly exposed. Next, the potential benefits of robotic assistance for providing surgeons with superior visualization, enhanced dexterity, precise movements and ergonomic comfort are discussed. Finally, an overview of the current technical challenges in the field of robotized cardiac surgery is presented.

I.1 The traditional CABG procedure

The first surgical interventions on the heart date from the end of the 19^{th} century, and have seen a considerable evolution in the past century, notably after the second world war [Rajan, 2004]. The technical knowledge acquired throughout the years associated with modern surgical instrumentation has reduced the mortality rates of most interventions on the heart to relatively low levels.

In the United States, the most common type of heart surgical intervention in adult patients is the CABG [NHLBI, 2007]. In France, around 80 000 CABG procedures are done each year [MSS, 2009]. The CABG is a mature technique, with a considerable number of related studies and statistics [Eagle and Guyton, 2004]. Basically, the procedure consists in bypassing an occlusion point on the coronary artery for reestablishing irrigation. The term 'anastomosys' stands for the connection between two organs or blood vessels. Grafts may be taken from the saphenous vein or the internal mammary artery (figure I.1) but the former source is preferred due to their higher longevity [Naik et al., 2003].

Currently, the reference CABG procedure involves a cardiopulmonary bypass (CPB), where blood circulation and oxygen are maintained by means of an extracorporeal circulation machine (also known as the heart-lung machine). The CPB is performed due mostly to the difficulty or impossibility of operating on the beating heart. For setting up a CPB, a median sternotomy is usually performed (a 16-20 cm incision in the thorax), allowing a



(a) (b)

Figure I.1 : (a) Illustration of two grafts for bypassing an occluded artery. (source: www.images.com) (b) The dashed line indicates the incision performed for a median sternotomy. (source: www.mcallenheart.com)

direct access to the heart (see illustration in figure I.1). In fact, the sternotomy and CPB are standard procedures for most types of heart interventions: Mitral valve replacements, congenital defect treatments, etc. In 2004, 646,000 open-heart procedures were performed in the United States. Among these, 427,000 bypass surgeries were performed on 249,000 patients [AHA, 2009].

Although a direct access to the heart through a sternotomy privileges the surgeon's dexterity and quick access to the target intervention area, open-chest surgery presents serious downsides to the patient. Significant trauma and the risk of infections due to the long duration of the procedure [Klesius et al., 2004] are the major downsides of this procedure. Whenever possible, less invasive solutions such as a mini-thoracotomy (8-10 cm incision between the ribs) are adopted, but they are limited to procedures on the frontal side of the heart.

Nevertheless, the greatest source of complications and post-operatory mortality for patients is due to the CPB. Problems such as inflammatory blood response to the heart-lung machine, the risk of microemboly, kidney disfunctions and neurological complications such as stroke during the clamping of the aorta have motivated new solutions that circumvent the use of extracorporeal circulation [Eagle and Guyton, 2004].

I.2 Off-pump CABG

Even though open-chest on-pump cardiac surgery is the standard for CABG procedures and high success rates are observed nowadays, the downsides of the extracorporeal circulation are not negligible. Since the CPB is performed mostly due to the difficulty of operating on a moving organ, mechanical stabilizers have been conceived for immobilizing a region of interest on the heart and performing the surgical gesture directly on the beating heart (see figure I.2).

The principle behind the commercially available mechanical stabilizers varies from suction (Medtronics) or local pression (Guidant-CTS). Statistical studies reveal a slighter decrease in the complications rate with respect to the on-pump scenario, suggesting fewer neurological complications [Patel et al., 2002], smaller risk of myocardial infarc-



(a) (b) (c) (d)

Figure I.2 : (a) The Octopus heart stabilizer based on suction (source: www.medtronic.com). (b-c) Solutions proposed by Guidant based on suction and pression respectively (source: www.bostonscientific.com). (d) The Cardiolock (source: eavr.u-strasbg.fr).

tus [Berdat et al., 2004], less convalescence time and less costs [Boyd et al., 1999]. For instance, off-pump CABG procedures have a death and/or complication risk of about 1% to 2% in a low-risk patient [Shekar, 2006].

However, studies on the performance of such stabilizers reveal a considerable residual heart motion due to insufficient immobilization, which ranges from 1-1.5 mm [Lemma et al., 2005] approximately. This is due specially to the greater flexibility of the narrow tools and the greater distance between mounting point and worksite [Riviere et al., 2006]. Another fact that must be taken into consideration is the physical impact of mechanical stabilizers, which may inflict severe tissue damage. Finally, interventions on the posterior part of the heart remain problematic, since the heart in these cases is considerably constrained while stabilized, reducing blood pressure [Dzwonczyk et al., 2005].

Recently, an active heart stabilizer which mechanically compensates for the residual heart motion has been proposed in [Bachta et al., 2007]. The Cardiolock stabilizer is based on the principle of vibration cancellation, where a piezo actuator compensates the efforts induced by the heart motion on the stabilizer shaft. The current prototype of the active stabilizer is illustrated in figure I.3. Initial studies reveal a considerable reduction in the residual heart motion (keeping a residual motion amplitude under 0.3 mm in the direction perpendicular to the heart).

I.3 Minimally invasive CABG

While the extracorporeal circulation can be avoided using mechanical stabilization, the access to the heart through a sternotomy or mini-thoracotomy remains an issue. In order to reduce the trauma inflicted on the patient, minimally invasive techniques are desirable, where the surgical gestures are performed through small incisions (1-2 cm) in the patient's body. The visual feedback is provided by an endoscopic camera and the surgeon operates specially designed surgical tools (see figure I.4). Inside each small incision, a trocart is placed, through which the surgical instruments are inserted. The trocarts seal the patient's interior, which is inflated with carbonic gas for increasing the operative workspace. Figure I.4 illustrates the surgical setup. The reduced invasiveness

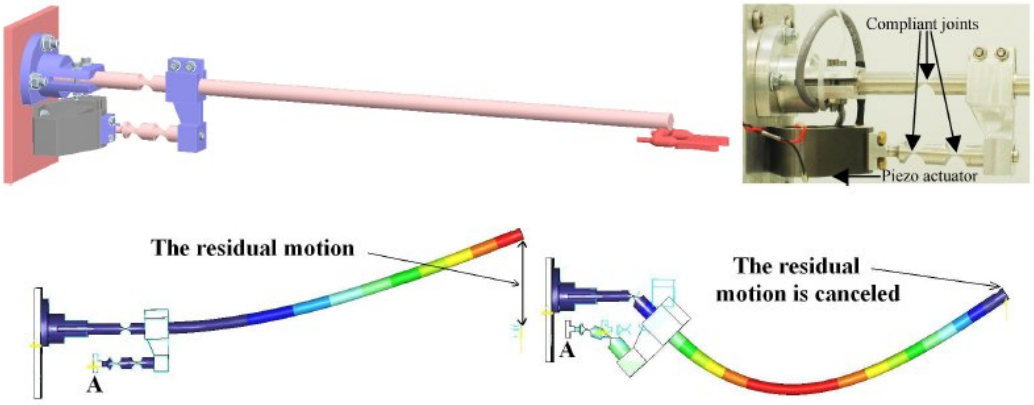
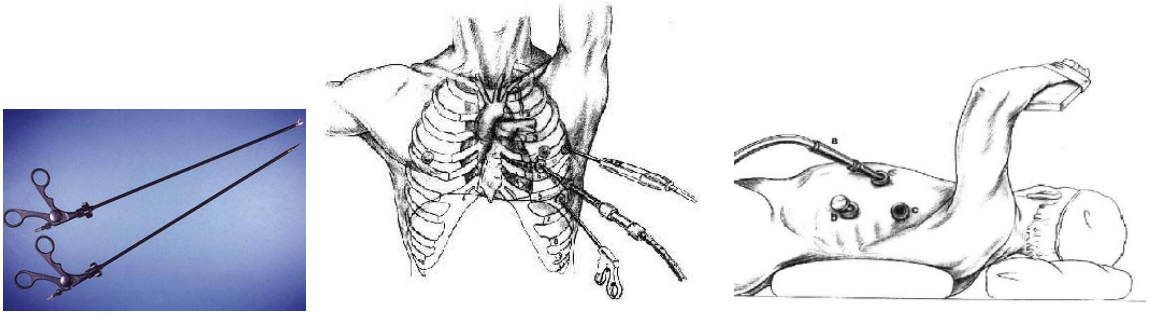


Figure I.3 : (Top) The CAD model of the stabilizer and a detail of the actuation mechanism. (Bottom) A finite element analysis of the compensation, with magnified displacements. (source: [Bachta et al., 2007].



(a) (b)

Figure I.4 : (a) Surgical tools for minimally invasive surgery (source: unknown). (b) Illustration of a minimally invasive intervention (source: unknown).

translates into smaller convalescence time and trauma for the patient and lower infection risks [Puskas et al., 1998]. Nevertheless, minimally invasive surgery still faces several technical limitations.

Although in one hand the minimally invasive procedure presents considerable advantages for the patient, the surgical intervention becomes considerably more difficult to surgeons, resulting in longer operation duration and more costs. The most important difficulties can be summarized as follows:

- *Mobility*: the passage of the surgical instruments by the trocart (the entry point on the surgeon's body) induces friction and imposes restrictions on the mobility inside the patient's body.

- *Fulcrum effect*: the trocart induces an inversion on the surgeon's movements, yielding a major detrimental influence on the performance of the surgical gesture [Gallagher et al., 1998]. Time and experience are required for surgeons to get automated to such effect.



Figure I.5 : A surgical team performing a minimally invasive surgery (source: www.gynlaparoscopy.com).

- *Vision*: interventions are normally performed using a single monocular endoscope and the depth perception loss due to the 2D visual feedback is a great drawback for minimally invasive surgery.

- *Haptic Feedback*: the friction between the surgical instruments and the internal organs reduces the tactile sensibility of the surgeon, making palpation of internal tissues impossible and increasing the risks of tissue damage.

- *Physiological Motion*: the respiratory and cardiac motions severely disturb delicate surgical tasks and the manual compensation demanded to the surgeon is tiring and risky. Even when mechanical stabilizers are employed, the residual motion due to insufficient immobilization is not negligible [Lemma et al., 2005].

- *Ergonomics*: poor hand-eye coordination, the complex manipulation of the endoscope and awkward posture during surgery make the surgical procedure tiring for surgeons, reducing their dexterity and prolonging the surgery duration (see figure I.5).

Due to the above mentioned limitations, minimally invasive procedures on the heart are still rare, despite the notable technical progress achieved recently. The manual challenge involved in this type of intervention associated with the time required to convert the procedure into a traditional one in case of an emergency are factors that reduce its acceptance by the medical community.

In the context of robotized cardiac surgery, robotic assistance can be conceived to overcome most of the limitations cited above, keeping the several benefits of minimally invasive surgery for patients while making the surgical act as intuitive as open surgery for the surgeons.

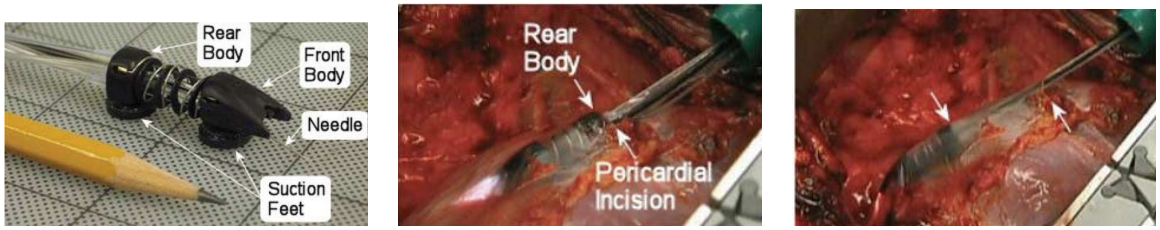


Figure I.6 : The HeartLander robot (source: [Riviere et al., 2006]).

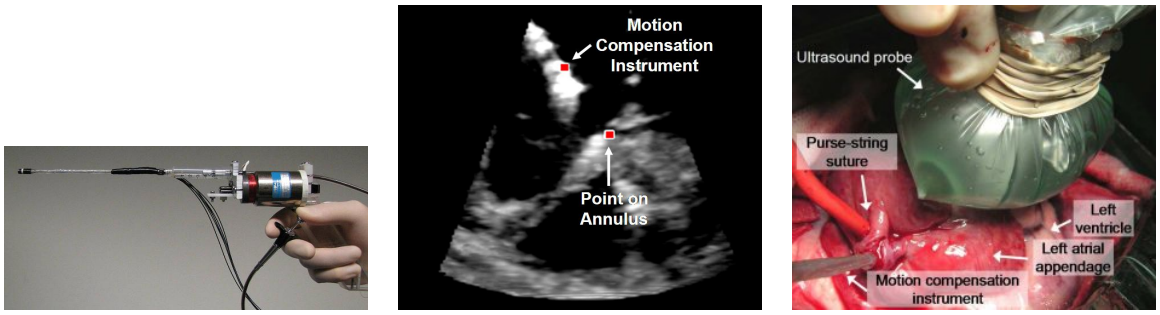


Figure I.7 : The hand-held robotic device for Mitral valve repairment surgery developed at the Harvard Biorobotics lab (source: [Yuen et al., 2009, HarvardBiorobotics, 2009]).

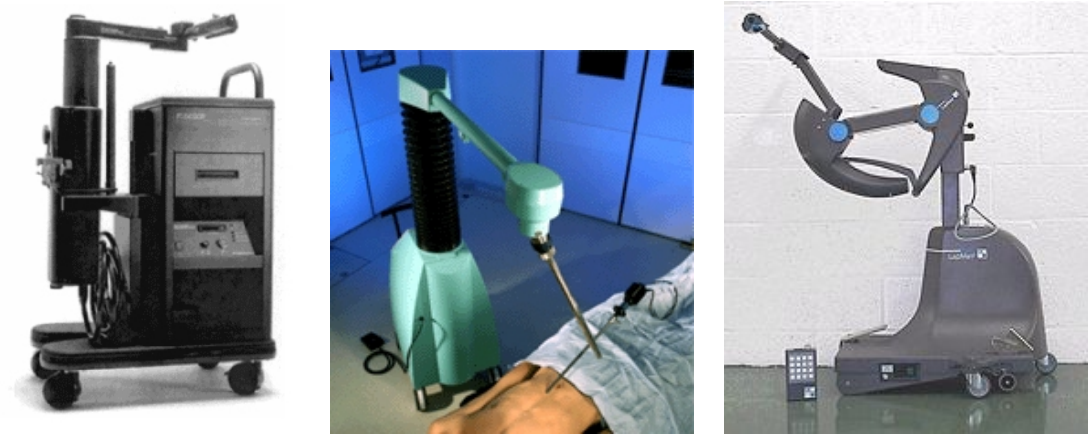
I.4 Robot assistance for cardiac surgery

Recently, prototype robotic tools have been developed for assisting surgeons on specific interventions on the heart with reduced invasiveness. That is the case of the HeartLander device, developed by researchers from Carnegie Mellon's robotic institute [Patronik, 2008]. It consists of a miniature mobile robot designed to be introduced into the pericardial sac and adhere to the epicardial surface by suction, moving to any desired location in an inchworm-like fashion (see illustration in figure I.6). The device can be used in procedures such as atrial ablation, epicardial electrode placement and myocardial injection of drugs.

Another example is the hand-held robotic instrument developed at the Harvard Biorobotics lab [HarvardBiorobotics, 2009] for Mitral valve repairment interventions. The device tracks the motion of the Mitral valve leaflets using ultrasound imaging and provides a one degree of freedom motion compensation for improved hand-eye coordination (see illustration in figure I.7). Active guidance based on ultrasound imaging was also the topic of a french project named GABIE [Dombre, 2005], which addressed the control aspects of guiding the surgical instruments during Mitral valve repair surgery [Sauvée et al., 2007].

I.4.1 Robotic assistance for minimally invasive surgery

Robotic assistance for minimally invasive surgery first appeared in the form of endoscope guidance systems, which are robotic arms that act as a "third arm", holding and positioning the endoscope according to the surgeons' commands. Three commercial systems were proposed (see figure I.8), differing on the way they are controlled: the Aesop (Computer Motion) which is controlled by voice, the EndoAssist (Armstrong Healthcare) which is controlled by the surgeons head movements and the LapMan (MedSys), which is con-



(a) (b) (c)

Figure I.8 : (a) Aesop (source: www.computermotion.com). (b) Endoassist (source: www.armstrong-healthcare.com). (c) LapMan (source: www.medsys.be).

trolled manually. In addition, lightweight solutions for endoscope holders which can be directly positioned over the patient have been conceived (figure I.9). An example is the LER (Light Endoscope Robot) system presented in Berkelman et al. [Berkelman et al., 2003].

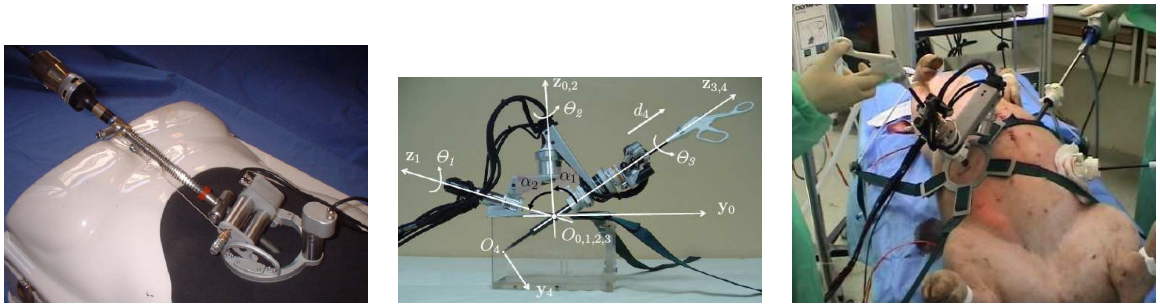
In the last decade, teleoperated robotic platforms have been proposed for performing minimally invasive surgery. Teleoperated surgical platforms consist of a master console that transmits the surgeon's gestures to the slave platform, which is the robot that holds the surgical tools and endoscope. Such platforms provide surgeons with superior visualization, enhanced dexterity, precise movements and ergonomic confort, making it possible for complex surgical procedures to be performed in a minimally invasive fashion. Two commercial platforms have been developed: the Zeus (Computer Motion) and DaVinci (Intuitive Surgical). The Zeus was the first surgical platform to be commercialized. The company that assembled the Zeus, Computer Motion, was later incorporated to Intuitive Surgical, producer of the DaVinci platform, which is nowadays the only commercially available surgical platform.

Alternative robotic designs for surgical platforms have also been developed in several universities and research centers (figure I.12). Examples are the RTW (Robotic Telesurgery Workstation) [Cavusoglu et al., 2001] developed by an association between Berkeley and the University of California at San Francisco, the ARTEMIS system (Advanced Robot and Telemanipulator system for Minimal Invasive Surgery) [Rininsland, 1999] from Karlsruhe, the DLR surgical platform [Hagn et al., 2008] in Munich, the Black Falcon [Madhani et al., 1998] developed at the MIT in Cambridge and the lightweight MC^{2}E robot designed by the ISIR[1] [Zemiti, 2005].

I.4.2 Limitations

Although current surgical platforms have considerably improved the ergonomics and mobility issues related to the minimally invasive procedure, the Zeus and DaVinci platforms are criticized by the medical community for their cost, volume and weight. On

[1]Institut des Systèmes Intelligents et de Robotique



(a) (b) (c)

Figure I.9 : (a) The LER (source: [Berkelman et al., 2003]). (b) The MC^{2}E (source: [Zemiti, 2005]). (c) The MC^{2}E robot mounted on a pig's abdomen illustrating its operational setup (source: [Zemiti, 2005]).



(a) (b) (c)

Figure I.10 : The Zeus surgical platform (a) The console (b) Robotic arms (c) The Zeus operation setup (source: www.computermotion.com)

the technical side, no solution for restoring tactile feedback to the surgeon exists and physiological motion compensation still needs to be manually canceled by the surgeon.

In minimally invasive CABG, miniature versions of mechanical stabilizers are available for performing off-pump interventions but the residual motion due to insufficient immobilization is still important. Statistics presented in [Kappert et al., 2001] show that 4% of surgeries using the DaVinci robot and mechanical stabilization are converted in a median sternotomy and 40% are converted in a mini-thoracotomy. This is due specially to the difficulty of performing delicate tasks such as suturing in the presence of physiological motion.

I.4.3 Towards robotic cardiac surgery with motion compensation

Technical challenges in minimally invasive surgery such as the restitution of haptic feedback and automatic physiological motion compensation have motivated the work of several research groups around the world. The primary goal of providing haptic feedback to surgeons is to lower the risks of tissue damage by robotic manipulation, while active motion compensation facilitates operations on moving organs.



(a)

Figure I.11 : The DaVinci surgical platform (a) The console (b) Robotic arms (source: www.intuitivesurgical.com)

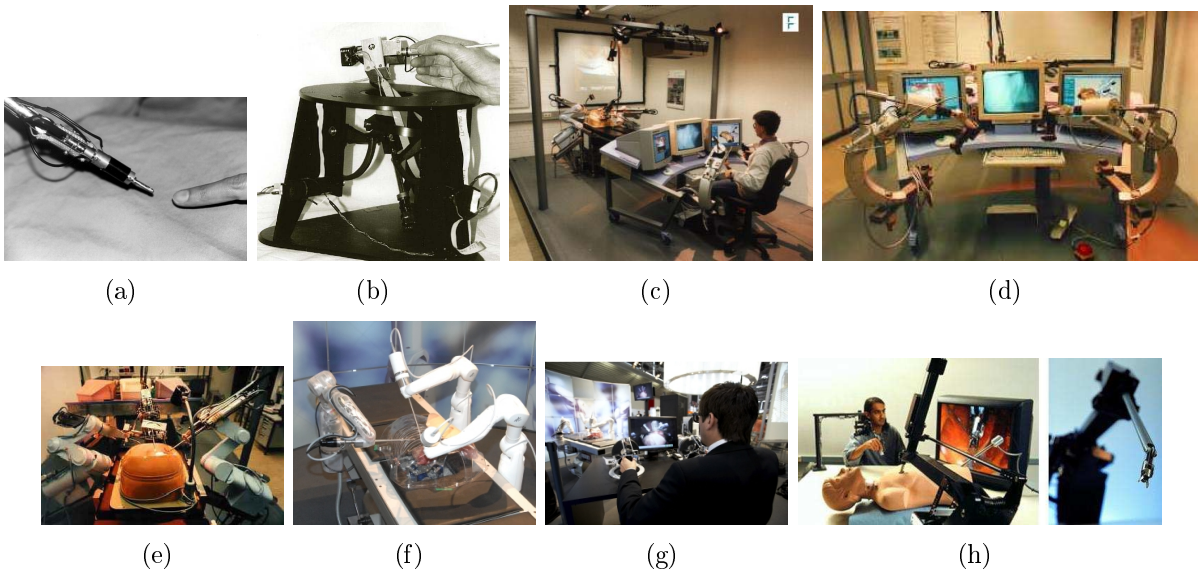
The challenge behind haptic feedback is mainly hardware-related, concerning aspects in the robot control architecture but mainly in the force sensor design, size, accuracy, biocompatibility and sterility [Zarrad, 2006]. Recent advances in the force sensing technology such as optical based force sensing has potential for attending the requirements of surgery [Peirs et al., 2004]. On the other hand, physiological motion compensation imposes challenges in several levels of the robot's design such as bandwidth and control architecture but most importantly it requires an accurate physiological motion estimation.

In the context of robotic assisted beating heart surgery, several works have been performed on the design and implementation of schemes for active motion cancellation. In the pioneer works of Nakamura [Nakamura et al., 2001], the concept of 'heartbeat synchronization' was proposed, which consists in canceling the relative motion between the surgical instruments and the beating heart (figure I.13). The goal is to give the surgeon the impression to operate on the heart as if it were stationary, maintaining the natural physiological motion unlike active motion canceling devices such as the previously mentioned Cardiolock system.

I.4.3.1 Modeling of the heart

Several research teams currently work on the study of the heart physiological phenomena. An example is the Cardiosense3D project [CardioSense3D, 2009] from INRIA, France. Their research effort is focused on proposing models of the heart's electro-physiology, contraction and relaxation mechanics, arterial circulation and perfusion [Delingette et al., 2006]. The models are intended to have a low number of parameters for facilitating its estimation from clinical data and intra-patient adaptation.

The development of models of the heart motion, which entails a in-depth study of the heart physiological properties, has direct applications in the development of reliable real-time motion estimation systems for beating heart surgery. Even though this type of research is still in its infancy, recent approaches illustrate the potential benefits in tasks



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure I.12 : (a-b) The RTW (source: robotics.eecs.berkeley.edu). (c-e) The ARTEMIS (source: [Rininsland, 1999]). (f-g) The DLR platform (source: www.dlr.de). (h) The Black Falcon (source: [Madhani et al., 1998]).

such as motion compensation for robotic assisted surgery [Bogatyrenko et al., 2009].

I.4.3.2 Robot control for beating heart surgery

At the control level, predictive control has been employed for coping with the beating heart dynamics, considering that the quasi-periodic heart motion can be successfully anticipated by motion prediction schemes. Bebek et al. [Bebek and Cavusoglu, 2007a] proposes the fusion of the heart motion measured by sonometric sensors and the electrocardiogram (ECG) signal and previous heart cycles are used as feedfoward signal to predict the future heart motion. Similarly, Ginhoux et al. [Ginhoux et al., 2003] proposes a system based on high-speed cameras for measuring the heart motion using artificial markers and future heart motion prediction based on a truncated Fourier series. However, if in one hand predictive control is a suitable control scheme for the heart motion dynamics, the heart motion prediction methods used in the works above display limited performance. In chapter III, a more detailed discussion on heart motion prediction is given.

In the literature, several motion compensation schemes that use force information have also been proposed. In Dominici et al. [Dominici et al., 2008], a linear predictive control scheme has been developed for maintaining a constant force on the heart surface, compensating the efforts induced by the heart motion. In a similar fashion, Cagneau et al. [Cagneau et al., 2007] uses a ILC (Iterative Learning Control) scheme on the control loop for compensating the respiratory motion.

I.4.3.3 Visual-based motion estimation

For estimating the heart motion - the crucial step for motion compensation - several types of sensors have been employed. A mechanical solution in the form of a Whisker sensor was proposed in [Bebek and Cavusoglu, 2007b], which consists in flexible metal-

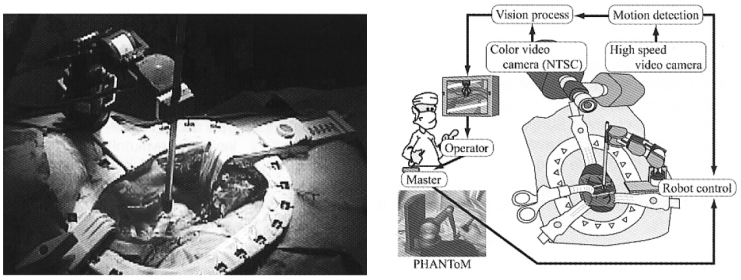


Figure I.13 : The 'heartbeat synchronization' concept proposed by Nakamura et al. (source: [Nakamura et al., 2001]).

lic shaft that measures the heart motion through the contact forces on the heart surface. In [Hoff et al., 2004], accelerometers were used for measuring the heart motion in open-surgery. Active vision techniques have also been proposed, such as the laser range finder prototype in [Hayashibe et al., 2006] conceived for compensating the liver motion and system for automatic positioning surgical robotic instruments proposed by [Krupa et al., 2004]. A technique based on structured light is currently being developed at LSIIT[2] and preliminary results reported in [Albitar et al., 2007] demonstrate the feasibility of using active vision for reconstructing intra-operative sites.

Nevertheless, from a practical point of view, passive vision-based techniques are preferred since only the visual feedback provided by the endoscope is used, avoiding contact with the heart surface and the introduction of additional sensors in the limited workspace. The first prototype of a motion compensation system based on vision was proposed in [Nakamura et al., 2001], where artificial markers were used to retrieve the heart motion. Although recent works still explore artificial markers [Sauvée et al., 2007] in order to simplify the visual tracking task, their use in surgical procedures is not practical. As an alternative, the heart motion estimation through natural landmarks on the heart surface was proposed by Tobias Ortmaier [Ortmaier, 2003]. Following this trend, in this thesis the focus is put on computer vision techniques for precisely estimating the 3D motion of the heart, relying solely on natural structures on the heart surface.

In the next chapter, the main challenges involved in estimating the heart motion using natural landmarks are presented. The different approaches proposed in the literature are also presented and discussed. The tracking method developed in this thesis is described in detail. In addition, *ex vivo* and *in vivo* experiments are provided for evaluating its performance.

[2]Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection - Université de Strasbourg

Chapter II

Estimating the heart motion

II.1 Introduction

The idea of compensating the beating heart motion for increasing the precision of the surgical gestures has motivated several works in the literature. Solutions vary from mechanical systems such as the moving support proposed by Trejos *et al.* [Trejos et al., 1999] to more sophisticated augmented reality systems [Stoyanov and Yang, 2007], where a virtually stable view is rendered for the surgeon for improved surgical performance. All systems report a significant increase in the precision and repeatability of the surgical gesture, attesting the practical value of a motion compensation system for robotized surgery.

Although the architecture of a complete motion compensation systems is still an open problem, a fundamental requirement in all systems is the accurate estimation of the beating heart motion. In this study, the aim is to estimate the heart surface motion with high accuracy and robustness based solely on the visual feedback provided by a stereo endoscope and the available natural landmarks on the heart surface.

II.2 Challenges

The challenges involved in the heart motion estimation start with the critical accuracy constraints of heart surgery, since fine cardiac structures are manipulated. Considering a common task such as performing suture points on a vessel 2 mm diameter wide, an accuracy under 20 μm is required.

The difficulties in tracking the beating heart using vision are start with the image acquisition system. Image quality (resolution, lens deformation), synchronization and acquisition speed are essentially hardware problems. Ginhoux *et al.* [Ginhoux et al., 2003] pointed out that the heart motion has very fast transients and with a slow acquisition rate, information loss due to aliasing is not negligible (see figure II.2). They also suggested an acquisition speed not smaller than 100 Hz for compensating the heart motion. This observation has direct implications on the design of a tracking algorithm.

For the investigations, *in vivo* images of a total endoscopic coronary bypass grafting captured by a DaVinci platform are used as a realistic clinical scenario[1]. Extracts of this image sequence are presented in figure II.1. Assuming that the heart motion can be

[1]The author would like to thank Danail Stoyanov and Guang-Zhong Yang for kindly providing this image sequence for our experiments.

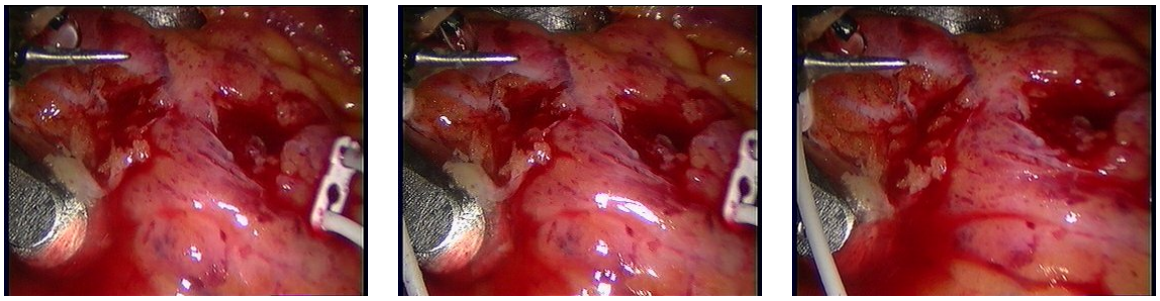


Figure II.1 : Images acquired by the left camera of a stereo endoscope during a total endoscopic coronary bypass grafting procedure using a DaVinci surgical platform.

extracted by tracking the motion of natural structures on the heart surface, the major problems encountered from the computer vision point of view can be grouped into three categories:

1. Illumination issues – The surface properties of soft-tissue gives rise to specular reflections, which are the direct reflection of the light source on the glossy, wet-like surface. These specular reflections work as occluders and considerably reduce the available visual information used by the tracking algorithm. Another important source of disturbance for the tracking algorithm are lighting changes. Due to the restricted workspace in MIS, the light source illuminates unevenly the operating site. Associated with the physiological motion, the brightness constancy assumption on which various tracking methods are based upon is violated, complicating the visual tracking task.

2. Appearance changes – Liquids and smoke which are present at the operating site often disturb visual tracking. It is also expected that as the surgeon manipulates the heart its appearance significantly changes. Due to the unpredictable nature of the changes, their modeling is very difficult.

3. Lack of visual information – Certain regions of the heart surface do not offer a set of identifiable and stable features or texture information, from which the heart motion can be inferred. Since the use of artificial markers is not practical, visual tracking is challenging.

Finally, since currently available image acquisition hardware offers limited acquisition rates, motion blur is an important problem. However, since it is a hardware-related problem, it is not cited above.

II.3 Previous works

This thesis work are inspired by the pioneer studies of Tobias Ortmaier [Ortmaier, 2003] on visual based estimation of the heart motion using natural landmarks on the heart surface. The recent advent of the stereo endoscope enabled stereo vision techniques to be applied for the problem of tracking the beating heart.

Although visual tracking is a classical problem in computer vision, traditional techniques have found limited success when dealing with soft-tissue deformations and complex

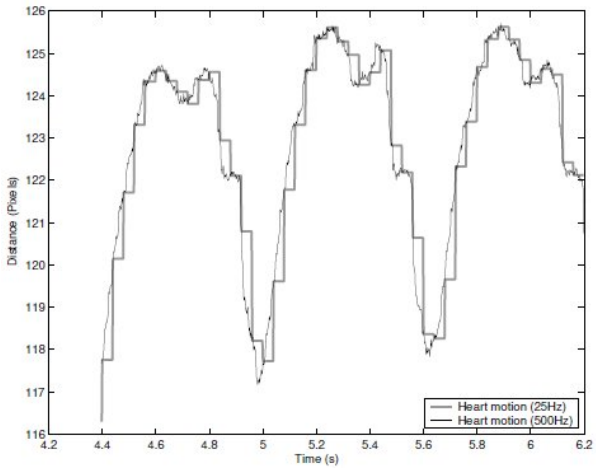


Figure II.2 : The heart motion sampled at 500 Hz and 25 Hz: information loss at 25 Hz due to aliasing is not negligible.

illumination conditions. Therefore, techniques must be tailored specifically for the challenging MIS scenario. In the literature, works can be coarsely divided into two classes: feature-based or region-based tracking.

II.3.1 Feature-based methods

Feature-based methods can be used to extract and track salient features on the heart surface such as blood vessels or other distinguishable structures for retrieving the heart motion. Feature tracking has already been applied in related problems such as intra-operative navigation [Burschka et al., 2005, Mourgues et al., 2001] and augmented reality [Hu et al., 2007, Rohl et al., 2009]. In the context of motion compensation for cardiac surgery, the first visual tracking method for retrieving the heart motion was proposed by Ortmaier [Ortmaier, 2003], which tracked salient features on the heart surface based on an affine motion model and correlation-based search. In his works, the motion estimation is restricted to the image plane but preliminary investigations on 3D tracking using stereo images are reported. Figure II.3 illustrates the 2D tracking method on an *in vivo* experiment.

A 3D tracking method was later proposed by Stoyanov *et al.* [Stoyanov et al., 2005b], which is based on the extraction and tracking of Maximally Stable Extremal Regions (MSER) [Matas et al., 2002] and Shi-Tomasi [Shi and Tomasi, 1994] features using a Lucas-Kanade tracker [Lucas and Kanade, 1981] extended to the stereo scenario. Figure II.4 illustrates the method. Although the proposed method suggests the feasibility of estimating the heart motion by tracking multiple features on the heart surface, problems concerning the lack of robustness of the feature tracking technique are still a major concern. More specifically, the large number of features extracted during the initialization process are quickly lost due to the inability of the method to cope with soft-tissue deformation and appearance changes.

In this context, several works have focused on performance issues regarding tracking features on soft-tissue. Notably, a study done by Mountney *et al.* [Mountney et al., 2007] evaluates existing feature descriptors in computer vision (which are normally tailored for

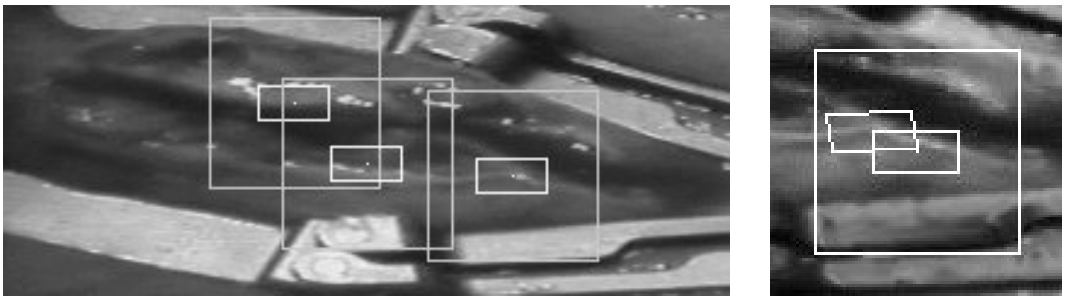


Figure II.3 : (Left) Features tracked on the heart surface using an affine model and correlation based search. (Right) A detail of the tracked feature (small windows) and the search zone (big window) (source: [Ortmaier et al., 2005]).

man-made rigid environments) in the specific context of deformable tissue tracking. More recently, Mountney *et al.* [Mountney and Yang, 2008] also proposed an on-line learning method for boosting the accuracy and persistence of feature tracking in MIS images. For similar purposes, the texture information present on the heart surface was studied in Noce *et al.* [Noce et al., 2007]. In their works, the texture of the heart surface is used to increase feature discriminability in a correlation-based tracking framework. Results suggest that although texture criteria helps to increase the discriminability between a salient feature and its neighborhood, the heart surface deformation is more significant than local minima problems. More recently, Giannarou *et al.* [Giannarou et al., 2009] performed a study on popular feature detectors and proposed an affine-invariant anisotropic feature detector appropriate for soft-tissue tracking. The proposed feature detector is capable of detecting a greater number of feature correspondences in the presence of geometric and photometric transformations. Finally, a method for modeling soft-tissue deformation through feature tracking on monocular images was proposed by Stoyanov *et al.* [Stoyanov and Yang, 2009]. The method infers surface deformation from tracking features on monocular images and uses the geometric modeling proposed by Pilet *et al.* [Pilet et al., 2007].

In conclusion, the feature based methods proposed in the literature demonstrate the possibility of retrieving the heart motion by tracking a set of salient features on the heart surface. Its main advantage is the low computational requirement which enables real-time performance. However, it is strongly dependent on the availability of stable features.

II.3.2 Region-based methods

Unlike feature-based methods where only a set of distinguishable image features is used, region-based methods use the intensity information of the whole image for estimating the heart surface motion. This class of methods, also called direct methods, are based on the assumption that the heart surface is smooth, continuous and sufficiently textured. If these assumptions stand, a parametric function can be used to describe the heart surface deformation and tracking can be formulated as an iterative registration problem.

Region-based methods first appeared as methods for estimating depth in the intra-operative field rather than 3D motion tracking. Lau et al [Lau et al., 2004] proposed a B-spline parametric model to retrieve depth from the disparity between stereo images of a beating heart. The method estimates the model's parameters iteratively using

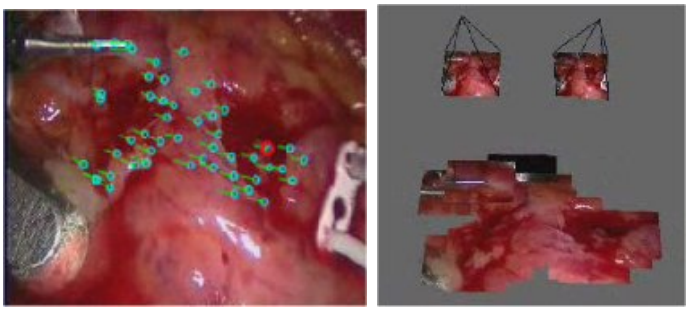


Figure II.4 : (Left) Salient features on the heart surface extracted for tracking. (Right) Resulting heart surface reconstructed in 3D (source: http://ubimon.doc.ic.ac.uk/dvs/m289.html).

Gauss-Newton minimization. Another approach for dense 3D depth recovery was proposed by Stoyanov *et al.* [Stoyanov et al., 2005a], where a piecewise bilinear mapping is applied for modeling soft-tissue deformation. In their implementation, they use the normalized cross-correlation (NCC) as a similarity measure and a quasi-Newton minimization technique. Figure II.5 illustrates reconstruction results using the two techniques mentioned above. Nevertheless, methods based on free-form deformation models [Schnabel et al., 2001] present serious issues concerning specular highlights and textureless regions of the heart surface. In Appendix D, a qualitative analysis of the two techniques cited above on the image sequence introduced in the previous section is presented in detail.

The issues raised above concern the complicated problem of finding a suitable deformation model for the heart surface. A possible solution could be a physical inspired model such as the membrane model proposed by Bader *et al.* [Bader et al., 2007] for describing the heart surface deformation. The temporal and spatial behavior of the membrane used to describe a partition of the heart surface is described by a linear partial differential equation. A Kalman filter is used to estimate the model's parameters and feature tracking is used to provide position measurements to the filter. However, convincing experimental results are still needed to demonstrate the validity of such model.

Finally, results of shape-from-shading (SfS) techniques for reconstructing the heart surface have been reported in the works by Lo *et al.* [Lo et al., 2008b, Lo et al., 2008a]. Compared to the computation stereo approaches presented earlier which perform well in regions with distinctive features and sufficient texture, shape-from-shading tend to perform well in regions with uniform albedo, little texture and smooth local curvature. Figure II.6 illustrates soft-tissue reconstruction results extracted from [Lo et al., 2008a] using SfS in the cardiac MIS scenario. Since SfS and computational stereo techniques are complementary, a probabilistic fusion framework based on Markov Random Fields (MRF) was proposed in [Lo et al., 2008a]. The integration of human perception through gaze-contigence in a similar fusion framework was recently proposed in Visentini-Scarzanella *et al.* [Visentini-Scarzanella et al., 2009]. Nevertheless, the uniform albedo and Lambertian surface requirement for SfS to provide reliable results are not often met in cardiac MIS images. Furthermore, the computational requirements are still an issue when real-time performance is envisaged.

