



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Estimating and understanding motion : from diagnostic to robotic surgery

Angélica Ivone Avilés Rivero

ADVERTIMENT La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del repositori institucional UPCommons (<http://upcommons.upc.edu/tesis>) i el repositori cooperatiu TDX (<http://www.tdx.cat/>) ha estat autoritzada pels titulars dels drets de propietat intel·lectual **únicament per a usos privats** emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei UPCommons o TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a UPCommons (*framing*). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del repositorio institucional UPCommons (<http://upcommons.upc.edu/tesis>) y el repositorio cooperativo TDR (<http://www.tdx.cat/?locale-attribute=es>) ha sido autorizada por los titulares de los derechos de propiedad intelectual **únicamente para usos privados enmarcados** en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio UPCommons. No se autoriza la presentación de su contenido en una ventana o marco ajeno a UPCommons (*framing*). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the institutional repository UPCommons (<http://upcommons.upc.edu/tesis>) and the cooperative repository TDX (<http://www.tdx.cat/?locale-attribute=en>) has been authorized by the titular of the intellectual property rights **only for private uses** placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading nor availability from a site foreign to the UPCommons service. Introducing its content in a window or frame foreign to the UPCommons service is not authorized (*framing*). These rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT POLITÈCNICA DE CATALUYA

DOCTORAL PROGRAM:
AUTOMATIC CONTROL, ROBOTICS AND VISION

**Estimating and Understanding Motion: From
Diagnostic to Robotic Surgery**

Angélica Ivone Avilés Rivero

ADVISOR:
ALICIA CASALS GELPÍ

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

APRIL 2017

Acknowledgements

This thesis was supported by a FPU national scholarship from the Spanish Ministry of Education with reference AP2012-1943 in a four years basis.

During these four years, I have been fortunate to work with very supportive and inspiring people that made me grow more than I could imagine both professionally and personally. Next paragraphs try, as much as words can, to express my gratitude to all the people that helped me during my PhD journey.

First and foremost I want to thank my advisor Alicia Casals for her endless support and motivation, and for giving me the freedom to go beyond my imagination. I admire her ability to manage research and appreciate all her time and contributions to make this journey productive and enjoyable. I also learned from her skills for improving significantly the writing and presentation of our papers. She is a clear example of how a scientist should be. Apart from the professional side, I enjoyed and learn a lot from her talks about what to expect after the PhD and politics.

I also want to thank Habib Ammari, firstly, for trusting me with the project related to the Ultrafast Ultrasound Imaging and for his very warm welcoming to the department of mathematics at the ETHz. Secondly, I am deeply grateful for his enthusiasm and always-positive comments that kept me motivated and encouraged for pursuing unique ideas. I am amazed for his ability to come with new solutions for any given problem. Within this project, I had also the opportunity to spend amazing months at the department of mathematics and their applications at the Ecole Normale Supérieure (ENS) where I worked with Thomas Widlak. He was an incredible mentor during my time at ENS and after ours meetings, I always walked away very motivated and positive, and with more clear ideas of how to shape our solution. Also, it could had not been possible to achieve a detailed proof of concept without Maartje Nillesen. She did a great job with the simulated data, this together, with her vast knowledge in the area of ultrasound imaging were the perfect complement to our project. I also appreciate

all her always positive feedbacks. Also, many thanks to Timothée, Wenlong and Matias for making my time at ENS enjoyable.

I was also fortunate to work with fascinating people at the George Washington University (GWU). Firstly, I want to thank James Hanh for hosting me during few months at the department of computer science at GWU. His ability to find the key questions to keep the right path is compelling. I also want to express my gratitude to John Philbeck from the Department of Psychology at GWU. His discussions allowed me to have a better understanding of how to handle a user study and its implications in terms of perception and cognition. I could not be more grateful to Naji Younes, from the department of Epidemiology and Biostatistics at GWU, for his incredible statistical explanations. Also, all guys that made my time there enjoyable: Wei, Yan, Yao, Xiao, Rehab and Shang. Particularly, I am deeply grateful to Samar Alsaleh for our amazing collaborations and all our rich scientific conversations that made us come up with unexpected solutions. Moreover, for showing me the computer graphics world, it was amazing to see her magic coding ability. Importantly, I am grateful to her for helping me to correct this thesis. Apart from the work side, I appreciate her invaluable and true friendship that taught me how to be a better person.

It was a pleasure to work with Pilar Sobrevilla and Eduard Montseny here at UPC. I appreciate the many coffees that we shared discussing work and for introducing me to the fuzzy theory world. They both created a perfect balance between theory and practice. Aside from the professional area, I am very thankful for their advices and care. I also want to thank Stella P. Raventos from the Josep Trueta University Hospital. I appreciate her medical explanations and her help in coordinating the user study at different hospitals. Also, to my lab mates at UPC with whom I shared infinite number of coffees: Xavi, Joan, Albert, Eduard, Oriol, and in particular, to my desk neighbour Narcís.

Finally yet importantly, I am infinity grateful to my sisters and best mates Viry and Karla. They have been a key support during my graduate studies. I also want to thank the person who encouraged me to pursue all my dreams and to do my best in all life matters, the person who taught me that my only limitation is myself, and from whom I always have received an unconditional support, my mother, Isabel. To her I dedicate this thesis.

Abstract

Estimating and understanding motion from an image sequence is a central topic in computer vision. The high interest in this topic is because we are living in a world where many events that occur in the environment are dynamic. This makes motion estimation and understanding a natural component and a key factor in a widespread of applications including object recognition, 3D shape reconstruction, autonomous navigation and medical diagnosis.

Particularly, we focus on the medical domain in which understanding the human body for clinical purposes requires retrieving the organs' complex motion patterns, which is in general a hard problem when using only image data. In this thesis, we cope with this problem by posing the question – *How to achieve a realistic motion estimation to offer a better clinical understanding?* We focus this thesis on answering this question by using a variational formulation as a basis to understand one of the most complex motions in the human's body, the heart motion, through three different applications: (i) cardiac motion estimation for diagnostic, (ii) force estimation and (iii) motion prediction, both for robotic surgery.

Firstly, we focus on a central topic in cardiac imaging that is the estimation of the cardiac motion. The main aim is to offer objective and understandable measures to physicians for helping them in the diagnostic of cardiovascular diseases. We employ ultrafast ultrasound data and tools for imaging motion drawn from diverse areas such as low-rank analysis and variational deformation to perform a realistic cardiac motion estimation. The significance is that by taking low-rank data with carefully chosen penalization, synergies in this complex variational problem can be created. We demonstrate how our proposed solution deals with complex deformations through careful numerical experiments using realistic and simulated data.

We then move from diagnostic to robotic surgeries where surgeons perform delicate procedures remotely through robotic manipulators without directly interacting with the patients. As a result, they lack force feedback, which is

an important primary sense for increasing surgeon-patient transparency and avoiding injuries and high mental workload. To solve this problem, we follow the conservation principles of continuum mechanics in which it is clear that the change in shape of an elastic object is directly proportional to the force applied. Thus, we create a variational framework to acquire the deformation that the tissues undergo due to an applied force. Then, this information is used in a learning system to find the nonlinear relationship between the given data and the applied force. We carried out experiments with in-vivo and ex-vivo data and combined statistical, graphical and perceptual analyses to demonstrate the strength of our solution.

Finally, we explore robotic cardiac surgery, which allows carrying out complex procedures including Off-Pump Coronary Artery Bypass Grafting (OPCABG). This procedure avoids the associated complications of using Cardiopulmonary Bypass (CPB) since the heart is not arrested while performing the surgery on a beating heart. Thus, surgeons have to deal with a dynamic target that compromise their dexterity and the surgery's precision. To compensate the heart motion, we propose a solution composed of three elements: an energy function to estimate the 3D heart motion, a specular highlight detection strategy and a prediction approach for increasing the robustness of the solution. We conduct evaluation of our solution using phantom and realistic datasets.

We conclude the thesis by reporting our findings on these three applications and highlight the dependency between *motion estimation* and *motion understanding* at any dynamic event, particularly in clinical scenarios.

Keywords: Motion Estimation, Motion Understanding, Cardiac Imaging, Robotic-Assisted Surgery, Topology Preservation, Supervised Learning

Resum

L'estimació i comprensió del moviment dins d'una seqüència d'imatges és un tema central en la visió per ordinador, el que genera un gran interès perquè vivim en un entorn ple d'esdeveniments dinàmics. Per aquest motiu és considerat com un component natural i factor clau dins d'un ampli ventall d'aplicacions, el qual inclou el reconeixement d'objectes, la reconstrucció de formes tridimensionals, la navegació autònoma i el diagnòstic de malalties.