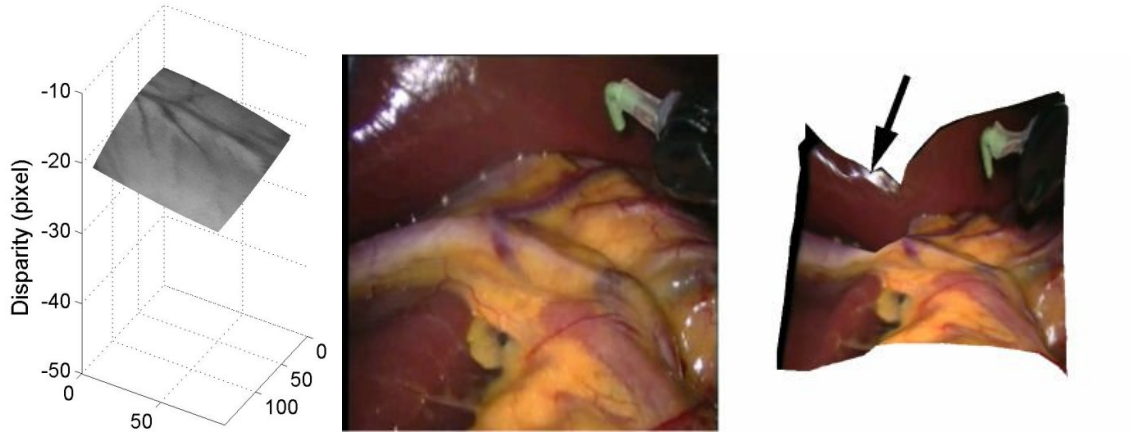


Figure II.5 : (Left) Disparity tracking using a B-spline deformable model (source: [Lau et al., 2004]). (Middle-Right) Reconstructed intra-operative site using a piecewise bilinear model (source: [Stoyanov et al., 2005a]).



Figure II.6 : Silicon heart reconstructed in 3D using shape-from-shading (source: [Lo et al., 2008a]).

II.3.3 Discussion

The works by Lo *et al.* [Lo et al., 2008a] presented above show an interesting route for increasing the robustness of visual motion tracking methods. After an analysis of the performance of each technique separately, it becomes evident that no method proposed in the literature *per se* is able to solve the problem of robustly and accurately retrieving the heart motion. Therefore, for overcoming the challenges of tracking the beating heart, multiple visual cues are needed.

This observation motivates the research for suitable computer vision techniques for 3D tracking and reconstruction in the MIS scenario since it represents an essential step for conceiving a practical motion compensation system for robotized surgical platforms. In conclusion, the analysis presented in this section suggests that region-based techniques cope better with the challenges of tracking soft-tissue, such as large tissue deformation and illumination variations.

II.4 Tracking the beating heart

Due to the complex nature of the heart surface, a suitable choice for a deformation model is complicated. The parametric models currently used to describe the heart sur-

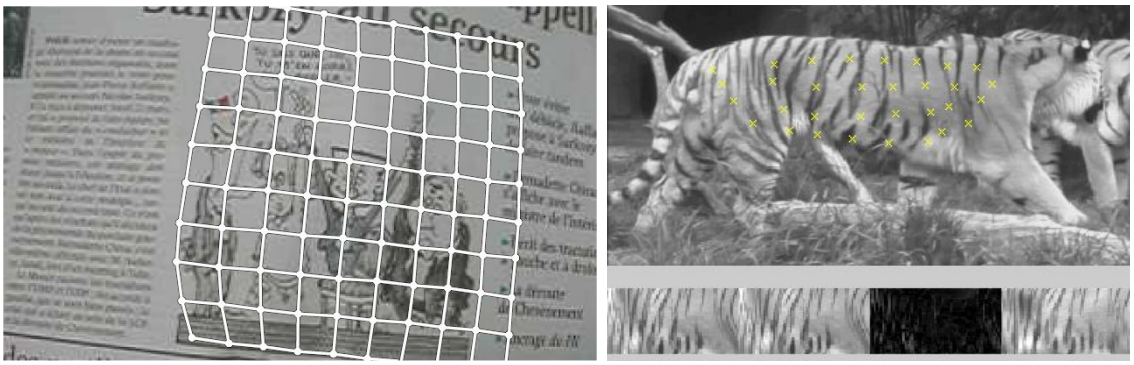


Figure II.7 : (Left) Tracking a newspaper cover (source: [Bartoli and Zisserman, 2004]) (Right) Tracking the deformation of a tiger's stripes (source: [Lim and Yang, 2005]).

face display problems with specular reflections and textureless (homogeneous) regions. For instance, the B-spline mapping in [Lau et al., 2004] does not employ any surface regularization, which may lead to erroneous shape estimates in regions without sufficient texture. Similarly, the bilinear mapping in [Stoyanov et al., 2005a] needs a very large number of control points to sufficiently approximate the heart surface shape.

For such reasons, a type of Radial Basis function called the Thin-Plate Splines (TPS) is chosen for modeling the heart surface deformations. The TPS has been successfully applied for modeling non-rigid tissue deformation, as in Bartoli *et al.* [Bartoli and Zisserman, 2004] and Lim *et al.* [Lim and Yang, 2005] (see illustrations of the techniques in figure II.7). Notably, one of the most interesting features of the TPS mapping is the flexibility in placing its control points. This represents the key feature that allows us to overcome the difficulties of previous methods concerning the lack of visual information when tracking textureless regions on the heart surface.

In recent works, extensions to the original affine formulation were proposed by Malis *et al.* [Malis, 2007] and Bartoli *et al.* [Bartoli et al., 2007], which incorporated the modeling of projective transformations. In this section, the TPS is introduced as a model for the heart surface deformations, following the parameterization proposed by Lim *et al.* [Lim and Yang, 2005]. Based on these previous works, a novel parameterization for tracking in 3D using the TPS in the stereo scenario is proposed. Next, the illumination compensation method proposed by Silveira *et al.* [Silveira and Malis, 2007] is incorporated to the tracking framework for increasing the robustness towards illumination variations and specular reflections. The methods presented in this chapter have been published in [Richa et al., 2010b].

II.4.1 The Thin-Plate Spline Warping

The Thin-Plate Spline is a radial basis function (RBF) that specifies an approximation function $f : \Re^2 \rightarrow \Re$ which minimizes its bending energy:

$$E_f = \int\int_{\Re^2} (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)\, dxdy \tag{II.1}$$

where f_{xx}, f_{yy} are second derivatives of f with respect to the spatial coordinates x and y, respectively. The function f of a point $\mathbf{x} = (x, y)$ in a plane is defined

by the Thin Plate Spline basis function $u(s) = s^2 log(s^2)$, a (α+3) parameter vector $\mathbf{t} = (w_1, ..., w_\alpha, r_1, r_2, r_3)^T$ and a set of α control points $\mathbf{c}_i = (\check{x}_i, \check{y}_i)$, such that:

$$f(\mathbf{x}) = r_1 + r_2 x + r_3 y + \sum_{i=1}^{\alpha} w_i \, u(||\mathbf{c}_i - \mathbf{x}||) \tag{II.2}$$

Also, side conditions must be considered in order for the function $f(\mathbf{x})$ to have square-integrable second derivatives:

$$\sum_{i=1}^{\alpha} w_i = \sum_{i=1}^{\alpha} w_i \, \check{x}_i = \sum_{i=1}^{\alpha} w_i \, \check{y}_i = 0 \tag{II.3}$$

To write a $\Re^2 \rightarrow \Re^2$ mapping $m(\mathbf{x})$ two RBFs f^x and f^y sharing their control points are stacked:

$$m(\mathbf{x}) = \begin{bmatrix} {}^x f \\ {}^y f \end{bmatrix} = \begin{bmatrix} {}^x r_2 & {}^x r_3 & {}^x r_1 \\ {}^y r_2 & {}^y r_3 & {}^y r_1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} + \sum_{i=1}^{\alpha} \begin{bmatrix} {}^x w_i \\ {}^y w_i \end{bmatrix} \, u(||\mathbf{c}_j - \mathbf{x}||) \tag{II.4}$$

The mapping m can be used to describe the positions of the pixel coordinates of a reference image on the current image. Bookstein [Bookstein, 1989] proposed a linear system to calculate the parameter vectors ${}^x\mathbf{t}$ and ${}^y\mathbf{t}$ of each Cartesian coordinate that define ${}^x f$ and ${}^y f$ based on the mapped positions of the control points $\mathbf{c}$ on the original image. This linear system can be written by stacking the mapped values $\mathbf{c}' = (\check{x}', \check{y}') = m(\mathbf{c})$ of each control point $\mathbf{c}$ into a matrix $\check{\mathbf{P}} = [\check{\mathbf{x}}' \, \check{\mathbf{y}}']$:

$$\begin{bmatrix} \mathbf{L} & \mathbf{P} \\ \mathbf{P}^{\mathrm{T}} & \mathbf{0}_{3,3} \end{bmatrix} \begin{bmatrix} {}^x\mathbf{t} & {}^y\mathbf{t} \end{bmatrix} = \begin{bmatrix} \check{\mathbf{P}} \\ \mathbf{0}_{3,2} \end{bmatrix} \tag{II.5}$$

where $L_{ij} = u(||\mathbf{c}_j - \mathbf{c}_i||)$, $\mathbf{P}$ are the stacked coordinates of the control points $(1, \check{x}, \check{y})$ on the original plane, $\mathbf{0}_{i,j}$ is a $i \times j$ zero matrix respectively. The mapped control points $\mathbf{c}'$ are called *control point correspondences*. Denoting the leftmost matrix as $\mathbf{K}$, the system can be inverted to solve for ${}^x\mathbf{t}$ and ${}^y\mathbf{t}$:

$$\begin{pmatrix} {}^x\mathbf{t} & {}^y\mathbf{t} \end{pmatrix} = \mathbf{K}^{-1} \begin{bmatrix} \check{\mathbf{P}} \\ \mathbf{0}_{2,3} \end{bmatrix} \tag{II.6}$$

Finally, since the correspondence between two images is considered, the transformed coordinates $\mathbf{x}' = (x', y')$ of the reference image as a function of control point correspondences $\mathbf{c}'$ can be calculated (see figure II.8(a)). For simplification purposes, the (α+3)$\times\alpha$ sub-matrix of $\mathbf{K}^{-1}$ is denoted as $\mathbf{K}_*$. This eliminates the need for adding zeros in the rightmost matrix in the above equation.

$$\begin{pmatrix} {}^x\mathbf{t} & {}^y\mathbf{t} \end{pmatrix} = \mathbf{K}_* \, \check{\mathbf{P}} \tag{II.7}$$

Given an image $\mathbf{I}$, the $g \times 2$ matrix $\mathbf{P}'$ of stacked transformed pixel coordinates $\mathbf{x}'$ such that $\mathbf{P}' = [\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_g]^{\mathrm{T}}$, where g is the number of pixels in the image $\mathbf{I}$ is computed as:

$$\mathbf{P}' = \begin{bmatrix} \mathbf{V} & \mathbf{W} \end{bmatrix} \mathbf{K}_* \, \check{\mathbf{P}} \tag{II.8}$$

where $V_{ij} = u(||\mathbf{x}_i - \mathbf{c}_j||)$ and $\mathbf{W}_i = [1, x_i, y_i]$, with $i \in \{1, ..., g\}$. The matrix $[\mathbf{V}\ \mathbf{W}]$ is denoted as $\mathbf{M}$. If the position of the control points in the reference plane does not change, the matrices $\mathbf{M}$ and $\mathbf{K}_*$ can be pre-computed.

From the formulation above, $w(\mathbf{x}_i, \check{\mathbf{P}})$ can be defined as the function that warps the i-th pixel $\mathbf{x}_i$ of the reference image based on control point correspondences $[\check{\mathbf{x}}'\ \check{\mathbf{y}}']$ stored in the matrix $\check{\mathbf{P}}$.

$$w(\mathbf{x}_i, \check{\mathbf{P}}) = [x'_i \quad y'_i] = [\ \mathbf{M}_i\, \mathbf{K}_* \check{\mathbf{x}}' \mid \mathbf{M}_i\, \mathbf{K}_* \check{\mathbf{y}}'\ ] \tag{II.9}$$

where $\mathbf{M}_i$ is the i-th row of the matrix $\mathbf{M}$.

II.4.2 Extension to 3D tracking

In the formulation given above, the TPS warping defines a mapping between each pixel position on $\mathbf{T}$ to the current image $\mathbf{I}$ of the surface, as defined in equation (II.8). Nevertheless the standard TPS warping is intrinsically affine and cannot capture projective deformations with a finite number of control points. Addressing this problem, Malis [Malis, 2007] proposed the parameterization of the projective depths of the tracked surface in order to represent projective deformations.

Adapting to the TPS formulation adopted from [Lim and Yang, 2005], the mapping m proposed in equation (II.4) can be extended to model surface depth by adding the necessary degrees of freedom with an additional TPS function ${}^{z}f$:

$$m(\mathbf{x}) = \begin{bmatrix} {}^{x}f \\ {}^{y}f \\ {}^{z}f \end{bmatrix} = \begin{bmatrix} {}^{x}r_2 & {}^{x}r_3 & {}^{x}r_1 \\ {}^{y}r_2 & {}^{y}r_3 & {}^{y}r_1 \\ {}^{z}r_2 & {}^{z}r_3 & {}^{z}r_1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} + \sum_{i=1}^{n} \begin{bmatrix} {}^{x}w_i \\ {}^{y}w_i \\ {}^{z}w_i \end{bmatrix} u(||\mathbf{c}_j - \mathbf{x}||) \tag{II.10}$$

Note that side conditions similar to those from equation (II.3) must be applied. The transformed pixel coordinates $\mathbf{x}' = (sx', sy', s)$ from the original image (now in projective coordinates) can be calculated in the same fashion as equation (II.8):

$$\mathbf{P}' = \mathbf{M}\, \mathbf{K}_* \begin{bmatrix} \check{\mathbf{P}} & \check{\mathbf{z}} \end{bmatrix} \tag{II.11}$$

where the additional column $\check{\mathbf{z}}$ represents the stacked depth positions for each control point, in pixels. For expanding this framework for the stereo case, first one must observe that, by definition, the position of the control points correspondences $\mathbf{c}'$ are the projections of 3D points onto the image plane:

$$\begin{bmatrix} \check{\mathbf{P}} & \check{\mathbf{z}} \end{bmatrix} = (\mathbf{C} \begin{bmatrix} \mathbf{X} \\ \mathbf{1}^{\mathrm{T}} \end{bmatrix})^{\mathrm{T}} = \begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \mathbf{C}^{\mathrm{T}} \tag{II.12}$$

where $\mathbf{C}$ is the 3×4 camera calibration matrix, $\mathbf{X}$ the $\alpha \times 3$ matrix of stacked 3D cartesian coordinates of the points that map to $\mathbf{c}'$ on the image and $\mathbf{1}$ a α column vector of ones. Following this interpretation, the equation (II.11) can be rewritten as a function of $\mathbf{X}$.

$$\mathbf{P}' = \mathbf{M}\, \mathbf{K}_* \begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \mathbf{C}^{\mathrm{T}} \tag{II.13}$$

In result, a linear relationship can be established between the reference image and its pixel's 3D coordinates. Assuming the world coordinate frame centered on the camera, the camera matrix $\mathbf{C}$ is composed uniquely by its intrinsic parameters $\mathbf{C}_*$ and the stacked

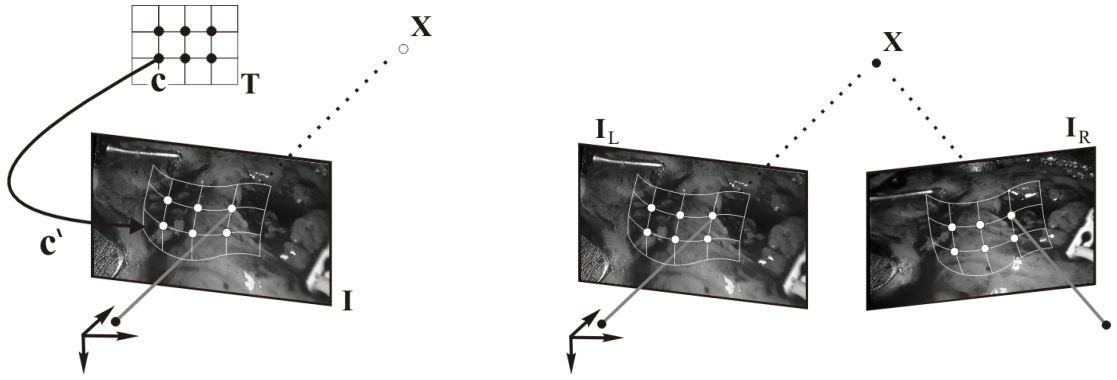


Figure II.8 : (a) The control points mapped on the current image are projections of 3D points in space onto the image. (b) If the alignment of the same reference image on both cameras of the stereo pair is sought, the TPS control points are necessarily the projections of the same 3D points in space onto both images. If a calibrated system is considered, these 3D points can be estimated from both images.

3D coordinates $\mathbf{X}'$ of every pixel from the reference image $\mathbf{T}$ (in the camera coordinate frame) can be found by inverting equation (II.13):

$$\mathbf{X}' = (\mathbf{C}_*^{-1}\mathbf{P}'^{\mathrm{T}})^{\mathrm{T}} = \mathbf{P}'\mathbf{C}_*^{-\mathrm{T}} \tag{II.14}$$

To fully constrain the estimation of the transformed pixel coordinates on equation (II.11) and consequently their 3D coordinates (equation (II.14)), a calibrated stereo rig as illustrated in figure (II.8)(b) is considered, with the world coordinate frame arbitrarily centered on the left camera.

Assuming that the TPS model can represent the heart surface shape, the tracking problem can be written as the problem of minimizing the alignment error between the reference image $\mathbf{T}$ in both left and right cameras $\mathbf{I}_l$ and $\mathbf{I}_r$. Redefining equation (II.9) for the 3D case, the new warping function $w_{3D}(\mathbf{x}_i, \mathbf{h}, \mathbf{C})$ is defined by the camera calibration matrix $\mathbf{C}$ and a 3α parameter vector $\mathbf{h} = ({}^x\mathbf{h}^{\mathrm{T}}, {}^y\mathbf{h}^{\mathrm{T}}, {}^z\mathbf{h}^{\mathrm{T}})^{\mathrm{T}}$ composed by the stacked columns of the matrix $\mathbf{X} = [{}^x\mathbf{h}, {}^y\mathbf{h}, {}^z\mathbf{h}]$:

$$w_{3D}(\mathbf{x}_i, \mathbf{h}, \mathbf{C}) = [x_i' \;\; y_i'] = [sx_i' \;\; sy_i' \;\; s_i] = \mathbf{M}_i\,\mathbf{K}_*\left[{}^x\mathbf{h} \;\; {}^y\mathbf{h} \;\; {}^z\mathbf{h} \;\; \mathbf{1}\right]\mathbf{C}^{\mathrm{T}} \tag{II.15}$$

where $\mathbf{M}_i$ is the i-th row of the matrix $\mathbf{M}$ which corresponds to $\mathbf{x}_i$.

II.4.3 Efficient Warp Estimation

Unlike the 2D case where the problem consists in estimating the control point correspondences between the reference and current images, in the stereo case the 3D coordinates of the points in space that map to control point positions that minimize the alignment error between the reference image and both stereo images are estimated simultaneously. This concept was illustrated in figure II.8(b). Letting $\mathbf{C}_l$ and $\mathbf{C}_r$ be the camera calibration matrices for the left and right cameras respectively and $\mathbf{I}_l$ and $\mathbf{I}_r$ their respective images, this problem can be formulated as:

$$\min_{\mathbf{h}} \; \epsilon = \sum_{\mathbf{x}\in A}\left[\left[\mathbf{I}_l(w_{3D}(\mathbf{x}, \mathbf{h}, \mathbf{C}_l)) - \mathbf{T}(\mathbf{x})\right]^2 + \left[\mathbf{I}_r(w_{3D}(\mathbf{x}, \mathbf{h}, \mathbf{C}_r)) - \mathbf{T}(\mathbf{x})\right]^2\right] \tag{II.16}$$

where A is the set of the template coordinates, $\mathbf{I}_l(w_{3D}(\mathbf{x},\mathbf{h},\mathbf{C}))$ and $\mathbf{I}_r(w_{3D}(\mathbf{x},\mathbf{h},\mathbf{C}))$ are the current left or right images transformed by the warping function $w_{3D}(\mathbf{x},\mathbf{h},\mathbf{C})$, respectively. Notice that the alignment error of the same reference image in both images from the stereo pair is minimized. This equation is based on the assumption that a given region of interest of the heart surface can be visualized on both stereo cameras (which is respected in the case of a very small baseline) and that the image transformation between stereo cameras can be sufficiently represented by the proposed TPS model (the validity of such assumption will be assessed later in the experiments in section II.5.1.1).

The minimization problem posed above is traditionally solved using Gauss-Newton or Levenberg-Marquardt techniques. In this method, the efficient second-order minimization (ESM) is employed, proposed by Benhimane *et al.* [Benhimane and Malis, 2004]. The ESM displays a significant improvement over the traditional template-based tracking techniques, since it has a faster convergence rate and larger convergence basin (see Appendix B for a detailed description of the ESM algorithm).

Let $\mathbf{h}_c$ be the current warping parameters, $\mathbf{h}_o$ the parameters at startup and $\delta\mathbf{h}$ the increment such that $\mathbf{h}+\delta\mathbf{h}$ minimizes equation (II.16). Since images from both stereo pairs $\mathbf{I}_l$ and $\mathbf{I}_r$ are considered, the increment $\delta\mathbf{h}$ can be calculated by stacking the left and right image Jacobians and error images:

$$\delta\mathbf{h} = -2\begin{bmatrix}\mathbf{J}(\mathbf{T},\mathbf{h}_o,\mathbf{x})+\mathbf{J}(\mathbf{I}_l,\mathbf{h}_c,\mathbf{x})\\ \mathbf{J}(\mathbf{T},\mathbf{h}_o,\mathbf{x})+\mathbf{J}(\mathbf{I}_r,\mathbf{h}_c,\mathbf{x})\end{bmatrix}^{+}\begin{bmatrix}\delta\mathbf{I}_l\\ \delta\mathbf{I}_r\end{bmatrix} \tag{II.17}$$

where $\delta\mathbf{I}_l$ and $\delta\mathbf{I}_r$ represent the stacked error images $(\mathbf{I}_l(w(\mathbf{x},\mathbf{h}_c))-\mathbf{T}(\mathbf{x}))$ and $(\mathbf{I}_r(w(\mathbf{x},\mathbf{h}_c))-\mathbf{T}(\mathbf{x}))$ for all $\mathbf{x}\in A$, respectively. The Jacobian matrix $\mathbf{J}$ is calculated by deriving the warping function (II.15) with respect to each parameter h_j of $\mathbf{h}$, with $j\in\{1,...,3\alpha\}$. The i-line of $\mathbf{J}$, corresponding to the pixel $\mathbf{x}_i$ is given as:

$$\mathbf{J}_i(\mathbf{I},\mathbf{h},\mathbf{x}_i) = \left[\frac{\partial\mathbf{I}(w_{3D}(\mathbf{x}_i,\mathbf{h}))}{\partial h_1} \dots \frac{\partial\mathbf{I}(w_{3D}(\mathbf{x}_i,\mathbf{h}))}{\partial h_{3\alpha}}\right] \tag{II.18}$$

A more detailed description of the elements in the Jacobian matrix above can be found in Appendix F.

Equation (II.17) is iterated to achieve higher precision. The whole procedure could also be performed at multiple scales to aid the convergence. When tracking through an image sequence, the inter-frame surface deformation is assumed to be sufficiently small. Hence, for every new frame, the previous surface parameters $\mathbf{h}_c$ can be used as initial guess in equation (II.17) for the current parameters.

In this framework for 3D tracking, no explicit matching between the stereo camera images is performed. Consequently no intermediate steps such as rectification are needed and tracking accuracy is increased.

II.4.4 Illumination Compensation

Due to the illumination conditions in MIS, lighting variations on the heart surface are not negligible. The lighting on the heart surface is uneven and its glossy nature gives rise to specular reflections that reduce the observable heart surface. For tracking to be robust, this phenomenon must be taken into account. Therefore, an illumination compensation step adapted to the MIS conditions is incorporated to the tracking method proposed in this section.

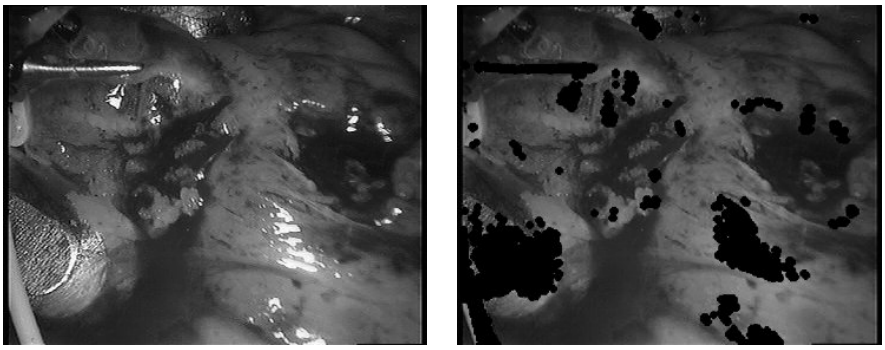


Figure II.9 : The proposed specular detection method. The pixels on the original image (left) affected by specular reflections are painted in black (right). The specular reflections are detected and removed from the computation of the warping parameters.

II.4.4.1 Specular highlights detection

Specular reflections on the heart surface considerably affect tracking performance because they act as occluders, reducing the available visual information. Most works propose to detect these effects based on simple thresholding of the intensity level, which is a practical solution since little computational effort is required.

The specular highlights detection step consists on finding saturated zones on the image by thresholding the intensity level for generating a specularity map. Based on the specularity map, the pixels corresponding to specular highlights on the images are removed from the estimation of the illumination and warping parameters during the optimization process (equation (II.24)). Figure II.9 shows some example results of the specular reflection detection. This basic detection approach is adopted since it provides satisfactory detection results (as verified in the experiments conducted in the next section).

II.4.4.2 Adopted strategy

In the specific context of soft-tissue tracking, feature-based methods use robust error measures for reducing tracking sensibility towards illumination variations [Mountney and Yang, 2008] or normalize the intensity level [Gröger et al., 2002]. However, these solutions have limited performance when applied for region-based methods due to the higher complexity of the illumination variations when tracking bigger regions of interest on the heart surface.

For modeling illumination variations over a large region of interest, the illumination compensation method proposed by Silveira and Malis [Silveira and Malis, 2007] is employed. The great advantage of the chosen method is that it does not require any knowledge of the reflectance properties of the target surface or the characteristics of the illumination source. Furthermore, the method adapts locally to illumination variations over the extent of the tracked region of interest, unlike affine (gain and bias) compensation methods such as [Bartoli, 2006].

The illumination compensation method proposed in [Silveira and Malis, 2007] can be formulated as the problem of finding the elementwise multiplicative lighting variation $\hat{\mathbf{I}}$ for each pixel $\mathbf{x}_i$ of the current image $\mathbf{I}$ and a global bias b such that $\mathbf{I}'$ matches the closest

the illumination conditions of the reference image $\mathbf{T}$:

$$\mathbf{I}'(\mathbf{x}_i) = \hat{\mathbf{I}}(\mathbf{x}_i, \mathbf{y})\mathbf{I}(\mathbf{x}_i) + b \tag{II.19}$$

The element $\hat{\mathbf{I}}$ can be modeled as a Thin-Plate spline surface surface $\hat{\mathbf{I}} = f(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ are pixel coordinates and $\mathbf{y} = [y_1, ..., y_{\beta+3}]$ is the vector containing the $\beta+3$ TPS surface parameters which vary in time:

$$\hat{\mathbf{I}}(\mathbf{x}, \mathbf{y}) \approx y_{\beta+1} + \mathbf{x}[y_{\beta+2}, y_{\beta+3}]^{\mathrm{T}} + \sum_{k=1}^{\beta} y_k \, u(||\mathbf{x} - \mathbf{q}_k||) \tag{II.20}$$

where the control points $\mathbf{q}$ are evenly spaced on the image to best capture the illumination changes. In practice, given the focal distance of the endoscope and the size of the tracked images, 4 control points are normally sufficient for modeling the illumination changes. Figure II.10 illustrates the placement of the illumination control points and TPS surface $\hat{\mathbf{I}}$. The only assumption made when using a TPS model is that the lighting changes are continuous over the reference image. In this application, the illumination transformations of the left and right cameras of the stereo pair are treated independently.

In this study, the illumination parametrization is developed based on the TPS using the same formulation as for the deformation warping presented in section II.4.1. The TPS surface of parameters $\mathbf{u}$ in equation (II.20) can be reparameterized as a function of $\mathbf{q}'$, which are the values of $\hat{\mathbf{I}}$ at each control point $\mathbf{q}$ (see illustration in figure II.10). Therefore, matrices ${}^{il}\mathbf{K}$ and ${}^{il}\mathbf{M}$ can be deduced similarly to equations (II.6) and (II.8), so that the g stacked values of $\hat{\mathbf{I}}$ for every pixel $\mathbf{x}$ on the image can be calculated as:

$$\hat{\mathbf{I}} = {}^{il}\mathbf{M} \, {}^{il}\mathbf{K}_* \mathbf{q}' \tag{II.21}$$

where ${}^{il}\mathbf{K}_*$ is the $\beta + 3 \times \beta$ submatrix of ${}^{il}\mathbf{K}^{-1}$:

$$\mathbf{y} = \underbrace{\begin{bmatrix} {}^{il}\mathbf{L} & \mathbf{q} \\ \mathbf{q}^{\mathrm{T}} & \mathbf{0}_{3,3} \end{bmatrix}^{-1}}_{{}^{il}\mathbf{K}^{-1}} \begin{bmatrix} \mathbf{q}' \\ \mathbf{0}_{3,1} \end{bmatrix} \tag{II.22}$$

with ${}^{il}L_{ij} = u(||\mathbf{q}_j - \mathbf{q}_i||)$ and ${}^{il}\mathbf{M} = [{}^{il}\mathbf{V} \quad {}^{il}\mathbf{W}]$, such that ${}^{il}V_{ij} = u(||\mathbf{x}_i - \mathbf{c}_j||)$ and ${}^{il}\mathbf{W}_i = [1, x_i, y_i]$, with $i \in \{1, ..., g\}$.

For estimating the illumination compensation parameters, the same optimization framework based on the ESM is used.

II.4.5 New minimization problem

Similarly to equation (II.17), a new parameter vector $\mathbf{h}' = [\mathbf{h}, \mathbf{q}'_l, \mathbf{q}'_r, b_l, b_r]$ can be defined, such that $\mathbf{h}'$ incorporates the warping and illumination parameters. Consequently, the previous minimization problem from equation (II.16) is redefined as:

$$\min_{\mathbf{h}, \mathbf{q}'_l, b_l, \mathbf{q}'_r, b_r} \epsilon = \sum_{\mathbf{x} \in A} \Big[\big[\hat{\mathbf{I}}(\mathbf{x}, \mathbf{q}'_l) \; \mathbf{I}_l(w_{3D}(\mathbf{x}, \mathbf{h})) + b_l - \mathbf{T}(\mathbf{x})\big]^2 + \tag{II.23}$$

$$+ \big[\hat{\mathbf{I}}(\mathbf{x}, \mathbf{q}'_r) \; \mathbf{I}_r(w_{3D}(\mathbf{x}, \mathbf{h})) + b_r - \mathbf{T}(\mathbf{x})\big]^2 \Big] \tag{II.24}$$

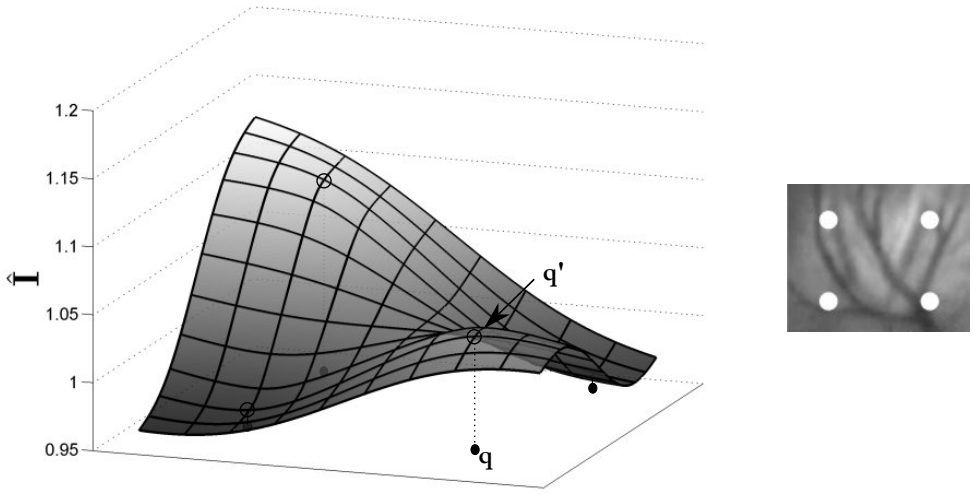


Figure II.10 : On the left, an illustration of the multiplicative gain $\hat{\mathbf{I}}$ modeled by a TPS surface, its corresponding control points $\mathbf{q}$ and their mapped positions $\mathbf{q}'$ on the surface. On the right, the distribution of the TPS control points $\mathbf{q}$ used in the experiments is illustrated

After the incorporation of the illumination compensation parameters, the Jacobian matrix $\mathbf{J}$ must be redefined. In this case, a new Jacobian matrix $\mathbf{J}^*$, composed of 3 sub-matrices ${}^w\mathbf{J}$,${}^q\mathbf{J}$,${}^b\mathbf{J}$:

$$\mathbf{J}^*(\mathbf{I},\mathbf{h},\mathbf{q}',b,\mathbf{x}) = \left[{}^w\mathbf{J}(\mathbf{I},\mathbf{h},\mathbf{q}',\mathbf{x}),\ {}^q\mathbf{J}(\mathbf{I},\mathbf{x}),\ {}^b\mathbf{J}(\mathbf{I},b,\mathbf{x})\right] \tag{II.25}$$

The matrix ${}^w\mathbf{J}(\mathbf{I},\mathbf{h},\mathbf{q}',\mathbf{x})$ is the Jacobian matrix corresponding to the warping parameters. It can be computed by replacing $\mathbf{I}$ by $\mathbf{I}'$ in equation (II.18):

$${}^w\mathbf{J}(\mathbf{I},\mathbf{h},\mathbf{q}',\mathbf{x}_i) = \frac{\partial \mathbf{I}'(w_{3D}(\mathbf{x}_i,\mathbf{h}))}{\partial \mathbf{h}} = \hat{\mathbf{I}}(\mathbf{x}_i,\mathbf{q}')\frac{\partial \mathbf{I}(w_{3D}(\mathbf{x}_i,\mathbf{h}))}{\partial \mathbf{h}} \tag{II.26}$$

The second sub-matrix, ${}^q\mathbf{J}(\mathbf{I},\mathbf{x})$ corresponds to the TPS surface that models $\hat{\mathbf{I}}$:

$${}^q\mathbf{J}(\mathbf{I},\mathbf{x}_i) = \mathbf{I}(\mathbf{x}_i)\,{}^{il}\mathbf{M}_i^{il}\mathbf{K}_* \tag{II.27}$$

where ${}^{il}\mathbf{M}_i$ is the i-th row of the matrix ${}^{il}\mathbf{M}$ which corresponds to $\mathbf{x}_i$. Notice that it is constant and does not depend on the illumination parameters $\mathbf{q}'$

The final third sub-matrix ${}^b\mathbf{J}(\mathbf{I},b,\mathbf{x})$ is a column of ones, which corresponds to the derivative of equation (II.19) with respect to b.

An interesting aspect about this illumination compensation technique is that little extra computational cost is needed, since only a few extra parameters are estimated per iteration. More details about this technique can be found in [Silveira and Malis, 2007].

II.5 Experimental results

In order to evaluate the proposed visual tracking method, two sets of experiments have been performed.

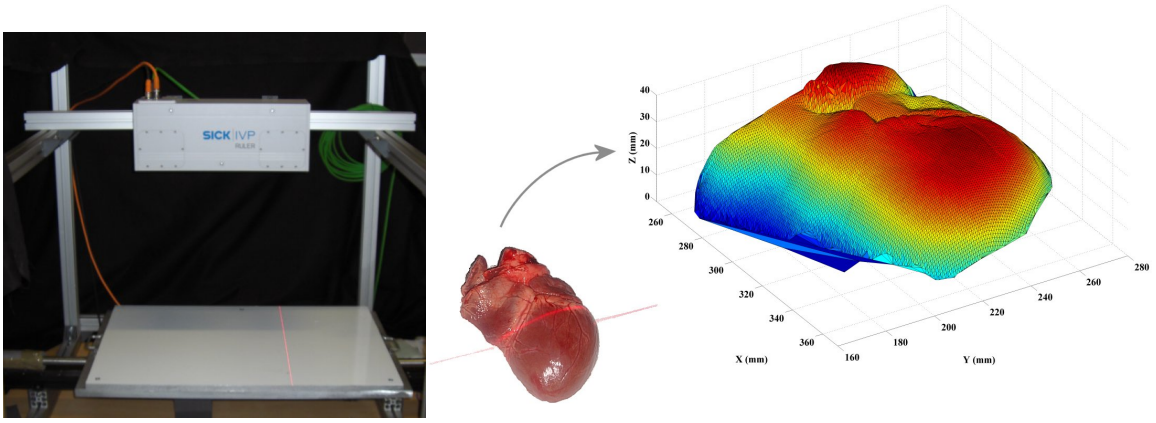


Figure II.11 : (left) The laser profilometer (Sick IVP Ruler, Sweeden) used for the experiment. (Right) The 3D shape of an *ex vivo* pig's heart with 0.2 mm depth precision. This data is used as ground truth for the experiment.

First, for justifying the choice for the TPS deformable model, experiments on *ex vivo* data are performed in order to assess the approximation quality of the TPS model for the heart surface in a static configuration. A comparative study between the TPS model and other deformable models proposed in the literature is presented in section II.5.1.1

Second, the dynamic tracking performance is analyzed through experiments conducted on two recorded *in vivo* images sequences. The tracking results are presented in section II.5.2.2 and the GPU implementation of the method is described in section II.5.2.3.

II.5.1 *Ex vivo* experiments

In order to assess the approximation quality of the TPS model for the heart surface in a static configuration, an experiment on an *ex vivo* porcine heart was performed. The objective was to evaluate the 3D heart surface shape reconstruction using TPS and the ground truth provided by a laser profilometer (Sick IVP Ruler, Sweeden) with 0.2 mm reconstruction precision. Figure II.11 contains a picture of the laser profilometer used in the experiment and the reconstructed 3D shape of the entire heart surface.

To approximate the shape of an arbitrarily defined 4x4 cm region of the heart surface, 25 control points evenly spaced on a grid were used. The region of interest is a point-cloud of 40x40 points, evenly spaced in the xy plane (the base plane of the profilometer). Figure II.12 displays in detail the selected region of interest and its approximation using a TPS surface, together with the corresponding approximation error. The TPS model was fitted to the heart surface using least-squares. The average error and standard deviation of the TPS representation with 25 control points is of 0.108 mm and 0.087 mm respectively, which is of the same order of magnitude of the measuring error.

II.5.1.1 Comparative analysis

For comparing the reconstruction quality between the proposed TPS model and other models proposed in the literature, the approximation results using two parametric models

Table II.1 : The approximation errors for the selected deformable models.

Model	**Average Error** (mm)	**Standard Deviation** (mm)	**Peak Error** (mm)
TPS	0.108	0.087	0.722
Bilinear	0.137	0.108	0.635
B-splines	0.108	0.084	0.541

for modeling the heart surface proposed in the literature are evaluated: the piecewise bilinear model [Stoyanov et al., 2005a] and the tensor B-splines [Lau et al., 2004] deformable models. For comparing the reconstruction quality of each model the same spatial resolution is used. For the TPS and bilinear models, 25 parameters are used, whereas 36 parameters are used for the B-spline model since for achieving the same spatial resolution as the other models, additional parameters are needed for estimating the borders of the surface. The approximated surfaces and reconstruction errors of each technique are plotted in figure II.13. The average approximation error and its standard deviation are also quantified in table II.1.

II.5.1.2 Discussion

The small approximation error indicated in figure II.12 suggests the heart surface analyzed in the experiment has an overall smooth shape and it can be well represented by a deformable model such as the TPS. Since most of the heart surface has a similar nature, the hypothesis made earlier about region-based approaches (section II.3) is reinforced.

A comparison between the three methods suggests that the TPS is more adequate model for reconstructing the heart surface shape. With a smaller number of parameters, the TPS approximation error is similar to that of the B-spline model and significantly lower than the piecewise bilinear model. Furthermore, the smallest peak error obtained with the B-spline model is due to the additional parameters of the B-spline model which provides a better estimation of the borders of the approximated surface where the peak errors occur. The piecewise bilinear model displays the worse reconstruction quality, due to its planar surface approximation. Finally, the large peak error using the TPS model on

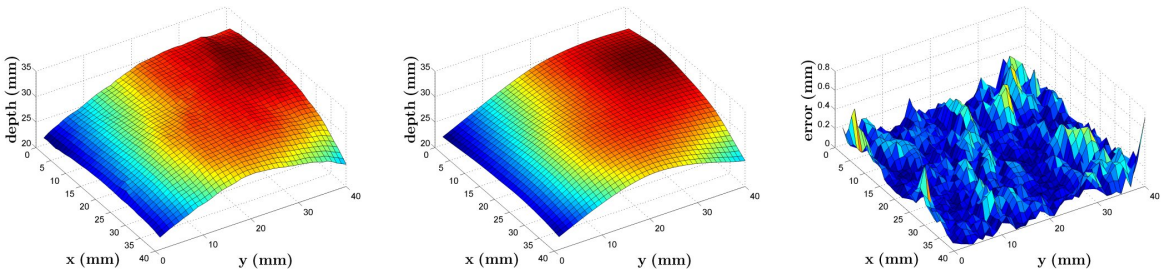


Figure II.12 : (Left) The detailed shape of the heart surface retrieved by the laser profilometer. (Middle) The approximation using the proposed TPS model. (Right) The approximation error.

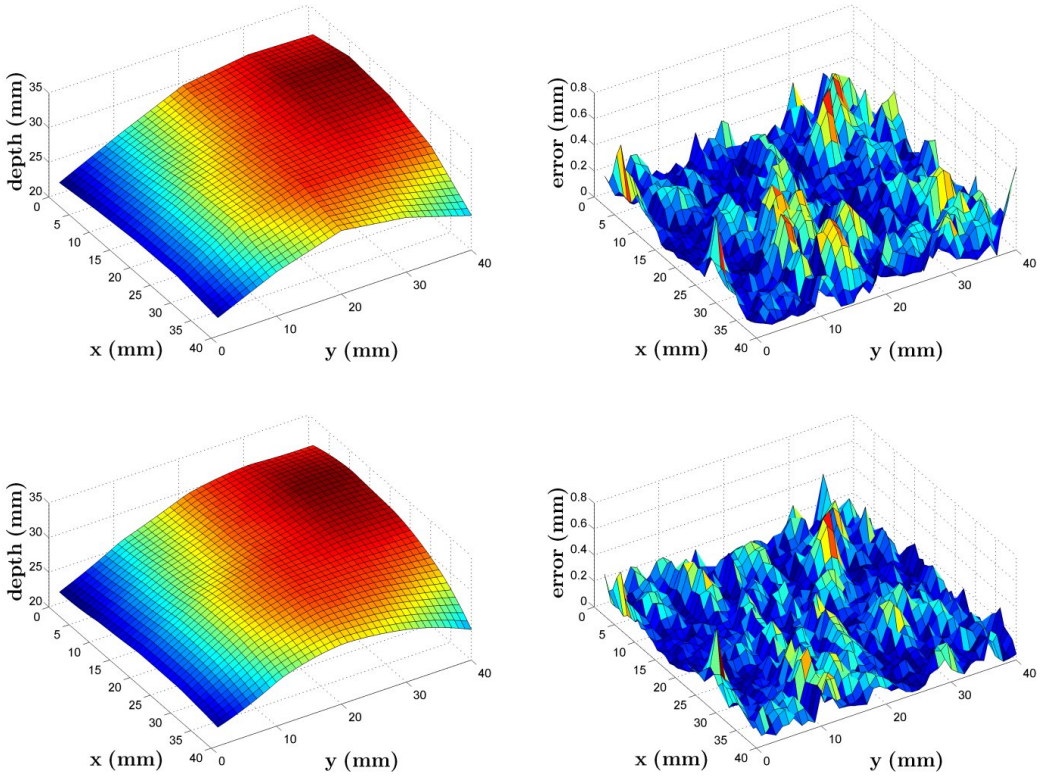


Figure II.13 : The approximated heart surface shape and the associated approximation error for (Top) the piecewise bilinear model (Bottom) the B-spline model.

one of the borders of the mapping is due to the regularization properties of the TPS, that minimizes the bending energy through the whole surface, unlike the previous mappings whose control points only affect their local support region. Unlike the B-spline and bilinear models, this issue can be circumvented using the TPS model by locally applying additional control points for increasing the spatial resolution of the mapping.

In this experiment, the control points are evenly placed on a grid. In practice, regions with a complex shape require additional control points for a better shape representation. It is important to remind that in the context of visual tracking the number of control points is utterly dependent on the texture information available on the heart surface.

II.5.2 *In vivo* experiments

After assessing the reconstruction quality using the TPS model through *ex vivo* experiments, the clinical value of the tracking method proposed in section II.4 is evaluated through experiments on two sets of recorded *in vivo* test images. Since it is not possible to retrieve the ground truth for the real 3D motion of the heart, results are evaluated based on the coherence of the retrieved 3D shape and motion of the heart surface.

The *in vivo* experimental data was first evaluated off-line using a MATLAB (Mathworks) implementation running at 0.1 frames per second on a standard PC with a 1.86 GHz Intel processor. An overview of the tracking algorithm is also presented in figure II.14. However, since higher tracking rates are required for deploying the proposed method

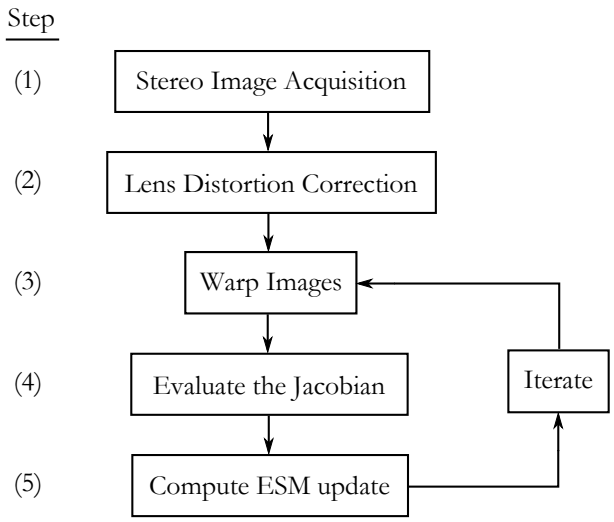


Figure II.14 : A schematic overview of the tracking algorithm.

in a robotic surgical platform, an implementation on a graphics processor unit (GPU) was developed. The details of the GPU implementation are given in section II.5.2.3.

II.5.2.1 Experimental data

The first *in vivo* sequence, dubbed DAVINCI, is an image sequence of a totally endoscopic coronary artery bypass graft (TECAB) operation using the DaVinci (Intuitive Surgical) surgical platform. This sequence was previously introduced for the analysis in section II.2. The sequence consists of 320x288 pixel color images of the heart captured at 50 Hz (see figure II.1) and contains 1600 images (32 seconds). The cameras of the stereo endoscope (figure II.16) were synchronized using a proprietary FPGA device and calibrated before the surgical procedure using a planar calibration object and standard techniques [Zhang, 2000]. A mechanical stabilizer was positioned on the patient's heart previous to the acquisition.

The second sequence, dubbed LIRMM1, is composed of 512x512 pixel grayscale images acquired from a porcine beating heart at 83.3 Hz[2]. As suggested in Ginhoux *et al.* [Ginhoux et al., 2003], a high acquisition rate is necessary for capturing the heart dynamics. However, no commercial solutions for high-speed stereo endoscopy are currently available. In order to perform a high speed endoscopic acquisition, two high speed 1M75 DALSA cameras attached to Storz endoscopes mounted with a small baseline are used for simulating a real stereo endoscope. The system was calibrated using MATLAB's calibration toolbox [Bouguet, 2009] (the resulting pixel reprojection error is below 0.15 pixel). The acquisition setup is illustrated in figure II.15. The heart was imaged for 60 seconds ($\approx$ 5000 images) in an open chest configuration with no mechanical stabilization, which yields a heart motion amplitude considerably larger than in the MIS scenario.

In both image sequences, the endoscopic lens induce considerable distortion in images of the operating site. Since this effect may severely distort the reconstructed 3D heart shape, a lens distortion correction step is required before the execution of the

[2]This specific frequency corresponding to a acquisition interval of 12 ms was chosen due to hardware limitations.

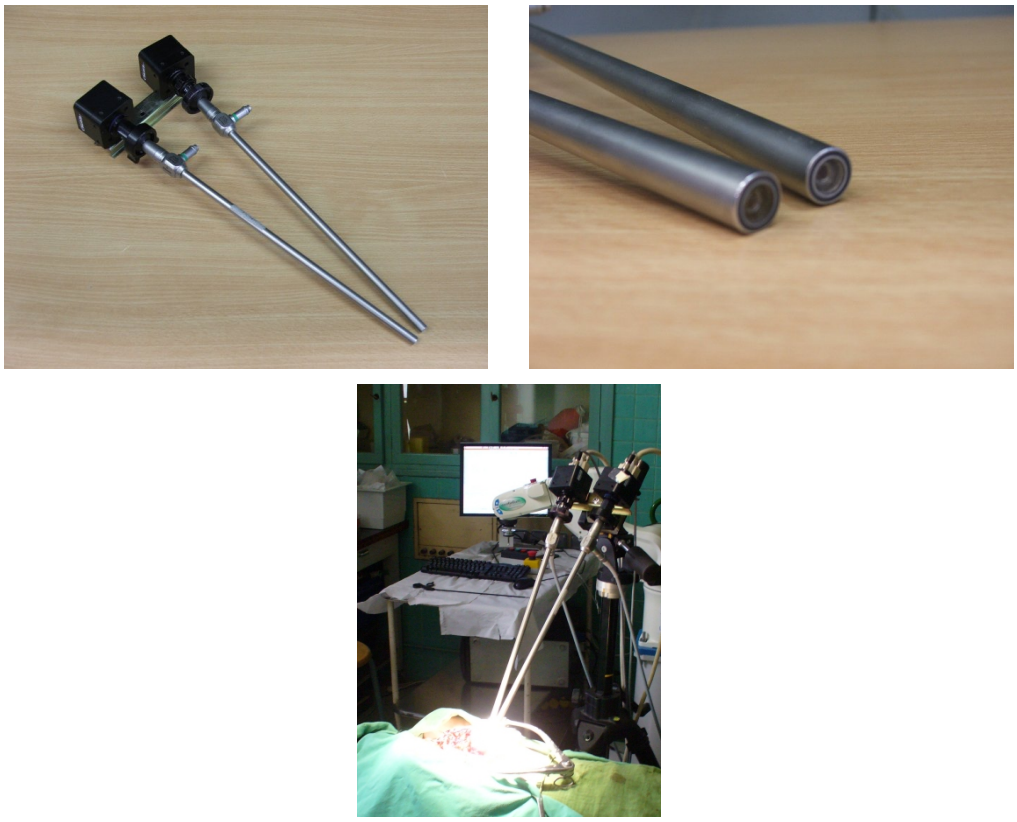


Figure II.15 : (Top Left) The acquisition setup for the LIRMM1 sequence using two Storz endoscopes connected to two high speed 1M75 DALSA cameras, mounted with a small baseline. (Top Right) Detail of the tip of the Storz endoscope used in the experiments. (Bottom) The acquisition setup in the operating room.

tracking algorithm. In this study, the lens distortion parameterization proposed by [Heikkilä and Silven, 1997] is adopted. Details on the lens distortion correction step are given in Appendix C.

II.5.2.2 Tracking results

II.5.2.2.1 The DAVINCI sequence

In this experiment, a 80x80 pixel image of a region on the heart surface using 6 control points is tracked. The number of control points is limited by the available texture on the heart surface and the inter-frame motion (the overlap of support regions of the control points due to large interframe motion leads to tracking instabilities). The reference image of the heart surface is manually selected and defines the region of interest to be tracked in 3D. The left camera image is arbitrarily used for defining the reference image. The control points are manually placed on regions with enough texture (their placement is illustrated in figure II.17(a)). Three main aspects of the tracking performance are analyzed: the estimated 3D heart shape, the illumination compensation and the required number of iterations.

1. **The estimated 3D heart shape** – Figure II.20 illustrates the target surface viewed by both left and right cameras for different instants of the heart cycle. The world

frame is centered at the left camera. The 3D displacement of the TPS control point highlighted by a white circle-dot in figure II.17(a) is plotted in figure II.19. The cardiac and respiratory motion for this interest point has an amplitude ranging from [-5.1, -3.8] mm, [1.7, 6.3] mm and [39.8, 42.6] mm for the xyz coordinates respectively. This matches the motion amplitude verified by the experiment reported in Stoyanov *et al.* [Stoyanov et al., 2005b] on the same image sequence.

2. **Illumination compensation** – In order to illustrate the significant illumination variations verified in the image sequence, the evolution in time of the illumination compensation parameters for both left and right images is plotted in figure II.22. In the experiments, a TPS surface with 4 control points was sufficient to model the illumination parameter $\hat{\mathbf{I}}$ (equation (II.20)). One can easily notice the correlation between the illumination parameters variation and the heart motion, as well as the illumination differences between left and right views. In average, the amplitude of $\mathbf{q'}_l$ and $\mathbf{q'}_r$ varies 20%. Concerning the parameters b_l and b_r that model the global intensity shift, the average amplitude variation is 5 gray levels (over 256 gray levels) for both left and right images. The large variations in the estimated illumination parameters confirms the illumination issues associated with the MIS scenario.

3. **The required number of iterations** – The maximum number of ESM iterations in



Figure II.16 : An example of a stereo endoscope integrated into the DaVinci surgical platform. Photo courtesy of Intuitive Surgical, Inc., 2008.

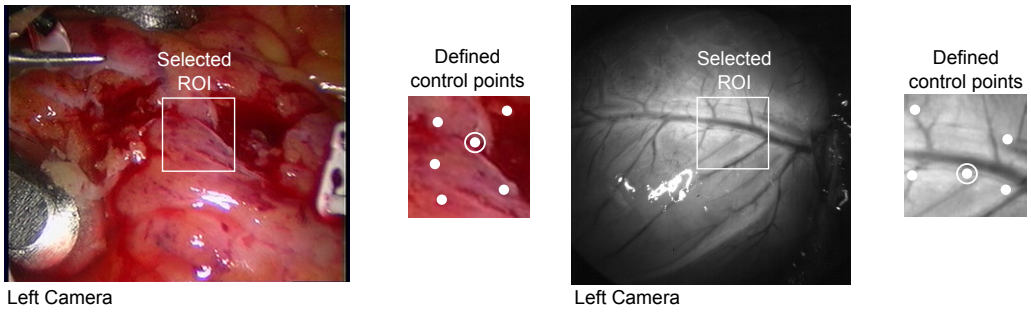


Figure II.17 : Selecting a region of interest from the left camera image from the (a) DAVINCI and (b) LIRMM1 sequences. The TPS control points are defined on the reference image (marked as white dots). The evolution of the highlighted control points (white circle dots) is plotted in figures II.19 and II.23 for the DAVINCI and LIRMM1 sequences, respectively.

this experiment was fixed to 30, even though in average only 8 iterations are needed for the tracking algorithm to converge (including the illumination parameters for both left and right images). In practice, the limitation of the number of iterations to 8 does not degrade the quality of the tracking results (as demonstrated in the next subsection, where the GPU implementation is discussed). The minimization loop converges if the sum of the elements of the increment vector drops below 0.01 ($\sum \delta\mathbf{h} < 0.01$). Figure II.21 (top) illustrates the number of iterations required by each frame. It is important to note the correlation between the peaks in number of iterations with disturbances due to specular reflections and large inter-frame motion (which induces considerable motion blur, since the images are acquired at a relatively low frequency). For illustrating this correlation, plots of the percentage of the back-warped image covered by specular reflections were included in figure II.21 and the speed of the point of interest plotted in figure II.19. Although the percentage of pixels affected by specular reflections is small (since the whole image is considered in the calculation), the increase in the required number of iterations for convergence in the presence of specular reflections is evident. Notice also that ESM convergence problems may also lead to peaks in the number of iterations.

II.5.2.2.2 The LIRMM1 sequence

In this experiment, a 128x128 pixel image of a region on the heart surface is tracked using 5 control points (the control point placement is illustrated in figure II.17(b)). The lower number of control points with respect to the previous experiment is due to the lack of available texture for placing additional control points. The reference image is also chosen from the left camera image. As for the previous test sequence, tracking performance is analyzed in three main aspects:

1. **The estimated 3D heart shape** – The deformation the target region on the heart surface undergoes is displayed in figure II.24 (the world frame is centered at the left camera). Figure II.23 displays the 3D displacement of the TPS control point highlighted by a white circle-dot in figure II.17(b). The motion amplitude for this interest point ranges from [4.5, 12.9] mm, [-5.5, 4.5] mm and [48.7, 56.6] mm for the xyz coordinates respectively.

2. **Illumination compensation** – Similarly to the previous experiment, 4 TPS control points were used for modeling the illumination parameter $\hat{\mathbf{I}}$. The evolution in time of the illumination compensation parameters for both stereo images is plotted in figure II.26. The amplitude of the illumination variations in this experiment are considerably larger than in the DAVINCI sequence, as can be visually verified in the back-warped images from different instants of the heart cycle displayed in figure II.26. This is due to a combination of the particular illumination conditions when acquiring at high-speed with endoscopes (a considerable illumination loss is verified on the borders of the images) and the large heart motion in the open-chest configuration. Certain regions display amplitude variations of over 30% (considering the parameters $\mathbf{q'}_l$ and $\mathbf{q'}_r$). The global intensity shifts b_l and b_r vary 10 gray levels in average.

3. **The required number of iterations** – Similarly to the previous experiment, the maximum number of ESM iterations in this experiment was fixed to 30, with the

same convergence criterion as the previous experiment. An analysis of the number of iterations during tracking plotted in figure II.21 lets us conclude that in average 6.3 iterations are needed for the tracking algorithm to converge. As in the experiment with the DAVINCI sequence, the limitation of the maximum number of iterations to 7 iterations does not degrade the quality of the tracking results. In addition, since the required number of iterations is related to the inter-frame heart motion, higher acquisition speeds require a smaller number of iterations for convergence.

II.5.2.3 Implementation on a Graphics Processor Unit (GPU)

The visual tracking algorithm must extract the heart motion on-line for an accurate synchronization with the robotic tools. This implies that tracking must run at high frame-rates. The first effort towards a fast implementation of the proposed tracking algorithm, an implementation using Intel Performance Primitives - IPP running at $\approx$ 5Hz, was far below the desired tracking frame-rate. Due to the larger computational requirements of the application, the computational power of recently released graphics processor units (GPU) is explored.

The market for real-time high-definition 3D graphics has given birth to high-parallel, multi-core processors with great computation power. Since GPUs are specially well-suited to address problems that can be expressed as data-parallel computations, an implementation of the tracking algorithm was developed in order to take advantage of the hardware efficient processing. For the experiments, a NVidia GTX280 (Santa Clara, EUA) graphics card programmed using CUDA (a programming extension to C) was used.

II.5.2.3.1 Improvements in computational efficiency

To illustrate the improvements in computational efficiency using the GPU, the improvements at each step of the tracking algorithm according to figure II.14 are summarized below:

- **Step 2** – The lens distortion on the endoscopic images is considerable and its correction is done by remapping the pixels using a pre-computed lookup table. The computation of the interpolated gray-values using bilinear interpolation is hardware accelerated on the graphics card. Therefore, the gain with respect to the CPU implementation where the intensity value of each pixel is computed sequentially is considerable.

- **Step 3** – For each camera, the mapped position of each pixel from the template is given by a matrix-vector multiplication (see equation (II.13)). The computational speed-up is obtained by computing each line of the matrix-vector multiplication on separate threads, which is far more efficient than the CPU implementation. Finally, the computation of the interpolated gray-values using bilinear interpolation is hardware accelerated.

- **Step 4** – For mounting the Jacobian matrix, each line of the matrix is handled by a separate thread. The low latency access to the graphics card internal memory allows for a significant speed-up compared to the CPU implementation. Second, the computation of the gradient of the warped images by convolution is highly

Table II.2 : Computational time in the GPU implementation for tracking in the DAVINCI sequence

Step	Required time (ms)
2	0.25
3	0.6
4	0.6
5	0.8
Total	2.25 (per iteration)

efficient on the graphics card, since the convolution with a kernel can be efficiently parallelized.

- **Step 5** – The computation of the ESM update is performed both on the CPU and the GPU. For the computation of the pseudo-inverse required by the minimization, the Moore-Penrose pseudo-inverse is used. Since the computation of a matrix inverse is required in the process and since its implementation on a GPU is still an open problem, the matrices are transfered to the CPU and use LAPACK C library functions to perform the inverse. As expected, this is the most time-costly operation required by the tracking algorithm. The computation of the Hessian matrix $\mathbf{J}^{\mathrm{T}}\mathbf{J}$, where $\mathbf{J}$ is the Jacobian matrix detailed in section II.4.2 as well as the multiplication of the Jacobian transpose with the error image is performed on the GPU. However, no significant speed-up is obtained with respect to the CPU implementation. This is due specially to the shape of the Jacobian matrix, which does not allow efficient access to the graphics card memory and the parallel computation using a larger number of threads.

II.5.2.3.2 Computational times

The time required for each step detailed above may vary depending on the size of the reference image, the number of spatial and illumination parameters. For exemplification purposes, the computation times per iteration required for tracking the DAVINCI sequence (using the same configuration presented earlier in this section) are detailed in table II.2. The time required by the memory transfers between the CPU and the GPU can be neglected. Considering that the time required to transfer the images from the cameras to the graphics card memory is around 0.2 ms, tracking speeds of 50 Hz can be achieved.

The most costly operation is the Hessian computation and the inverse in step 5. This is due to the impossibility of parallelizing the computation using a large number of threads. As previously mentioned, it is important to note that the number of iterations required for the tracking algorithm to converge depends on the inter-frame motion, which is directly related to the image acquisition speed. Therefore, if the acquisition speed is increased, the computational requirement decreases. An example of this relation can be found in the LIRMM1 experiment, where even though the heart motion amplitude is large (since the beating heart is not mechanically stabilized), the number of iterations for convergence is smaller than in the DAVINCI experiment (6.3 against 8 iterations).

II.5.2.3.3 Trade-off between accuracy and number of iterations

Due to the computation time restrictions, the number of iterations of the minimization loop are limited. Consequently, concerns regarding the accuracy of the estimated heart motion with restricted iterations arise. For example, considering the tracking experiments on the LIRMM1 sequence, for achieving frame-rate (tracking at the acquisition frequency) the number of iterations must be limited to 8 (computation time per frame of 18 ms). The effects of the limited number of iterations on the accuracy of the retrieved heart motion are measured in figure II.18, where a plot shows the difference between the heart trajectory of the control point highlighted by a white circle-dot in figure II.17 tracked using a maximum of 8 iterations per frame and the previous heart motion plotted in figure II.19 using a limit of 30 iterations per frame.

The error plots indicate an average error of 0.05 mm, which lets us conclude that limitation on the number of iterations has little influence on the retrieved heart motion, with the exception of periods of large inter-frame motion and large specular reflections where error peaks of 0.3 mm were detected. However, in these periods the estimation of the tracking parameters does not converge even with a 30 iteration limit (see figure II.21) and therefore the estimated heart surface is less reliable.

II.5.2.4 Discussion and perspectives

The experimental results suggest that the proposed tracking method is able to accurately retrieve the heart surface motion in spite of illumination variations and the complex heart motion. However, certain factors limit the performance of the algorithm. First, only a restricted number of control points can be used in the TPS mapping due to the lack of texture information and the large inter-frame motion. This constraint limits the spatial resolution of the algorithm and the accuracy of the heart surface reconstruction. Second, the current design of the tracking method is very sensitive to specular reflections, which significantly disturb the estimation of the warping parameters.

Since the TPS control points must be placed in regions on the heart surface with sufficient texture information, the shape of regions with homogeneous texture is interpolated by the TPS surface, reducing the spatial resolution of the TPS mapping, consequently affecting the accuracy of the heart surface reconstruction. A possible solution to circumvent this problem is to use pre-operative information for designing a deformable model that better describes the heart surface deformation in the absence of visual information.

Since the method is formulated as an iterative minimization algorithm, the estimation of the tracking parameters may fall into false minima if the inter-frame displacement of the region of interest is too high. The problem can be partially solved using a coarse-to-fine approach similar to the hierarchical estimation proposed in [Stoyanov et al., 2005a]. However, similarly to the previously discussed problem with textureless regions, the ideal solution is the incorporation of biomechanical constraints to the motion of the control points to better translate the physical properties of the heart surface during deformation.

The current tracking method relies on the manual placement of the TPS control points. The problem lies on the absence of specific criteria for determining if a region is sufficiently textured. Traditionally, the image gradient [Baker and Matthews, 2002] is the most common criterion for selecting textured regions but recent studies suggest that gradient *per se* can even diminish tracking performance [Benhimane et al., 2007]. A possible solution to the problem is an adaptive parameterization technique similar to the work of Hansen

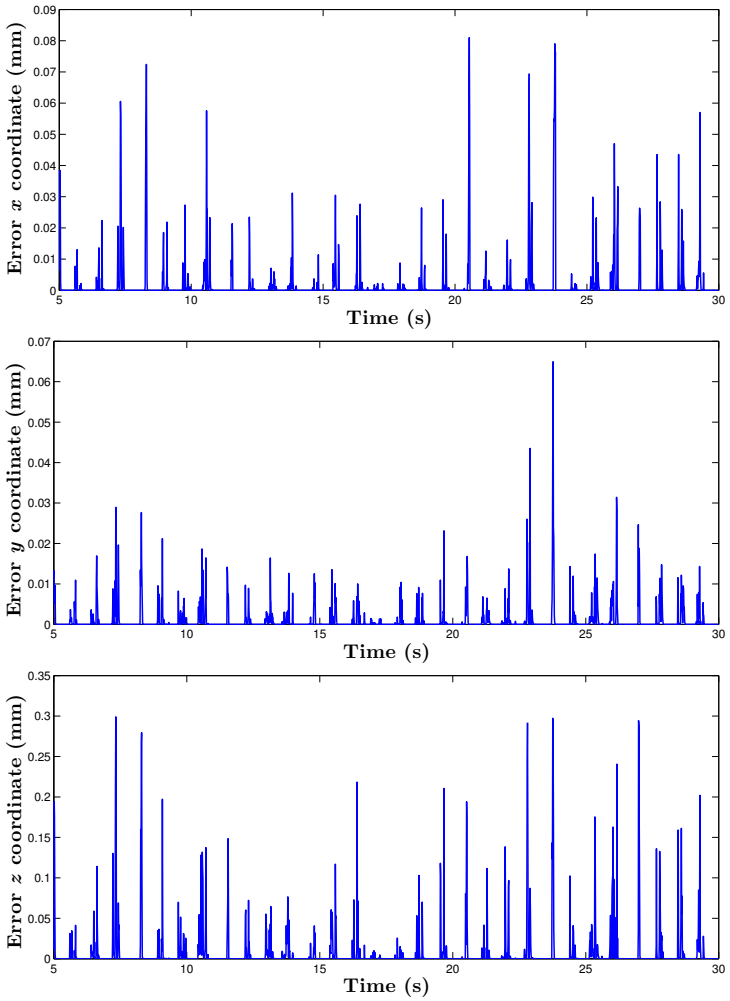


Figure II.18 : The absolute difference between the tracked heart trajectories in all Cartesian coordinates using a maximum of 8 and 30 iterations per frame when tracking a ROI in the DAVINCI sequence.

et al. [Hansen et al., 2008], which describes a method for refining a control point mesh of a multivariate B-spline mapping in order to provide a better representation of image deformations.

Finally, the correlation between the presence of specular reflections and the high number of iterations required for the minimization loop to convergence displayed in figures II.21 and II.25 suggests that specular reflections considerably disturb tracking. For handling such events the simple detection and removal method proposed earlier is not sufficient for correctly estimating the heart surface deformation because not enough visual information is available. A possible solution to this problem is the use of motion prediction methods for modeling the position of TPS control points whose support region is more severely affected by such phenomenon. In the next chapter, it will be shown how the prediction of the future heart motion can be used to bridge tracking problems related to specular reflections.

II.6 Conclusion

In this chapter, a visual tracking method for estimating the 3D deformation of a region of interest on the heart surface is proposed. A thin-plate spline model is used for representing the heart surface deformations. The possibility of arbitrarily placing control points offered by the TPS mapping allows us to easily circumvent problems with the estimation of the heart surface deformation due to lack of texture. In addition, the TPS mapping compares favorably against other deformable models proposed in the literature in terms of reconstruction quality, as demonstrated by the *ex vivo* experimental results.

The main contribution consists in the efficient parameterization for 3D tracking using stereo endoscopic images. The novel parameterization of the TPS for 3D tracking in the stereo scenario allows an efficient formulation of the tracking problem, without requiring disparity search or other intermediary steps that may reduce the tracking accuracy.

The practical value of the proposed tracking method is demonstrated by the *in vivo* experiments. The illumination compensation algorithm applied in MIS scenario displays good performance, greatly increasing tracking robustness against lighting variations. Finally, the high tracking frame-rates achieved using the GPU implementation offer the possibility of deploying the proposed visual tracking method in a prototype surgical platform for synchronizing the motion of the surgical tools with the beating heart.

However, the current design of the algorithm is still not robust to occlusions (either by surgical instruments or large specular reflections). For tackling this problem, the motion dynamics of the beating heart can be used as an important source of information. In the next section, a description of how the quasi-periodic beating heart motion can be used for increasing the robustness of visual tracking is given.

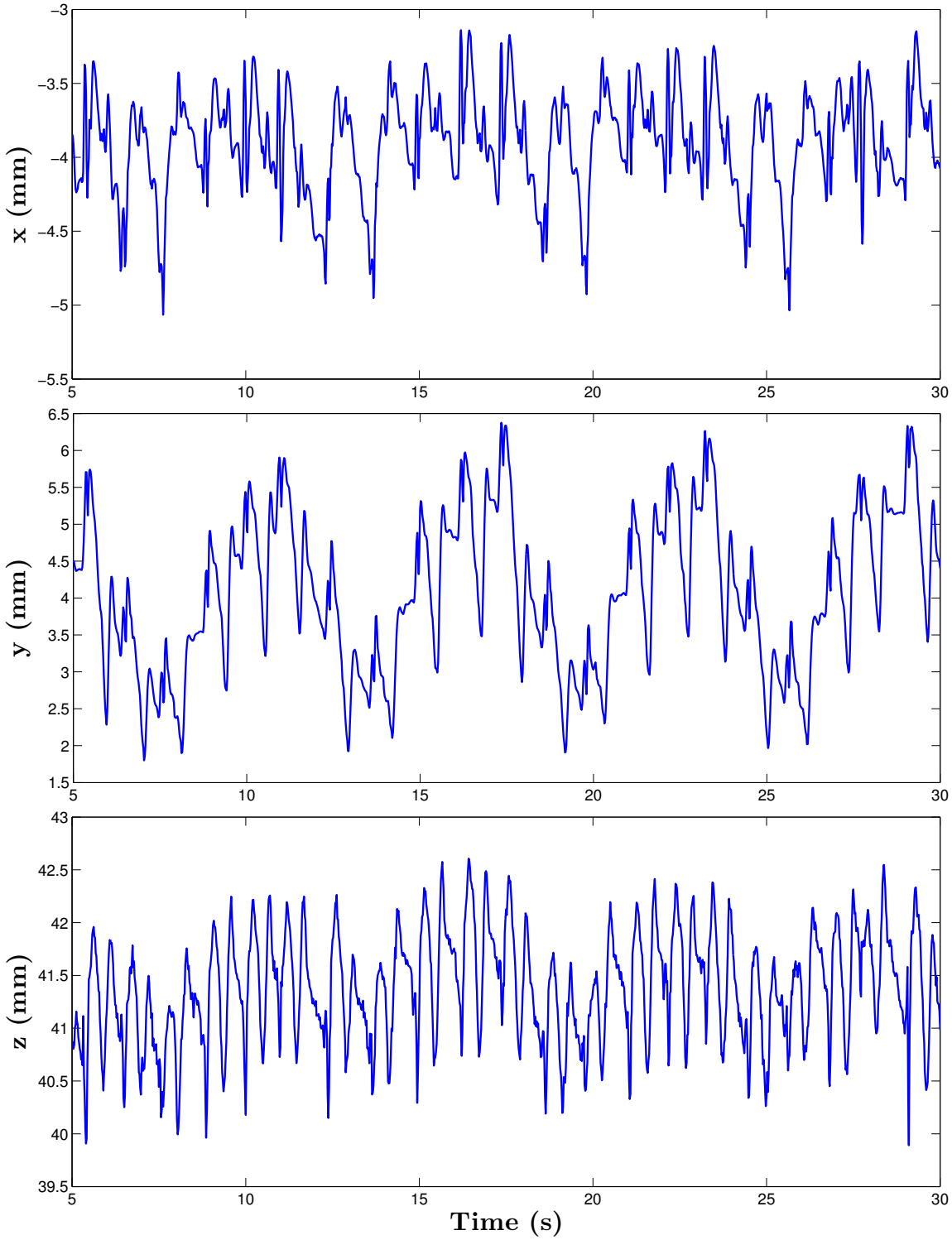


Figure II.19 : The DAVINCI sequence – The retrieved 3D coordinates of a point of interest on the heart surface.

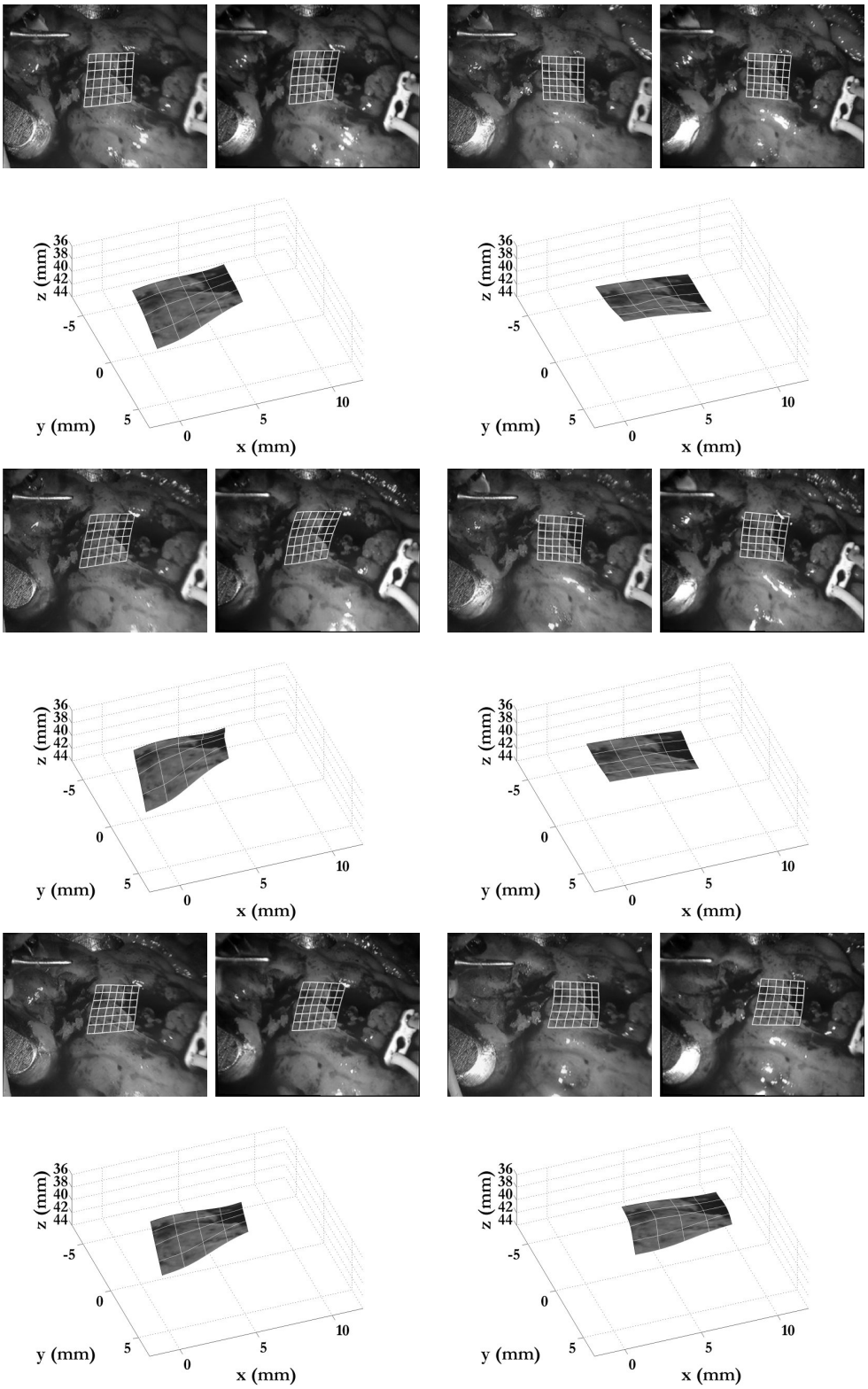


Figure II.20 : The DAVINCI sequence – For different instants of the heart cycle, we illustrate the target region tracked on the left and right images of the stereo pair and the corresponding TPS surface approximation. The reconstruction is presented using the left camera as world coordinate frame.

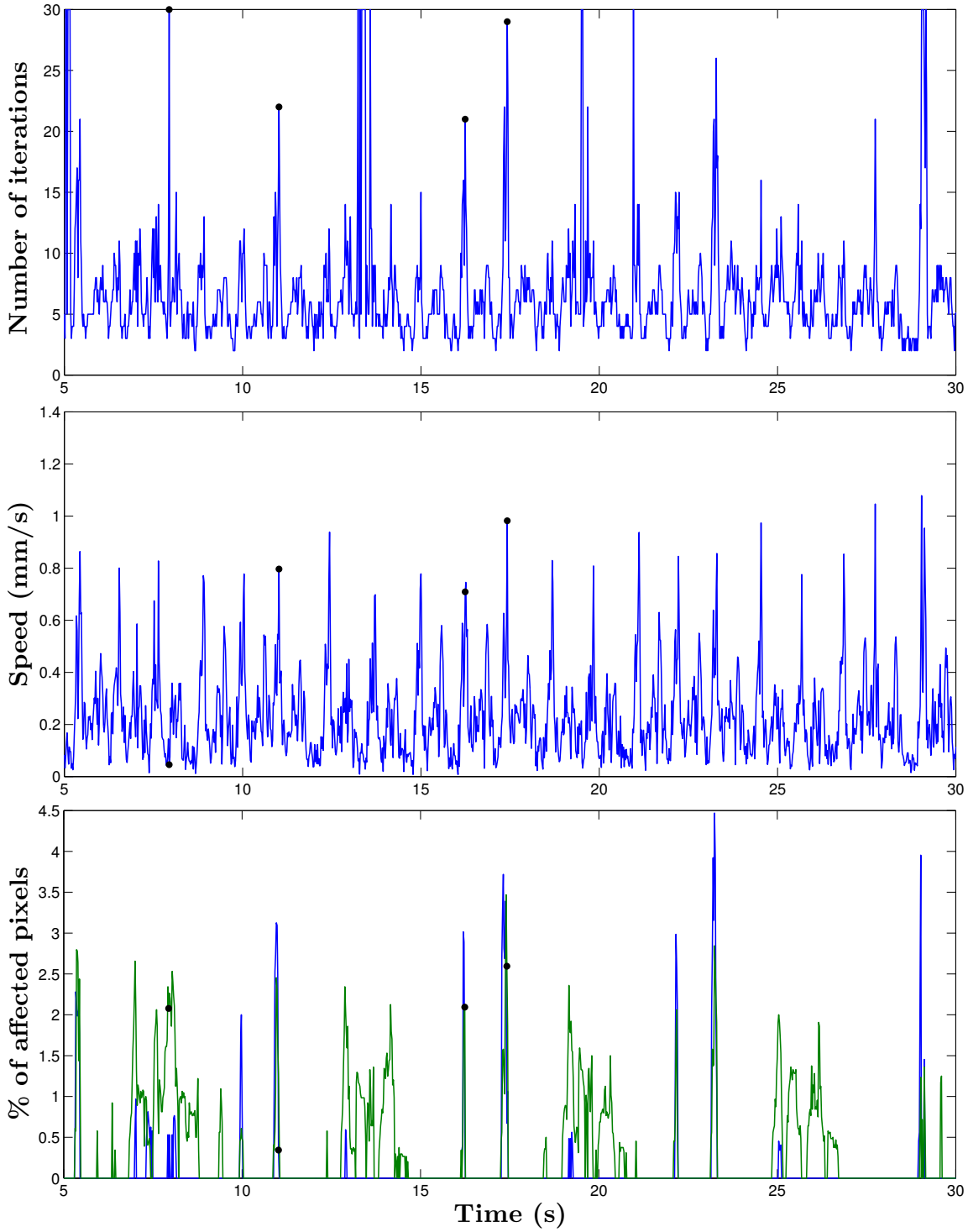


Figure II.21 : The DAVINCI sequence – (Top) The required number of iterations (Middle) The speed of a given point of interest on the heart surface (Bottom) Percentage of pixels affected by specular reflections on both left and right images (in green and blue respectively). The black dots correspond to peaks in the number of iterations at times $t = 7.94, 11.02, 16.24, 17.42$s. Notice the correlation between the peaks in number of iterations with disturbances due to specular reflections and large inter-frame motion. ESM convergence problems also lead to peaks in the number of iterations.