En particular, ens situem en l'àmbit mèdic en el qual la comprensió del cos humà, amb finalitats clíniques, requereix l'obtenció de patrons complexos de moviment dels òrgans. Aquesta és, en general, una tasca difícil quan s'utilitzen només dades de tipus visual. En aquesta tesi afrontem el problema plantejant-nos la pregunta - *Com es pot aconseguir una estimació realista del moviment amb l'objectiu d'oferir una millor comprensió clínica?* La tesi se centra en la resposta mitjançant l'ús d'una formulació variacional com a base per entendre un dels moviments més complexos del cos humà, el del cor, a través de tres aplicacions: (i) estimació del moviment cardíac per al diagnòstic, (ii) estimació de forces i (iii) predicció del moviment, orientant-se les dues últimes en cirurgia robòtica.

En primer lloc, ens centrem en un tema principal en la imatge cardíaca, que és l'estimació del moviment cardíac. L'objectiu principal és oferir als metges mesures objectives i comprensibles per ajudar-los en el diagnòstic de les malalties cardiovasculars. Fem servir dades d'ultrasons ultraràpids i eines per al moviment d'imatges procedents de diverses àrees, com ara l'anàlisi de baix rang i la deformació variacional, per fer una estimació realista del moviment cardíac. La importància rau en que, en prendre les dades de baix rang amb una penalització acurada, es poden crear sinergies en aquest problema variacional complex. Mitjançant acurats experiments numèrics, amb dades realístiques i simulades, hem demostrat com les nostres propostes solucionen deformacions complexes.

Després passem del diagnòstic a la cirurgia robòtica, on els cirurgians realitzen procediments delicats remotament, a través de manipuladors robòtics, sense

interactuar directament amb els pacients. Com a conseqüència, no tenen la percepció de la força com a resposta, que és un sentit primari important per augmentar la transparència entre el cirurgià i el pacient, per evitar lesions i per reduir la càrrega de treball mental. Resolem aquest problema seguint els principis de conservació de la mecànica del medi continu, en els quals està clar que el canvi en la forma d'un objecte elàstic és directament proporcional a la força aplicada. Per això hem creat un marc variacional que adquireix la deformació que pateixen els teixits per l'aplicació d'una força. Aquesta informació s'utilitza en un sistema d'aprenentatge, per trobar la relació no lineal entre les dades donades i la força aplicada. Hem dut a terme experiments amb dades in-vivo i ex-vivo i hem combinat l'anàlisi estadístic, gràfic i de percepció que demostren la robustesa de la nostra solució.

Finalment, explorem la cirurgia cardíaca robòtica, la qual cosa permet realitzar procediments complexos, incloent la cirurgia coronària sense bomba (off-pump coronary artery bypass grafting o OPCAB). Aquest procediment evita les complicacions associades a l'ús de circulació extracorpòria (Cardiopulmonary Bypass o CPB), ja que el cor no s'atura mentre es realitza la cirurgia. Això comporta que els cirurgians han de tractar amb un objectiu dinàmic que compromet la seva destresa i la precisió de la cirurgia. Per compensar el moviment del cor, proposem una solució composta de tres elements: un funcional d'energia per estimar el moviment tridimensional del cor, una estratègia de detecció de les reflexions especulars i una aproximació basada en mètodes de predicció, per tal d'augmentar la robustesa de la solució. L'avaluació de la nostra solució s'ha dut a terme mitjançant conjunts de dades sintètiques i realistes.

La tesi conclou informant dels nostres resultats en aquestes tres aplicacions i posant de relleu la dependència entre l'estimació i la comprensió del moviment en qualsevol esdeveniment dinàmic, especialment en escenaris clínics.

Paraules clau: Estimació del Moviment, Comprensió del Moviment, Imatge Cardíaca, Cirurgia Robòtica, Preservació de la Topologia, Aprenentatge Supervisat

Table of contents

List of figures	xiii
List of tables	xxi
1 Introduction	1
1.1 Contributions	5
1.2 Publications	7
1.3 Thesis Overview	9
2 Robust Cardiac Motion Estimation using Ultrafast Ultrasound Data: A Low-Rank-Topology-Preserving Approach	13
2.1 Ultrafast Ultrasound Imaging: Beyond the Human Eye	16
2.2 Preserving Diffeomorphic Features	18
2.3 A Low-Rank-Topology-Preserving Approach	20
2.4 Deformation Recovery	23
2.5 Topology Preservation	26
2.6 Experimental Results	29
2.6.1 Subjects and acquisition	29
2.6.2 Validation scheme	30
2.6.3 Results	31
2.7 Conclusions and Future Work	45
3 Towards Retrieving Force Feedback in Robotic-Assisted Surgery: A Supervised Neuro-Recurrent-Vision Approach	47
3.1 Vision-based Force Estimation	51
3.2 3D Deformable Shape Recovery	52
3.2.1 Robust 3D Shape Recovery	55
3.3 Retrieving Force Feedback	58
3.3.1 Force Estimation: Supervised Recurrent Learning	58

3.4	Experimental Results	66
3.4.1	Data Description	66
3.4.2	Tasks Description	67
3.4.3	Evaluation Scheme	68
3.4.4	Results and Discussion	70
3.5	Conclusions and Future Work	80
4	From Motion Estimation to Clinical Evaluation: A Perception	
	Experimental Study	83
4.1	Sensory Substitution in Teleoperation	86
4.2	Aim of this Work	89
4.3	Perceptual Study	90
4.3.1	Subjects Description	91
4.3.2	Visualizations Description	91
4.3.3	Experimental Procedure	95
4.4	Experimental Results	97
4.4.1	Evaluation Scheme	97
4.4.2	Analysis and Results	98
4.5	Conclusion	105
5	Sliding to Predict: Improving Vision-Based Beating Heart	
	Motion Estimation by Modeling Temporal Interactions	107
5.1	Challenges in Vision-Based Beating Heart Motion Estimation	110
5.2	Towards Cardiac Motion Estimation	112
5.2.1	Specular Reflection Elimination	113
5.2.2	Cardiac Motion Estimation	114
5.2.3	Sliding to Predict: Improving Cardiac Motion Estimation	117
5.3	Experimental Results	121
5.3.1	Cardiac Data Description	121
5.3.2	Evaluation Scheme	121
5.3.3	Results and Discussion	123
5.4	Conclusions	129
6	Concluding Remarks	131
6.1	Future Work	133
6.2	Beyond Medical Applications	134
	References	137

Table of contents	xi
-------------------	----

Appendix A Mathematical Proofs	157
---------------------------------------	------------

Appendix B Estimation Theory	159
-------------------------------------	------------

List of figures

1.1	Estimating and understanding motion is necessary in different applications. (a) Surface reconstruction is improved by taking temporal features which includes specular highlights. (b) Object recognition of dynamic objects is possible by knowing the object changes over time.	2
1.2	Top row displays everyday deformable objects in which the environment conditions are less restrictive in comparison with those illustrated in the middle row that comes from a clinical setup. Bottom row shows difficulties associated with clinical data.	3
2.1	Typical ultrasound acquisition setup: High-frequency waves allow capturing the view of inner organs such as the heart. Cardiac motion can be estimated by computing the spatial correspondence between time frames.	14
2.2	Overview of our proposed approach. (From left to right) an ultrafast ultrasound cardiac sequence is acquired, this data is then represented in low-rank in order to speed up the solution and reduce noise. Later on, cardiac motion is computed enforcing topology preservation which allows keeping the anatomical structure of the heart. Finally, an analysis of the results is offered.	15
2.3	Evolution of ultrasound imaging techniques over time, from real-time up to ultrafast imaging. Figure reproduced from [33]. . .	17
2.4	Comparison between Conventional and Ultrafast Ultrasound acquisition. Left-side shows conventional acquisition in which a full image is generated for each transmitted pulse whereas the right-side shows an image is generated in a single transmission by computing multiple lines in parallel.	18

2.5	Left: Singular values $\sigma_{k+1} = \ \mathbf{C} - \mathbf{C}_k\ ^2$. Right: CPU time to compute the rank- k approximation \mathbf{C}_k	22
2.6	Top row shows an original and denoised frame after applying low-rank process along with the rejected space. Bottom row, A and B, show zoom in views of the same frames in which we can see that both noise and some artifacts were removed.	23
2.7	When topology-preserving is not enforced, unrealistic transformations can appear in the result. A way to ensure topology preservation is by checking the Jacobian determinant $ J $. When $ J $ is equal to 1 then the volume is preserved. Small positive or large positive numbers of $ J $ result in contractions or expansions. But having $ J \in (-\infty, 0)$ can result in distortions, overlapping, and creation of new structures.	27
2.8	Sample frames of the raw data extracted from the two datasets used for evaluating our approach.	29
2.9	The first row shows the convergence history of the complete sequence (1000 frames) using three different pre-processing techniques, while the second row shows the number of iterations it took to find the minimum for few of those frames. Box-plot at the right side shows the CPU time comparison of the different techniques.	32
2.10	(A) Resulted transformations, during complex deformations, without applying topology preservation. Highlighted areas denote structure violations that are more clearly displayed in the zoom-in views (A.1). The resulted Jacobian determinant are shown in (A.2). Resulted transformations after applying topology preservation are shown in (B) and can be compared with (A), in which (B.1) and (B.2) show that they keep the mesh structures with most of the Jacobian determinant staying at 1.	34
2.11	(From right to left) Accumulated displacement for the apical view of the left ventricle. Few samples of the approximated axial and lateral displacements (top and bottom) are compared against the ground truth. Left side shows the Jacobian determinant of the same sample frames which reflects preservation of the anatomy.	37
2.12	Numerical comparison (in mm) between the real and estimated displacement values using Root-Mean-Square Error (RMSE).	38

2.13	Mean accumulated displacement of the seven segments of the left ventricle. Blue circles make reference to the axial displacement while red squares refer to the lateral displacement.	39
2.14	(Top) Radial and longitudinal strain profiles of the left ventricle. These profiles are evaluated by their sign where negative values reflect shortening and positive ones reflect stretching. (Bottom) Few frames of the cardiac cycle showing the radial strain. . . .	40
2.15	Noise reduction achieved by low-rank representation. Part (I) at first column shows two noisy input frames while the next three columns show the denoised sequences, the removed noise, and the space where the eliminated noise lies. Part (II) shows the error and computational time of the rank-k approximation.	41
2.16	Part A shows surface deformation without topology preservation while the same deformation is shown in B after enforcing topology preservation via our penalization term. Part C shows the estimated displacements of different parts of the heart along with the strain magnitude.	43
3.1	(A) shows tool-tissue interaction during Robotic-Assisted Surgery which lacks force feedback that informs the surgeon about the amount of applied force. (B) shows the observable displacements after applying a force, which we obtain using a sensorless approach that relies, in part, on computing the 3D shape of the tissue over time.	48
3.2	Flowchart of our approach for estimating applied forces in robotic surgical systems. We first propose a visual approach to compute the deformation structure over time. Then, the available information is used as input to an artificial neural network which accurately estimates the applied force.	50
3.3	(a) The 3-dimensional tissue surface is reconstructed from the projections of homologue points on the left and right lattices. (b) Illustration of how tissue deformation is directly proportional to the applied force.	53
3.4	Specular highlights cause major tracking disturbance. We deal with this issue using a real-time detection and inpainting approach that accurately recovers a specular-free image.	55

3.5	Surgical tools can partially occlude the tracked region of interest which affects the 3D shape recovery over time (Left side). Right side shows a side view of occluded lattice regions from different views.	56
3.6	3D deformation structure of the tissue, obtained by our vision approach, plotted at different time instants.	57
3.7	Left-side shows the structure of a biological neuron of the human brain while the right-side shows an artificial neuron that imitates the functioning of the biological one.	59
3.8	Left side a simple recurrent neural network while right side shows its unfolded version through time.	60
3.9	Estimation of the applied forces is achieved by means of a RNN in which three types of output units can be identified (zoom in the upper row). Those units with delayed feedback save past information that helps to increase accuracy. Additionally, at the right side a visualization of the network over time is displayed. . .	61
3.10	Three dimensional illustration of the activation functions used in the architecture.	62
3.11	In order to estimate the applied force, we use an architecture based on LSTM-RNN (part A) which combines basic units with cells. Part B shows a single cell block in detail and shows that each of the cells is composed of a set of units that enforce constant error flow which helps stabilizing force estimation over time. Additionally, part C shows an illustration of the hidden layer with 10 cells over time.	65
3.12	The realistic surgical setting, with typical RAMIS surgical setup, used to obtain the two ex-vivo datasets. The force sensor is used to obtain the ground truth to validate our estimation.	67
3.13	Raw data of the three different datasets used to evaluate our proposal (one in-vivo and two ex-vivo).	68
3.14	(a) Typical way to access the patient during a RAMIS. (b) Illustration of the palpation and exploration surgical tasks used to test the efficiency of our solution.	69

3.15	Tissue deformation that result from applying a force at different time instants is illustrated in parts (A) and (B) along with the recovered 3D deformable structure using our proposed visual approach. Finally, plots at part (C) show a comparison between the computed displacement (at contact point) in X,Y,Z directions against the reference measurements given by the geometry of motion of the robot from dataset II. The zoom-in views demonstrate the high estimation accuracy of our approach even during complex deformation as it can capture small (I-II) and large displacements (III). It also eliminates the noise in the geometry of motion as shown in (IV).	71
3.16	Illustration of tissue deformation that result from applying force at different time instants along with the 3D deformable structure recovered using our proposed visual approach. Our proposal was tested under different variation of illumination, occlusions and complex deformation.	72
3.17	Optimization plots resulted from our energy functional for different cases in which retrieving the 3D shape is challenging including complex deformations and change of illumination.	72
3.18	These linear regression plots show the associated strength between the real (target) and estimated force (output) measurements of both training and test datasets. In both sets, the points fit a line showing a tight relationship between the measurements and demonstrating the accuracy of the force estimation.	74
3.19	Plots in top part show the real force measures, in X,Y and Z directions, and those estimated by our approach. Bottom plots illustrate the RMSE results in all directions.	75
3.20	Stability criteria is shown in these plots using the ex-vivo datasets where the estimated and real force measures are plotted at different time intervals of a longer period of time.	76
3.21	Retrieved displacements of the four immediate neighboring points are plotted first without voting process correction (top) then with voting process (bottom).	79
3.22	(From left to right) Comparison of displacements, at contact point, between real-geometric measure and visual approach with and without V-ANFIS. Zoom-In displacements are also shown in order to observe the improvement when V-ANFIS is applied.	80

4.1	A typical teleoperated robotic surgical system using a master-slave configuration. At the master side, surgeon is provided with a 3D patient view and is able to perform the procedure using finger controls and foot pedals. All surgeon's actions are reproduced by the slave which holds the surgical instruments.	84
4.2	Examples of surgical tasks where knowing the applied force is relevant and helps decreasing the procedure completion time and avoiding injuries.	85
4.3	(From left to right) The four visualizations used to carry out our experiments at different time instants. The color-coding used to indicate the level of risk according to the magnitude of the applied force.	92
4.4	Sample frames from four datasets with the force feedback visual cue displayed.	94
4.5	Expert and Novices preference level per human factor. Each plot shows the percentage of acceptance of each system.	99
4.6	Global view of the obtained results showing experts vs novices responses of each systems.	100
4.7	Plots show the percentage of positive responses that each system received from the complete population per human factor.	100
4.8	(From left to right) Total responses evaluating each system received from the whole population. Results obtained from a post-hoc test for multiple comparison.	101
4.9	(From left to right) Population in our experiments comes from four specialities which we divided in two subgroups: experts and novices. Distribution of the users preference in which 95% of the novices and 100% of the experts preferred visual feedback over no feedback.	101
4.10	The advantages and disadvantages of each visualization system as reported by the users.	103
5.1	Left side shows a mechanical stabilizer called Octopus TM Nuvo (figure reproduced from [137]) for heart motion cancellation. Compared with these type of devices, the right side illustrates an alternative approach that tracks the area of interest using the endoscopic camera.	108
5.2	Top row displays samples of typical endoscopic images while middle row highlight.	111

5.3	Overview of our proposed approach composed of three main parts. (From left to right) The image sequence acquired from the endoscope at the slave side is passed to the first step which eliminates the specular highlight artifact. The specular-free images then go through our cardiac motion estimation step which recovers a 3D deformable surface of the region of interest. Finally, the last step guarantees information at all time by predicting the motion in cases where there is occlusion.	112
5.4	Specular highlights detection and inpainting results of our proposed algorithm on four different datasets. From left to right: input image; detected specular regions and information retrieval via Sobelev inpainting.	113
5.5	From left to right. Results obtained from our energy function with different number of control points. CPU time reported during the optimization process.	115
5.6	3D Diffeomorphic surface reconstruction from the projection of the lattice points defined at each stereo-pair image.	116
5.7	A toy example that illustrates how we restructure our sequential data to be used with a standard supervised learning approach .	118
5.8	Top part illustrates the architectures of both RBM and CRBM. The left-bottom part shows illustration of how we use the reconstructed heart motion as an input for CRBM while the right side shows the accumulated lattice points over time. . . .	119
5.9	Sample frames of the raw data from the two datasets used in our experimentation	121
5.10	Top part shows results from our specularity elimination approach on three different medical datasets while the bottom part shows zoom-in views of the inpainting results along with signal-to-noise ratio SNR plots	122
5.11	For each dataset from top to bottom: example frames of the input raw data, accumulated displacement of the reconstructed 3D heart at different time instances, and visualization of the recovered region of interest.	123