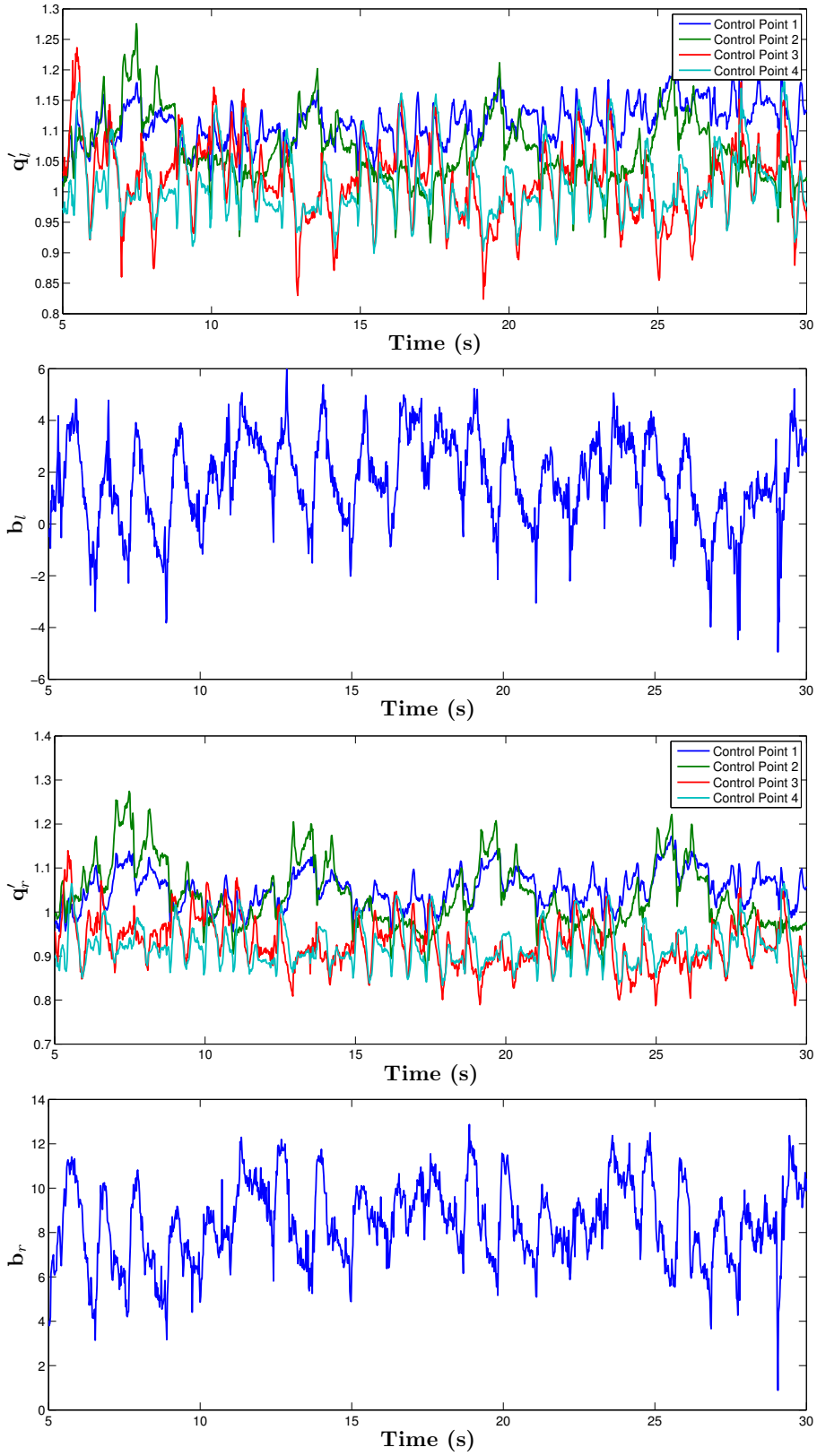


Figure II.22 : The DAVINCI sequence – The evolution of the illumination compensation parameters $\mathbf{q}'_l, b_l$ and $\mathbf{q}'_r, b_r$ for the left and right views, respectively.

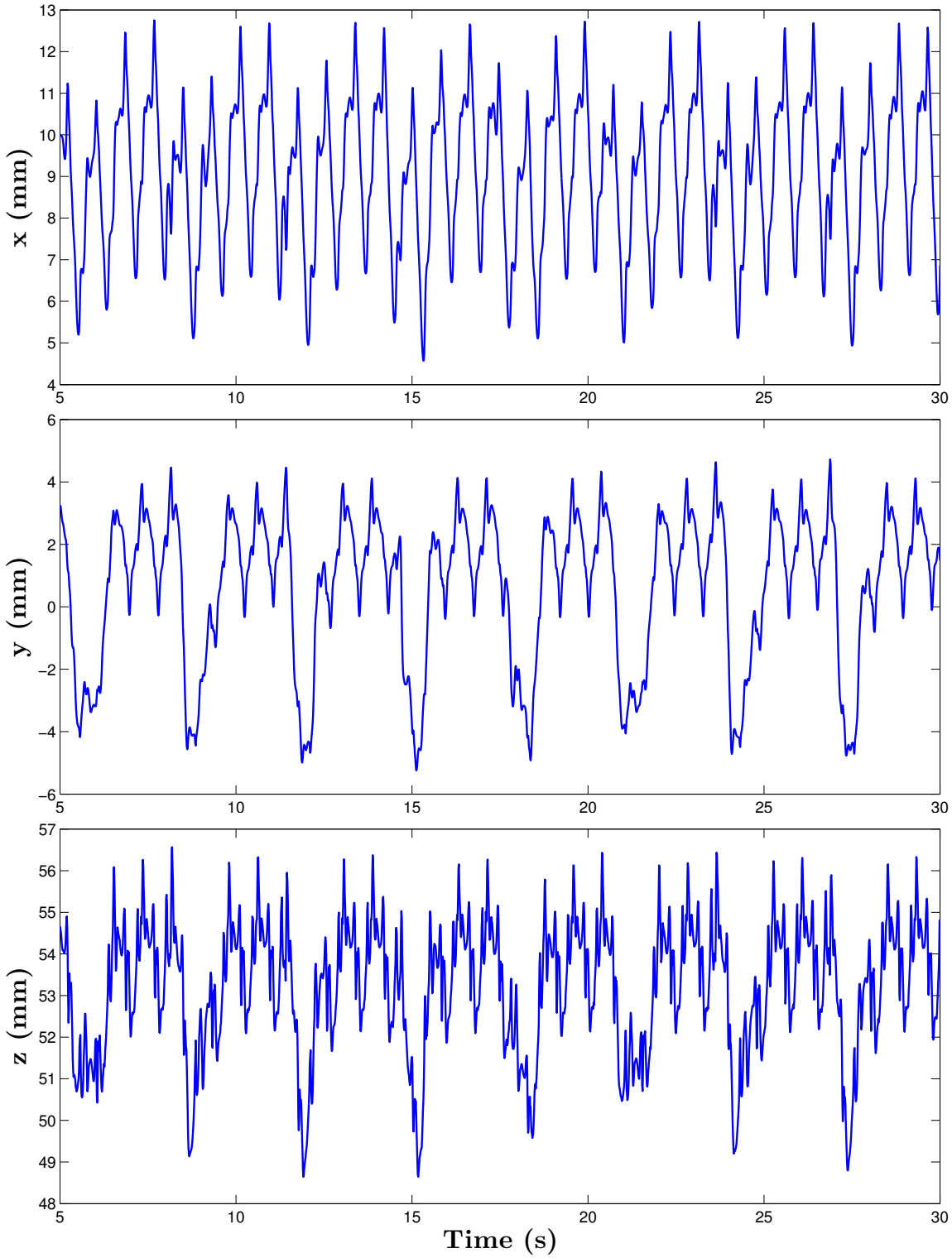


Figure II.23 : The LIRMM1 sequence – The retrieved 3D coordinates of a point of interest on the heart surface.

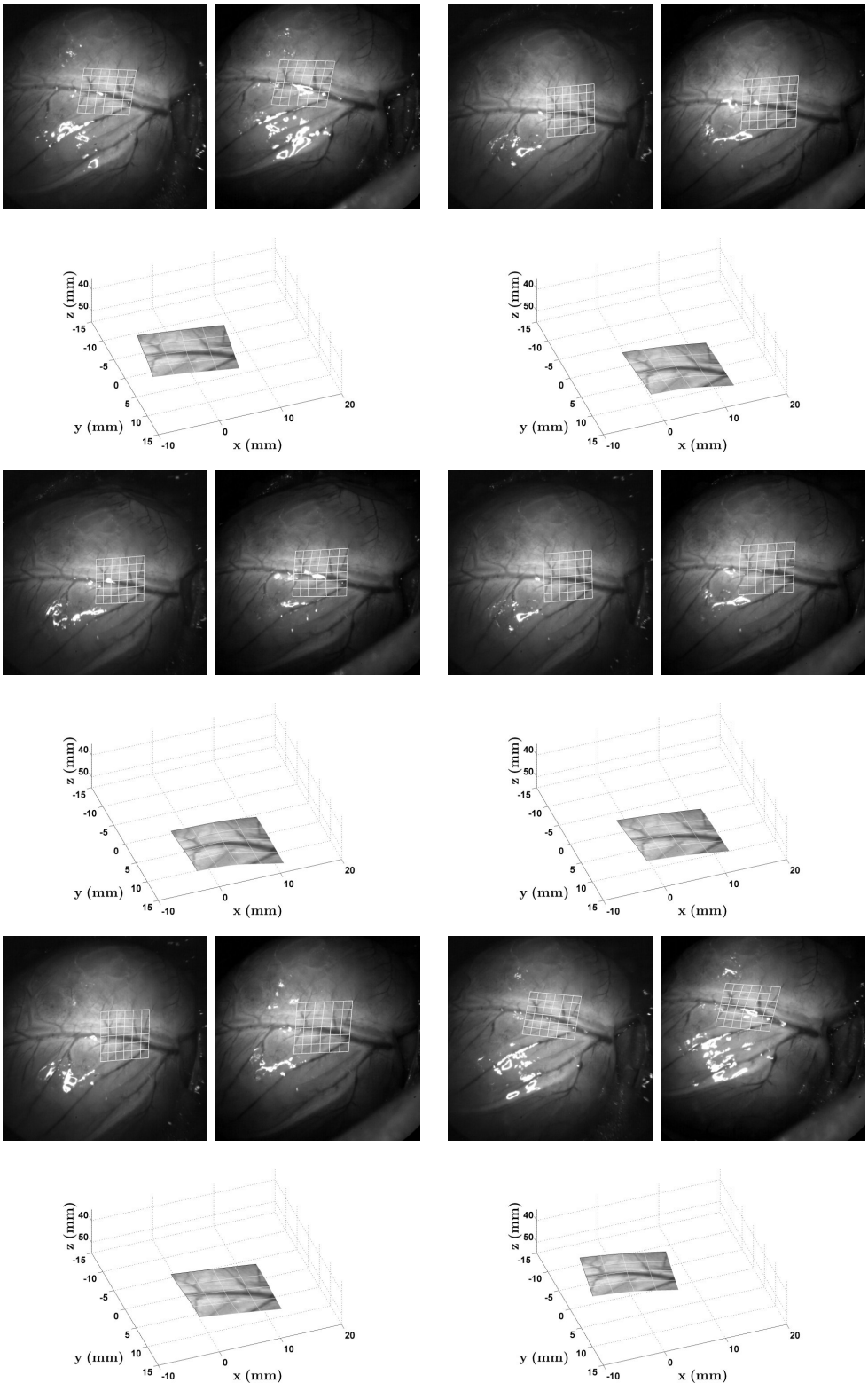


Figure II.24 : The LIRMM1 sequence – For different instants of the heart cycle, the target region tracked on the left and right images of the stereo pair and the corresponding TPS surface approximation are illustrated. The reconstruction is presented using the left camera as world coordinate frame.

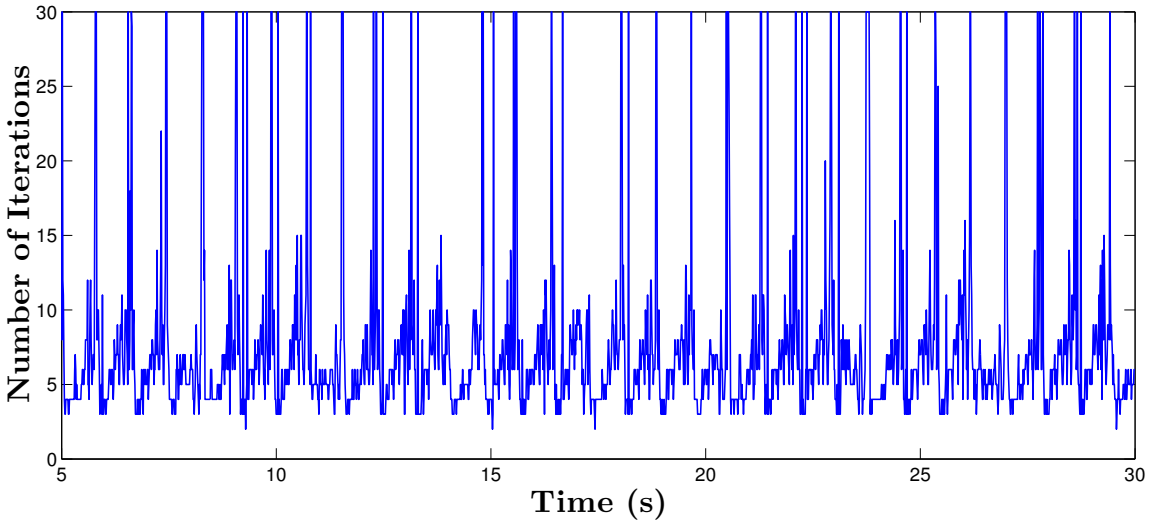


Figure II.25 : The LIRMM1 sequence – The required number of iterations during tracking.

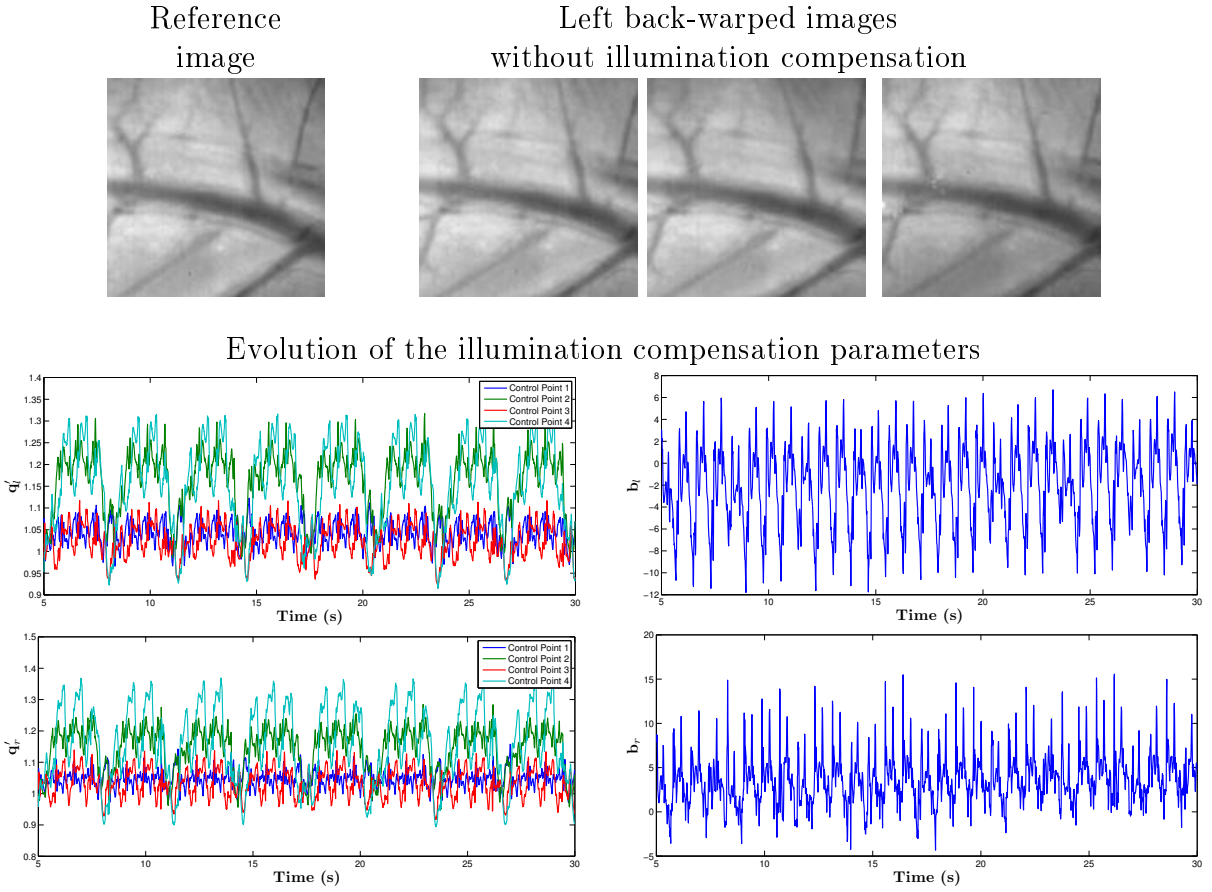


Figure II.26 : The LIRMM1 sequence – On top, left back-warped images of the tracked region of interest in different instants for illustrating the illumination variations during tracking. The variations on both left and right back-warped images are compensated for avoiding tracking loss. On the bottom, the evolution of the illumination compensation parameters $\mathbf{q}'_l, b_l$ and $\mathbf{q}'_r, b_r$ for the left and right views, respectively.

Chapter III

Robust visual tracking

III.1 Introduction

In the previous chapter, a visual tracking method for estimating the 3D motion of the heart surface was proposed. Tracking is solved as an iterative registration method: at every frame, the previous estimated tracking parameters are used to initialize the minimization procedure. Furthermore, no regularization or constraint is applied to the TPS model deformation. Thus, even though this formulation gives full freedom for the deformable model to adapt to complex shapes, it does not take into account the quasi-periodicity of the beating heart motion.

Clearly, valuable information can be extracted from the beating heart motion dynamics for improving the accuracy and robustness of visual tracking. More specifically, if the heart dynamics can be modeled, the future heart motion can be anticipated and tracking disturbances such as occlusions (e.g. surgical instruments, smoke, blood, specular reflections, etc) or tracking failures can be bridged. In this chapter, the problem of modeling and estimating the beating heart motion for predicting its future behavior in the context of visual tracking is studied. Two major novelties are introduced in this chapter: first, a time-varying dual Fourier series for modeling the quasi-periodic beating heart motion is proposed. For estimating the parameters of the Fourier series, a probabilistic framework based on the Extended Kalman filter (EKF) is used. Second, the heart motion prediction method is integrated to the visual tracking, creating a unified framework for estimating the temporal motion and spatial deformation of the heart surface for improved visual tracking robustness. The performance of the proposed heart motion prediction method is assessed through several experiments using *in vivo* data. The resulting improvements in the tracking robustness facing tracking failures and occlusions are presented and discussed. Preliminary results have been published at [Richa et al., 2010a].

This chapter is organized as follows. In the next section, the proposed prediction scheme based on the dual Fourier series model and the EKF estimation framework is described in detail. Experiments on *in vivo* data and a comparative study with similar works in the literature are also presented for attesting the good performance of the proposed prediction method. In section III.3, the robust heart motion tracking framework is presented in detail. In order to evaluate the improvements in the tracking robustness, experiments using the LIRMM1 image sequence are performed. Finally, a conclusion and a discussion on future work perspectives are provided.

III.2 Predicting the heart motion

In this section, a method for predicting the heart motion is described. The method is based on a dual Fourier series that explicitly models the cardiac and respiratory motions. For estimating the parameters of the Fourier series, a probabilistic framework based on the Kalman filter is used. The details of the filter design are presented. First, for a better understanding of the heart motion dynamics, a spectral analysis of the recorded heart motion database is performed. The state of the art in the literature is also presented and discussed.

III.2.1 Beating heart motion dynamics

For modeling the heart motion, a more profound knowledge of its characteristics is necessary. Towards this goal, spectrograms of the beating heart motion data collected in the tracking experiments from the previous chapter (subsection II.5.2.2) and recorded heart motion data using artificial markers from the experiments reported in Sauvée *et al.* [Sauvée et al., 2007] are performed. The objective is to analyze the frequency distribution of the beating heart motion and visualize the slow heartbeat frequency variations (frequency shifts are verified mostly in the cardiac motion since a breathing machine is used in all experiments).

The recorded heart motion data from [Sauvée et al., 2007], dubbed LIRMM2 experimental data, was acquired using a high-speed camera connected to an endoscope and an artificial marker fixed on the surface of a porcine beating heart. The heart was imaged in an open chest configuration with no mechanical stabilization for 15 seconds (1900 samples) at 125 Hz. Figure III.1 illustrates the artificial marker used to measure the heart surface motion. The 3D motion of the center of the marker is plotted in Figure III.1. The amplitude of the heart motion ranges from [-2.9, 4.6] mm, [-1.8, 1.3] mm and [37.5, 49.8] mm for the xyz coordinates respectively.

Figures III.2, III.3 and III.4 display the spectrograms of the DAVINCI, LIRMM1 and LIRMM2 experimental data respectively (the DAVINCI and LIRMM1 experimental data consist of the motion collected from the tracking experiments presented in the previous chapter). For computing the spectrogram, a Fast-Fourier Transform (FFT) is computed on a window of the heart motion signal using a Hanning window of one forth of the signal length. As the window slides in the temporal axis, the power spectral density of each frequency segment is stored (see color bar on the right of the figures). For a convenient frequency resolution, the maximum frequency is set to 10 Hz.

The plots from all experimental data indicate that the dominant frequencies in all axes are situated between 0 - 5Hz, with significant energy up to 8Hz. While the frequency of the respiratory and cardiac motions remains practically constant in the LIRMM1 experimental data, considerable frequency variations can be detected in the DAVINCI and LIRMM2 data.

Although the two prominent peaks in the FFT of the heart motion signal correspond to the respiratory and cardiac motions, these two components of the heart motion cannot be separated by simple filtering. As described in the work of Cuvillon [Cuvillon et al., 2006], the cardiac motion is modulated by the volume inside the thoracic cage, which is determined by the volume of the lungs. Thus, a coupling between the two sources of motion should exist. However, analyzing the spectrogram of the DAVINCI, LIRMM1

and LIRMM2 experimental data, no spectral energy could be detected in combinations of the respiratory and cardiac frequencies and therefore, this coupling can be considered negligible in our experimental database.

It is important to note that in the same way mechanical stabilizers can be used to attenuate the heart motion amplitude, pacemakers can be used for circumventing the heart frequency variations in time and maintaining a more stable heart cycle [Ortmaier et al., 2005].

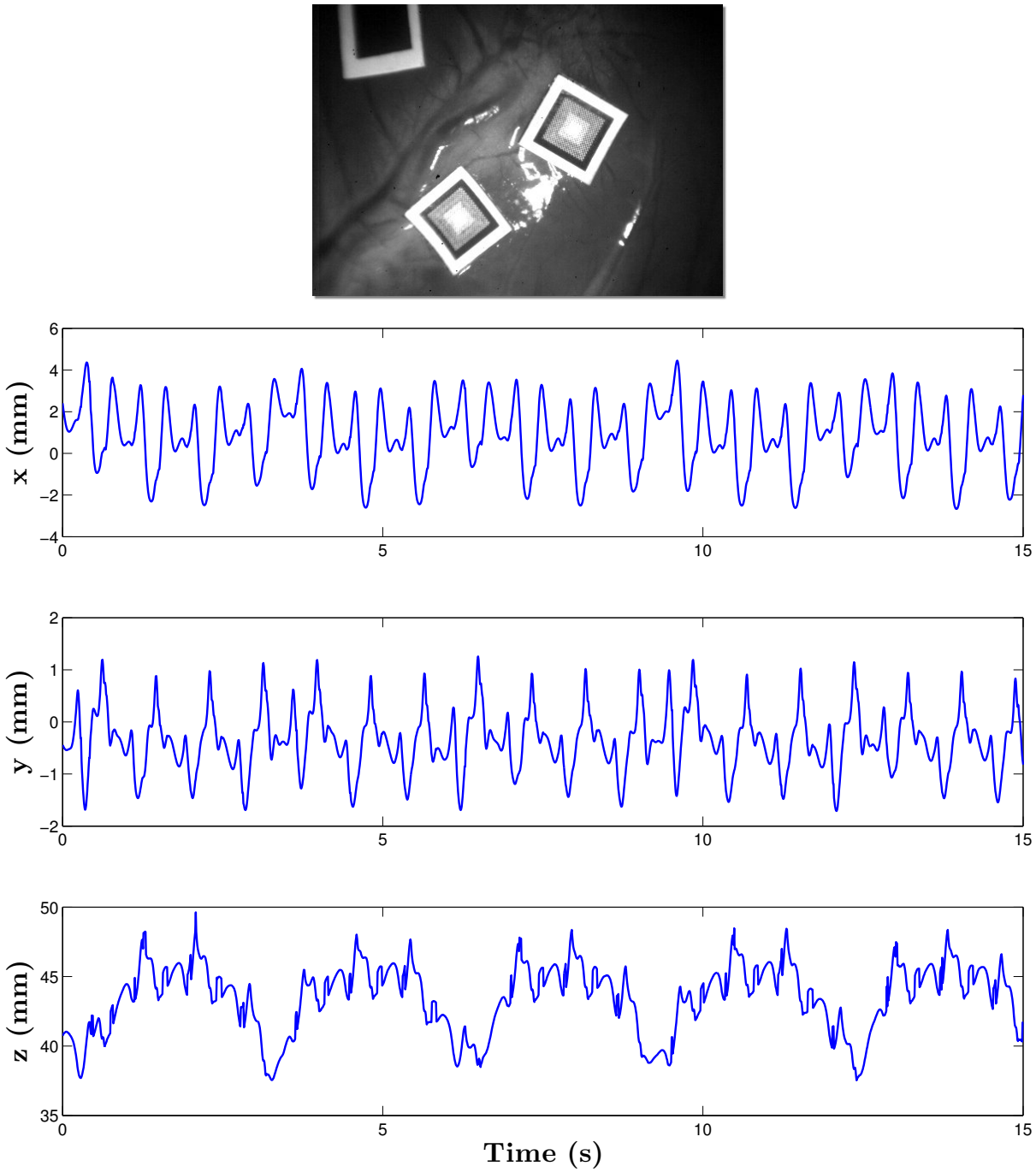


Figure III.1 : The LIRMM2 experimental data. On the top, the artificial marker placed on the heart surface used for estimating the heart motion (for more details, refer to Sauvée *et al.* [Sauvée et al., 2007]). The 3D motion of the center of the artificial marker is recorded for a duration of 15 seconds.

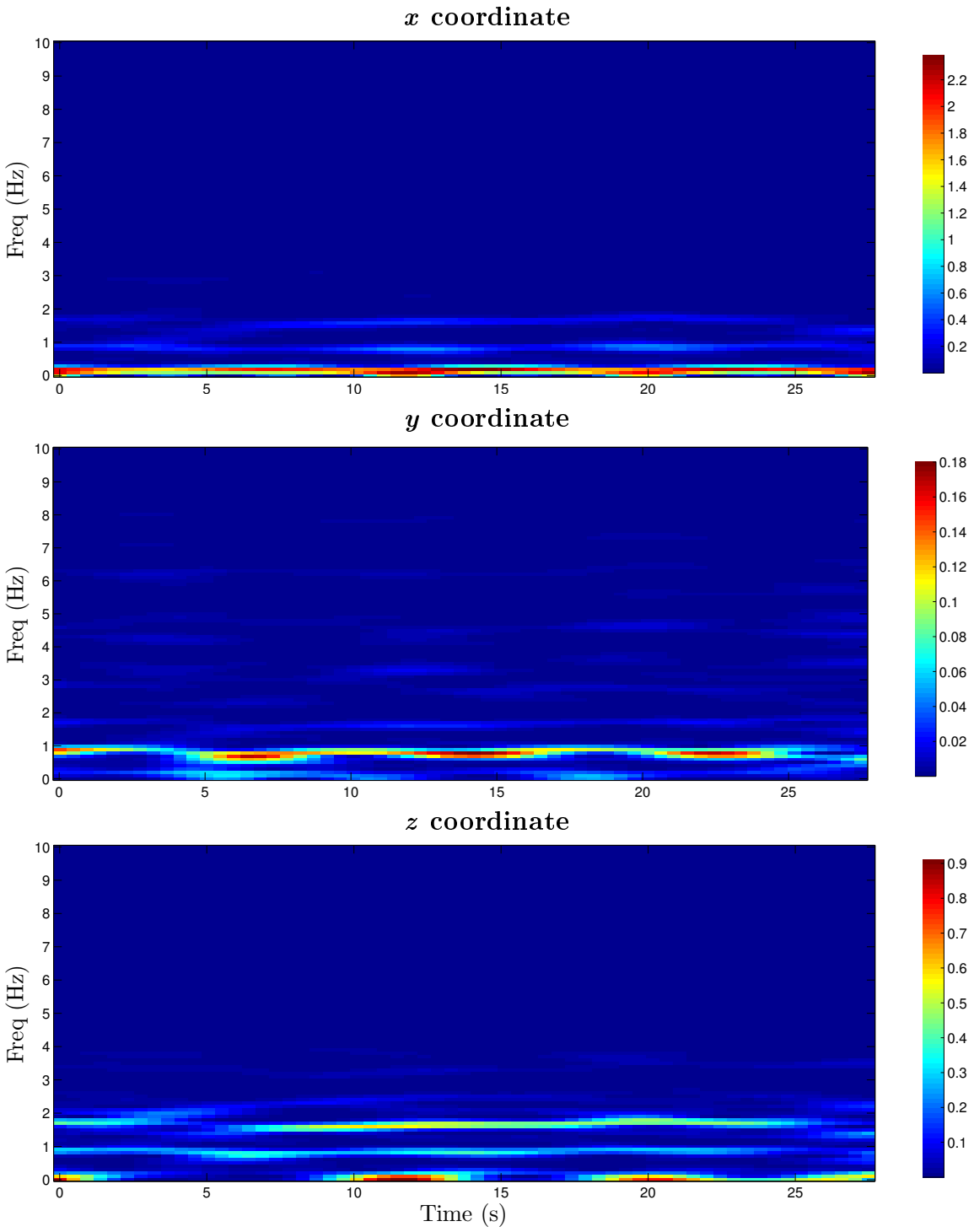


Figure III.2 : The frequency spectrum of the 3D motion recovered from the experiment on the DAVINCI sequence plotted in Figure II.19. The intensity of the frequency components can be determined by the color mapping on the right.

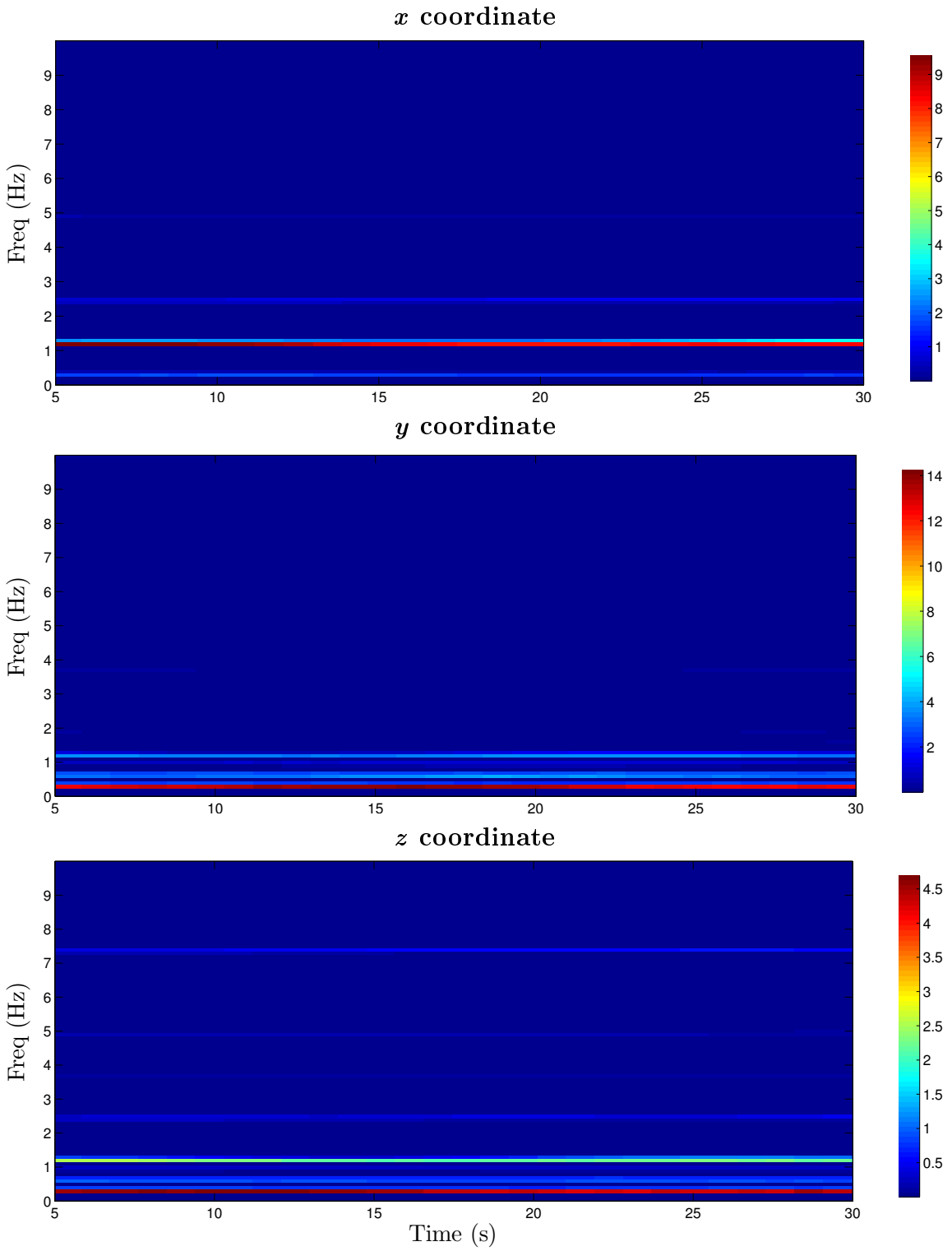


Figure III.3 : The frequency spectrum of the 3D motion recovered from the experiment on the LIRMM1 sequence plotted in Figure II.23. The intensity of the frequency components can be determined by the color mapping on the right.

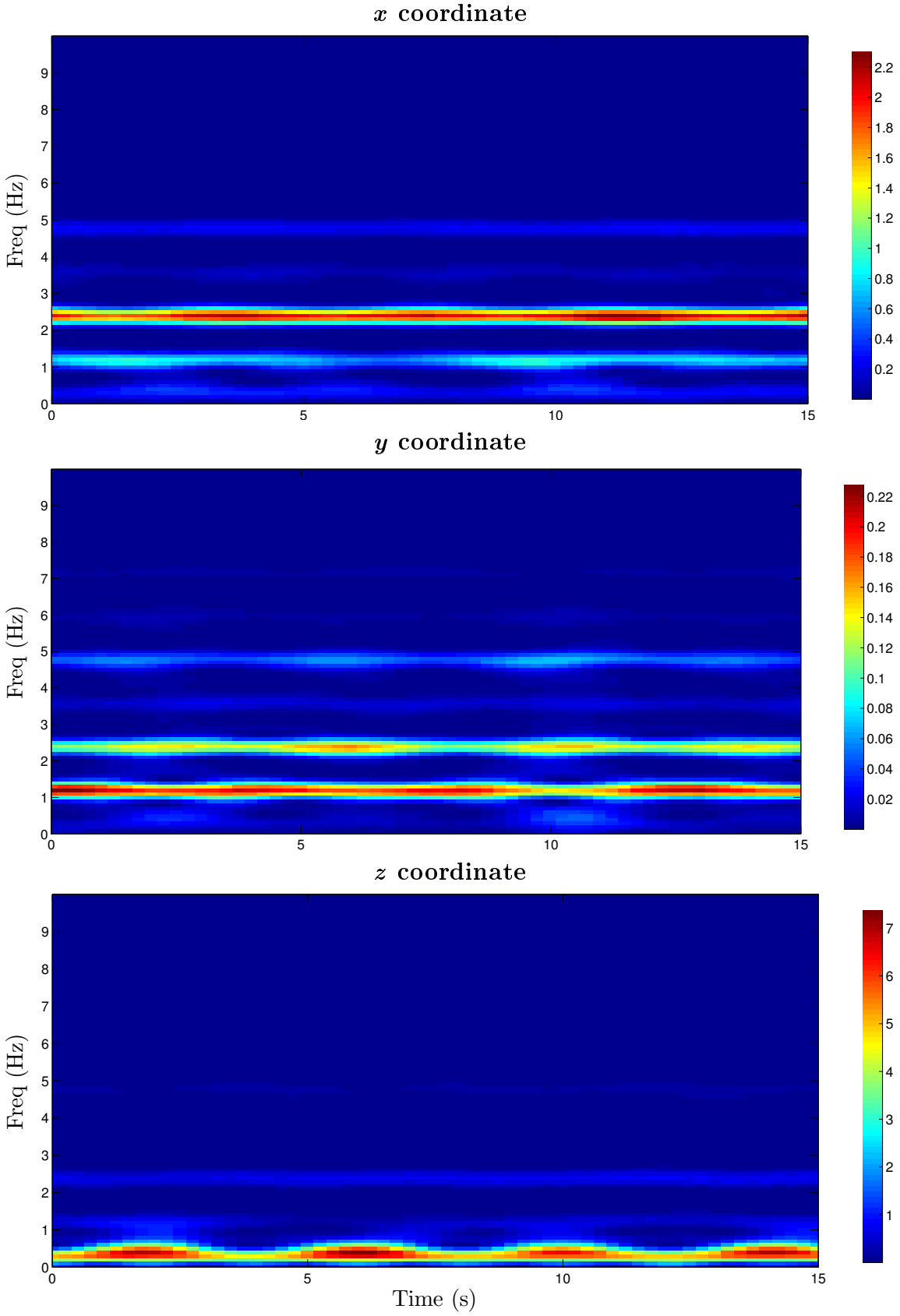


Figure III.4 : The frequency spectrum of the 3D heart motion from the LIRMM2 experimental data plotted in Figure III.1. The intensity of the frequency components can be determined by the color mapping on the right.

III.2.2 Previous works

The modeling and prediction of physiological motion have been an active topic of research, mainly in the domains of radiology [Seibert et al., 2007, Ernst et al., 2008], magnetic resonance imaging [Stegmann and Larsson, 2003] and pathological tremor characterization [Bó et al., 2009]. In the past decades, more attention has been given to the problem of modeling and predicting the beating heart motion.

In the context of robotized cardiac surgery, the model of the heart motion can be very useful at several levels of a surgical robotic assistant design. The central problem consists in modeling and predicting future positions of a point of interest (POI) or region of interest (ROI) on the heart surface. For this purpose, several paradigms have been proposed in the literature. Bebek *et al.* [Bebek and Cavusoglu, 2007a] used copies of previous heartbeat cycles synchronized with an electrocardiogram (ECG) signal for predicting the following heartbeat cycle. The experiment used recorded porcine heart motion data obtained with the sonometric sensor system described in Cavusoglu *et al.* [Cavusoglu et al., 2005]. In this method, the respiratory motion is suppressed by a low-pass filter and only the cardiac motion is considered. The prediction method was used in a robot tracking task for reducing the tracking error.

In a similar fashion, Franke *et al.* [Franke et al., 2008] employed a generalized linear predictor updated using recursive least squares for providing future position estimates of the heart position in order to reduce the robot tracking error. The linear estimator is a generalization of previous works reported in [Franke et al., 2007]. Although only the cardiac motion can be predicted with a reasonable number of parameters, the experimental results using the same experimental data acquired in Cavusoglu *et al.* [Cavusoglu et al., 2005] displayed a significant improvement in the prediction accuracy in comparison with [Bebek and Cavusoglu, 2007a], which used copies of the previous heartbeat cycle for predicting the following cycle.

In the context of visual tracking, Ortmaier *et al.* [Ortmaier et al., 2005] used embedded vectors of previous heart cycles for increasing feature tracking resilience. The prediction method is capable of bridging tracking failures and occlusions when tracking the beating heart using endoscopic images. However, results are restricted to displacements in the image plane and the method requires a large history of recorded heart motion for providing reliable estimates.

A thorough investigation of the heart motion was performed by Cuvillon [Cuvillon et al., 2006], who proposed a motion prediction algorithm based on a linear parameter variant model that is a function of the ECG and respiratory signals. The heart motion is measured in high-speed using fast cameras and artificial markers placed on the heart surface and the ECG and airflow signals are recorded. In his investigations, a coupling between the respiratory and cardiac motions was detected: the respiratory motion has a crucial influence on the volume inside the thoracic cage and therefore the heartbeat is modulated by the volume of air inside the lungs.

Other approaches have been proposed in the literature but still lack experiments on real heart motion data. In Bader *et al.* [Bader et al., 2007], a membrane model of the heart surface based on partial differential equations using position measurements from a stereo camera system is used for predicting the future heart motion. In Duindam *et al.* [Duindam and Sastry, 2007], a motion model based on a combination of several basis functions for a high fidelity representation of the heart dynamics is proposed. Although

simulations prove their feasibility for modeling quasi-periodic motion, their performance facing the complex heart dynamics still need to be evaluated.

In Ginhoux *et al.* [Ginhoux et al., 2003], the respiratory motion is filtered out from the heart motion signal (it was treated as perfectly periodic since the patient is on a breathing machine) and a Fourier series with 5 harmonics is used for modeling the cardiac motion. After the series coefficients are identified, the cardiac rhythm is assumed to be steady and the controller uses the predicted heart motion as feed-forward signal for reducing the robot tracking error.

In Thakral *et al.* [Thakral et al., 2001], a non-stationary Fourier series is used for modeling the quasi-periodic beating heart motion and its coefficients are estimated using least mean squares. In their works, the respiratory and cardiac motions that comprise the heart motion are estimated separately. The method was tested on a rat chest wall using a fiber optic displacement sensor.

More recently, Bachta *et al.* [Bachta et al., 2009] proposed a heart motion model based on three Fourier series. Two series model the respiratory and cardiac components separately and another series models the coupling between the respiratory and cardiac components. In their experiments, the cardiac and respiratory frequencies are measured using external signals (ECG and airflow signals) and the series coefficients are estimated using recursive least-squares. A comparative study shows the superior performance of the method against existing methods in the literature.

In a probabilistic framework, Yuen *et al.* [Yuen et al., 2008] used an extended Kalman filter for modeling and predicting the mitral valve annulus motion for overcoming the problems imposed by the slow frame-rate of ultrasound imaging and reduce the synchronization error between the surgical tool and the mitral valve. Although the mitral valve motion has simpler dynamics in comparison with the heart surface motion and only one degree of freedom is considered, a comparative study showed the superior estimation quality using Kalman filtering in comparison with recursive least squares and the previous heart cycle (proposed in [Bebek and Cavusoglu, 2007a]).

However, in most methods described above, the respiratory and cardiac motions that comprise the heartbeat are treated separately or only the cardiac motion is modeled. Another drawback of most strategies is the filtering of the respiratory motion, which induces delays in the system. An analysis of the techniques cited above suggests that the time-varying Fourier series is a straightforward and flexible model for describing the quasi-periodic heart motion. Next, a heart motion model based on a time-varying dual Fourier series that explicitly models the cardiac and respiratory components is described in detail.