5.12	Left side shows the Jacobian Determinant results of our vision based approach, with and without applying our topology preservation term, in two different cases: the retrieval of complex deformation and the under illumination variation. The right side shows the convergence results of our optimization process on the two datasets while using the topology preservation term	124
5.13	The motion of a point of interest over time used in the prediction stage.	125
5.14	Estimated vs predicted motions in the three direction using NARX predictor over 200 frames	126
5.15	Estimated vs predicted motions in the three direction using EKF predictor over 200 frames	127
5.16	Estimated vs predicted motions in the three direction using CRBM predictor over 200 frames	128
5.17	Numerical comparison of the three prediction models in mm between the target and predicted values using RMSE	129
6.1	(a) Our proposed approach for achieving a specular-free object was evaluated with synthetic data. (B) It first creates a superpixel representation of the image domain to reduce computational time. (C) By restricting the searching to only key areas in the temporal dimension, we can efficiently obtain a specular-free object. . . .	134

List of tables

2.1	Decomposition of the Casorati matrix \mathbf{C} and the rank 100-approximation \mathbf{C}_{100}	22
2.2	Jacobian determinant conditions for topology preservation . . .	28
2.3	Performance comparison between our proposed approach and other state of the art approaches	32
2.4	Performance analysis: low-rank vs. full-rank data and their reaction to different degrees of topology preservation (see text for discussion).	35
2.5	Performance analysis of full vs low rank for different cases of our variational framework	44
3.1	Summary of Activation Functions used in this Chapter	62
3.2	Residual Error evaluation of our deformation approach	70
3.3	Statistical nonparametric analysis of our proposal to estimate the applied forces. It takes into consideration the ex-vivo datasets and the real measure.	75
3.4	The description of the architecture used in our V-ANFIS approach	79
3.5	Performance analysis of existing and proposed approaches. . . .	80
4.1	Color-coding used at each visualization to indicate the level of risk depending on the force applied	95
4.2	Questionnaire used to evaluate each visualization option based on five human factor criteria	96
4.3	Statistical nonparametric analysis of the results obtained from the experts and novices preferences. Left side shows the p-values while right side shows the adjusted ones.	99

“Above all, don’t fear difficult moments. The best comes from them.”

Rita Levi-Montalcini

1

Introduction

Human’s biological vision system is adept at dealing with the complex and dynamic changes produced in the world around us. For many years, scientists have attempted to achieve similar performance with computer vision systems. Many efforts have been put to enable computers to *estimate and understand* objects’ motion to be aware of surrounding events. *Motion estimation* can be defined as the process of determining the necessary transformations that describe the deformation of the image domain as a result of changes produced in adjacent frames of an image sequence. Once this process is carried out, a natural question that arises is – What do those sets of transformations tell us about the scene? The response to this question comes from *motion understanding* as it gives an explicit representation of the estimated motion that allows creating an interpretation of the dynamic scene to realize what is happening.

Estimating and understanding motion from an image sequence still remains a challenging problem in computer vision and is an essential component in widespread of applications, including object recognition (e.g. [141, 152]), surface reconstruction (e.g. [78, 159]), autonomous exploration (e.g. [197, 175]), image guided surgeries (e.g. [140, 100]), and diagnostic imaging (i.e. [127, 23]). An illustration of two applications can be seen in Fig. 1.1. At the top part, surface reconstruction of a dynamic synthetic torso is improved by taking into account temporal features such as specular highlights. The bottom part illustrates an

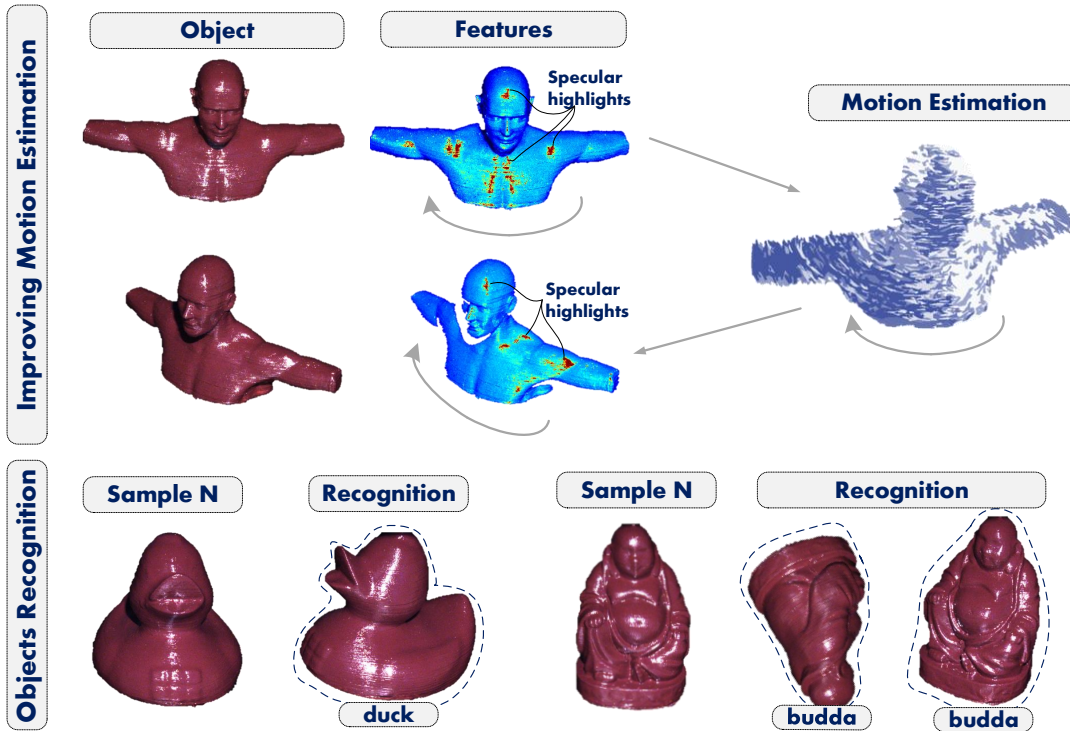


Fig. 1.1 Estimating and understanding motion is necessary in different applications. (a) Surface reconstruction is improved by taking temporal features which includes specular highlights. (b) Object recognition of dynamic objects is possible by knowing the object changes over time.

object recognition task in which given a set of samples (a video sequence), the object at hand can be recognized independently of the variance in rotation and scaling. A common factor in these objects is that they can be described with less restrictive assumptions due to their rigidity, which is not the case when one has to deal with non-rigid (deformable) objects.

It is very common to find deformable objects in our surrounding environment (examples can be seen at top part of Fig. 1.2). In order to retrieve the motion of such non-rigid objects, complex transformations are needed to describe them. This difficult task has been recognized in different works such as [220, 2, 94]. Particularly in medical scenarios, complex deformations are not the only difficulty that computer vision systems have to face while trying to estimate the motion of deformable objects. The middle row of Fig. 1.2 shows examples of deformable objects typically seen in clinical setups while the bottom row highlights some difficulties associated with medical data. Generally speaking, the performance of any vision based solution is highly dependent on the available visual information. Thus, *it is very important to handle any source of error that might affect the*

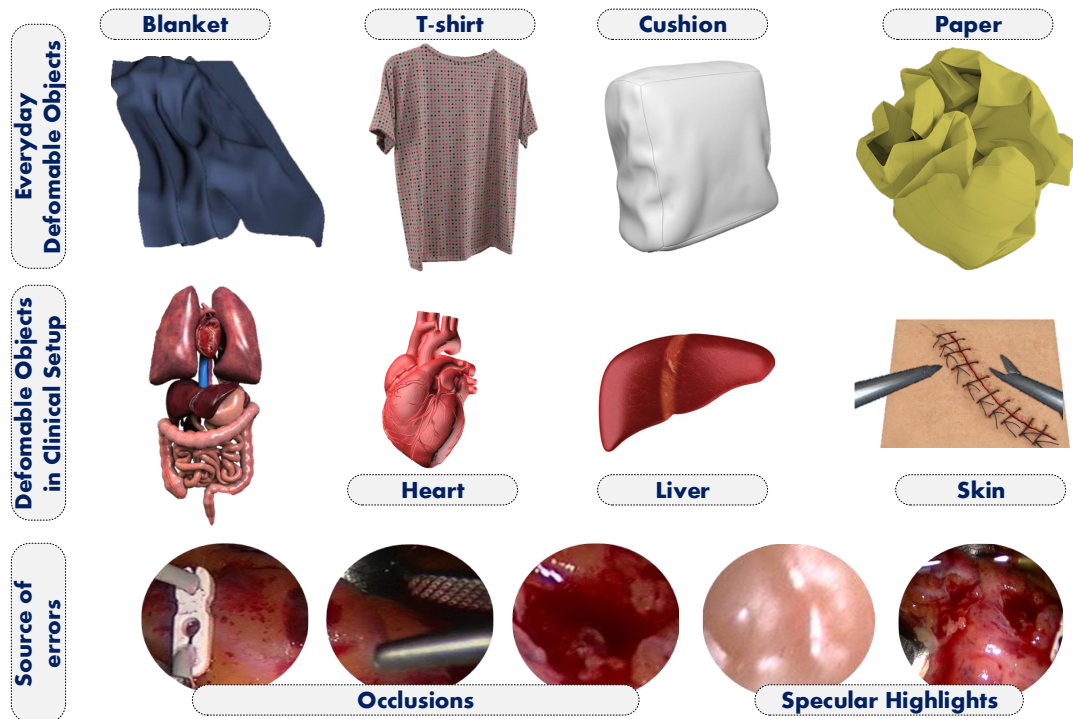


Fig. 1.2 Top row displays everyday deformable objects in which the environment conditions are less restrictive in comparison with those illustrated in the middle row that comes from a clinical setup. Bottom row shows difficulties associated with clinical data.

solution's robustness and precision. In this thesis, we also recognize the problems that exist in surgical settings and should be taken into account in order to get a realistic solution:

Glossy Organs' Surface The internal organs often have glossy surfaces with strong reflectivity, which results on having specular highlights on the targeted surface. Specular highlights are white bright spots that hinder the performance of a vision-based solution as they partially occlude the targeted surface, appear as additional features, generate discontinuities in the images, or cause loss of texture or color information. This well-known issue has been recognized as a challenging task in many works such as [9, 46, 5].

Occlusions In medical scenarios, dealing with cases in which two or more objects that are spatially separated interfere with each other is a common fact. These undesirable singularities compromise the motion estimation precision since they eliminate partial information of the target. Just like specular highlights, occlusions have been identified by the scientific community as

a task that must be handled in vision-based solutions. Examples of such works can be found in [111, 114, 123]

We focus this thesis on the medical domain in which understanding the human's body for clinical purposes requires retrieving the organs' complex motion patterns, which is in general a hard problem when using only image data. We tackle this challenging task by posing the question – *How to achieve a realistic motion estimation to offer a better clinical understanding?* We address this thesis on responding to this question by using a variational formulation as a basis to understand one of the most complex motions in the human's body, the heart's motion, through three applications (i) cardiac motion estimation for diagnostic, (ii) force estimation and (iii) motion estimation and prediction, both for robotic surgery.

To extract the complex motion patterns from the three aforementioned applications, we utilize useful information coming from two main medical modalities: ultrafast ultrasound (UUS) and endoscopic imaging. The particular characteristics of each modality enforce the employment of different tools while developing the solution. While UUS is able to extract the internal structure of the heart with a good temporal resolution using non-invasive techniques (sound waves), one should take into account the inherent artifacts such as the granular texture (speckles) and the lack of well-defined borders. Endoscopic image sequences on the other hand allow capturing external structure of the heart's surface but have the complication of dealing with the strong homogeneous texture of the acquired images. Taking into account the nature and characteristics of these modalities, we optimize our variational framework to meet with their unique set of advantages and constraints to achieve our objective of having robust and efficient motion estimation.

Although these three applications emerge from different clinical roots, they do share two significant common denominators between them:

1. The need to construct models that are capable of capturing the complex dynamics that tissues undergo either by their inherent motion (for example the beating heart) or by external events (for example an applied force over the tissue's surface).
2. The need to generate an explicit interpretation and analysis of the visual information from which one can gain more understanding about the environment and subsequently make better decisions.

With these two factors in mind, through this thesis we set the basis for achieving realistic motion estimation with the main goal of offering better clinical understanding.

1.1 Contributions

Through three applications, we set the basis for achieving a realistic motion estimation with the aim of offering a better clinical understanding. We go beyond existing solutions from the state of the art making the following key contributions:

(i) Cardiac Motion Estimation for Diagnostic

We focus on a central topic in cardiac imaging that is the estimation of the cardiac motion. The main aim is to offer objective and understandable measures to physicians for helping them in the diagnostic of cardiovascular diseases. We employ ultrafast ultrasound data and tools for imaging motion drawn from diverse areas such as low-rank analysis and variational deformation to perform a realistic cardiac motion estimation. The significance is that by taking low-rank data with carefully chosen penalization, synergies in this complex variational problem can be created.

Contributions

- We promote low-rank data representation. As a stand-alone tool, this kind of representation offers several advantages, such as speeding up the global solution and decreasing the noise in the image sequence.
- Another key point is topology preservation. A penalization term for the Jacobian determinant is used in order to guarantee a diffeomorphic transformation. We use a regularizer to rule out distortions while at the same time control the magnitude of expansion and compression.
- The combination of the two previous tools turns out to be synergistic and powerful as it allows computing an accurate displacement field that is mathematically well-motivated and computationally efficient.

(ii) Vision-Based Force Estimation in Robotic-Assisted Surgeries

In Robotic-Assisted Minimally Invasive Surgeries (RAMIS), surgeons perform delicate procedures remotely through robotic manipulators without directly interacting with the patients. As a result, they lack force feedback, which is an important sense for increasing surgeon-patient transparency and avoiding injuries and high mental workload. To cope with this problem, we describe a novel approach to estimate the applied forces during Robotic-Assisted Surgery. Since all RAMIS settings include a videoscopic view of the operation, we can employ the available visual information of the tool-tissue interaction and relate it directly to the applied force. From the conservation principles of continuum mechanics it is clear that the change in shape of an elastic object is directly proportional to the force applied. Following this principle, we propose a novel approach that is based on a variational framework that allows computing the observable deformation after a force is applied. Then, this information is used in a learning system that finds the nonlinear relationship between the given data, force and deformation, and use it to estimate the applied force. In particular, our contributions to this field are:

Contributions

- A new energy functional to compute the 3D tissue deformation. We prove through careful numerical results that it offers a better minima, with respect to other methods, and a low computational cost.
- We propose the use of a powerful supervised learning system that allows finding the optimal nonlinear relationship between the given data and the applied force. We demonstrate the adaptability across subjects and the stability of our solution during long periods of time based on in-vivo and ex-vivo datasets.
- The topic of designing a proper visual display of force feedback has not been sufficiently discussed yet. Through a perceptual study we demonstrate the potential of using sensory substitution for transmitting the force information. Based on a careful statistical, graphical, and perceptual analysis, we also provide user-centered recommendations for the design of visual displays for robotic surgical systems.

(iii) Physiological Cardiac Motion Cancellation and Prediction

In the last decades, robotic surgical systems have allowed performing complex procedures including Off-Pump Coronary Artery Bypass Grafting (OPCABG). This procedure avoids the associated complications of using Cardiopulmonary Bypass (CPB) since the heart is not arrested while performing the surgery. Thus, surgeons have to deal with a dynamic target, which compromises their dexterity and precision. Towards cancelling the cardiac motion, we propose a solution based on a variational framework formulated in both L^1 and L^2 and we then increase robustness in term of delays and occlusions by adding a prediction stage. While this is an important part of our solution the main contributions are:

Contributions

- We propose a diffeomorphic variational framework which is able to deal with the inherent complex deformation of a beating heart. It also incorporates a preprocessing stage for dealing with specular highlights.
- A key point is our prediction stage which is different from existing approaches where well-known algorithms from estimation theory, such as the Extended Kalman Filter, are used. We propose to restructure the given sequential data to formulate a standard supervised learning problem.

1.2 Publications

The following is a list of the publications derived from this thesis:

Journal Publications

- ▷ **A.I Aviles**, S.M. Alsaleh, J. Hanh and A. Casals *Sliding to Predict: Improving Vision-Based Cardiac Motion Cancellation by Modeling Temporal Interactions*, Submitted to The International Journal for Computer Assisted Radiology and Surgery, 2017.
- ▷ **A.I Aviles**, S.M. Alsaleh, S.P. Raventos, N. Younes, J. Philbeck, J. Hanh and A. Casals *Sensory Substitution for Force Feedback Recovery: A Perception Experimental Study*, Submitted to ACM Transactions on Applied Perception, 2017.
- ▷ **A.I Aviles**, T. Widlak, A. Casals, M.M. Nillesen and H. Ammari *Robust Cardiac*

Motion Estimation for Ultrafast Ultrasound Data: A Low-Rank-Topology-Preserving Approach, Conditionally Accepted to Physics in Medicine and Biology, 2017.

▷ **A.I. Aviles**, S.M. Alsaleh, J.K. Hahn and A. Casals, *Towards Retrieving Force Feedback in Robotic-Assisted Surgery: A Supervised Neuro-Recurrent-Vision Approach*, IEEE Transactions on Haptics, 2016.

▷ **A.I. Aviles**, P. Sobrevilla and A. Casals, *An approach for physiological motion compensation in robotic-assisted cardiac surgery*, Experimental & Clinical Cardiology, 2014.

Conference Publications

▷ **A.I. Aviles**, S.M. Alsaleh and A. Casals, *3D Diffeomorphic Deformation with Mixture Components as Visual Stimuli for Perceiving Interaction Forces in Robotic-Assisted Surgery*, submitted to IEEE International Conference on Intelligent Robots and Systems (IROS), 2017.