III.2.3 Non-stationary dual Fourier series model

The heart motion can be considered as the sum of the respiratory and cardiac motions, which can be represented as a dual non-stationary Fourier series. The concept is illustrated in Figure III.5. Given the 3D coordinates $\mathbf{p} = [{}^{x}p\ {}^{y}p\ {}^{z}p]$ of a POI on the heart surface, the motion dynamics p of a Cartesian coordinate at a given instant t can be parameterized

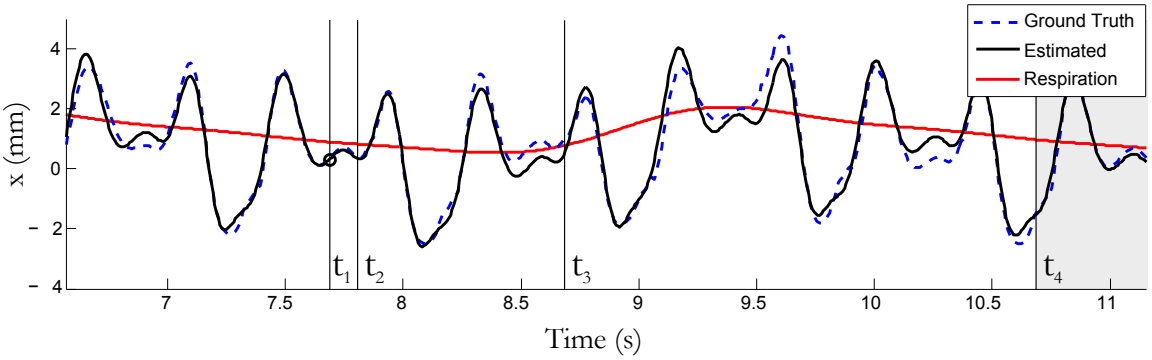


Figure III.5 : Example of the beating heart motion estimation using the dual Fourier series. The estimated Fourier series at an instant t_1 is plotted. Prediction horizons of 0.18, 1 and 3s (marked at t_2, t_3, t_4 respectively) are illustrated.

as the sum of two Fourier series, such that:

$$p(t,\mathbf{f}) = \sum_{h=1}^{H_r}\Big[a_h \sin\Big(h\sum_{k=t_0}^{t}\omega_r(k)\Big) + b_h \cos\Big(h\sum_{k=t_0}^{t}\omega_r(k)\Big)\Big] + c_r + \sum_{h=1}^{H_c}\Big[d_h \sin\Big(h\sum_{k=t_0}^{t}\omega_c(k)\Big) + e_h \cos\Big(h\sum_{k=t_0}^{t}\omega_c(k)\Big)\Big], \quad \text{(III.1)}$$

where H_r and H_c are the number of harmonics for modeling the respiratory and cardiac components respectively, ω_r and ω_c are the respiratory and cardiac frequencies, $\sum_{t_0}^{t}\omega$ is the sum of all estimated ω starting from t_0 in discrete steps of Δt and $\mathbf{f}$ is the corresponding vector containing the Fourier series parameters:

$$\mathbf{f} = [a_1, ..., a_{H_r}, b_1, ..., b_{H_r}, c_r, d_1, ..., d_{H_c}, e_1, ..., e_{H_c}]^{\mathrm{T}} \quad \text{(III.2)}$$

Consequently, a POI $\mathbf{p}$ has $\gamma = 3 \cdot [2 \cdot ({}^{x}H_r + {}^{x}H_c) + 1 + 2 \cdot ({}^{y}H_r + {}^{y}H_c) + 1 + 2 \cdot ({}^{z}H_r + {}^{z}H_c) + 1]$ parameters plus the respiratory and cardiac frequencies, which are shared among all coordinates and points. The number of harmonics H_r and H_c among the xyz directions may vary due to differences in their motion complexity. At a given instant t, the computation of future position estimates using equation (III.1) is straightforward, considering a stationary system within the prediction horizon.

III.2.4 The Extended Kalman Filter

In the proposed formulation, the Kalman Filter (KF) is employed for the recursive estimation of the Fourier series parameters. The KF offers several advantages, such as the explicit modeling of the stochastic uncertainties associated with the proposed motion model and position measurements. Since the Fourier series parameter estimation is a nonlinear problem, the extended KF (EKF) is used (see [Simon, 2006] for more details on the EKF).

The EKF state vector $\mathbf{y}$ for estimating the trajectory of φ POIs $\mathbf{p} = [{}^{x}p\ {}^{y}p\ {}^{z}p]$ is composed of $(\varphi \cdot \gamma + 2)$ parameters, such that:

$$\mathbf{y} = [{}^{1}\mathbf{f}_x, {}^{1}\mathbf{f}_y, {}^{1}\mathbf{f}_z, {}^{2}\mathbf{f}_x, {}^{2}\mathbf{f}_y, {}^{2}\mathbf{f}_z, ..., {}^{\varphi}\mathbf{f}_x, {}^{\varphi}\mathbf{f}_y, {}^{\varphi}\mathbf{f}_z, \omega_r, \omega_c]^{\mathrm{T}}; \quad \text{(III.3)}$$

where $[{}^i\mathbf{f}_x, {}^i\mathbf{f}_y, {}^i\mathbf{f}_z]$ are the parameter vectors of the i-th estimated POI $\mathbf{p}_i$.

The KF is divided between the prediction and correction phases. All parameters are modeled as random walk processes. In the prediction phase, a stationary system is assumed, which implies that the *a priori* estimate of the filter state $\hat{\mathbf{y}}^-$ at an instant t is equal to the state value from the previous instant $(t - \Delta t)$:

$$\hat{\mathbf{y}}^-(t) = \hat{\mathbf{y}}(t - \Delta t) = \begin{cases} \mathbf{f}(t) = \mathbf{f}(t - \Delta t) & \forall \mathbf{f} \\ \omega_r(t) = \omega_r(t - \Delta t) \\ \omega_c(t) = \omega_c(t - \Delta t) \end{cases} \tag{III.4}$$

In the prediction phase, the state covariance matrix $\mathbf{S}$ is propagated according to:

$$\mathbf{S}^-(t) = \mathbf{A}\mathbf{S}(t - \Delta t)\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \tag{III.5}$$

where $\mathbf{Q}$ is the process covariance matrix and $\mathbf{A}$ is the system transfer matrix. Since the filter state at an instant t is initialized with the values from $(t - \Delta t)$, $\mathbf{A}$ is defined as an identity matrix $\mathcal{I}$ of size $\varphi\gamma + 2$:

$$\mathbf{A} = \mathcal{I}_{\varphi\gamma+2} \tag{III.6}$$

In the correction phase, the available position measurements $\mathbf{h}$ are used to update the initial estimates $\hat{\mathbf{y}}^-$, yielding the *a posteriori* estimate $\hat{\mathbf{y}}^+$. The updated state is computed as:

$$\mathbf{y}^+(t) = \mathbf{y}^-(t) + \underbrace{\mathbf{S}^-(t)\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{S}^-(t)\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}}_{\mathbf{G}(t)}(\mathbf{h}(t) - p(t, \mathbf{y}^-(t))) \tag{III.7}$$

where $\mathbf{R}$ is a diagonal matrix containing the measurement error covariance values and $\mathbf{H}$ is the observation matrix. The matrix $\mathbf{G}(t)$ indicated above is called the Kalman gain matrix. In order to compute the observation matrix $\mathbf{H}$, a linearization of the Fourier series is done around the current filter estimate $\hat{\mathbf{y}}^-(t)$, such that $\mathbf{H} = \left.\frac{\partial p(t,\mathbf{y})}{\partial \mathbf{y}}\right|_{\mathbf{y}=\hat{\mathbf{y}}^-(t)}$. For example, in the first line of $\mathbf{H}_{3\varphi,\ \varphi\gamma+2}$, each column corresponds to the derivative of ${}^1p_x$ (the first coordinate of the first POI) with respect to $\mathbf{y}$:

$$\mathbf{H}_{1,\ \varphi\gamma+2} = \begin{bmatrix} \sin(\sum_{k=t_0}^{t} \hat{\omega}_r(k)) \\ \sin(2\sum_{k=t_0}^{t} \hat{\omega}_r(k)) \\ \vdots \\ \sin(H_r \sum_{k=t_0}^{t} \hat{\omega}_r(k)) \\ \cos(\sum_{k=t_0}^{t} \hat{\omega}_r(k)) \\ \cos(2\sum_{k=t_0}^{t} \hat{\omega}_r(k)) \\ \vdots \\ \cos(H_r \sum_{k=t_0}^{t} \hat{\omega}_r(k)) \\ \sin(\sum_{k=t_0}^{t} \hat{\omega}_c(k)) \\ \sin(2\sum_{k=t_0}^{t} \hat{\omega}_c(k)) \\ \vdots \\ \sin(H_r \sum_{k=t_0}^{t} \hat{\omega}_c(k)) \\ \cos(\sum_{k=t_0}^{t} \hat{\omega}_c(k)) \\ \cos(2\sum_{k=t_0}^{t} \hat{\omega}_c(k)) \\ \vdots \\ \cos(H_r \sum_{k=t_0}^{t} \hat{\omega}_c(k)) \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \sum_{h=1}^{H_r} h\ [a_h \cos(h \sum_{k=t_0}^{t} \hat{\omega}_r(k)) - b_h \sin(h \sum_{k=t_0}^{t} \hat{\omega}_r(k))] \\ \sum_{h=1}^{H_c} h\ [d_h \cos(h \sum_{k=t_0}^{t} \hat{\omega}_c(k)) - e_h \sin(h \sum_{k=t_0}^{t} \hat{\omega}_c(k))] \end{bmatrix}^{\mathrm{T}} \tag{III.8}$$

where $\sum_{k=t_0}^{t} \hat{\omega}(k)$ is the sum of all estimated $\hat{\omega}$ starting from $t = t_0$.

The updated error covariance matrix is computed as:

$$\mathbf{S}^{+}(t) = (\mathcal{I} - \mathbf{G}(t)\mathbf{H})\mathbf{S}^{-}(t) \tag{III.9}$$

A brief overview of the predictive EKF is given below. When initializing the filter, all state vector values are set to zero except the two frequencies, for which initial values are normally available in practice (from ECG and ventilation machine). Any *a priori* knowledge of the signal can be useful to aid the convergence of the filter during initialization and to better choose the number of harmonics of the Fourier series according to the heart motion complexity.

Algorithm *The EKF algorithm*

1. t ←0
2. Initialize **y** and **S**
3. **while** *system is running*
4. **do** t ←t + Δt

5. compute the *a priori* state estimate $\hat{\mathbf{y}}^-$ using equation (III.4)
6. compute the *a priori* state covariance matrix $\mathbf{S}^-$ using equation (III.5)
7. compute the Kalman gain matrix $\mathbf{G}$ from equation (III.7)
8. compute the observation matrix $\mathbf{H}$ according to equation (III.8)
9. compute the *a posteriori* state estimate $\hat{\mathbf{y}}^+$
10. compute the *a posteriori* state covariance matrix $\mathbf{S}^+$ using equation (III.9)

The only parameters that need tuning in the filter are the process covariance matrix $\mathbf{Q}$ and the measurement error covariance matrix $\mathbf{R}$. In our formulation, both matrices are diagonal and their values are empirically chosen such that more confidence is put in the model than the measurements. This ensures that the Fourier series locks on the coarse heart trajectory and accurately predicts the future heart motion for long prediction horizons even in the presence of small heart motion fluctuations.

III.2.5 Experiments

For evaluating the performance of the proposed prediction method, the tracking results from the DAVINCI and LIRMM1 sequences from section II.5.2.2 (which henceforth will be dubbed DAVINCI and LIRMM1 experimental data) and the LIRMM2 motion data introduced in section III.2.1 were used. As previously explained in section III.2.3, the heart motion is predicted using equation (III.1), considering a stationary system within the prediction horizon. This idea was previously illustrated in Figure III.5, featuring the estimated Fourier series at an instant t_1.

The experiments are divided in three groups. In the first group, the prediction errors for short and long prediction horizons are evaluated using the available experimental data. In the second group, the predictive filter is tested against simulated amplitude changes in the cardiac motion. In the third group, a disturbance in the heart motion in all Cartesian coordinates is simulated for testing the predictive filter performance for handling an abnormal heart behavior such as an arrhythmia.

The prediction error at a given time step is computed as the Euclidean distance $||\mathbf{d} - \mathbf{v}||$, where $\mathbf{d}$ and $\mathbf{v}$ are the predicted and true 3D positions of the POI, respectively. At every sample, the root mean square and the absolute peak prediction errors are measured within a given prediction horizon.

III.2.5.1 EKF parameter configuration

For all experiments, EKF state vector is initialized with zeros, with the exception of the cardiac and respiratory frequencies. The only *a priori* information about the motion signal is the choice of the number of harmonics for modeling respiratory H_r and cardiac H_c components of the heart motion. For all experiments, the respiratory component H_r was set to 3 and the cardiac component H_c was set to 5 (for the xyz directions). Therefore, the EKF state vector contains $3\gamma + 2 = 3 \cdot [2 \cdot ({}^xH_r + {}^x H_c) + 1 + 2 \cdot ({}^yH_r + {}^y H_c) + 1 + 2 \cdot ({}^zH_r + {}^z H_c) + 1] + 2 = 161$ elements in all experiments in this section.

For initializing the process covariance matrix $\mathbf{Q}$, the filter state noises are considered as independent and the diagonal elements of the matrix corresponding to the respiratory components σ_r^2, the offset σ_o^2, the cardiac components σ_c^2 and the frequencies variance $\sigma_{\omega r}^2, \sigma_{\omega c}^2$ are set using the values given in Table III.1. The matrix $\mathbf{R}$ that contains the

Table III.1 : The values of the diagonal elements of the process covariance matrix **Q**.

Parameters	DAVINCI experimental data	LIRMM1 experimental data	LIRMM2 experimental data
σ_r^2	10^{-5}	10^{-5}	10^{-7}
σ_c^2	10^{-4}	10^{-4}	10^{-6}
σ_o^2	10^{-6}	10^{-5}	10^{-7}
$\sigma_{\omega r}^2$	10^{-11}	10^{-11}	10^{-10}
$\sigma_{\omega c}^2$	10^{-10}	10^{-10}	10^{-10}

measurement error variance is a diagonal matrix with values arbitrarily set to 10^{-2} for all xyz coordinates. The values for the matrix **R** are the same for all experimental datasets.

III.2.5.2 Experiments description

Prediction errors for short and long prediction horizons
For investigating the prediction performance in time for the available experimental data, the prediction error is evaluated at every instant for 0.18, 1 and 3 second prediction horizons (which correspond to 15, 83 and 250 future steps respectively for the LIRMM1 experimental data, 25, 125 and 375 future steps respectively for the LIRMM2 experimental data and 9, 50 and 150 future steps for the DAVINCI experimental data). The three prediction horizons are chosen according to the visual tracking application: 0.18s is the average duration of disturbances caused by specular reflections and 3s is considered as the assumed duration of an occlusion caused by a surgical instrument.

For the LIRMM1, LIRMM2 and DAVINCI experimental data, the prediction errors for the 0.18, 1 and 3 second prediction horizons are plotted in Figures III.6, III.7 and III.8 respectively, and quantified in Table III.2. The identification error, which is the difference between the true position and the position estimated by the filter at a given instant is also included in the error plots for illustrating how well the motion model fits the heart motion at a given instant. For illustrating the evolution of the Fourier series coefficients, the amplitude of each harmonic of the Fourier series that model the x component of the LIRMM1 experimental data is plotted in Appendix E. The estimated cardiac and respiratory frequencies are also displayed.

Amplitude changes
The prediction filter must be robust to changes in the heart rhythm. In order to investigate the filter's behavior to amplitude changes, an artificial amplitude variation in the heart motion is simulated. The experiment consists in slowly damping the original heart motion from the LIRMM1 experimental data with respect to its mean to 50% of its original value in all Cartesian coordinates. The behavior of the filter and the evolution of the prediction errors for a prediction horizon of 1 second are analyzed. Results are plotted in Figure III.9.

Simulated arrhythmia
Cardiac arrhythmia may also cause irregular heart motion. For studying the EKF performance under such events, a disturbance in the heart motion signal in all Cartesian

coordinates is simulated as illustrated in Figure III.10. Again, the LIRMM1 experimental data is used for the experiment. The effects on the prediction error for a 1 second prediction horizon is plotted in Figure III.10.

III.2.5.3 Discussion on the predictive filter performance

Prediction errors for short and long prediction horizons
In general, the prediction errors presented in Figures III.6, III.7 and III.8 indicate a very low RMS error value and maximum peak errors for all experiments under 20% of the respective signal amplitudes. These experimental results attest the good performance of the predictive filter, which successfully captures the true heart dynamics. This is reinforced by the prediction error values given in Table III.2, which indicate that the RMS and peak errors do not increase significantly when the prediction horizon is extended, suggesting the proposed motion model successfully approximates the heart motion.

Comparing the prediction quality among the experiments with the three datasets, the prediction quality using the DAVINCI experimental data displayed the poorest results (for instance, even though the prediction errors are similar, the amplitude of the DAVINCI experimental data, which is acquired from a mechanically stabilized heart is much smaller in comparison with the amplitude of the LIRMM1 data). The poor results can be explained by the large frequency variations detected in the frequency analysis (see Figure III.2). Since the cardiac frequency is not stable during the sequence, the prediction quality degrades quickly with the increase of the prediction horizon.

Amplitude changes
The capacity of the filter to adapt to amplitude variations was properly displayed by the extreme case presented in Figure III.9. From the error plots, one can verify that the filter takes approximately 1 second to re-adapt to the signal changes, later stabilizing in a lower level since the signal amplitude was reduced by 50%.

Simulated arrhythmia
Next, the simulated arrhythmic heart behavior presented in Figure III.10 is the most challenging event for the prediction filter. Although the error plots show a considerable large error during the event, the filter successfully copes with the disturbance, converging quickly past the event. It is also important to remark that although the simulated arrhythmia generates a major disturbance in the predictive filter, such abnormal heart behaviors can be easily detected from the heart electric activity and therefore in practice they do not represent a critical problem.

General issues
The relatively high identification error verified in the experiments with both sequences indicates that the heart motion model only describes the coarse heart trajectory. This represents a limitation of the Fourier series model, which is insufficient to represent the full complexity of the heart activity based only on the heart motion history. Although the identification error could be reduced by decreasing the values of the measure error matrix $\mathbf{R}$, instabilities in the estimation of the Fourier series parameters could occur and the quality of the future heart motion prediction for larger prediction horizons is reduced.

The increase of the number of Harmonics of the Fourier series does not significantly increase the prediction quality. Therefore, their number is kept small to avoid possible

Table III.2 : Predictive filter performance on *in vivo* data

DAVINCI experimental data

Horizon	**average RMS error**(mm)	**average peak error** (mm)
0.18 s	0.73	1.00
1 s	0.86	1.55
3 s	1.00	2.03

LIRMM1 experimental data

Horizon	**average RMS error**(mm)	**average peak error** (mm)
0.18 s	0.80	1.18
1 s	0.87	1.67
3 s	1.02	2.04

LIRMM2 experimental data

Horizon	**average RMS error**(mm)	**average peak error** (mm)
0.18 s	0.72	1.21
1 s	0.83	1.91
3 s	0.80	2.36

instabilities in their estimation and an unnecessary computational burden.

Another aspect of Kalman filtering is the adequate choice for the filter covariance matrices. Although the values proposed in section III.2.4 were not finely tuned for the experimental data, an adaptive update of the filter uncertainty parameters could better estimate the heart motion, hence producing better future estimates.

The performance increase obtained with a Kalman filtering framework has been analyzed in previous works [Yuen et al., 2008]. However, direct comparisons with experimental results of similar techniques reported in the literature is not possible since the used experimental data highly influences the prediction errors, as clearly demonstrated when analyzing the amplitude changes in the input signal shown in the experimental section. In fact, the prediction error is dependent on the amplitude of the heart motion since the RMS error measure is not linear and the prediction errors are proportional to the motion amplitude.

The heart motion prediction can be further improved using the proposed formulation by integrating the ECG and respiratory signals directly in the filter design, as suggested in section III.2.4. For instance, the heart electric activity precedes the mechanical motion and the ECG waves may help predict more accurately the heartbeat contraction and relaxation cycles. The EKF framework adopted provides an elegant framework for fusing different sources of information in a straightforward fashion. Furthermore, the ECG can also be used for detecting abnormal cardiac behavior (e.g. arrhythmia).

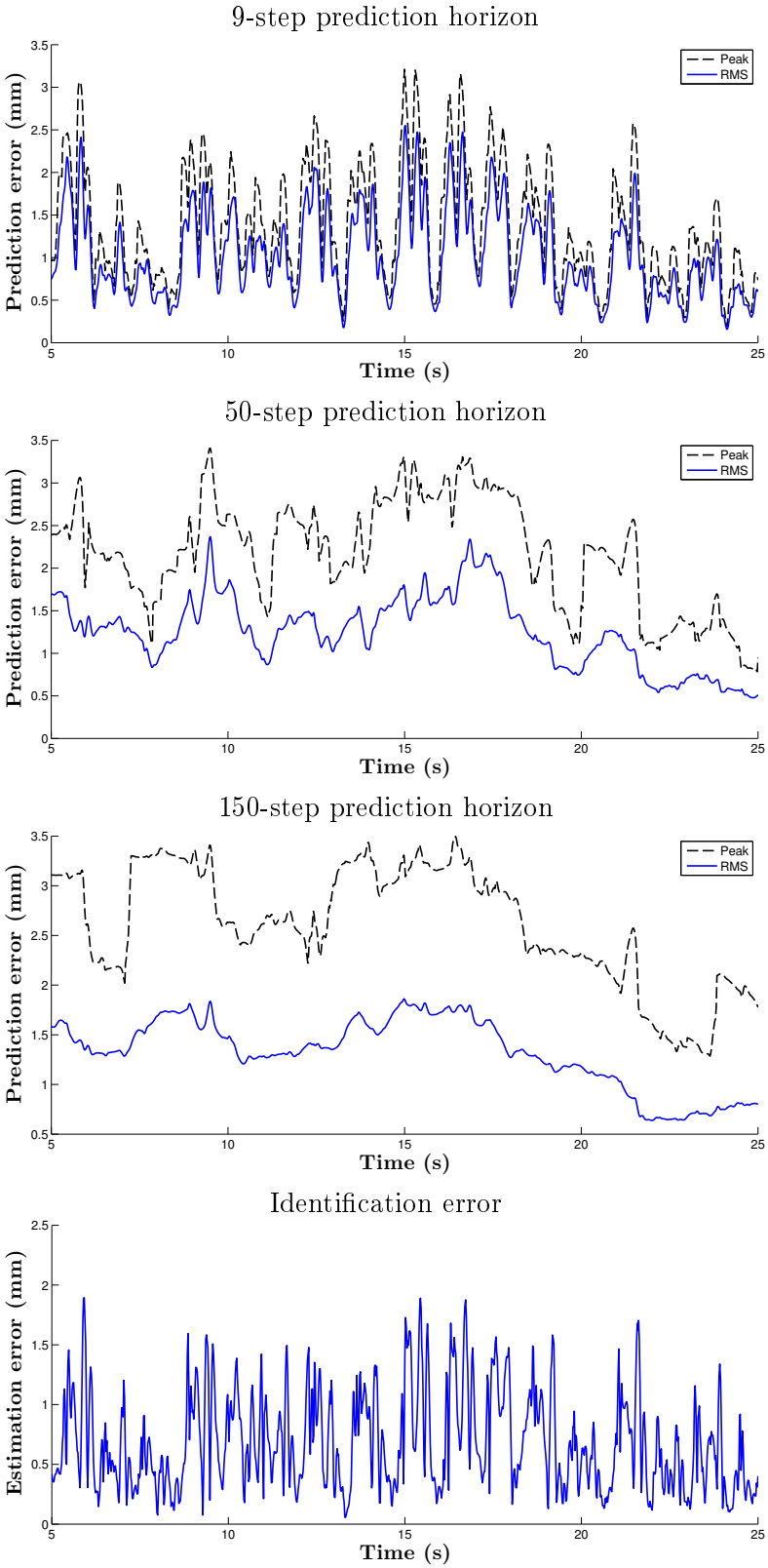


Figure III.6 : Prediction errors using the DAVINCI experimental data - From top to bottom, the RMS and peak prediction errors for a 9-step, 50-step, 150-step prediction horizon and the identification error, respectively.

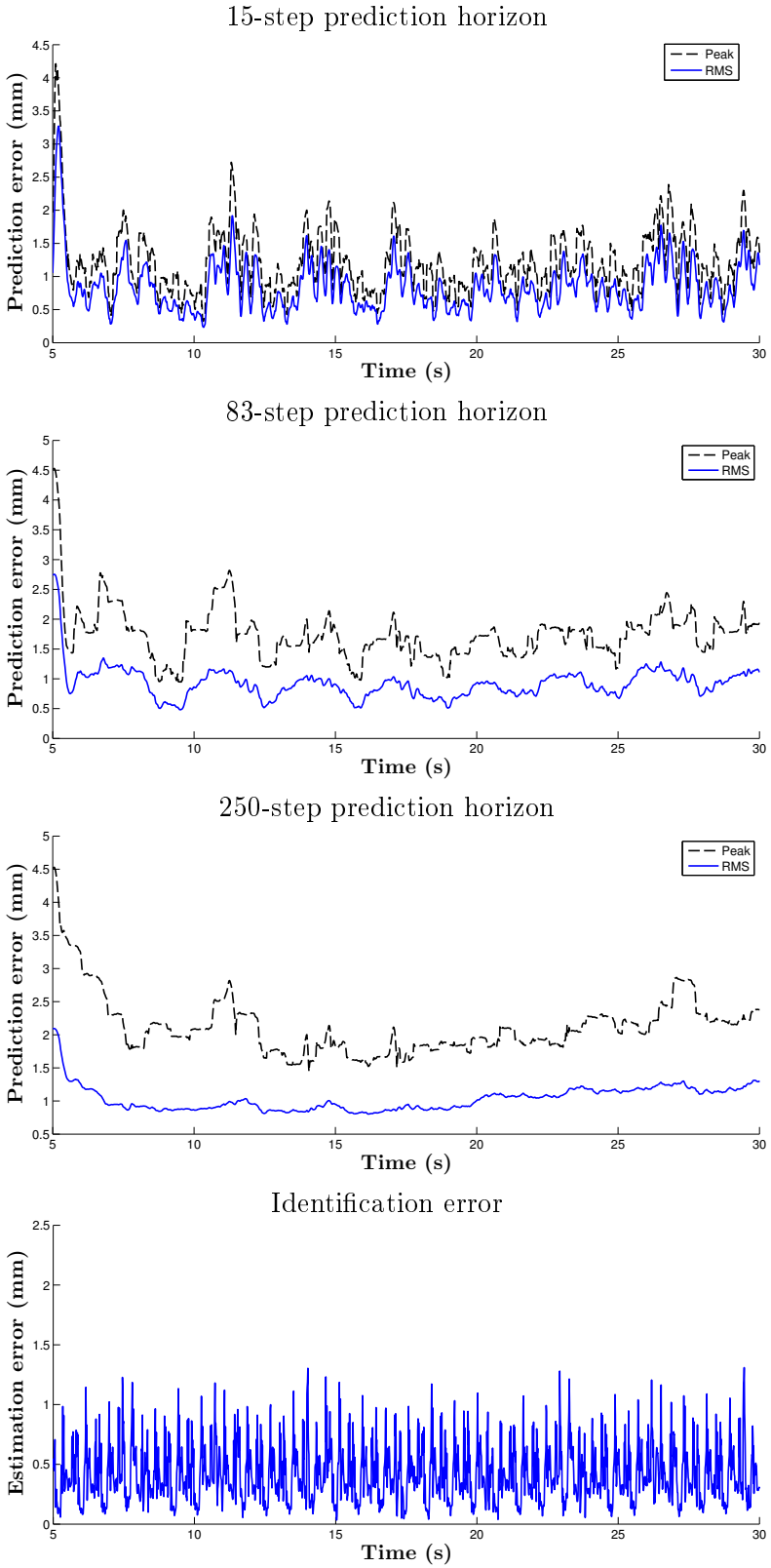


Figure III.7 : Prediction errors using the LIRMM1 experimental data - From top to bottom, the RMS and peak prediction errors for a 15-step, 83-step, 250-step prediction horizon and the identification error, respectively.

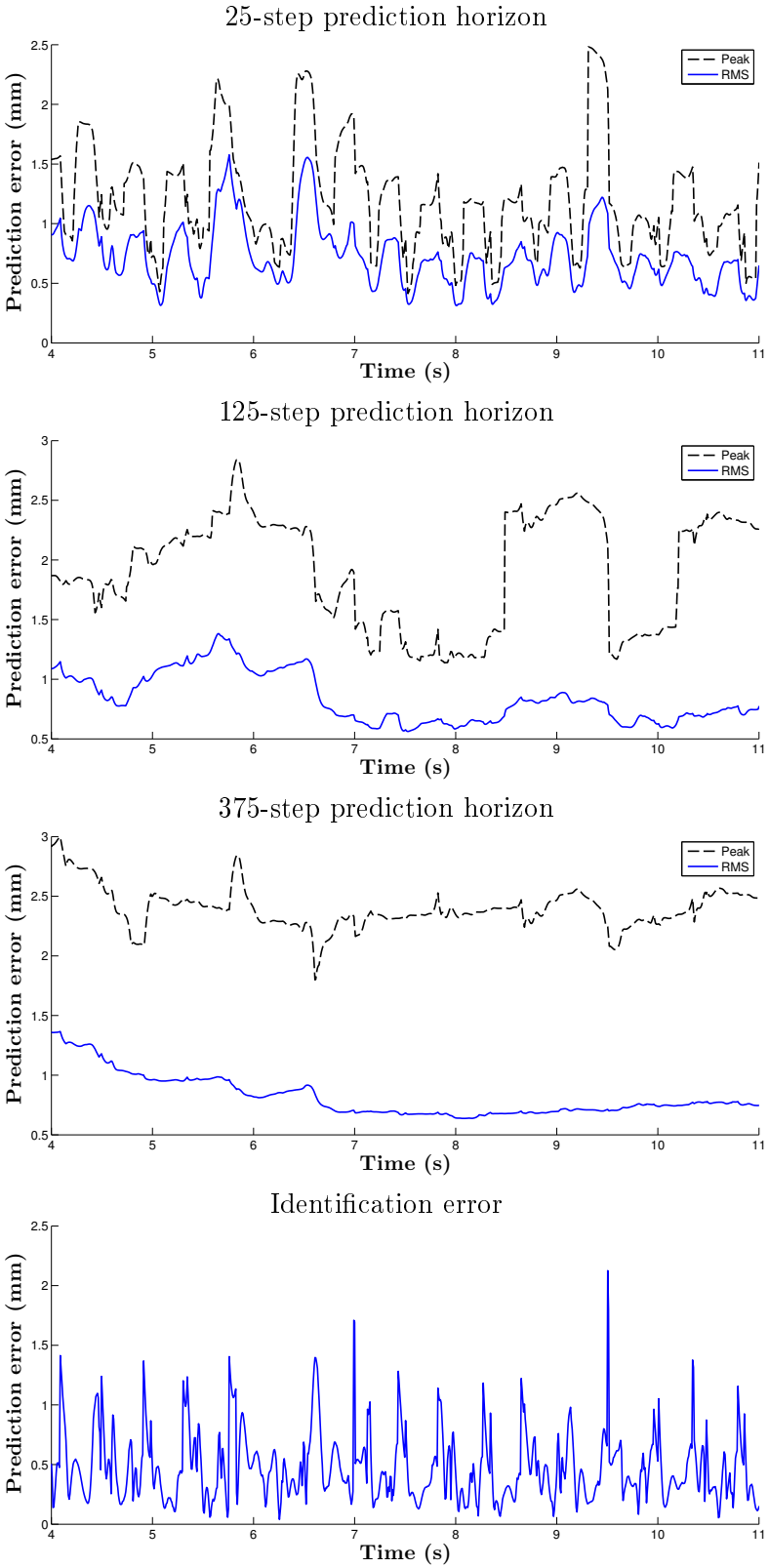


Figure III.8 : Prediction errors using the LIRMM2 experimental data - From top to bottom, the RMS and peak prediction errors for a 25-step, 125-step, 375-step prediction horizon and the identification error, respectively.

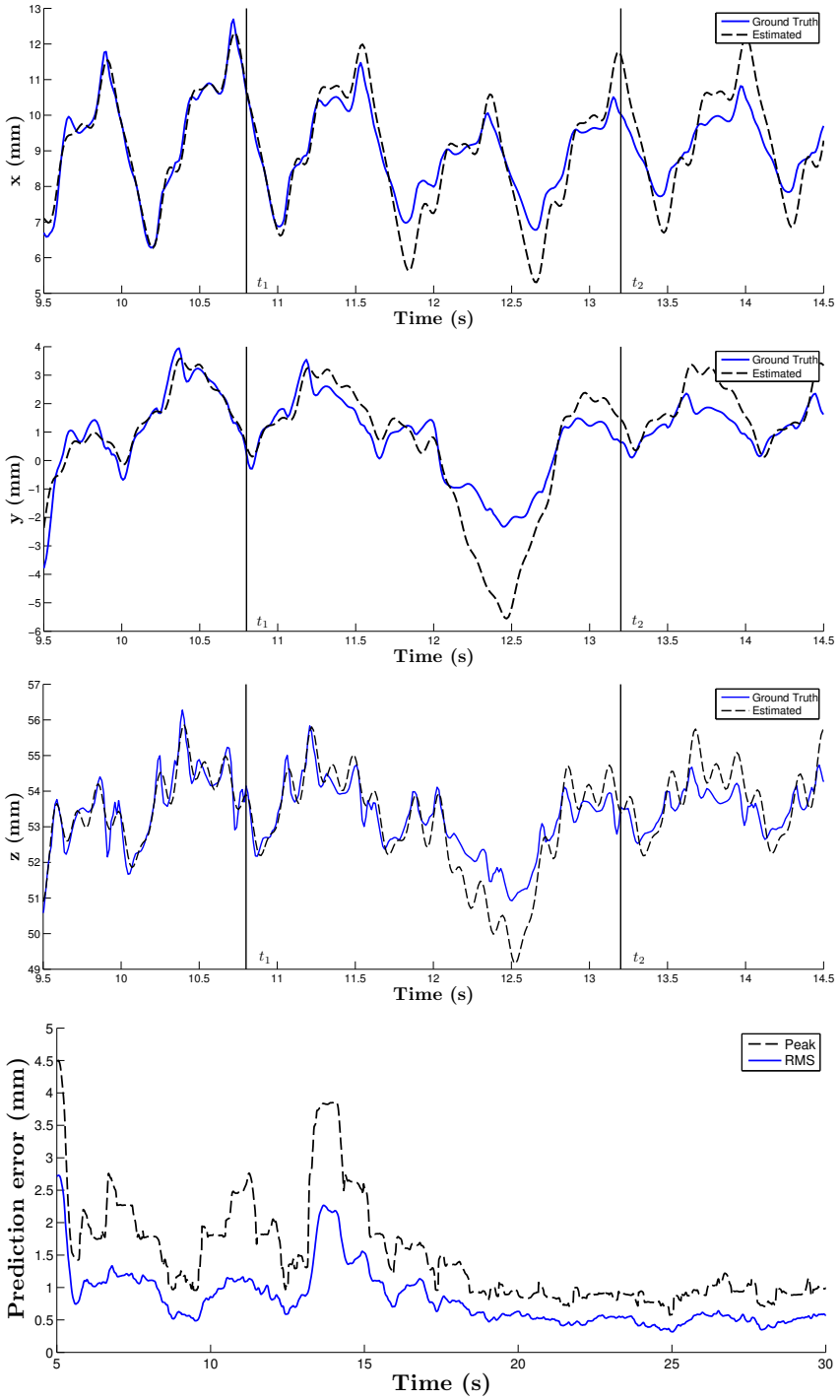


Figure III.9 : (Top) The original heart motion is slowly damped from $t_1 = 10.8s$ to $t_2 = 13.2s$ to 50% of its original amplitude for all Cartesian coordinates. (Bottom) The prediction error plot for a 1 second prediction horizon indicates the fast filter adaptation to the amplitude change.

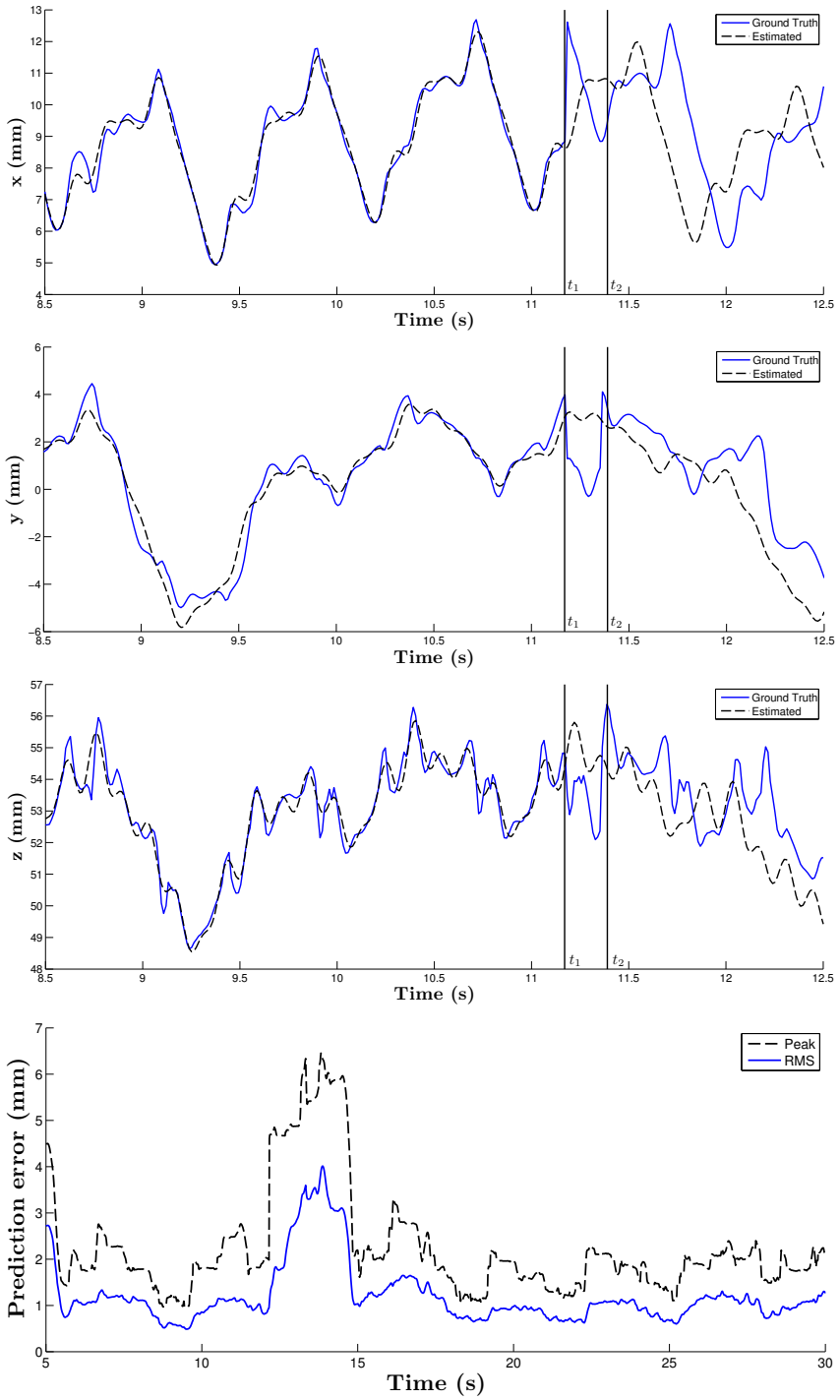


Figure III.10 : A signal disturbance for simulating an arrhythmia is induced from $t_1 = 11.17s$ to $t_2 = 11.34s$ in all Cartesian coordinates of the LIRMM1 experimental data. The estimated Fourier series at t_1 is plotted. On the bottom, the prediction error plot for a 1 second horizon indicate high prediction errors during the disturbance. Nevertheless, the prediction quality is quickly reestablished after the disturbance.

III.2.5.4 Comparative Study

In this section, the proposed prediction method is compared with similar techniques in the literature. Only methods capable of real-time operation are considered. In addition, only the heart motion history is used for estimating the future heart motion (techniques such as [Cuvillon et al., 2006] which use external signals like the ECG and ventilator airflow are not considered in this study).

For the experiments, the LIRMM1 experimental data is used. This experimental data is chosen since it does not display significant frequency variations and facilitates the tuning of the different methods selected for this comparative study. Although in the previous sections all Cartesian coordinates of the POI predicted position are estimated simultaneously, for the comparative study the prediction results are evaluated separately, allowing for a more independent analysis of the prediction results.

III.2.5.4.1 Methods chosen for the study

In this study, the following methods are chosen for the comparison:

1. The method proposed by Ortmaier *et al.* [Ortmaier et al., 2005].
2. The amplitude modulation method proposed in Bachta *et al.* [Bachta et al., 2009].
3. The two-stage prediction method proposed by Thakral *et al.* [Thakral et al., 2001].

The AR model proposed in Franke *et al.* [Franke et al., 2008] is kept outside the scope of this study since only the cardiac motion can be modeled with a reasonable number of parameters.

III.2.5.4.2 Tuning

Each method is tuned based on the parameters proposed by the authors in their respective articles and adjusted manually for an optimized performance. The parameters used in each method are plotted in Table III.3. For each method, the parameters that require tuning are summarized below:

- In the prediction method proposed by Ortmaier *et al.*, four parameters have to be defined. First, the length of the heart motion history o_p from which the embedded vectors of previous heart cycles are estimated. Second, the temporal distance between elements of the embedded vector o_h, which controls the prediction accuracy. Third, the length of the embedded vector o_l (o_h and o_l should be as long as at least one respiratory cycle). Finally, the number of embedding vectors o_n used for computing the average future position estimate controls the robustness with respect to noise.

- The amplitude modulation method proposed by Bachta *et al.* is composed of the dual Fourier model described above and an additional term for modeling the coupling between respiratory and cardiac motions, which is as a product of two Fourier series. Therefore, only the number of harmonics of each Fourier series that compose the model has to be defined.

Table III.3 : The parameter values for each experimental dataset.