▷ S.M. Alsaleh, **A.I. Aviles**, A. Casals and J.K. Hahn, *Let Specular Highlights Perform!: Temporal Image Priors for Specular-Free Image Recovery*, Submitted to ACM SIGGRAPH, 2017.

▷ **A.I. Aviles**, S.M. Alsaleh, E. Montseny, P. Sobrevilla and A. Casals, *A Deep-Neuro-Fuzzy Approach for Estimating the Interaction Forces in Robotic Surgery*, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2016.

▷ **A.I. Aviles**, T. Widlak, A. Casals and H. Ammari, *Towards Estimating Cardiac Motion Using Low-Rank Representation and Topology Preservation for Ultrafast Ultrasound Data*, International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

▷ S.M. Alsaleh, **A.I. Aviles**, P. Sobrevilla, A. Casals and J.K. Hahn, *Adaptive segmentation and mask-specific Sobolev inpainting of specular highlights for endoscopic images*, International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

▷ S.M. Alsaleh, **A.I. Aviles**, A. Casals and J.K. Hahn, *Toward robust specularities detection and inpainting in cardiac images*, SPIE Medical Imaging, 2016.

▷ **A.I. Aviles**, S.M. Alsaleh, P. Sobrevilla and A. Casals, *Sensorless Force Estimation using a Neuro-Vision-Based Approach for Robotic-Assisted Surgery*, IEEE EMBS Neural Engineering Conference, 2015.

▷ **A.I. Aviles**, S.M. Alsaleh, P. Sobrevilla and A. Casals, *Force-Feedback Sensory Substitution using Supervised Recurrent Learning for Robotic-Assisted Surgery*, International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.

▷ S.M. Alsaleh, **A.I. Aviles**, P. Sobrevilla, A. Casals and J.K. Hahn, *Automatic and*

Robust Single-Camera Specular Highlight Removal in Cardiac Images, International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.

▷ **A.I. Aviles**, S.M. Alsaleh, E. Montseny and A. Casals, *V-ANFIS for Dealing with Visual Uncertainty for Force Estimation in Robotic Surgery*. Joint International Fuzzy Systems Association World Congress and European Society of Fuzzy Logic and Technology Conference (IFSA-EUSFLAT), 2015.

▷ **A.I. Aviles**, A. Marban, A. Sobrevilla, P. Fernandez, and A. Casals *A Recurrent Neural Network Approach for 3D Vision-Based Force Rstimation*. International Conference on Image Processing Theory, Tools and Applications (IPTA), 2014.

▷ **A.I. Aviles**, P. Sobrevilla and A. Casals. Unconstrained $L1$ —regularized minimization with interpolated transformations for heart motion compensation. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2014.

1.3 Thesis Overview

Chapter 2: Robust Cardiac Motion Estimation using Ultrafast Ultrasound Data: A Low-Rank-Topology-Preserving Approach.

This chapter is centered on improving diagnostic of cardiovascular diseases by estimating the heart’s motion using a relatively new modality *Ultrafast Ultrasound* (UUS) imaging. We first guide the attention to the advantages that UUS imaging offers over conventional modalities, and then, we move to describing the significance of preserving the anatomical structure of the organs, followed by a well-detailed explanation about our novel solution to improve clinical diagnostic. We then demonstrate how our proposed variational solution deals with complex deformations through careful numerical experiments. We conclude the chapter by highlighting the synergy between low-rank and topology preservation and remarking that our technique promises to be useful for analyzing organs experiencing complex motion other than the heart, as for example the movement of lungs in respiration.

Chapter 3: Towards Retrieving Force Feedback in Robotic-Assisted Surgery: A Supervised Neuro-Recurrent-Vision Approach.

We focus this chapter on one of the major problems in medical robotics that is the lack of force feedback, which restricts the surgeon’s sense of touch and might reduce precision during a procedure. We open this chapter by doing an exhaustive review of the state-of-the-art on the topic to find

the major issues of existing solutions. We then present our approach that combines visual and geometric information in a deep neural architecture that estimates the applied force. Our approach is carefully evaluated on phantom and realistic tissues in which we report an average root-mean square error of 0.02 N. We conclude this chapter by remarking that our solution avoids the drawbacks usually associated with force sensing devices, such as biocompatibility and integration issues.

Chapter 4: From Motion Estimation to Clinical Evaluation: A Perception Experimental Study. Once force information is obtained a natural question is –how to provide this information to the surgeon? This chapter extends the approach presented in Chapter 3 through an experimental study with the aim of responding to this question. We begin by explaining the existing options for transmitting the interaction forces to the surgeon. With this in mind, we highlight an attractive alternative called sensory substitution which allows transcoding information from one sensory modality to another. Afterwards, we describe three relevant aspects for the scope of our study: the subjects, the visualizations and the experimental procedure. We conclude this chapter by proving the potential of sensory substitution, particularly vision modality, in robotic surgical systems. Based on a careful statistical, graphical, and perceptual analysis, we provide user-centered recommendations for the design of visual displays for robotic surgical systems

Chapter 5: Sliding to Predict: Improving Vision-Based Cardiac Motion Cancellation by Modeling Temporal Interactions. We start by offering an overview of the existing solutions related to cardiac motion cancellation. We then set up the key factors to keep in mind for achieving a realistic solution when one chooses a vision-based approach. Then we follow with the description of our approach for canceling cardiac motion. We point out the robustness of our solution by taking into consideration undesirable perturbations such as visual occlusions and specular highlights. We then move to an exhaustive evaluation using phantom and in-vivo datasets. Finally, we close this chapter by presenting the conclusion and future work.

Chapter 6: Conclusions and Future Work. The last chapter concludes the thesis with a detailed discussion of our contributions in the clinical field. We highlight the adaptability of our solutions in other areas such as

computer graphics and present an example of our work in this area. We close this thesis by posing a set of questions that allow sketching directions for future work.

*“If at first the idea is not absurd, then there will be
no hope for it”*

A. Einstein

2

Robust Cardiac Motion Estimation using Ultrafast Ultrasound Data: A Low-Rank-Topology-Preserving Approach

According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death in the world. A key factor for early detection and prevention of these diseases is to analyze the cardiac motion to diagnose, for example, valve conditions or motion abnormality. The cardiac mechanics can be studied and analyzed through the heart deformation [60]. In most of the medical laboratories, diagnosis is based on the visual inspection of the heart’s motion [91]. However, the results are conditioned to the experience of the expert, and in consequence, they are highly variable and subjective. Thereby, the need of having objective and understandable measures emerged, and up-to-date cardiac motion estimation is a central topic in cardiac imaging [128].

The estimation of cardiac motion is a challenging problem to be tackled. In search of achieving a good motion estimation, different authors have used various biomedical imaging modalities including: magnetic resonance imaging (MRI),

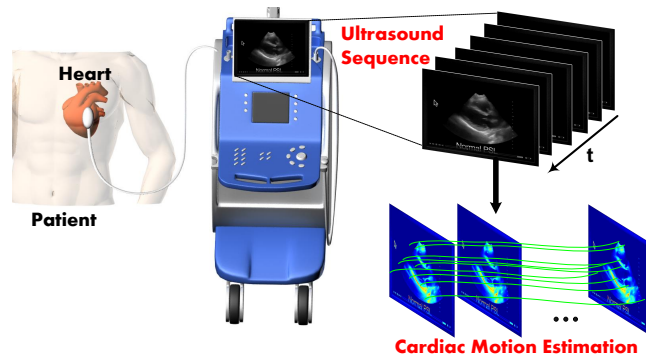


Fig. 2.1 Typical ultrasound acquisition setup: High-frequency waves allow capturing the view of inner organs such as the heart. Cardiac motion can be estimated by computing the spatial correspondence between time frames.

computed tomography (CT), single photon emission computed tomography (SPECT), and positron emission tomography (PET) (e.g. [179, 151, 227, 188]). However, the lack of resolution in modalities such as PET and SPECT ($\approx 4\text{--}7$ mm [150]), together with the exposure to radiation with CT and PET/SPECT, and the magnetic interference and cost of MRI make their use unsuitable in many applications.

An alternative modality to estimate cardiac motion is ultrasound imaging (US)(see Fig. 2.1). US is very popular due to its low cost, high accessibility, real-time interaction, non-ionization, portability and rapid assessment [48, 207]. It has become routinely used in multiple clinical scenarios including diagnostic and prevention of heart diseases. US allows capturing, for example, the heart’s size and shape, strain rate, ventricular deformation, and abnormal motions. US has shown its feasibility for tissue tracking and estimated motion analysis [51, 202].

Despite these benefits, ultrasound has disadvantages related to the presence of noise and occasional artifacts and its limited acquisition speed. The poor temporal resolution of conventional US hinders the retrieval of different mechanical events of the heart [47], [214]. Nonetheless, recent advances in ultrafast ultrasound (UUS) imaging have overcome some of these drawbacks, particularly temporal resolution, thanks to their higher frame rate (greater than 1000fps) [213], which is advantageous for cardiac motion estimation. An UUS system is capable of computing many lines in parallel, generating in this way a full image from one single transmitting event. Different applications of UUS emerge, such as tissue and blood motion estimation, imaging of micro bubbles or neurovascular coupling. Moreover, UUS facilitates advancements in disease prevention, diagnosis, and therapeutic monitoring [214]. In this chapter, we study the application of using UUS to estimate cardiac motion.