Method	**Parameters**	**LIRMM1 experimental data**
Ortmaier	o_p	1000
	o_h	2
	o_l	100
	o_n	3
Bachta	H_r	3
	H_c	6
	H'_r (coupling)	2
	H'_c (coupling)	4
Thakral	H_r	3
	H_c	6
Richa	H_r	3
	H_c	6

- The two-stage method proposed by Thakral *et al.* consists of estimating the Fourier series corresponding to the respiratory and cardiac motions separately. Therefore, the respiratory motion and frequency are estimated in an early step and the motion residue is modeled by another Fourier series centered on the cardiac frequency. Since its formulation is similar to the dual Fourier series, the number of harmonics that compose the model is defined as in the previous experiments in this section.

For identifying the parameters of models based on the Fourier series, the EKF framework described in section III.2.4 is used. The parameters of the EKF used in the comparative study are included in Table III.4.

III.2.5.4.3 Results

For each technique, the accuracy of the predicted heart position, 1 second in the future, is computed (83 future steps for the LIRMM1 experimental data). The first 8 seconds from the beginning of each sequence are used for the initialization of the prediction methods.

Figure III.12 displays the predicted heart motion results. The mean and absolute peak error values are illustrated in Figure III.11 and quantified in Table III.5 (the smallest prediction values for each Cartesian coordinate are highlighted). The lowest prediction errors were obtained with the amplitude modulation method proposed by Bachta *et al.* but the proposed dual Fourier series has shown comparable performance. The method proposed by Ortmaier *et al.* produced low prediction results for the x Cartesian coordinate but in general, the technique displayed the highest peak errors (an increase in the parameter o_n did not reduce the peak errors). Finally, the two-stage prediction method proposed by Thakral *et al.* did not successfully lock to the heart motion. Consequently, the technique displays the worse prediction performance among the methods chosen in this study.

Table III.4 : The Extended Kalman Filter setup for the comparative study.

Method	**Parameters**	**LIRMM1 experimental data**
Bachta	σ_r^2	10^{-6}
	σ_c^2	10^{-5}
	σ_o^2	10^{-7}
	σ_{cr}^2 (coupling)	10^{-10}
	σ_{cc}^2 (coupling)	10^{-10}
	$\sigma_{\omega r}^2$	10^{-11}
	$\sigma_{\omega c}^2$	10^{-11}
Thakral	σ_r^2	10^{-7}
	σ_c^2	10^{-7}
	σ_o^2	10^{-8}
	$\sigma_{\omega r}^2$	10^{-10}
	$\sigma_{\omega c}^2$	10^{-10}
Richa	σ_r^2	10^{-6}
	σ_c^2	10^{-7}
	σ_o^2	10^{-5}
	$\sigma_{\omega r}^2$	10^{-10}
	$\sigma_{\omega c}^2$	10^{-10}

Table III.5 : The prediction errors for each method.

Method	**Coord.**	**LIRMM1 experimental data**	
		Mean	Peak
Bachta	x	0.36	1.52
	y	**0.53**	**2.36**
	z	0.36	2.04
Ortmaier	x	**0.29**	**1.37**
	y	0.66	5.13
	z	0.41	2.85
Thakral	x	0.77	2.62
	y	0.87	4.68
	z	0.54	2.73
Richa	x	0.40	**1.37**
	y	0.72	3.30
	z	**0.32**	**1.65**

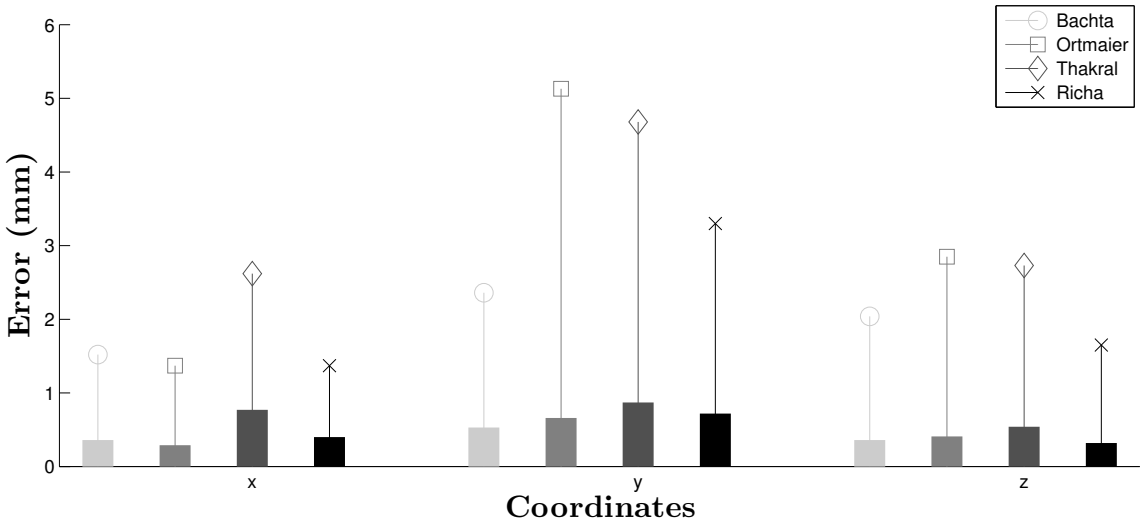


Figure III.11 : The mean and peak prediction error values for each method analyzed in the comparative study for all Cartesian coordinates of the LIRMM1 experimental data.

III.2.5.4.4 Discussion

Although the coupling between the respiratory and cardiac motions could not be detected in the experimental data, when present, this coupling has a significant influence on the overall heart motion. In this case, the amplitude modulation method may produce superior results compared to the dual Fourier series since the coupling is considered in the model. Nevertheless, the amplitude modulation method requires a much higher number of parameters in comparison with the dual Fourier series (twice the number of parameters in comparison with the dual Fourier series model, as shown in Table III.3). Consequently, additional experiments are required for evaluating how the additional parameters affect the prediction performance and the robustness against frequency and amplitude changes.

As indicated by the error plots, the prediction method proposed by Ortmaier *et al.* produces prediction results comparable to the Fourier based techniques, even though it displayed the highest peak prediction errors. However, the major drawback of this approach is the poor performance facing frequency variations, once changes in the cardiac frequency can only be handled if the motion history (normally limited to 10 seconds due to computation limitations) contains similar motion within the frequency range of the variations.

Contrary to the results reported in [Bachta et al., 2009], the two-stage motion estimation and prediction display significantly worse results than in the case when the parameters of the motion model are estimated simultaneously, as suggested by the high prediction errors in Table III.5. This highlights a very attractive feature of the EKF identification framework, which allows for all model parameters to be estimated simultaneously, unlike the approach using least mean squares described in [Thakral et al., 2001]. Although the experiments reported in [Yuen et al., 2008] suggest a superior performance using the KF framework in comparison with the least mean squares, further experiments focusing on the parameters identification are required.

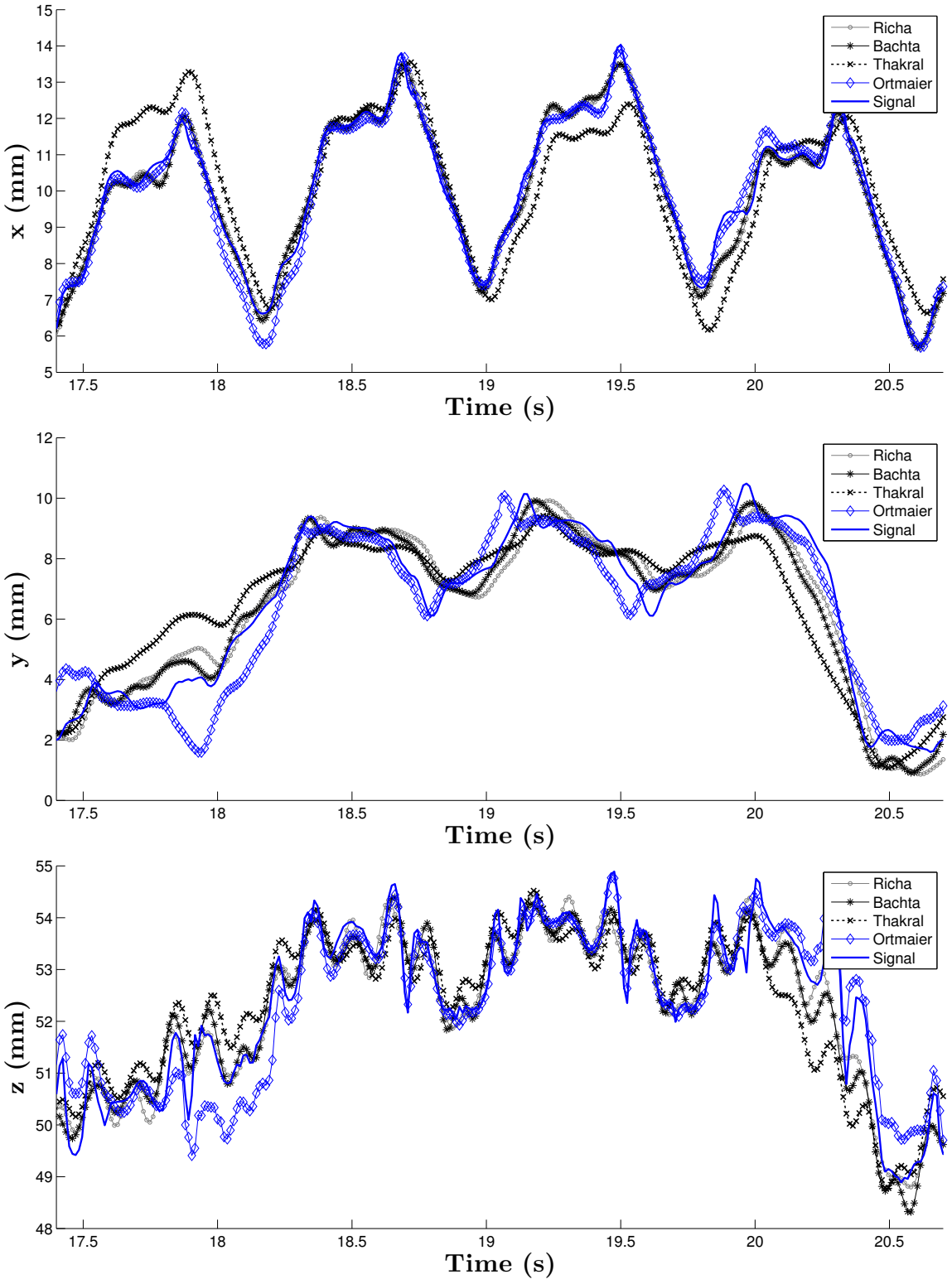


Figure III.12 : The predicted heart motion, 1 second in the future, for all Cartesian coordinates of the LIRMM1 experimental data.

III.3 Robust Visual Tracking

Similarly to the majority of visual tracking methods proposed in the literature, the tracking method formulated in the previous chapter uses the immediate previous tracking parameters to initialize the minimization procedure at every frame. Consequently, tracking is not robust to disturbances caused by large specular reflections or occlusions by surgical instruments. Therefore, for coping with these problems, the prediction framework is incorporated to the tracking method described in the previous chapter. The objective is to use the heart motion estimated by the predictive EKF for bridging periods of disturbances and occlusions, where no tracking information is available. The integration of the predictive EKF and the tracking method creates a unified framework for estimating the temporal motion and spatial deformation of the heart surface for improved visual tracking robustness.

An important remark is that although the predicted motion is accurate enough to restart tracking, it cannot be used for motion compensation since the prediction error is too large for the required specifications of cardiac surgery. Hence, prediction is used in the context of visual motion compensation for the only purpose of bridging occlusions and tracking failures. Therefore, when synchronizing the robotic tools with the heart motion, if the visual tracking results are not reliable, the robot retreats to a safe position and waits for the tracking quality to be normalized.

III.3.1 A unified heart motion estimation framework

A diagrammatic overview of the unified tracking framework is illustrated in Figure III.13. The idea is to estimate the temporal motion of the 3D points in space that parameterizes the tracked ROI on the heart surface (see the concept of the TPS parameterization illustrated in Figure II.8) by the predictive EKF in parallel with the visual tracking task. The Cartesian coordinates of the 3D points stacked in the TPS parameter vector $\mathbf{h}$ (equation II.15) are considered as different POIs in the prediction method from section III.2.4.

At every frame, the POI positions are updated based on their corresponding Fourier series model (prediction step). Images from the stereo cameras are acquired and the current heart deformation is evaluated by the tracking algorithm presented in section II.5.2.2. Next, the quality of the tracking results is evaluated based on analysis of the image alignment error and the estimated 3D heart shape (section III.3.2). If the tracking quality is sufficiently high, the KF correction step uses the parameter vector $\mathbf{h}$ estimated by the tracking algorithm as a measure of the current shape of the tracked ROI and the Fourier series parameters $[\mathbf{f}_x, \mathbf{f}_y, \mathbf{f}_z]$ of each POI that parameterize the tracked TPS surface are updated.

In the event of a tracking failure or occlusion (either by specular reflections or surgical tools), the estimated 3D heart shape may not be reliable and since no measurements are available the EKF correction step is not executed. In this case, the predicted heart motion using the estimated Fourier series parameters at the moment when tracking was suspended is used for restarting tracking in the subsequent frames.

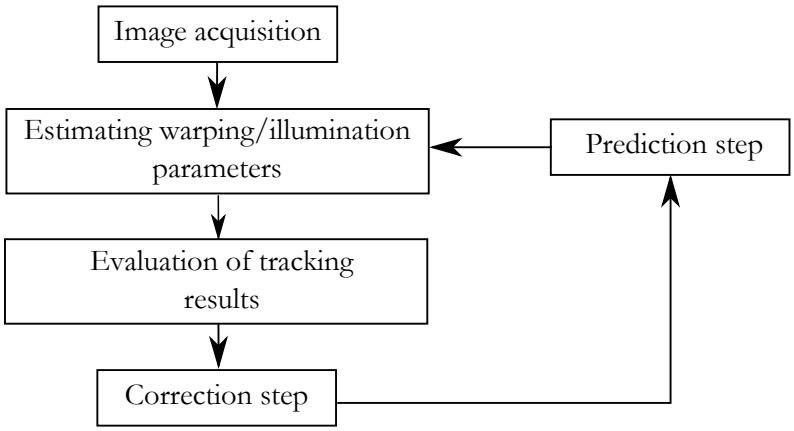


Figure III.13 : A diagrammatic overview of the unified tracking framework.

III.3.2 Evaluating tracking quality

Due to perturbations of various sources (specular reflections, occlusions, motion blur, etc.) and tracking convergence problems (*e.g.* due to local minima), tracking quality may vary drastically in time. Therefore, for detecting tracking problems, some quality measure from the visual tracking step is necessary. In this study, two criteria are used for determining if the output of the visual tracking step is reliable at every frame:

- The alignment error – The tracking quality is fundamentally a function of the alignment error between the reference image $\mathbf{T}$ and the warped left and right images $\mathbf{I}_l(w(\mathbf{x}, \mathbf{h}, \mathbf{C}_l))$ and $\mathbf{I}_r(w(\mathbf{x}, \mathbf{h}, \mathbf{C}_r))$.
- The estimated heart shape – Eventually, the alignment error alone is not able to efficiently detect convergence problems due to local minima. Therefore, an analysis of the estimated heart surface shape is performed for detecting inconsistent tracking results.

III.3.2.1 The alignment error

Tracking quality is considered as a function of the alignment error between the reference image $\mathbf{T}$ and the warped left and right images $\mathbf{I}_l(w(\mathbf{x}, \mathbf{h}))$ and $\mathbf{I}_r(w(\mathbf{x}, \mathbf{h}))$. For measuring the alignment error between two images, 3 similarity measures were evaluated in this study: the normalized cross correlation (NCC), the metric based on the Bhattacharyya coefficient used in Bourger *et al.* [Bourger et al., 2007] and mutual information (MI).

A detailed description of the similarity functions above can be found in appendix A. For choosing the most convenient error measure, the alignment errors during the tracking experiments from the previous chapter (section II.5.2.2) were measured. From the analysis of the error plots for both LIRMM1 and DAVINCI tracking experiments, the NCC proved to be the most discriminant measure of alignment error, since it displays a smaller error output variability when tracking different image sequences and invariability to illumination shifts. Therefore, the NCC is adopted in this study as a similarity measure for evaluating the alignment error.

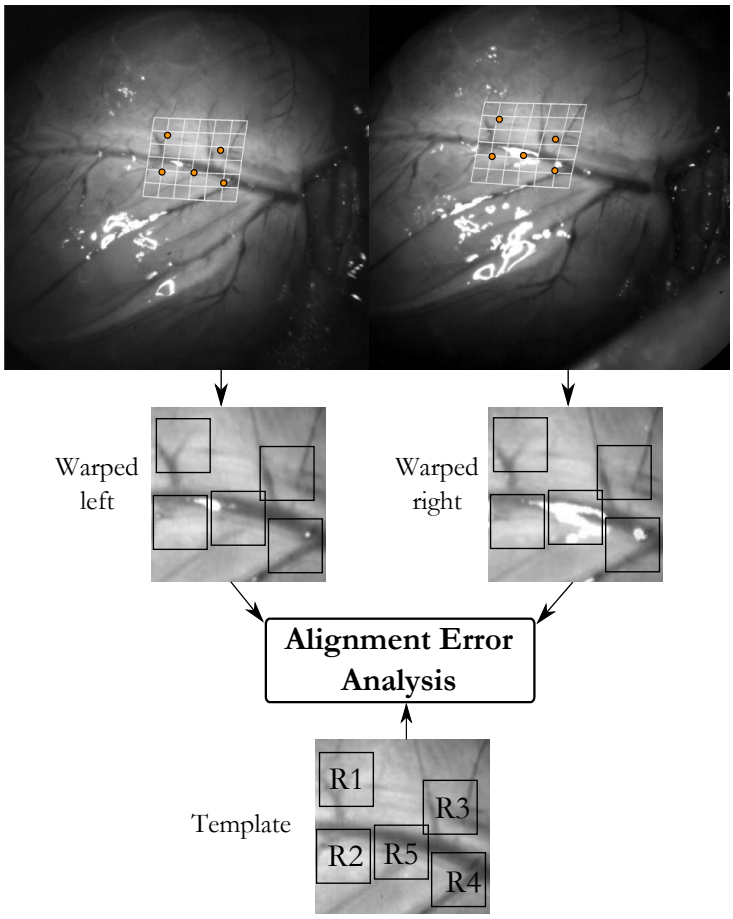


Figure III.14 : For evaluating the tracking quality at a given frame, the alignment error between the warped images and the reference image is measured for several subsets of the image (R1 to R5), corresponding to the support regions of the TPS control points (marked by dots on the left and right stereo images on the top).

Since the TPS model is a deformable model, the alignment error over the reference image $\mathbf{T}$ is not uniform (certain control points may be better estimated than others). Consequently, we perform a separate analysis for each TPS control point that models the heart deformation. This is done by measuring the NCC coefficients e_l and e_r of a 40x40 pixel region centered around each control point on the left and right stereo images respectively and the reference subregion on the reference image $\mathbf{T}$. Although this method only provides a coarse approximation of the alignment error, it is an attractive solution since little computational effort is required (see illustration in Figure III.14).

In case the tracked region on the heart surface is affected by large specular reflections or if surgical instruments occlude the operating site, tracking quality decreases significantly. When an occlusion occurs, a large number of control points may be poorly estimated. For this purpose, we detect if the smallest NCC coefficients e (where $e = min(e_l, e_r)$ is the smallest NCC coefficient between e_l and e_r of a given region) of more than three regions drop below a defined threshold τ. Although this criterion is defined empirically, it translates well the periods when the available visual information is insufficient for providing reliable tracking. In the event of an occlusion, visual tracking is suspended

completely and it is only restarted when all control points become visible again.

III.3.2.2 Shape analysis

Although the alignment error is straightly related to the tracking quality, the method described earlier cannot efficiently detect convergence problems in the ESM. This issue is illustrated in Figure III.15, where the tracking results of a given frame of the tracking experiment with the LIRMM1 sequence feature an erroneous estimate of the heart shape due to tracking convergence problems. Notice that according to the occlusion detection criteria, the NCC coefficients are not sufficiently low for the ROI to be considered occluded.

In this context, the estimated heart surface shape can serve as a good indicator of the tracking quality and the bending energy of the estimated TPS surface can be used to detect inconsistent tracking results. The TPS bending energy for one coordinate axis can be computed using equation (II.1) from the previous chapter. As demonstrated in Bookstein *et al.* [Bookstein, 1989] a value proportional to the bending energy, called l, can be used in order to simplify the calculations. The value l of a TPS surface $\mathbf{P}$ as in equation (II.13) can be computed as the sum of the bending energies of all Cartesian coordinates:

$$l = ({}^{x}\mathbf{t}_*)^{\mathrm{T}}\mathbf{L}({}^{x}\mathbf{t}_*) + ({}^{y}\mathbf{t}_*)^{\mathrm{T}}\mathbf{L}({}^{y}\mathbf{t}_*) + ({}^{z}\mathbf{t}_*)^{\mathrm{T}}\mathbf{L}({}^{z}\mathbf{t}_*) \tag{III.10}$$

where $\mathbf{t}_*$ is the TPS parameter vector $\mathbf{t}$ of a given Cartesian coordinate axis containing only the weights $(w_1, ..., w_n)$ of the corresponding TPS function and $\mathbf{L}$ is the matrix defined in equation (II.5). Notice that the TPS parameter vectors $[{}^{x}\mathbf{t}\ {}^{y}\mathbf{t}\ {}^{z}\mathbf{t}]$ from which $[{}^{x}\mathbf{t}_*\ {}^{y}\mathbf{t}_*\ {}^{z}\mathbf{t}_*]$ can be computed from the 3D coordinates $\mathbf{X}$ stacked in the parameter vector $\mathbf{h}$ as:

$$\begin{bmatrix} {}^{x}\mathbf{t} & {}^{y}\mathbf{t} & {}^{z}\mathbf{t} \end{bmatrix} = \mathbf{K}_* \begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \mathbf{C}^{\mathrm{T}} \tag{III.11}$$

where $\mathbf{C}$ is the camera matrix. Although the left camera was previously defined as the world coordinate frame, the analysis is invariant to the view choice.

Plotting the value l for the estimated TPS surface in the tracking experiment with the LIRMM1 image sequence, one can notice that the spikes in the bending energy plot can be used to detect tracking errors (see Figure III.16). For example, the largest peak in Figure III.16 corresponds to the tracking result previously illustrated in Figure III.15. In general, the bending energy does not vary significantly between heart cycles and therefore an inconsistent tracking result can be detected by thresholding: if the bending energy value l exceeds a threshold ϵ, tracking is suspended and the heart motion predicted by the EKF is used for restarting tracking.

III.3.3 Experimental results

For illustrating the performance of the unified tracking method, tracking experiments using the LIRMM1 sequence have been performed. The LIRMM1 sequence was chosen because large specular reflections are present during tracking. The tracking performance facing false minima (ESM convergence) problems, occlusions by specular reflections is analyzed in details. In addition, a simulated 3 second tracking occlusion attests the capacity for overcoming problems with eventual occlusions by surgical instruments.

III.3.3.1 System initialization

Concerning the tracking initialization, the same parameters from the experiments described in section II.5.2.2 (the same 128×128 reference image with 5 TPS control points as in Figure II.17) are used.

The EKF is initialized similarly to the experiments presented earlier in section III.2.5 with the state vector sets as zeros, except for the cardiac and respiratory frequencies. When tracking starts, the EKF acquires sufficient heart motion data for the filter parameters to convergence. Once the EKF has converged, the steps illustrated in Figure III.13 are executed at each frame.

In the experiments, tracking is suspended if any of the two following criteria are met:

1. The NCC coefficient $e = min(e_l, e_r)$ of more than 3 TPS control points (of the 5 that comprise the whole mapping) drops below a threshold $\tau = 0.6$

2. The value l (proportional to the TPS bending energy) exceeds a threshold $\epsilon > 0.007$.

When tracking is suspended, the EKF is used to predict the future heart motion in order to restart tracking in the subsequent frames.

III.3.3.2 Specular reflections

As described in the previous chapter, the regions affected by specular reflections are removed from the estimation of the warping parameters. Nevertheless, large specular reflections can occlude large portions of the heart surface and due to the lack of visual information, certain control points of the TPS surface may be poorly estimated. An example containing the tracked ROI on the heart surface when tracking under large specular reflections (extracted from the LIRMM1 experiment from section II.5.2.2) is illustrated in Figure III.17. As indicated by the corresponding NCC error plots, the tracking quality when tracking under specular reflections is considerably low. Although tracking is not lost in the given example, the available visual information used for estimating the heart motion is poor. Using the unified tracking framework, occlusions due to specular reflections can be detected and the predicted heart motion can be used to bridge such periods of disturbances, avoiding tracking loss.

Occlusions by a specular highlights are often present during tracking and such disturbances may last up to 0.12 s (10 frames at 83.3 Hz). Figure III.19 illustrates the number of occlusion events detected during tracking. From the occlusion detection plot, it is evident that specular reflections are the major source of tracking disturbances in the sequence. Figure III.18 contains the measured NCC coefficients during tracking for the 5 support regions corresponding to each TPS control point (the support regions are defined in Figure III.14).

III.3.3.3 Long occlusions

The unified tracking framework method also allows us to tackle the problem of occlusion by surgical tools. Surgical instruments eventually occlude the operating site for considerably long periods of time and the proposed prediction scheme offers a solution for automatic tracking reinitialization in such cases. In Figure III.20, a simulated 3-second tracking occlusion is presented, together with the successful tracking reestablishment after

the event. The occlusion is simulated by suspending the correction step of the Kalman filter at an arbitrary instant t_1. For restarting tracking, the heart motion estimated at t_1 is used to predict the heart position at t_2, 3 seconds later. For visualizing the accuracy of the predicted heart motion, tracking results (the motion plots of one TPS control point) are also presented throughout the whole sequence for comparison purposes.

III.3.3.4 Discussion

In the current formulation, the system presents three important drawbacks, which are detailed below:

- The system initialization – The system is only operational once the dual Fourier series parameters have converged. Since no *a priori* knowledge of the heart motion is used in the experiments, tracking must remain stable during the system initialization. This can be circumvented by pre-initializing the motion parameters using a 3D feature tracker, tracking a set of features on the heart for learning the motion parameters.

- Prediction quality – Tracking cannot be restarted if the prediction quality is poor. Except cases where the ECG can be used for detecting an abnormal heart behavior (such as arrhythmias, etc), the heart motion prediction quality assessing is a very difficult task.

- Occlusion detection – Another weak point of the robust tracking method is the empirical choice of the threshold values for the occlusion detection and the arbitrarily chosen size of the subregions on the image for the computation of the NCC coefficients. Although these are not critical issues, the size of the subregions depends on the heart structures present on the reference image and a fixed window size may lead to suboptimal detection results.

III.4 Conclusion

In this chapter, a robust visual tracking method which uses the heart motion prediction for bridging occlusions and tracking failures is proposed. Two novelties are introduced:

1. A time-varying dual Fourier series for modeling the quasi-periodic beating heart motion. For estimating the parameters of the Fourier series, a probabilistic framework based on the Extended Kalman filter (EKF) is used.

2. A unified framework for estimating the temporal motion and spatial deformation of the heart surface for robust visual tracking. By integrating the estimation of the heart dynamics and the deformable tracking algorithm formulated in the previous chapter, tracking disturbances such as occlusions (e.g. by surgical instruments, smoke, blood, specular reflections, etc) or tracking failures can be bridged.

The good quality of the heart motion prediction, as well as the increased robustness achieved with the unified tracking framework is demonstrated through several experiments on *in vivo* data.

As previously said, the heart motion prediction can be useful at several levels of the surgical robot assistant design. Using the proposed motion prediction scheme, an important aspect in the motion compensation system – the latency from the image acquisition to the communication with the robot – can also be compensated. In addition, in the context of visual tracking, several tracking methods could also profit from a motion prediction scheme. For instance, a motion prediction method could be used for increasing feature tracking resilience, bridging tracking disturbances and occlusions.

Recalling the discussion on multiple visual cue fusion for increased tracking quality from the previous chapter, it is evident that no visual tracking method *per se* is able to solve the problem of robustly and accurately retrieving the heart motion. In this chapter, the temporal heart motion dynamics are proposed as an additional support for the visual tracking task. However, there is large room for improvement in the proposed unified motion estimation framework, starting from the incorporation of external signal sources such as ECG and ventilation airflow for increasing prediction quality to more accurate methods for evaluating tracking quality, which could allow "switching" between multiple visual cues and overall higher tracking quality.

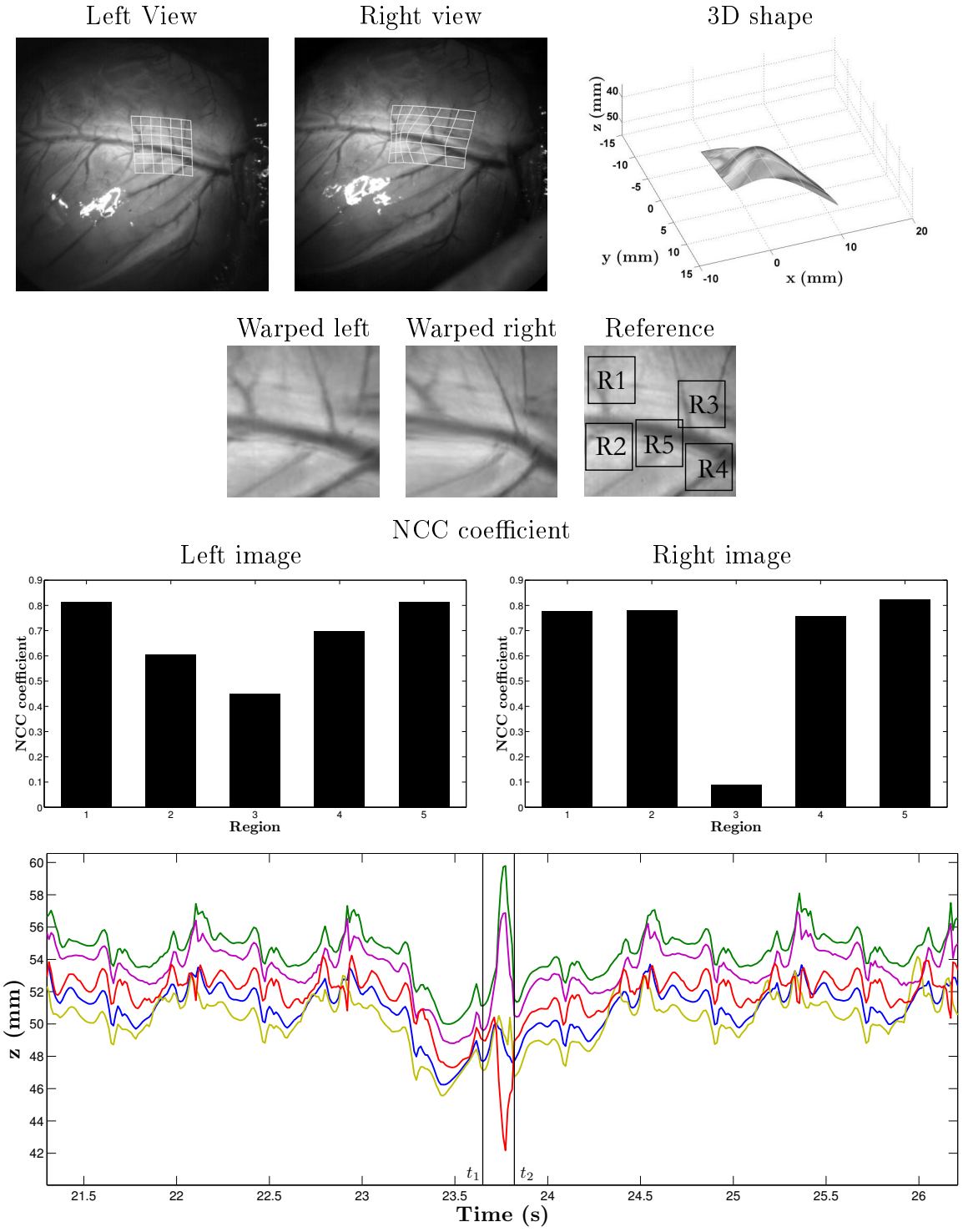


Figure III.15 : On top, an example of a erroneous shape estimation due to tracking convergence problems. According to the occlusion detection criterion, the alignment error is not sufficiently low for the ROI to be considered occluded since the NCC coefficient of only one region is below the occlusion threshold ($\tau = 0.4$). On the bottom, the effects of the convergence problem that takes place between t_1=23.65s and t_2=23.82s is revealed in the motion plots of the z component of all 5 control points.

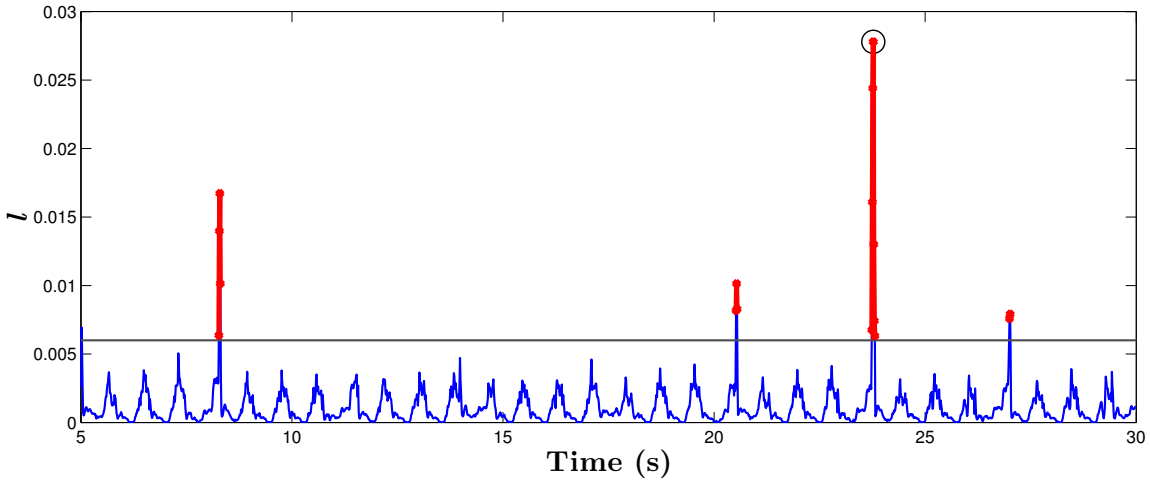


Figure III.16 : The value l (proportional to the TPS bending energy) computed for the estimated heart surface shape during the tracking experiments with the LIRMM1 sequence from the previous chapter. The estimated surfaces with energy l superior to ϵ (ϵ is defined by the line on the plot) are highlighted. Notice that the largest energy peak highlighted with a circle corresponds to the tracking results illustrated in Figure III.15 (stereo images and reconstructed 3D surface on the top).

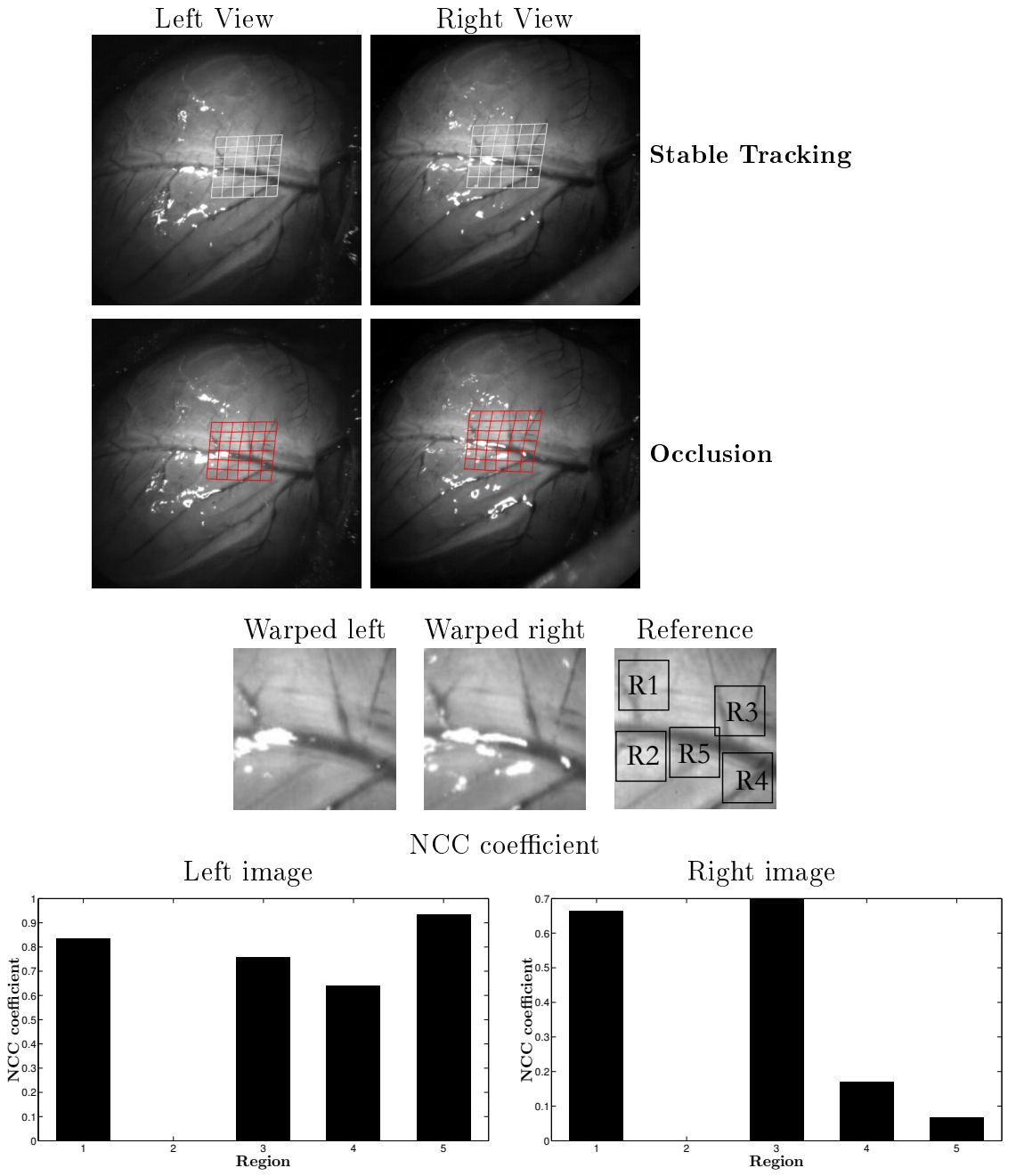


Figure III.17 : An occlusion due to large specular reflections. The poor visual information available at the moment of the occlusion can be verified by the NCC plots for the left and right warped images, which indicate that more than 2 regions are below the occlusion threshold $\tau = 0.4$.

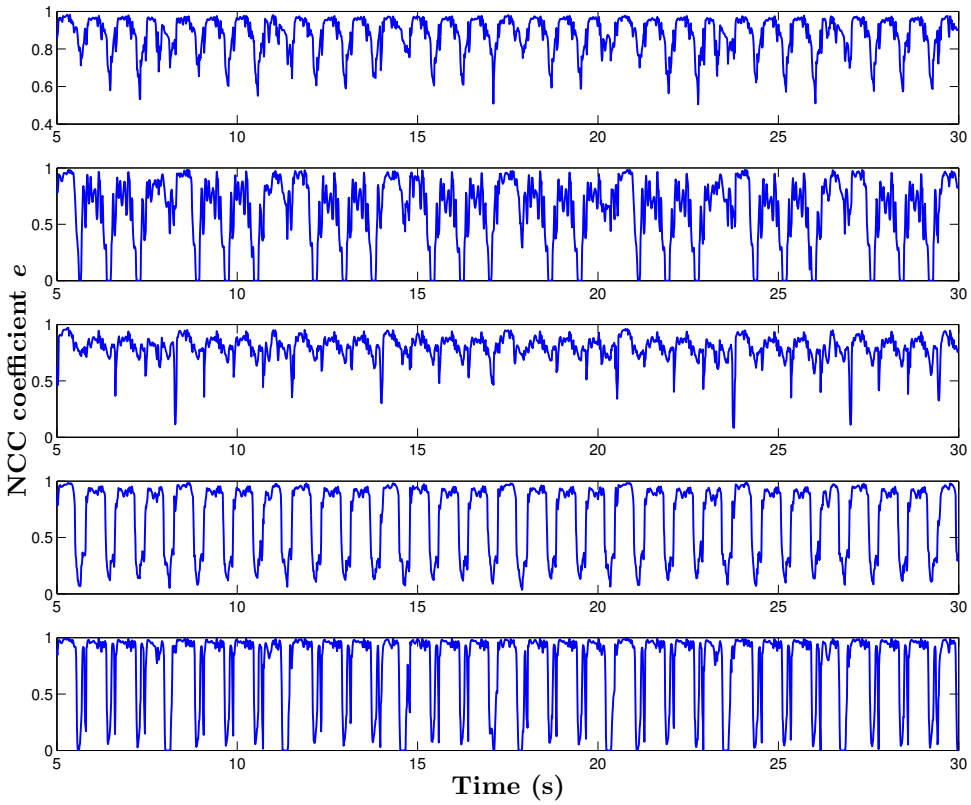


Figure III.18 : The evolution in time of the NCC coefficients during the robust tracking experiment using the LIRMM1 sequence. From top to bottom, the regions 1 to 5 defined in Figure III.14.

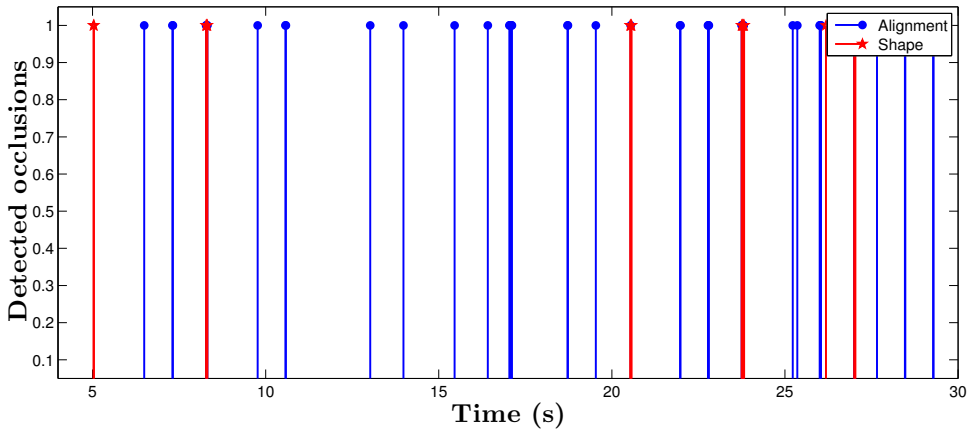


Figure III.19 : Detected occlusions during the robust tracking experiment using the LIRMM1 sequence (the value 1 indicates an occlusion event). The occlusions detected by the alignment error analysis and shape analysis are displayed separately. Note that specular reflections are the major source of occlusions.

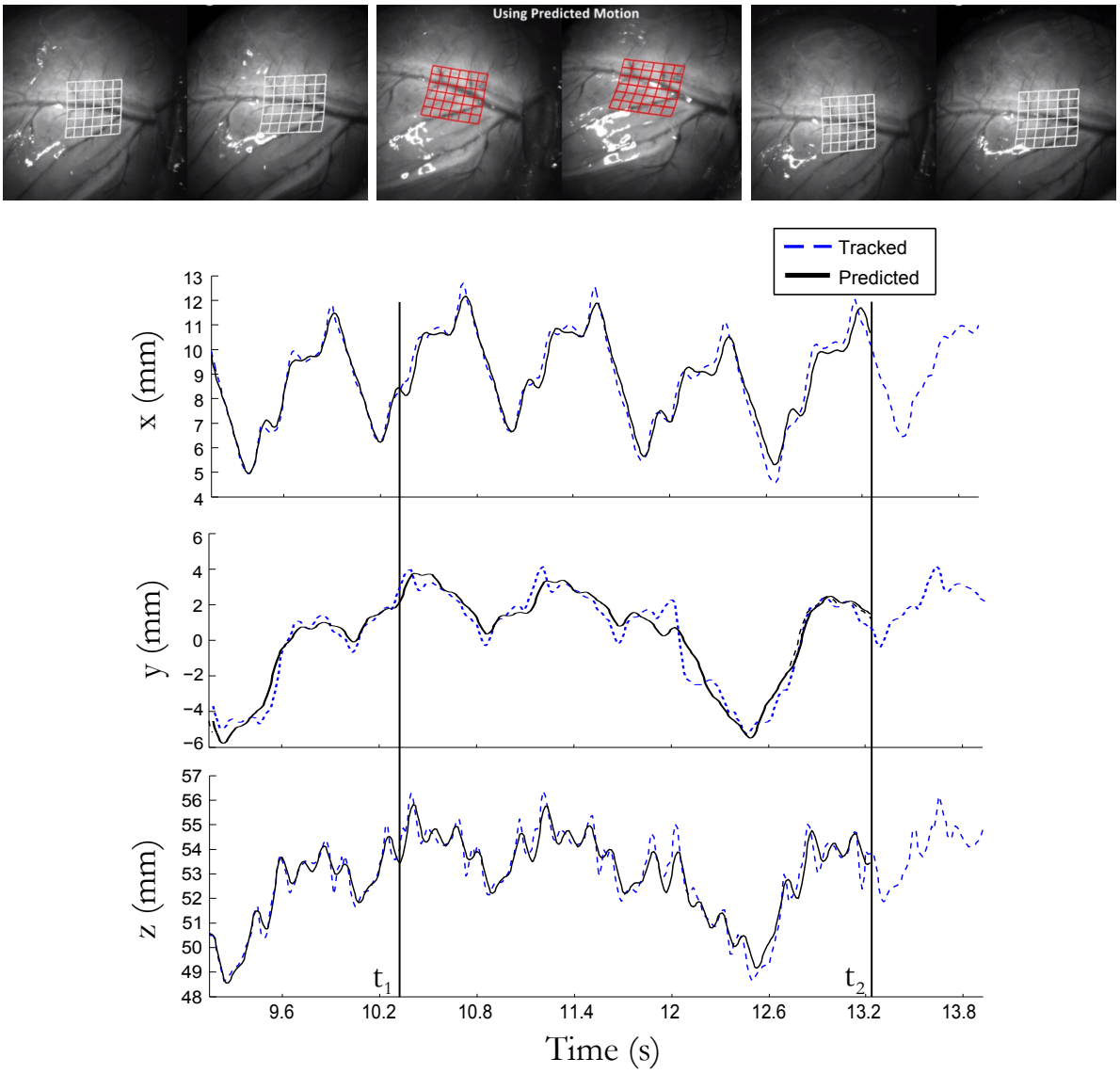


Figure III.20 : (Top Left) The mesh illustrates the tracked region of interest on both left and right endoscopic images. (Top Middle) During the simulated occlusion, tracking is suspended and the predicted trajectory of the several POI that comprise the mesh is displayed as the mesh in red. (Top Right) Tracking is successfully reestablished once the occlusion is over. (Bottom) The tracked and the predicted heart motion at an instant t_1 (the heart motion for $t > t_1$ is computed using equation (III.1)) for all Cartesian components of a given POI.

Chapter IV

General Conclusion

IV.1 Conclusion

Robotic assistance has the potential to offer significant improvements to the medical practice in terms of precision, safety and comfort. In the domain of cardiac MIS, technical improvements such as force feedback and active motion compensation aim to make the surgical act more intuitive for surgeons.

This thesis is focused on the problem of active motion compensation for robotized cardiac MIS. For estimating the heart motion, only the visual feedback from the endoscope is used, which avoids the introduction of additional sensors in the limited workspace in MIS. Furthermore, only natural landmarks on the heart surface are used for estimating the 3D motion of the heart.

However, tracking the beating heart imposes several challenges from the computer vision point of view. During this thesis, a robust visual tracking method based on a thin-plate spline parametric model has been proposed for describing the heart surface deformation. Since the lighting on the heart surface is uneven and specular reflections considerably reduce the observable heart surface, an efficient illumination compensation step is incorporated to the tracking method. Finally, for increasing the tracking robustness, a motion prediction method based on a time-varying dual Fourier series for modeling the quasi-periodic beating heart motion is developed. By integrating the visual tracking and motion prediction methods, tracking robustness facing tracking disturbances and occlusions is significantly improved.

In a motion compensated environment, the surgical tools move in synchrony with the physiological motion. This imposes a very short computational time window for tracking the heart motion. For executing the tracking method at high frame rates, the computational power of recently released graphics processor units (GPU) is exploited. Current results demonstrate the feasibility of tracking the heart motion at frequencies up to 80 Hz.

For assessing the performance of the proposed methods, *ex vivo* and *in vivo* experiments were performed in order to validate the choice for the TPS deformable model and evaluate the tracking results. Since a high acquisition rate is necessary for capturing the heart dynamics and no commercial solutions for high-speed stereo endoscopy are available, a stereo rig with two high-speed cameras and endoscopes was assembled for capturing the *in vivo* data used in the experiments.

IV.2 Perspectives

In order to design a reliable visual-based motion compensation system for a practical use in real surgical scenarios, the following issues need to be addressed:

- *Test data* – A larger image database of surgical interventions are required in order to gain more knowledge of the intra-operative environment and guide the development of motion compensation techniques. Once the role of the motion compensation system during the surgical act is defined, the system specifications can be specified.

- *Modeling* – The accurate modeling and identification of soft-tissue properties, as well as the surgical tools and devices present in the surgical workspace can significantly improve the estimation accuracy of the heart surface deformations and the quality of the heart motion prediction.

- *Real-time operation* – In order to maintain a minimum delay between the image acquisition and the actuation of the surgical tools, the physiological motion estimation must be performed very quickly. For coping with the increasing complexity of the motion estimation algorithms, the use of new computational resources such as GPUs is a possible solution.

In short term perspectives, certain issues concerning the integration of the motion compensation system in a robotic assistant require special attention. A schematic view of the robotic assistant equipped with a motion compensation system is displayed in figure IV.1. The stereoscopic image acquisition and tracking are done in a separate computer in order to avoid long delays when transferring the images from the cameras to the graphics card. Next, the 3D motion of the ROI on the heart is estimated and transfered to the master device, which synchronizes the robotic tools with the beating heart motion. Finally, the superposition of the surgeon's commands and the estimated physiological motion in a tele-operation scheme is a key issue for the successful integration of the works developed in this thesis in a surgical robotic platform.

In the proposed integration scheme, two important issues must be addressed:

- *System requirements* – In order to sufficiently capture the heart dynamics, the surgical platform should be equipped with stereo endoscopes capable of acquiring images at rates superior to 100 Hz (with no compromise to the visual feedback for the surgeon). In addition, the image resolution should be sufficiently high for capturing the details of fine heart structures. Thus, these requirements concern the choice of cameras, illumination (extra lighting is needed when acquiring at higher speeds) and acquisition devices.

- *Delays* – If the processing delays which start from the image acquisition by the endoscope, tracking, control and the actuation stages are not sufficiently low, critical errors can be induced in the motion compensation system. Even though palliative solutions such as motion prediction can be adopted, the compensation errors due to latency should be kept within safe levels required by cardiac surgery (under 200μm [Cavusoglu et al., 2005]).

Finally, in a longer-term perspective, the integration of the proposed motion compensation system in an Augmented Reality (AR) system [Fuchs et al., 1998] for intra-operative

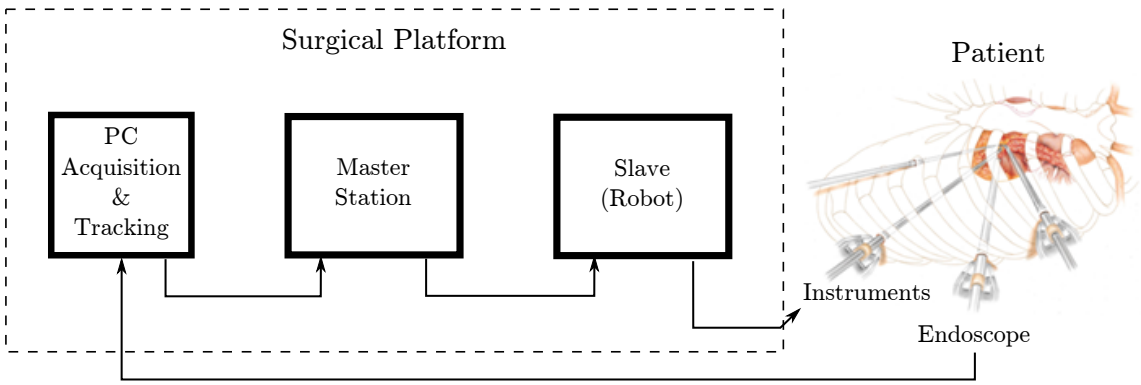


Figure IV.1 : The schematic overview of a surgical robotic platform with a motion compensation system.

navigation is envisaged. The accurate estimation of the deformations of internal organs is a key requirement for merging pre-operative and intra-operative data. Patient-specific information such as the internal organ structures can be used for increased surgical safety and surgical guidance.

Publications

Journal

1. Richa R., Poignet P., Liu C., 3D Motion Tracking for Beating Heart Surgery Using a Thin-Plate Spline Deformable Model, *The International Journal of Robotics Research (IJRR) - Special Issue on Robot Vision* 29 (2-3) (2010) pp. 218-230.

Book Chapter

1. Richa R., Bo A. P. L., Poignet P., Motion tracking for beating heart surgery, *Submitted to "Surgical Robotics - Systems, Applications, and Visions", Editors: J. Rosen, B. Hannaford and R. Satava, Springer.*

International Conferences

1. Richa R., Bo, A. P. L., Poignet P., Beating Heart Motion Prediction for Robust Visual Tracking, *In Proceedings of IEEE Conference on Robotics and Automation (ICRA '10)*, To appear, Anchorage, USA, 2010.

2. Richa R., Poignet P., Liu C., Efficient 3D Tracking for Motion Compensation in Beating Heart Surgery, *In Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI '08)*, volume 5242-II, pages 684-691, New York, USA, 2008.

3. Richa R., Poignet P., Liu C., Deformable Motion Tracking of the Heart Surface, *In Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '08)*, pp. 3997-4003, Nice, France, 2008.

4. Richa R., Bo, A. P. L., Poignet P., Motion Prediction for Tracking the Beating Heart *In Proceedings of IEEE International Conference of the Engineering in Medicine and Biology Society (EMBS '08)*, pp. 3261-3264, Vancouver, Canada, 2008.

Appendice

Appendix A

Error Measures

A.1 Quantifying the alignment error

In the stereo tracking framework proposed in chapter II, tracking quality can be considered as a function of the alignment error between the reference image $\mathbf{T}$ and the warped left and right images $\mathbf{I}_l(w(\mathbf{x}, \mathbf{h}))$ and $\mathbf{I}_r(w(\mathbf{x}, \mathbf{h}))$. For measuring the alignment error between two images, 3 similarity measures were evaluated in our studies: the normalized cross correlation (NCC), the metric based on the Bhattacharyya coefficient used in Bourger *et al.* [Bourger et al., 2007] and mutual information (MI).

A.2 The normalized cross-correlation (NCC)

The cross-correlation can be derived from the squared Euclidean distance $d^2(x, y)$ between the intensity values of two images $\mathbf{I}_1(x, y)$ and $\mathbf{I}_2(x, y)$:

$$d^2(x, y) = \sum_{x,y \in A} [\mathbf{I}_1(x, y) - \mathbf{I}_2(x, y)]^2 \tag{A.1}$$

where A is the set of all pixels in the image $\mathbf{I}_1$. Expanding the equation above:

$$d^2(x, y) = \sum_{x,y} [\mathbf{I}_1^2(x, y) - 2\mathbf{I}_1(x, y)\mathbf{I}_2(x, y) + \mathbf{I}_2^2(x, y)] \tag{A.2}$$

From the equation above the cross-correlation term can be derived, which is the only term that is a function of both images $\mathbf{I}_1$ and $\mathbf{I}_2$:

$$\sum_{x,y} [\mathbf{I}_1(x, y)\mathbf{I}_2(x, y)] \tag{A.3}$$

Although the cross-correlation term above is a measure of similarity between images, if the case of matching a reference image over an image sequence is considered, the cross-correlation is not invariant to changes in image amplitude such as those caused by changing lighting conditions. In order to overcome such difficulties, the images $\mathbf{I}_1$ and $\mathbf{I}_2$ can be normalized to unit length, yielding the normalized cross-correlation (NCC) coefficient:

$$\mathcal{C}(x, y) = \frac{\sum_{x,y} [\mathbf{I}_1(x, y) - \bar{\mathbf{I}}_1][\mathbf{I}_2(x, y) - \bar{\mathbf{I}}_2]}{\sqrt{\sum_{x,y} [\mathbf{I}_1(x, y) - \bar{\mathbf{I}}_1]^2 \sum_{x,y} [\mathbf{I}_2(x, y) - \bar{\mathbf{I}}_2]^2}} \tag{A.4}$$



(a) The Lena image (b) Region 1 (c) Region 2

Figure A.1 : The Lena picture - For testing the error functions, two regions of interest on the image are selected (region 1 is marked in black and region 2 in white).

The NCC coefficient may vary from [-1,1] but in this study, only the range [0,1] where 1 is a perfect match is considered (results in the interval [-1,0] are rounded up to 0).

A.3 A metric based on the Bhattacharyya coefficient

First, let us consider the η bin histogram $\mathbf{g}$ of an image $\mathbf{I}$, where:

$$\mathbf{g} = g_l|_{l=1,\ldots,\eta} \quad \text{where} \quad \sum_{l=1}^{\eta} g_l = 1 \tag{A.5}$$

The statistical distance d between two histograms ${}^1\mathbf{g}$ and ${}^2\mathbf{g}$ of images $\mathbf{I}_1$ and $\mathbf{I}_2$ respectively can be defined as:

$$d = \sqrt{1 - \sum_{l=1}^{\eta} {}^1g_l \; {}^2g_l} \tag{A.6}$$

Minimizing the distance d is equivalent to maximizing the Bhattacharrya coefficient. Therefore, the inverse of the distance d is used as a similarity measure between two images:

$$\mathcal{D} = \frac{1}{d} \tag{A.7}$$

The output of the $\mathcal{D}$ may vary depending on the histograms of the input images. In this study, only 8-bit grayscale images are used, with intensity values ranging from [0,255] gray values. For avoiding sensitivity problems regarding small intensity shifts, the number of histogram bins η is reduced to 32. The histogram for regions 1 and 2 from Figure A.1 are plotted in Figure A.2.

A.4 Mutual Information

For quantifying the amount of information in an image, the image entropy can be used. It measures the dispersion in the probability distribution of an image, yielding high

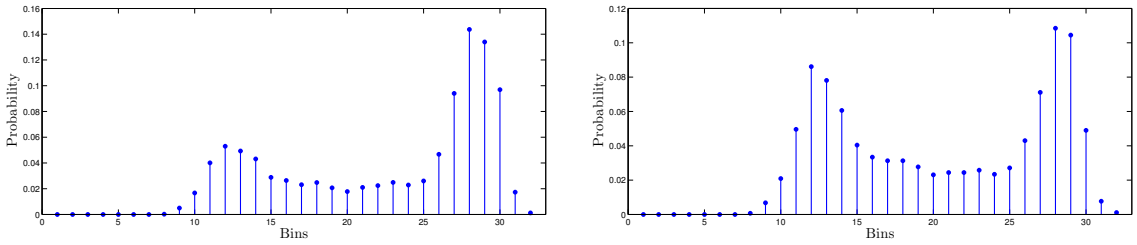


Figure A.2 : 32-bin histograms for region 1 (left) and region 2 (right).

values for images with many gray levels and lower values to images with a lower histogram dispersion.

For measuring the amount of mutual information between two images $\mathbf{I}_1(x, y)$ and $\mathbf{I}_2(x, y)$, their joint probability distribution must be defined. First, a 2D histogram where each axis is the number of possible grayscale values in the image is generated. For each and same pixel position on both images, a histogram cell is incremented. The joint entropy can then be defined as:

$$j(\mathbf{I}_1, \mathbf{I}_2) = -\sum_{u,v} p(u, v) \cdot log(p(u, v)) \tag{A.8}$$

where $p(u, v)$ is the joint probability obtained from the 2D histogram. Notice that the entropy of an image $\mathcal{E}(\mathbf{I})$ can be computed as $\mathcal{E}(\mathbf{I}) = j(\mathbf{I}, \mathbf{I})$, which is a measure of histogram dispersion. Figure A.3 illustrates an example of a joint probability distribution function for a given pair of images in Figure A.2. Note that if two images are perfectly aligned, the histogram is highly focused. As the images mis-align, the dispersion grows. The mutual information between two images can be defined as:

$$\mathcal{M}(\mathbf{I}_1, \mathbf{I}_2) = j(\mathbf{I}_1) + j(\mathbf{I}_2) - j(\mathbf{I}_1, \mathbf{I}_2) \tag{A.9}$$

From the definition above, one can observe that maximizing the mutual information is equivalent to minimizing the joint entropy $j(\mathbf{I}_1, \mathbf{I}_2)$. In this study, the inverse of the joint entropy is used as a similarity measure:

$$\mathcal{J} = \frac{1}{j(\mathbf{I}_1, \mathbf{I}_2)} \tag{A.10}$$

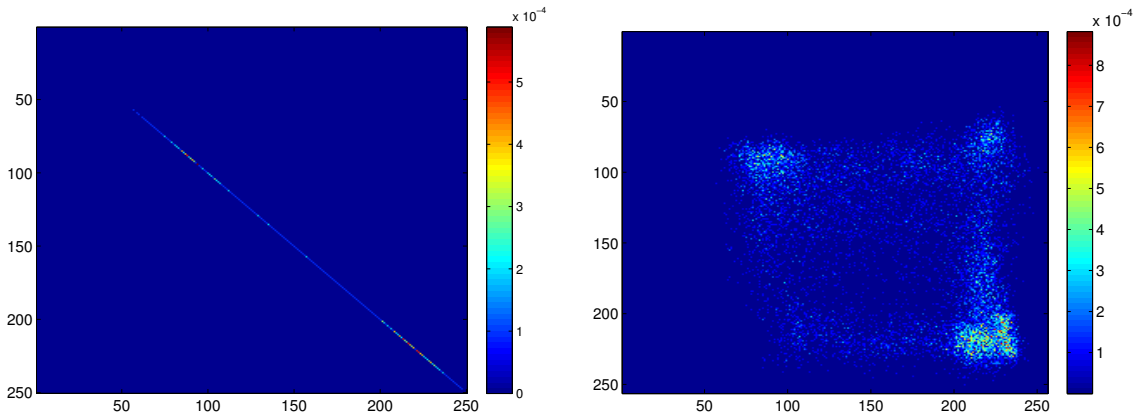


Figure A.3 : (Left) The joint probability distribution of region 1 (from Figure A.1) with itself. (Right) The joint probability distribution of regions 1 and 2.

A.5 Experiments

For choosing the most convenient error measure for tracking the beating heart, the alignment errors during the tracking experiments from the previous chapter (the tracking experiments on the LIRMM1 and DAVINCI experimental data discussed in section II.5.2.2) have been analyzed for the three techniques described above.

For illustration purposes, we plot in Figure A.4 the error outputs using the three similarity measures chosen in this study for the frame extracted from the LIRMM1 tracking experiment illustrated in Figure III.14. Notice that the NCC measure translates the best the presence of specular reflections on the images.

This is due mainly to a smaller error output variability when tracking different image sequences, the NCC has shown to be the most convenient measure of alignment error. The metric based on the Bhattacharyya coefficient and mutual information display large error output variations between different sets of reference images and are also very sensitive to motion blur.

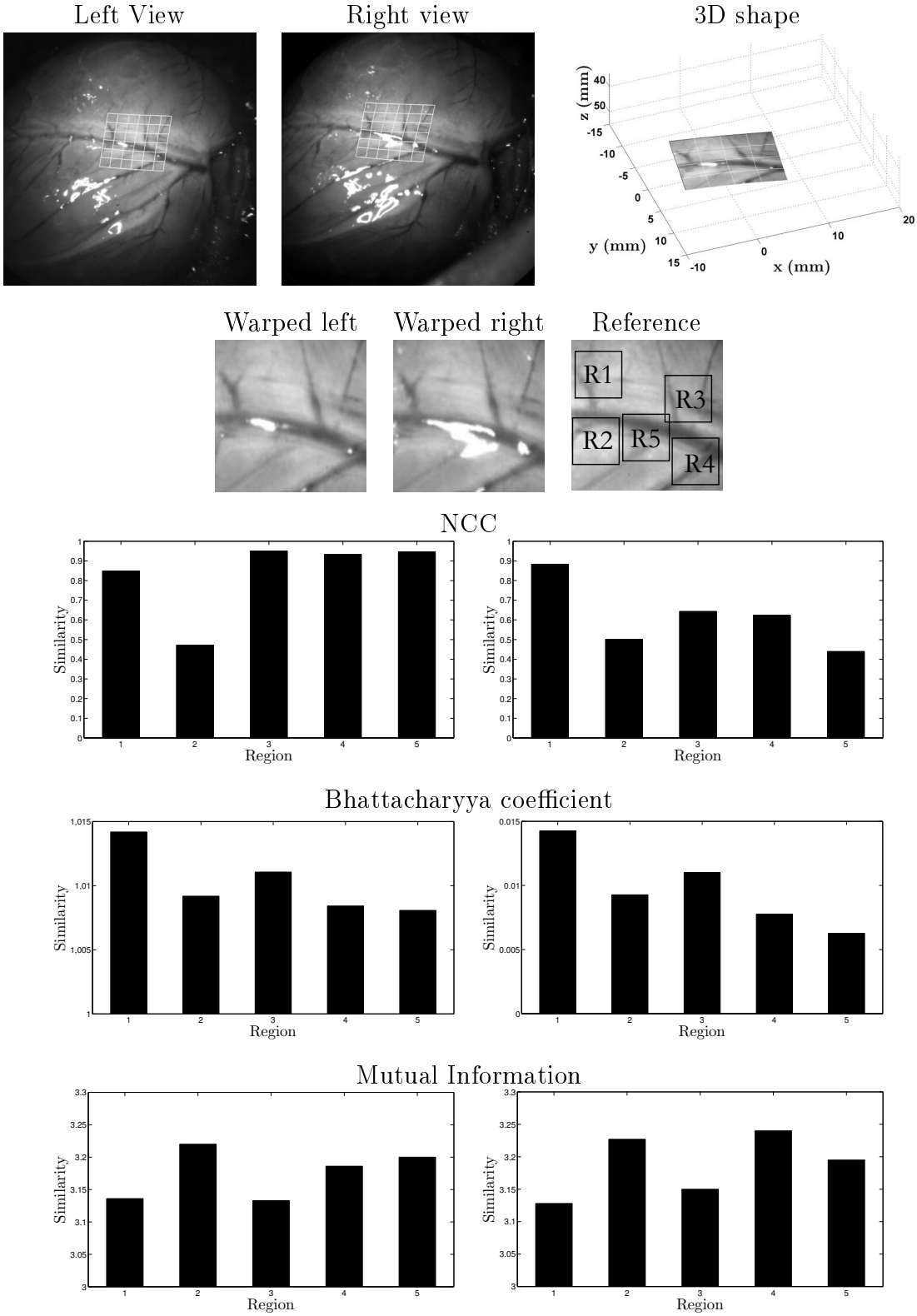


Figure A.4 : The image alignment error for both left and right views is measured using three similarity measures. The output of each similarity measure used in our studies for the tracking results illustrated on top is computed (the left and right columns correspond to the left and right warped stereo images).

Appendix B

The Efficient Second-order Minimization (ESM)

B.1 The template tracking problem

The ESM algorithm was presented in Benhimane *et al.* [Benhimane and Malis, 2004]. Let us consider a $(h \times w)$ image template $\mathbf{T}$ with pixel coordinates $\mathbf{x} = (x, y)$. For simplification purposes, the columns of $\mathbf{T}$ are stacked to obtain a vector of intensities of all pixels $\mathbf{x}_i$ (with $i = 1, 2, ..., w \cdot h$). In addition, $w(\mathbf{x}, \mathbf{h})$ can be defined as an image transformation operation (such as a translation, affine or projective transformation, etc) where $\mathbf{x} \mapsto w(\mathbf{x}, \mathbf{h}) = \mathbf{x}'$.

The goal of template-based tracking is to find the transformation $\mathbf{h}$ that minimizes the alignment error between the mapped pixels of the current image $\mathbf{I}$ and the reference template $\mathbf{T}$. Traditionally, minimizing the alignment error is equivalent to minimizing the cost function $f(\mathbf{h})$ which is the sum of squared differences (SSD) between $\mathbf{T}$ and $\mathbf{I}(w(\mathbf{x}, \mathbf{h}))$, which is the current image $\mathbf{I}$ transformed by the mapping of parameters $\mathbf{h}$:

$$f(\mathbf{h}) = \frac{1}{2} \sum_{i \in A} [\mathbf{I}(w(\mathbf{x}_i, \mathbf{h})) - \mathbf{T}(\mathbf{x}_i)]^2 = \frac{1}{2} ||\mathbf{I}(w(\mathbf{x}_i, \mathbf{h})) - \mathbf{T}(\mathbf{x}_i)||^2 \tag{B.1}$$

where A is the set of pixels positions $\mathbf{x}$ in the reference image $\mathbf{T}$. The minimization of the non-linear equation above is done iteratively. At each iteration, an update $\Delta\mathbf{h}$ is computed in order to find the global minimum of the cost function. The minimum is reached when the derivative of the cost function with respect to the warping parameters is zero:

$$\frac{\partial f(\mathbf{h})}{\partial \mathbf{h}} = 0 \tag{B.2}$$

For linearizing the above equation, the derivative of f is approximated about the current warping parameters $\mathbf{h} = \mathbf{h}_c$:

$$\frac{\partial f(\mathbf{h})}{\partial \mathbf{h}} \approx \Delta\mathbf{I}\, \mathbf{J}(\mathbf{h}) + \Delta\mathbf{h}^T \mathbf{S} \tag{B.3}$$

where $\Delta\mathbf{I} = \mathbf{I}(w(\mathbf{x}, \mathbf{h}_c)) - \mathbf{T}(\mathbf{x})$, $\mathbf{J}$ is the derivative of $\mathbf{I}(w(\mathbf{x}, \mathbf{h}))$ with respect to $\mathbf{x}$ and $\mathbf{S}$ is the Hessian matrix:

$$\mathbf{S} = \left. \frac{\partial^2 f(\mathbf{h})}{\partial \mathbf{h}^2} \right|_{\mathbf{h}=\mathbf{h}_c} \tag{B.4}$$

To overcome convergence problems and reduce the computational cost of the problem, several methods can be used to approximate $\hat{\mathbf{S}}$, such as:

$$\begin{array}{ll} \hat{\mathbf{S}} = \alpha\mathbf{I} & \text{Steepest Descent} \\ \hat{\mathbf{S}} = \mathbf{J}^{\mathrm{T}}\mathbf{T} & \text{Gauss-Newton} \\ \hat{\mathbf{S}} = \mathbf{J}^T\mathbf{T} + \alpha(\mathbf{J}^T\mathbf{T}) & \text{Levenberg-Marquardt} \end{array}$$

The standard Newton minimization method solves equation (B.2) using an approximation for the Hessian matrix listed above:

$$\Delta\mathbf{h} = -\mathbf{S}^{-1}\mathbf{J}(\mathbf{h})^T\Delta\mathbf{I} \tag{B.5}$$

B.2 The second-order approximation

Based on the second order minimization introduced by Ezio Malis in [Malis, 2004], a second-order Taylor series of the image $\mathbf{I}$ evaluated at $\mathbf{h}_c$ centered at $\mathbf{h}_0$ (which represent the identity warp $w(\mathbf{x}, \mathbf{h}_0) = \mathbf{x}$) yields:

$$\mathbf{T}(\mathbf{x}) \approx \mathbf{I}(w(\mathbf{x}, \mathbf{h}_c)) + \mathbf{J}(\mathbf{h}_c)\Delta\mathbf{h} + \frac{1}{2}\mathbf{D}(\mathbf{h}_c, \Delta\mathbf{h})\Delta\mathbf{h} \tag{B.6}$$

where $\mathbf{D}(\mathbf{h}_c, \Delta\mathbf{h})$ is a matrix that depends on $\mathbf{h}$ and on the Hessian matrices of $\mathbf{I}(w(\mathbf{x}, \mathbf{h}_c))$. Using the first order series of $\mathbf{J}(\mathbf{h})$ about $\mathbf{h}_c$ evaluated at $\mathbf{h}_0$, one obtains:

$$\mathbf{J}(\mathbf{h}_0) \approx \mathbf{J}(\mathbf{h}_c) + \mathbf{D}(\mathbf{h}_c, \delta\mathbf{h}) \tag{B.7}$$

Replacing the above equation in (B.6):

$$\mathbf{T}(\mathbf{x}) \approx \mathbf{I}(w(\mathbf{x}, \mathbf{h}_c)) + \frac{1}{2}(\mathbf{J}(\mathbf{h}_0) + \mathbf{J}(\mathbf{h}_c))\Delta\mathbf{h} \tag{B.8}$$

which is equivalent to:

$$\Delta\mathbf{I} \approx \frac{1}{2}(\mathbf{J}(\mathbf{h}_0) + \mathbf{J}(\mathbf{h}_c))\Delta\mathbf{h} \tag{B.9}$$

The above equation is the second order approximation of $\Delta\mathbf{I}$ for a small $\Delta\mathbf{h}$. In order to obtain the increment $\Delta\mathbf{h}$ for a given frame, the pseudo-inverse of the mean of the Jacobians is computed:

$$\Delta\mathbf{h} \approx 2(\mathbf{J}(\mathbf{h}_0) + \mathbf{J}(\mathbf{h}_c))^+\Delta\mathbf{I} \tag{B.10}$$

It is important to note that results equivalent to the ESM method described above have also been proposed in the works of Averbuch *et al.* [Averbuch and Keller, 2004], based on a similar formulation of the image alignment problem. Nevertheless, the ESM is more efficient than the bidirectional gradient method in [Averbuch and Keller, 2004] since the latter requires the inverse of a Jacobian matrix of twice the size as the one in the ESM.



Figure B.1 : Lenna image used for the experiments. The chosen region of interest is displayed on the right.

B.3 Comparative analysis

In this section, the three methods described in this appendix (Gauss-Newton, Levenberg-Marquandt and the ESM) are compared. The Lenna image is chosen for the experiments. A 100 $\times$100 template has been chosen from the image as shown in Figure B.1. The template image was warped 1000 times using random homography transformations, adding Gaussian noise to the original coordinates of the 4 corners of the template. The standard deviation of this Gaussian noise σ is increased from 1 to 12. The percentage of results that converged among all 1000 tests are illustrated in Figure B.2(top). From the experiments, the ESM presents the best performance, with a percentage of convergence around 45% at $\sigma = 12$. Figure B.2(down) illustrates the average convergence speeds for the three methods with a fixed $\sigma = 10$, illustrating the superior performance of the ESM.

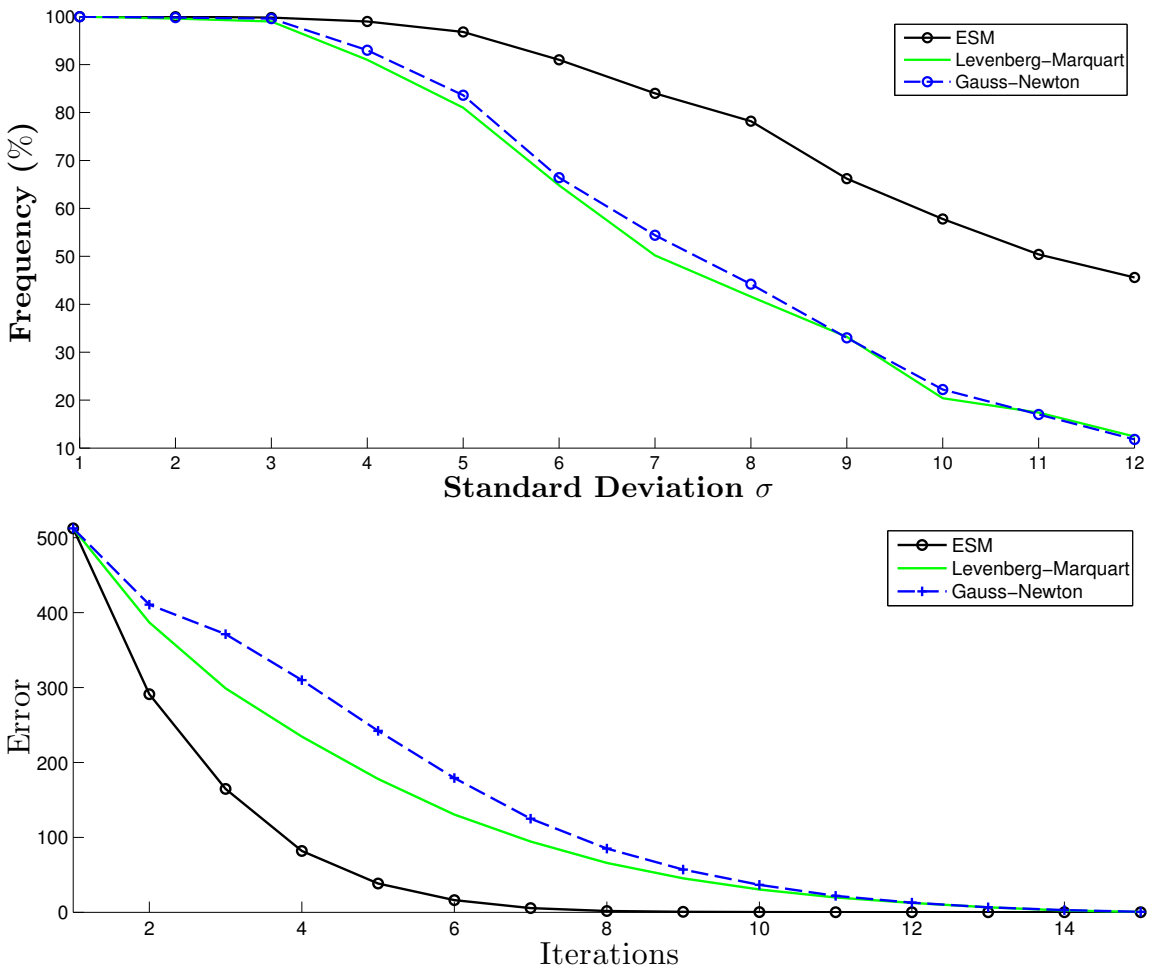


Figure B.2 : Comparative analysis of the convergence frequency (left) and convergence rate (right).

Appendix C

Lens distortion correction

Let a point in space $\mathbf{X}$ of coordinates $[X, Y, Z]$ in the camera reference frame. Assuming a calibrated camera, the focal length $\boldsymbol{\kappa} = [\kappa_1, \kappa_2]$, the principal point $\boldsymbol{\upsilon} = [\upsilon_1, \upsilon_2]$ and the radial and tangential distortion parameters $\boldsymbol{\rho} = [\rho_1, ...\rho_5]$ are known. Projecting the point $\mathbf{X}$ onto the image plane, one obtains the normalized (pinhole) point $\mathbf{x}_n$ such that:

$$\mathbf{x}_n = \begin{bmatrix} X/Z \\ Y/Z \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} \tag{C.1}$$

For including the lens distortion effect, the normalized distorted point coordinate $\mathbf{x}_d$ is computed as:

$$\mathbf{x}_d = \begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + \rho_1 r^2 + \rho_2 r^4 + \rho_5 r^6)\mathbf{x}_n + \mathbf{o} \tag{C.2}$$

where $r = ||\mathbf{x}_n||^2$ and $\mathbf{o}$ is the tangential distortion vector given as:

$$\mathbf{o} = \begin{bmatrix} 2\ \rho_3 x_n y_n + \rho_4(r^2 + 2x_n^2) \\ \rho_3(r^2 + 2y_n^2) + 2\ \rho_4 x_n y_n \end{bmatrix} \tag{C.3}$$

Finally, the pixel coordinates $\mathbf{x}_p$ of a point $\mathbf{X}$ in space on the image plane is given as:

$$\mathbf{x}_p = \begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} \kappa_1 x_d + \upsilon_1 \\ \kappa_2 y_d + \upsilon_2 \end{bmatrix} \tag{C.4}$$

For correcting the original distorted image $\mathbf{I}_d$ acquired by the camera, the 3D points of each pixel of the ideal undistorted image $\mathbf{I}$ are projected onto the image plane using the distortion model above. For computing the intensity values of $\mathbf{I}$, a simple bilinear interpolation of the intensity values of the distorted image $\mathbf{I}_d$ for each distorted pixel values $\mathbf{x}_p$ is sufficient. An example of the steps above is given in figure C.1. The calibration parameters used for computing the undistorted image are given in Table C.1.

Focal Length	Principal Point	Distortion Parameters
$\boldsymbol{\kappa} = \begin{bmatrix} 661.67001 \\ 662.8285 \end{bmatrix}$	$\boldsymbol{\upsilon} = \begin{bmatrix} 306.09590 \\ 240.78987 \end{bmatrix}$	$\boldsymbol{\rho} = \begin{bmatrix} -0.26425 \\ 0.22645 \\ 0.00020 \\ 0.00023 \\ 0 \end{bmatrix}$

Table C.1 : The calibration data for the example in figure C.1



Figure C.1 : A lens distortion correction example. On the left, the original distorted image $\mathbf{I}_d$ acquired by the camera and on the right, the undistorted image $\mathbf{I}$ (source: [Bouguet, 2009]).

Appendix D

Parametric deformable models

D.1 Depth from disparity

For evaluating the deformable models for representing the heart surface proposed in the literature, the performance of different techniques for retrieving a dense 3D depth map of the operating site in MIS is compared. Depth is estimated from the disparity between the rectified stereo pair [Fusiello et al., 2001]. Although rectification is not strictly necessary, it greatly simplifies the formulations and the transformation from disparity to Cartesian coordinates.

In the next section, a qualitative analysis of the reconstruction results using a TPS model, tensor B-Splines [Lau et al., 2004] and Piecewise Bilinear mappings [Stoyanov et al., 2005a] are given. For estimating the parameters of each deformable model, the ESM algorithm is used. It is also important to remark that no illumination compensation techniques are used in the comparison.

D.2 Experimental results

D.2.1 Setup

For this experiment, the DAVINCI image sequence (see section II.5.2.1) is used. For comparing reconstruction results, a 150x150 pixel region of interest defined in figure D.1 with 25 evenly spaced control points is considered for the TPS and bilinear models, whereas 36 parameters are used for the B-spline model for achieving the same spatial resolution as the other models. The number of ESM iterations was limited to 50, except for the hierarchical piecewise bilinear mapping estimation. The minimization loop converges if the sum of elements of the increment vector drops below 0.01 ($\sum \delta \mathbf{h} < 0.01$).

D.2.2 Discussion

Examples of reconstructed surfaces using each parametric model are given in figure D.2. The experimental results relative to each deformable model are summarized below:

- The B-spline model: since no regularization is enforced, the B-spline surface can deform freely and fits better the workspace shape. Since the B-spline basis functions are estimated locally, it converges faster than all methods (an average of 15.57

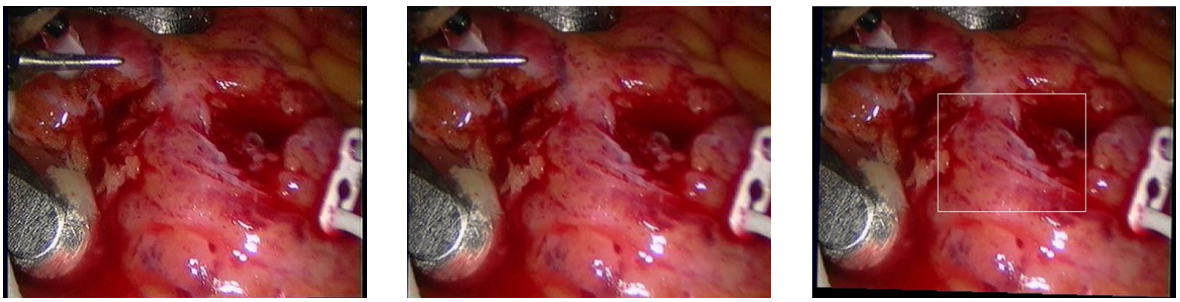


Figure D.1 : (Left) Original left stereo image extracted from the DAVINCI sequence (Middle) The original image after the correction of the lens distortion (Right) The undistorted image rectified using the algorithm proposed by Fusiello *et al.* [Fusiello et al., 2001]. The region of interest is delimited by the white square.

iterations are required for convergence). However, the estimation of the mapping borders is problematic and depth recovery may become unstable.