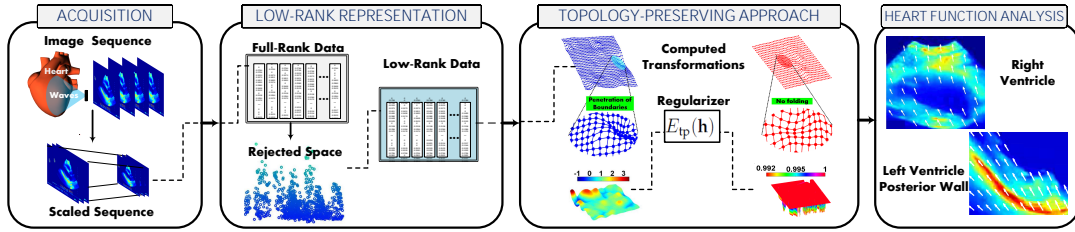


Fig. 2.2 Overview of our proposed approach. (From left to right) an ultrafast ultrasound cardiac sequence is acquired, this data is then represented in low-rank in order to speed up the solution and reduce noise. Later on, cardiac motion is computed enforcing topology preservation which allows keeping the anatomical structure of the heart. Finally, an analysis of the results is offered.

Different works for US cardiac motion estimation have been reported in the literature. Most of them use conventional ultrasound (e.g. [171, 118, 242, 86, 243]), while few works refer to the use of ultrafast ultrasound imaging (e.g. [156, 193]). With the use of both, conventional or ultrafast ultrasound, the approaches proposed to retrieve cardiac motion can be classified into those using the characteristics of the radio frequency (RF) signal, and those using an image sequence and applying computer vision techniques.

Motion estimation techniques in the first category use the natural acoustic reflections of the radio frequency (RF) signal. In this case, cardiac motion can be computed by either applying speckle tracking techniques, which use the amplitude of the signal, or radio-frequency-based correlation techniques, which use the phase information (e.g. [62, 127, 156, 193]). Another promising approach to estimate the cardiac motion relies on capturing the RF ultrasonic signals and then processing them to obtain some relevant information, such as the heart's characteristics.

In the second category of motion estimation techniques, we can find solutions based on Optical Flow (OF), in which the relative motion of the heart is computed from the velocities of patterns' brightness (examples can be found in [59, 218, 224, 73]). Although different authors have proved the feasibility of OF for motion estimation, its main drawback is that it only works for small and relatively non-complex deformations. Another common solution, usually adapted for more complex motions, is non-rigid registration, in which displacements of the tissue can be tracked by computing the spatial correspondences between frames [118, 138, 163, 142].

Another vision based technique for estimating the dynamics of the heart relies on tracking the heart's borders using deformable models (e.g. [3, 95, 194, 139]). However, cardiac motion estimation can be inaccurate when the displacements

are parallel to the edge or when a well-defined border is missing, which is a common problem in US images [118].

In this chapter, we describe a new approach to estimate cardiac motion from ultrafast ultrasound modality (see Fig. 2.2). Based on a variational formulation for non-rigid registration in L^2 , we include a maximum likelihood type estimator to increase the robustness of the solution in the sense of being able to deal with outliers. While this is an important part of the solution, the main contribution is: combining a low-rank data representation with a topology-preserving approach. Particularly:

- We promote low-rank data representation. As a stand-alone tool, low-rank data representation offers several advantages, such as speeding up the global solution by reducing the computational time and decreasing the noise in the image sequence.
- Another key point is topology preservation. A penalization term for the Jacobian determinant is used in order to guarantee a diffeomorphic transformation. We use a regularizer to rule out distortions while at the same time control the magnitude of expansion and compression.
- The combination of the two previous tools turns out to be synergistic and powerful as it allows computing an accurate displacement field that is mathematically well-motivated and computationally efficient.

The remainder of this chapter is organized as follows. An introduction to Ultrafast Ultrasound modality is described in Section 2.1. Section 2.2 presents previous literature related to the topology preserving problem. In Section 2.3, we present our low-rank data representation strategy, while in Subsection 2.4 we describe our variational framework to recover the deformation over time. In Section 2.5, we describe the penalization term we used to achieve topology preservation. In Section 2.6, we validate our proposal offering experiments on simulated and real datasets. Finally, section 2.7 provides a final conclusion and directions for future works.

2.1 Ultrafast Ultrasound Imaging: Beyond the Human Eye

The basic principle of conventional ultrasound is based on the use of sound waves with frequencies higher than 20 KHz. To generate a 2D image, pulses of sound

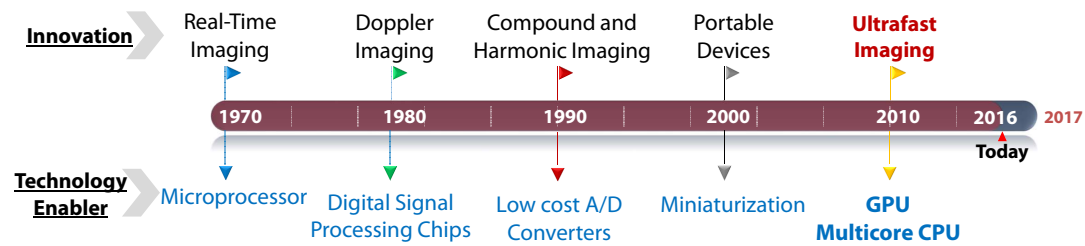


Fig. 2.3 Evolution of ultrasound imaging techniques over time, from real-time up to ultrafast imaging. Figure reproduced from [33].

are sequentially propagated from the transducer to the tissue and then returned to the transducer as reflected echoes within a plane of 90° . A line image is then generated for each transmitted pulse having as a result ~ 28 frames per second [47] (see left-side of Fig. 2.4 for illustration).

Despite the fact that ultrasound imaging is commonly used everyday in hospitals, its main limitation appears as soon as high frame rate is required. High temporal resolution is very important to assess diverse events of the human body that cannot be seen at low frame rates, such as tissue motion or blood flow. During more than 30 years, numerous researchers have attempted to overcome this problem by developing what is called *Ultrafast Ultrasound (UUS)* (e.g. [52, 200, 155, 166]). From Fig. 2.3, we can see the evolution of the Ultrasound modality, from the introduction of real-time imaging in the 70's, passing by including doppler measurements in the 80's, and building a better image quality through compound and harmonic imaging in the 90's, or the challenging miniturization design, until the recent innovations in ultrafast ultrasound imaging.

Although the theoretical concept of *Ultrafast* was first introduced by Bruneel et al. [40] in 1977, it was not until recently that thanks to the advent of GPUs technology, the use of Ultrafast innovation became possible in clinical scenarios. The main idea behind UUS is the parallel computation of multiple lines, which results in entire images being transmitted per event and, in consequence, in a high temporal resolution (see right-side of Fig. 2.4). This can be achieved by using either of the following approaches: gating techniques or plane-/diverging-wave imaging [47, 214].

When *gating techniques* are used, a large imaging part is first separated into small subparts, then, each subpart is imaged at a high frequency. To obtain the electrocardiogram (ECG) of the full image, gating is used to combine all subparts. Despite the potentials of gating techniques (demonstrated in [177, 170] for example), the main disadvantage appears when the ECGs are highly different during the cardiac cycles. To tackle this problem, some authors proposed to use

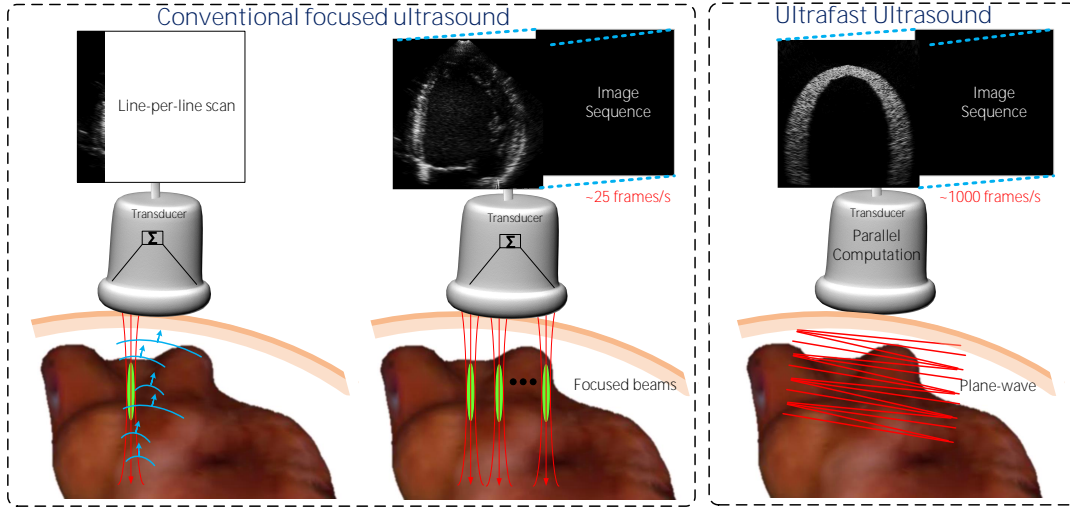


Fig. 2.4 Comparison between Conventional and Ultrafast Ultrasound acquisition. Left-side shows conventional acquisition in which a full image is generated for each transmitted pulse whereas the right-side shows an image is generated in a single transmission by computing multiple lines in parallel.

plane-/diverging wave imaging (e.g. [214, 178]) in which unfocused beams are used, that is, changing the aperture of the transducer. In this case, the main drawback is given by the unfocused beam that degrades the spatial resolution. However, this can be solved by using coherent plane-wave which allows generating high-quality images.