- The Piecewise Bilinear Model: bilinear mapping does not provide a smooth approximation of the heart surface shape, which translates in a high alignment error. Moreover, without the hierarchical mapping estimation proposed in the original paper, depth recovery is inefficient and tracking eventually diverges. In average, 18.73 iterations are required for convergence.

- TPS model: a visual inspection indicates that the most stable results are obtained using the TPS mapping. Due to its intrinsic regularization properties, the TPS model adapts well to the heart surface shape and converges with a small number of iterations per frame (12.5 iterations in average), as indicated in figure D.3.

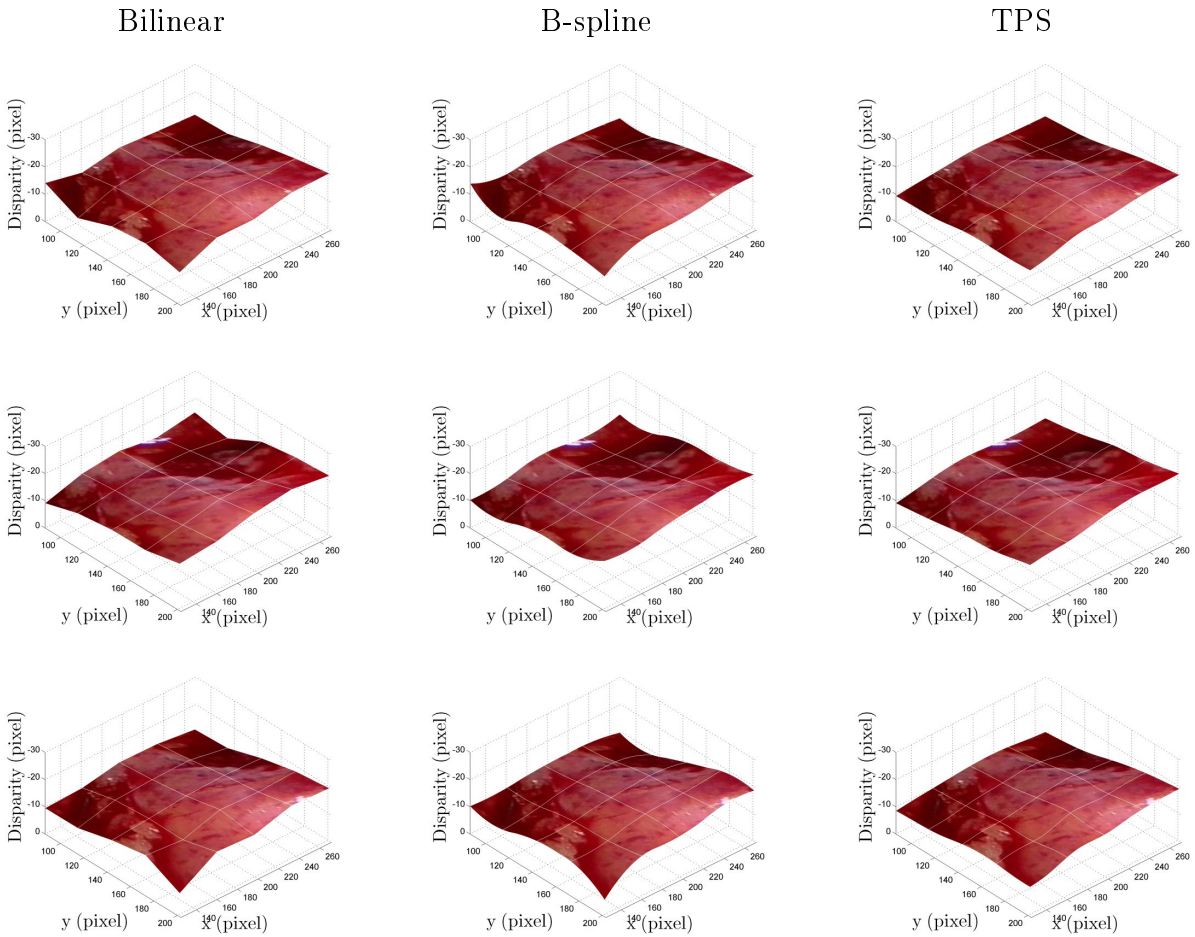


Figure D.2 : Comparative Analysis – Reconstructed ROI on the heart surface at a given instant (left column) Piecewise bilinear model (middle column) B-spline model (right column) adaptive TPS model

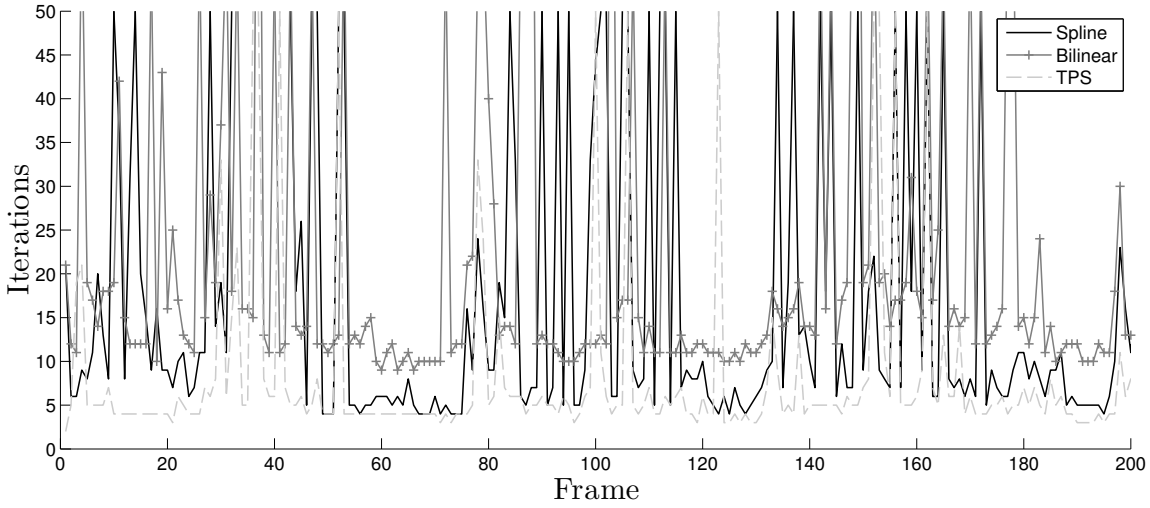


Figure D.3 : The number of ESM iterations for convergence.

Appendix E

Estimated Fourier series coefficients

E.1 The estimated Fourier series harmonics

For illustrating the evolution of the Fourier series coefficients, the amplitude of each harmonic of the Fourier series that model the x component of the LIRMM1 experimental data is plotted in Figure E.2. Notice the transient during the initialization, since the coefficients are initialized as zeros. The respiratory and cardiac frequencies are plotted in Figure E.1. Notice that no significant frequency variations can be verified in the LIRMM1 experimental data.

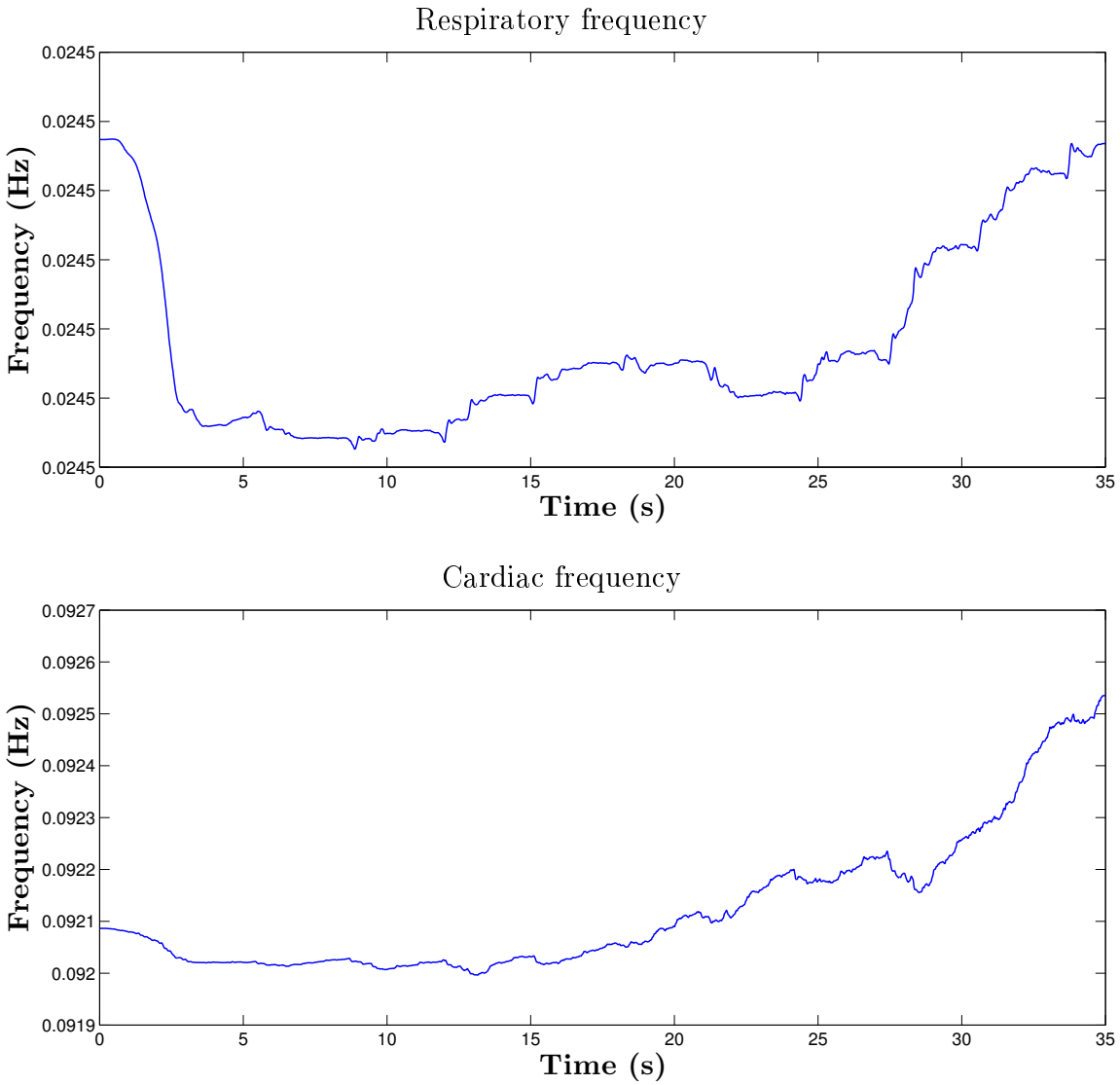


Figure E.1 : The estimated respiratory and cardiac frequencies in the experiments with the LIRMM1 data. Notice that no significant frequency variations can be verified.

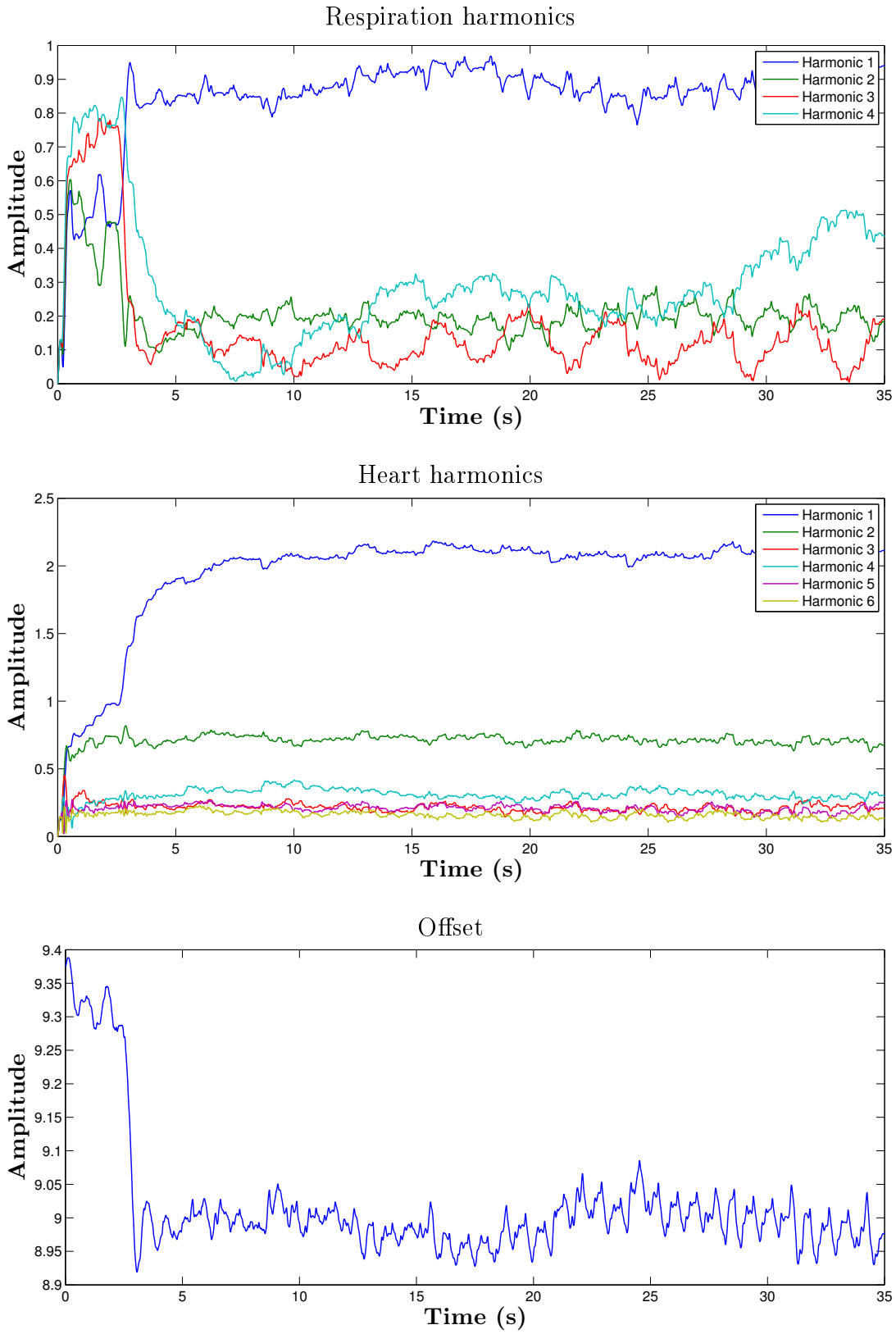


Figure E.2 : The amplitude of the Fourier series coefficients that model the LIRMM1 experimental data.

Appendix F

The Jacobian Matrix

F.1 Building the Jacobian matrix

Given the α-vector $\mathbf{w} = w_1, w_2, ..., w_i$, resulting from the multiplication between matrices $\mathbf{M}_i\mathbf{K}_*$ in equation (II.15), we can compute the terms of the Jacobian $\mathbf{J}$ corresponding to the Cartesian coordinates of the i-th control point in $\mathbf{h}$ as follows:

$$\frac{\partial \mathbf{I}(w_{3D}(\mathbf{x}_i, \mathbf{h}))}{\partial h_i} = (\nabla I_x w_i C_{1,1} + \nabla I_y w_i C_{2,1})/s_i \tag{F.1}$$

$$\frac{\partial \mathbf{I}(w_{3D}(\mathbf{x}_i, \mathbf{h}))}{\partial h_{i+\alpha}} = (\nabla I_x w_i C_{1,2} + \nabla I_y w_i C_{2,2})/s_i \tag{F.2}$$

$$\frac{\partial \mathbf{I}(w_{3D}(\mathbf{x}_i, \mathbf{h}))}{\partial h_{i+2\alpha}} = w_i\Big(\frac{C_{1,3}\nabla I_x + C_{2,3}\nabla I_y}{s_i} - C_{3,3}\frac{\nabla I_x s x_i' + \nabla I_y s y_i'}{s_i^2}\Big) \tag{F.3}$$

where α is the number of TPS control points and $C_{i,j}$ is the element in i-th line and j-th column from the corresponding camera calibration matrix $\mathbf{C}$. Notice that the elements $[sx_i' \;\; sy_i' \;\; s_i]$ are computed using equation (II.15).

Acronyms

AR	Augmented Reality
CABG	Coronary Artery Bypass Grafting
CUDA	Compute Unified Device Architecture
ECG	Electrocardiogram
EKF	Extended Kalman Filter
ESM	Efficient Second-order Minimization
FFT	Fast Fourier Transform
GPU	Graphics Processing Unit
IPP	Intel Performance Primitives
LMS	Least Mean Squares
MIS	Minimally Invasive Surgery
NCC	Normalized Cross-Correlation
POI	Point Of Interest
Sfs	Shape from Shading
TECAB	Totally Endoscopic Coronary Artery Bypass

Bibliography

[AHA, 2009] AHA (2009). American heart association. `http://www.americanheart.org`.

[Albitar et al., 2007] Albitar, I. C., Graebling, P., and Doignon, C. (2007). Robust Structured Light Coding for 3D Reconstruction. In *Proceedings of IEEE International Conference on Computer Vision (ICCV '07)*, pages 1–6, Rio de Janeiro, Brazil.

[Averbuch and Keller, 2004] Averbuch, A. and Keller, Y. (2004). Fast motion estimation using bidirectional gradient methods. *Proceedings of IEEE Transactions on Image Processing*, 13:1042–1054.

[Bachta et al., 2009] Bachta, W., Renaud, P., Cuvillon, L., Laroche, E., Forgione, A., and Gangloff, J. (2009). Motion prediction for computer-assisted beating heart surgery. *IEEE Transactions on Biomedical Engineering*, 56(11):2551–2563.

[Bachta et al., 2007] Bachta, W., Renaud, P., Laroche, E., Gangloff, J., and Forgione, A. (2007). Cardiolock: an active cardiac stabilizer, first in vivo experiments using a new robotized device. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, volume 4791 of *Lecture Notes in Computer Science (LNCS)*, pages 78–85, Brisbane, Australia. Springer.

[Bader et al., 2007] Bader, T., Wiedermann, A., Roberts, K., and Hanebeck, U. D. (2007). Model-based motion estimation of elastic surfaces for minimally invasive cardiac surgery. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '07)*, pages 871–876, San Diego, USA.

[Baker and Matthews, 2002] Baker, S. and Matthews, I. (2002). Lucas-kanade 20 years on: A unifying framework: Part 1. Technical report, Robotics Institute, Carnegie Mellon University, Pittsburgh, USA.

[Bartoli, 2006] Bartoli, A. (2006). Direct image registration with gain and bias. In *Topics in Automatic 3D Modeling and Processing Workshop*, Verona, Italy.

[Bartoli et al., 2007] Bartoli, A., Perriollat, M., and Chambom, S. (2007). Generalized thin-plate spline warps. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pages 1–8, Minneapolis, USA.

[Bartoli and Zisserman, 2004] Bartoli, A. and Zisserman, A. (2004). Direct estimation of non-rigid registrations. In *Proceedings of the 15th British Machine Vision Conference*, volume 2, pages 899–908, Kingston, UK.

[Bebek and Cavusoglu, 2007a] Bebek, O. and Cavusoglu, M. C. (2007a). Intelligent control algorithms for robotic-assisted beating heart surgery. *IEEE Transactions on Robotics*, 23(3):468–480.

[Bebek and Cavusoglu, 2007b] Bebek, O. and Cavusoglu, M. C. (2007b). Whisker sensor design for three dimensional position measurement in robotic assisted beating heart surgery. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '07)*, pages 225–231, San Diego, USA.

[Benhimane et al., 2007] Benhimane, S., Ladikos, A., Lepetit, V., and Navab, N. (2007). Linear and quadratic subsets for template-based tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pages 1–6, Minneapolis, USA.

[Benhimane and Malis, 2004] Benhimane, S. and Malis, E. (2004). Real-time image-based tracking of planes using efficient second-order minimization. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '04)*, pages 943–948, Sendai, Japan.

[Berdat et al., 2004] Berdat, P. A., Müller, J. S. K., Kipfer, B., Eckstein, F., Immer, F. F., and Carrel, T. (2004). Totally arterial off-pump vs. on-pump coronary revascularization : comparison of early outcome. *Interactive Cardiovascular and Thoracic Surgery*, 3:176–181.

[Berkelman et al., 2003] Berkelman, P., Boidard, E., Cinquin, P., and Troccaz, J. (2003). Ler: The light endoscope robot. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '03)*, pages 2835–2840, Las Vegas, EUA.

[Bó et al., 2009] Bó, A. P. L., Poignet, P., and Geny, C. (2009). Filtering voluntary motion for pathological tremor compensation. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '09)*, pages 55–60, St. Louis, USA.

[Bogatyrenko et al., 2009] Bogatyrenko, E., Hanebeck, U. D., and Szabó, G. (2009). Heart surface motion estimation framework for robotic surgery employing meshless methods. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '09)*, pages 67 – 74, St. Louis, USA.

[Bookstein, 1989] Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(6):567–585.

[Bouguet, 2009] Bouguet, J.-Y. (2009). Camera calibration toolbox for matlab. `http://www.vision.caltech.edu/bouguetj/calib_doc/index.html`.

[Bourger et al., 2007] Bourger, F., Doignon, C., Zanne, P., and de Mathelin, M. (2007). A model-free vision-based robot control for minimally invasive surgery using esm tracking and pixels color selection. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '07)*, Rome, Italie.

[Boyd et al., 1999] Boyd, W. D., Desai, N. D., Rizzo, D. F. D., Novick, R. J., McKenzie, F. N., and Menkis, A. H. (1999). Off-pump surgery decreases post-operative complications and resource utilization in the elderly. In *Annals of Thoracic Surgery*, volume 68, pages 1490–1493.

[Burschka et al., 2005] Burschka, D., Li, M., Ishii, M., Taylor, R. H., and Hager, G. D. (2005). Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery. *Medical Image Analysis*, 9(5):413 – 426.

[Cagneau et al., 2007] Cagneau, B., Zemiti, N., Bellot, D., and Morel, G. (2007). Physiological motion compensation in robotized surgery using force feedback control. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '07)*, pages 1881–1886, Rome, Italy.

[CardioSense3D, 2009] CardioSense3D (2009). The Cardiosense3D project. `http://www-sop.inria.fr/CardioSense3D/`.

[Cavusoglu et al., 2005] Cavusoglu, M. C., Rotella, J., Newman, W. S., Choi, S., Ustin, J., and Sastry, S. S. (2005). Control algorithms for active relative motion cancelling for robotic assisted off-pump coronary artery bypass graft surgery. In *Proceedings of the 12th International Conference on Advanced Robotics (ICAR)*, pages 431–436, Seattle, USA.

[Cavusoglu et al., 2001] Cavusoglu, M. C., Williams, W., Tendick, F., and Sastry, S. S. (2001). Robotics for telesurgery : Second generation berkeley/ucsf laparoscopic telesurgical workstation and looking towards the future applications. In *Proc. of the 39th Allerton Con. on Communication, Control and Computing*.

[Cuvillon et al., 2006] Cuvillon, L., de Mathelin, M., and Forgione, A. (2006). Towards robotized beating heart TECABG: Assessment of the heart dynamics using high-speed vision. *Computer Aided Surgery*, 11(5):267–277.

[Delingette et al., 2006] Delingette, H., Pennec, X., Soler, L., Marescaux, J., and Ayache, N. (2006). Computational models for image guided, robot-assisted and simulated medical interventions. *Proceedings of the IEEE*, 94(9):1678– 1688.

[Dombre, 2005] Dombre, E. (2005). Projet gabie: Guidage actif basé sur l'imagerie echographique. In *ROBEA'05: Robotique et Entités Artificielles*, pages 49–56.

[Dominici et al., 2008] Dominici, M., Poignet, P., and Dombre, E. (2008). Compensation of physiological motion using linear predictive force control. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '08)*, pages 1173–1178, Nice, France.

[Duindam and Sastry, 2007] Duindam, V. and Sastry, S. (2007). Geometric motion estimation and control for robotic-assisted beating-heart surgery. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '07)*, pages 871–876, San Diego, USA.

[Dzwonczyk et al., 2005] Dzwonczyk, R., del Rio, C., Sun, B., Michler, R., and Howie, M. (2005). Devices used to expose the posterior coronary artery in opcabg surgery may cause ischemia. In *Proceedings of the 31th Bioengineering conference*, volume 2-3, pages 148–149.

[Eagle and Guyton, 2004] Eagle, K. A. and Guyton, R. A. (2004). Acc/aha 2004 guideline update for coronary artery bypass graft surgery. Technical report, American College of Cardiology/American Heart Association.

[Ernst et al., 2008] Ernst, F., Schlaefer, A., Dieterich, S., and Schweikard, A. (2008). A fast lane approach to lms prediction of respiratory motion signals. *Biomedical Signal Processing and Control*, 3(4):291 – 299.

[Franke et al., 2007] Franke, T. J., Bebek, O., and Cavusoglu, M. C. (2007). Improved prediction of heart motion using an adaptive filter for robot assisted beating heart surgery. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '07)*, pages 509–515, San Diego, USA.

[Franke et al., 2008] Franke, T. J., Bebek, O., and Cavusoglu, M. C. (2008). Prediction of heartbeat motion with a generalized adaptive filter. In *Proceedings of IEEE Conference on Robotics and Automation (ICRA '08)*, pages 2916–2921, Pasadena, USA.

[Fuchs et al., 1998] Fuchs, H., Livingston, M. A., Raskar, R., D. Colucci, K. K., State, A., Crawford, J. R., Rademacher, P., Drake, S. H., and Mayer, A. A. (1998). Augmented reality visualization for laparoscopic surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '98)*, volume 1496 of *Lecture Notes in Computer Science (LNCS)*, pages 934–943, Cambridge, USA. Springer.

[Fusiello et al., 2001] Fusiello, A., Trucco, E., and Verri, A. (2001). A compact algorithm for rectification of stereo pairs. *Machine Vision Applications*, (1):16–22.

[Gallagher et al., 1998] Gallagher, A. G., McClure, N., McGuigan, J., Ritchie, K., and Sheehy, N. P. (1998). An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy*, 30(7):617–620.

[Giannarou et al., 2009] Giannarou, S., Visentini-Scarzanella, M., and Yang, G. Z. (2009). Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI '09)*, pages 1059–1062, Boston, USA.

[Ginhoux et al., 2003] Ginhoux, R., Gangloff, J., de Mathelin, M., Soler, L., Sanchez, M. A., and Marescaux, J. (2003). Active filtering of physiological motion in robotized surgery using predictive control. *IEEE Transactions on Robotics*, 21(1):67–79.

[Gröger et al., 2002] Gröger, M., Ortmaier, T., Sepp, W., and Gerd, H. (2002). Tracking local motion on the beating heart. In Imaging, S. M., editor, *Visualization, Image-Guided Procedures and Display*, volume 4681, pages 233–241. SPIE.

[Hagn et al., 2008] Hagn, U., Nickl, M., Jörg, S., Tobergte, A., Kübler, B., Passig, G., Gröger, M., Fröhlich, F., Seibold, U., Konietschke, R., Le-Tien, L., Albu-Schäffer, A.,

Grebenstein, M., Ortmaier, T., and Hirzinger, G. (2008). Dlr mirosurge – towards versatility in surgical robotics. In *7. Jahrestagung der Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie e.V.*, pages 143 – 146.

[Hansen et al., 2008] Hansen, M. S., Larsen, R., Glocker, B., and Navab, N. (2008). Adaptive parametrization of multivariate b-splines for image registration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pages 1–8, Washington, USA.

[HarvardBiorobotics, 2009] HarvardBiorobotics (2009). 3D Ultrasound-Guided Robotic Motion Compensation for Beating Heart Intracardiac Surgery. `http://biorobotics.harvard.edu/`.

[Hayashibe et al., 2006] Hayashibe, M., Suzuki, N., and Nakamura, Y. (2006). Laser-scan endoscope system for intraoperative geometry aqusition and surgical robot safety management. *Medical Image Analysis*, (10):509–519.

[Heikkilä and Silven, 1997] Heikkilä, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pages 1106–1112, San Juan, Puerto Rico.

[Hoff et al., 2004] Hoff, L., Elle, O. J., Grimnes, M. J., Halvorsen, S., Alker, H. J., and Fosse, E. (2004). Measurements of heart motion using accelerometers. In *Proceedings of IEEE International Conference of the Engineering in Medicine and Biology Society (EMBS '04)*, pages 2049–2051, San Francisco, USA.

[Hu et al., 2007] Hu, M., Penney, G. P., Edwards, P. J., Figl, M., and Hawkes, D. J. (2007). 3D Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, volume 4791 of *Lecture Notes in Computer Science (LNCS)*, pages 68–77, Brisbane, Australia. Springer.

[Kappert et al., 2001] Kappert, U., Cichon, R., Schneider, J., Gulielmos, V., Ahmadzade, T., Nicolai, J., Tugtekin, S.-M., and Schueler, S. (2001). Technique of closed chest coronary artery surgery on the beating heart. *European Journal of Cardio-thoracic Surgery*, 20:765–769.

[Klesius et al., 2004] Klesius, A. A., Dzemali, O., Simon, A., Kleine, P., Abdel-Rahman, U., Herzog, C., Wimmer-Greinecker, G., and Moritz, A. (2004). Successful treatment of deep sternal infections following open heart surgery by bilateral pectoralis major flaps. *European Journal of Cardio-thoracic Surgery*, 25:218–223.

[Krupa et al., 2004] Krupa, A., Morel, G., and de Mathelin, M. (2004). Achieving high precision laparoscopic manipulation through adaptive force control. *Advanced Robotics*, 18(9):905–926.

[Lau et al., 2004] Lau, W., Ramey, N. A., Corso, J. J., Thakor, N. V., and Hager, G. D. (2004). Stereo-based endoscopic tracking of cardiac surface deformation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, volume 3217

of *Lecture Notes in Computer Science (LNCS)*, pages 494–501, Saint Malo, France. Springer.

[Lemma et al., 2005] Lemma, M., Mangini, A., Redaelli, A., and Acocella, F. (2005). Do cardiac stabilizers really stabilize? experimental quantitative analysis of mechanical stabilization. *Interactive CardioVascular and Thoracic Surgery*, 4:222–226.

[Lim and Yang, 2005] Lim, J. and Yang, M. (2005). A direct method for modeling non-rigid motion with thin plate spline. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pages 1196–1202, Washington, USA.

[Lo et al., 2008a] Lo, B., Chung, A. J., Stoyanov, D., Mylonas, G., and Yang, G. Z. (2008a). Real-time intra-operative 3D tissue deformation recovery. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI '08)*, pages 1387–1390, Paris, France.

[Lo et al., 2008b] Lo, B., Scarzanella, M., Stoyanov, D., and Yang, G. Z. (2008b). Belief propagation for depth cue fusion in minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '08)*, Lecture Notes in Computer Science (LNCS), pages 104–112, New York, USA. Springer.

[Lucas and Kanade, 1981] Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision (ijcai). In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pages 674–679.

[Madhani et al., 1998] Madhani, A. J., Niemeyer, G., and Jr., J. K. S. (1998). The black falcom : A teleoperated surgical instrument for minimally invasive surgery. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '98)*, pages 936–944, Victoria, Canada.

[Malis, 2004] Malis, E. (2004). Improving vision-based control using efficient second-order minimization techniques. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '04)*, Barcelona, Spain.

[Malis, 2007] Malis, E. (2007). An efficient unified approach to direct visual tracking of rigid and deformable surfaces. In *Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS '07)*, pages 2729–2734, San Diego, USA.

[Matas et al., 2002] Matas, J., Chum, O., Urba, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the 15th British Machine Vision Conference*, pages 384–396, Cardiff, UK.

[Mountney et al., 2007] Mountney, P., Lo, B., Thiemjarus, S., Stoyanov, D., and Yang, G. Z. (2007). A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, volume 4792 of *Lecture Notes in Computer Science (LNCS)*, pages 34–41, Brisbane, Australia. Springer.

[Mountney and Yang, 2008] Mountney, P. and Yang, G. Z. (2008). Soft tissue tracking for minimally invasive surgery: Learning local deformation online. In *Medical Image*

Computing and Computer-Assisted Intervention (MICCAI '08), volume 5242 of *Lecture Notes in Computer Science (LNCS)*, pages 364–372, New York, USA. Springer.

[Mourgues et al., 2001] Mourgues, F., Devernay, F., and Coste-manière, E. (2001). 3D reconstruction of the operating field for image overlay. In *IEEE and ACM International Symposium on Augmented Reality*, pages 191–192, New York, USA.

[MSS, 2009] MSS (2009). Ministère de la santé et des sports. `http://www.sante.gouv.fr`.

[Naik et al., 2003] Naik, M. J., Abu-Omar, Y., Alvi, A., Wright, N., Henderson, A., Channon, K., Forfar, J. C., and Taggart, D. P. (2003). Total arterial revascularisation as a primary strategy for coronary artery bypass grafting. *Postgraduate Medical Journal*, 79:43–48.

[Nakamura et al., 2001] Nakamura, Y., K., and Kawakami, H. (2001). Heartbeat synchronization for robotic cardiac surgery. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '01)*, volume 2, pages 2014–2019, Seoul, Korea.

[NHLBI, 2007] NHLBI (2007). National heart lung and blood institute. `http://www.nhlbi.nih.gov/health/dci/Diseases/chd/chd_what.html`.

[Noce et al., 2007] Noce, A., Triboulet, J., and Poignet, P. (2007). Efficient tracking of the heart using texture. In *Proceedings of IEEE International Conference of the Engineering in Medicine and Biology Society (EMBS '07)*, pages 4480–4483, Lyon, France.

[Ortmaier et al., 2005] Ortmaier, T., Groger, M., Boehm, D. H., Falk, V., and Hirzinger, G. (2005). Motion estimation in beating heart surgery. *IEEE Transactions on Biomedical Engineering*, 52(10):1729–1740.

[Ortmaier, 2003] Ortmaier, T. J. (2003). *Motion Compensation in Minimally Invasive Robotic Surgery*. PhD thesis, Technischen Universitat Munchen - TUM.

[Patel et al., 2002] Patel, N. C., Deodhar, A. P., Grayson, A. D., Pullan, D. M., Keenan, D. J. M., Hasan, R., and Fabri, B. M. (2002). Neurological outcomes in coronary surgery : Independent effect of avoiding cardiopulmonary bypass. In *Annals of Thoracic Surgery*, volume 74, pages 400–406.

[Patronik, 2008] Patronik, N. (2008). *A Miniature Mobile Robot for Precise and Stable Access to the Beating Heart*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[Peirs et al., 2004] Peirs, J., Clijnen, J., Reynaerts, D., Brussel, H. V., Herijgers, P., Corteville, B., and Boone, S. (2004). A micro optical force sensor for force feedback during minimally invasive robotic surgery. *Sensors and Actuators A: Physical*, 115(2-3):447 – 455.

[Pilet et al., 2007] Pilet, J., Lepetit, V., and Fua, P. (2007). Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision*, 2(76):109–122.

[Puskas et al., 1998] Puskas, J. D., Wright, C. E., Ronson, R. S., Brown, W. M., Gott, J. P., and Guyton, R. A. (1998). Off-pump multi-vessel coronary bypass via sternotomy is safe and effective. In *Annals of Thoracic Surgery*, volume 66, pages 1068–1072.

[Rajan, 2004] Rajan, B. R. S. (2004). History of cardiac surgery in the armed forces. *Indian Journal of thoracic and Cardiovascular Surgery*, 20:42–44.

[Richa et al., 2010a] Richa, R., Bó, A. P. L., and Poignet, P. (2010a). Beating heart motion prediction for robust visual tracking. In *Proceedings of IEEE Conference on Robotics and Automation (ICRA '10) - to appear*, Anchorage, USA.

[Richa et al., 2010b] Richa, R., Poignet, P., and Liu, C. (2010b). Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model. *The International Journal of Robotics Research (IJRR) – Special Issue on Robot Vision*, 29(2-3):218–230.

[Rininsland, 1999] Rininsland, H. (1999). Artemis: A telemanipulator for cardiac surgery. *European Journal of Cardio-thoracic Surgery*, 16:106–111.

[Riviere et al., 2006] Riviere, C. N., Gangloff, J., and de Mathelin, M. (2006). Robotic compensation of biological motion to enhance surgical accuracy. *Proceedings of the IEEE*, 94(9):1705–1716.

[Rohl et al., 2009] Rohl, S., Speidel, S., Sudra, G., Gehrig, T., Muller-Stich, B. P., Gutt, C., and Dillmann, R. (2009). A surface model for intraoperative soft tissue registration. In *Proceedings of the International Congress on Computer Assisted Radiology and Surgery (CARS '09)*, pages 106–107, Berlin, Germany.

[Sauvée et al., 2007] Sauvée, M., Noce, A., Poignet, P., Triboulet, J., and Dombre, E. (2007). Three-dimensional heart motion estimation using endoscopic monocular vision system: From artificial landmarks to texture analysis. *Biomedical Signal Processing and Control*, 2(3):199 – 207.

[Schnabel et al., 2001] Schnabel, J., Rueckert, D., Quist, M., Blackall, J., Castellano-Smith, A., Hartkens, T., Penney, G., Hall, W., Liu, H., Truwitt, C., Gerritsen, F., Hill, D., and Hawkes, D. (2001). A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, Lecture Notes in Computer Science (LNCS), pages 573–581, Utrecht, The Netherlands. Springer.

[Seibert et al., 2007] Seibert, R., Ramsey, C., Harris, C., Garvey, D., and Hines, W. (2007). Respiration motion prediction using time-delay kernel regression modeling. *International Journal of Radiation Oncology/Biology/Physics*, 69(3, Supplement 1):S128 – S129. Proceedings of the American Society for Therapeutic Radiology and Oncology 49th Annual Meeting.

[Shekar, 2006] Shekar, P. S. (2006). On-pump and off-pump coronary artery bypass grafting. *Circulation*, 113:51–52.

[Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pages 593–600, Seattle, USA.

[Silveira and Malis, 2007] Silveira, G. and Malis, E. (2007). Real-time visual tracking under arbitrary illumination changes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pages 1–6, Minneapolis, USA.

[Simon, 2006] Simon, D. (2006). *Optimal State Estimation: Kalman, "H_∞" and Nonlinear Approaches*. John Wiley & Sons, Inc.

[Stegmann and Larsson, 2003] Stegmann, M. B. and Larsson, H. B. W. (2003). Motion-compensation of cardiac perfusion mri using a statistical texture ensemble. In *Proceedings of the international conference on Functional Imaging and Modeling of the Heart (FIMH '03)*, volume 2674, Lyon, France.

[Stoyanov et al., 2005a] Stoyanov, D., Darzi, A., and Yang, G. Z. (2005a). A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery. *Computer Aided Surgery*, 4(10):199–208.

[Stoyanov et al., 2005b] Stoyanov, D., Mylonas, G. P., Deligianni, F., Darzi, A., and Yang, G. Z. (2005b). Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, volume 3750 of *Lecture Notes in Computer Science (LNCS)*, pages 139–146, Palm Springs, USA. Springer.

[Stoyanov and Yang, 2007] Stoyanov, D. and Yang, G. Z. (2007). Stabilization of image motion for robotic assisted beating heart surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, volume 4791 of *Lecture Notes in Computer Science (LNCS)*, pages 417–424, Brisbane, Australia. Springer.

[Stoyanov and Yang, 2009] Stoyanov, D. and Yang, G. Z. (2009). Soft-tissue deformation tracking for robotic assisted minimally invasive surgery. In *Proceedings of IEEE International Conference of the Engineering in Medicine and Biology Society (EMBS '09)*, pages 254–257, Boston, USA.

[Thakral et al., 2001] Thakral, A., Wallace, J., Tomlin., D., Seth, N., and Thakor, N. V. (2001). Surgical motion adaptive robotic technology (s.m.a.r.t): Taking the motion out of physiological motion. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, Lecture Notes in Computer Science (LNCS), pages 317–325, Utrecht, The Netherlands. Springer.

[Trejos et al., 1999] Trejos, A. L., Salcudean, S. E., Sassani, F., and Lichtenstein, S. (1999). On the feasibility of a moving support for surgery on the beating heart. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '99)*, volume 1679 of *Lecture Notes in Computer Science (LNCS)*, pages 1088–1098, Cambridge, UK. Springer.

[Visentini-Scarzanella et al., 2009] Visentini-Scarzanella, M., Mylonas, G. P., Stoyanov, D., and Yang, G. Z. (2009). *i*-BRUSH: A Gaze-Contigent Virtual Paintbrush for Dense

3D Reconstruction in Robotic Assisted Surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '09)*, Lecture Notes in Computer Science (LNCS), pages 353–360, London, UK. Springer.

[WHO, 2009] WHO (2009). World health organization. `http://www.who.int/en/`.

[Yuen et al., 2008] Yuen, S. G., Novotny, P. M., and Howe, R. D. (2008). Quasiperiodic predictive filtering for robot-assisted beating heart surgery. In *Proceedings of IEEE Conference on Robotics and Automation (ICRA '08)*, pages 1050–1057, Pasadena, USA.

[Yuen et al., 2009] Yuen, S. G., Yip, M. G., Vasilyev, N. V., Perrin, D. P., del Nido, P. J., and Howe, R. D. (2009). Robotic force stabilization for beating heart intracardiac surgery. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI '09)*, volume 5761 of *Lecture Notes in Computer Science (LNCS)*, pages 26–33, London, UK. Springer.

[Zarrad, 2006] Zarrad, W. (2006). *Télé-opération avec retour d'effort pour la chirurgie mini-invasive*. PhD thesis, LIRMM - Université Montpellier 2.

[Zemiti, 2005] Zemiti, N. (2005). *Commande en Effort des Systèmes Robotiques pour la Chirurgie Mini-Invasive*. PhD thesis, Université Paris 6.

[Zhang, 2000] Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(11):1330–1334.