UUS is a breakthrough in the clinical domain since it has opened a wide range of clinical applications, since most of the physical events in the human body occurs in milliseconds, such as functional imaging of the brain, characterization of tumors, and blood flow (for example see [165, 215]). Particularly, in cardiac imaging UUS has allowed improving the estimation of deformation and motion as well as the analysis of strain and ventricular walls to name some applications. Due to the aforementioned reasons, in this work we use the temporal resolution advantages of the UUS to improve the estimation of the heart's motion.

2.2 Preserving Diffeomorphic Features

During cardiac motion estimation, unpreserved topologies result in violations of region convexity and are reflected in penetration of boundaries and overlapped or distorted mesh elements. Thereby, topology preservation is important in order to ensure connectivity between the structures, maintain the relations between neighboring elements, and avoid distortion of existing structures. When the

deformations are small, topology is preserved by the smoothness offered by the regularization term, but this is not the case when large deformations appear, as it happens when the complex dynamics of the heart is to be retrieved.

From the conservation principles of continuum mechanics, it is clear that the elastic displacement field in cardiac motion can be modeled as an isomorphism resp. diffeomorphism. Necessary and sufficient conditions to achieve this are that its deformation gradient tensor exists and is nonsingular at every point in the object [203]. In terms of the Jacobian J , this means that J exists and that the $\det(|J|) \neq 0$ at every point in the body. For a positive volume, it is required that $|J| > 0$ throughout the body [203, 3.2].

A well-known approach to achieve topology preservation is by controlling the Jacobian determinant. Dacorogna in [50] presented a detailed discussion related to the Jacobian determinant equation in which he demonstrated its ability to achieve topology preservation. Jacobian determinant has been used in problems involving deformable structures in order to achieve more realistic transformations (e.g. [45, 11, 192]). However, in this section we cover only those works that have relation with modeling deformable objects. In a multidimensional elastic registration framework, authors in [115] explored a barrier function for penalizing locally non-invertible functions.

The problem of achieving topology preservation has been reported in different works addressing MRI and US data. The authors in [45] ensured topology preservation by defining a threshold of 0.5 for the Jacobian determinant. Then, for the resulting values lower than the threshold, they generated a new template, equal to the previous deformed template, to continue with the registration process. Similarly, Ashburner in [11] and [10] penalized singular values of the Jacobian having lognormal distribution. To enforce the Jacobian positivity, Musse and colleges [147] proposed a 2D parametric approach based on the constraint of the continuous hierarchical modeling of the deformation field.

Later on, Noblet et al. [157] reported an extension of Musse's work in the three-dimensional space. They presented a hierarchical deformation field model in which the Jacobian determinant was conditioned when it had negative values. The main difference between the two works is the nontrivial optimization problem obtained in the 3D space. Topology preservation was treated as a hard constraint in [102], where they restricted the Jacobian determinant by a set of intervals in a grid region. If those conditions were not met, then topology preservation was enforced in terms of gradients. Explicit control of the deformation in terms of the determinant of the Jacobian was reported in [79]. In comparison with

similar works, here, authors proposed the use of point-wise inequality constraints (i.e. they achieved topology preservation by controlling voxel by voxel instead of using integral measures).

Another approach was presented in [237], in which topology was preserved by quantifying the magnitude of deformations and examining the statistical distributions of Jacobian maps in the logarithmic space. A two-step solution was proposed in [117]. Authors first corrected the gradient vectors of the deformation and then reconstructed the deformation based on a minimization problem on a convex subset of the underlying Hilbert space. As a result, they achieved a well-defined Jacobian on the image domain. An extension of that work was presented in [192], in which authors proposed a solution based on independent problems of small dimension that allow parallel computation.

Zhang et al. [242, 243] developed a temporally diffeomorphic motion estimation approach for conventional cardiac ultrasound sequences. In that work, the authors addressed the topology-preservation problem by using the smooth velocity field with a differential operator in a Sobolev space. The resulting transformation defines a group of diffeomorphisms. In [129], authors used the Beltrami coefficient (BC) to represent an orientation-preserving diffeomorphism. To deal with the computational cost of the BC method, they presented a splitting algorithm, one part solves the BC whereas the other involves the quasi-conformal map. However, the BC was reduced to constrain the Jacobian. And last but not least, a biophysically constrained framework for large deformation diffeomorphic image registration was proposed in [131]. They achieved topology preservation by controlling the Jacobian determinant and the amount of shear in the deformation map using a nonlinear Stroke regularization scheme.

2.3 A Low-Rank-Topology-Preserving Approach

As a recent promising tool, low-rank data representation has been promoted in a variety of areas, including computer vision, machine learning for fitting problems, and computing the low-rank approximation of a matrix among others [43, 44, 153]. A low-rank representation has been useful for processing big data because in many datasets, the relevant information lies in a low-dimensional space [81]. Moreover, it became more attractive since it has been proved that when a matrix A has rank r , a small subset is enough to reconstruct it exactly

[44]. In particular, low-rank representations have been used in motion recovery in optical flow, e.g. in [74] to recover the motion in videos of facial movements.

Our motivation for using low-rank data is threefold: First, we aim to increase the computational speed of the solution. Second, it is a way to denoise the ultrasound data and avoid artifacts in the recovered deformation field. Third, we aim to investigate the synergy of the low-rank representation with the preservation of the topology.

For a precise description of our low-rank data representation, consider an ultrafast ultrasound image sequence, $F = \{f_s\}_{s=0}^{S-1}$, with S frames of size $M * N$. Then, a new structure of the data is given in the form of a single matrix, called Casorati matrix \mathbf{C} , which columns are the S frames in a vectorized way:

$$\mathbf{C} = \mathbf{C}(F) = \begin{bmatrix} f_0(1, 1) & \cdots & f_{S-1}(1, 1) \\ \vdots & & \vdots \\ f_0(M, N) & \cdots & f_{S-1}(M, N) \end{bmatrix} \quad (2.1)$$

where $f_s(m, n)$ is the scalar value of the sequence in frame s at a given pixel location (m, n) .

Let us now turn to the theoretical prerequisites to produce a low rank representation of the Casorati representation \mathbf{C} , exploiting the high correlation in the columns of the matrix.

Theorem 1 (*Singular Value Decomposition, SVD*) For any real matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ there exist orthogonal matrices $\mathbf{U} \in \mathbb{R}^{M \times M}$ and $\mathbf{V} \in \mathbb{R}^{N \times N}$, and a diagonal matrix $\mathbf{S} = (\sigma_1, \dots, \sigma_r, 0, \dots, 0)$ with $r \leq \min(M, N)$ such that

$$\mathbf{A} = \sum_{y=1}^r \sigma_y u_y v_y^T = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (2.2)$$

where the positive numbers $\sigma_1 \geq \dots \geq \sigma_r > 0$ are unique and are called the singular values of \mathbf{A} . Then, the triplet $\mathbf{U} \mathbf{S} \mathbf{V}$ is called singular value decomposition (SVD). The value $r \leq \min(M, N)$ is equal to the rank of \mathbf{A} . For proof, see Appendix A.

A major issue in medical applications is the large amount of data to be processed, which is the case when the cardiac motion is estimated. Thus, instead of computing the Singular Value Decomposition (SVD) of a large and dense matrix, it is enough to compute the set of dominant singular values. This allows keeping the most relevant information in a subspace which is smaller than the original one. Thus, the problem of building a low-rank representation can be given by finding the k -dominant singular values of \mathbf{A} . Mathematically, finding the rank- k approximation of matrix \mathbf{A} can be described as:

	U	S	V
C	13824x13824	13824x1814	1814x1814
C₁₀₀	13824x100	100x100	1814x100

Table 2.1 Decomposition of the Casorati matrix **C** and the rank 100-approximation **C₁₀₀**

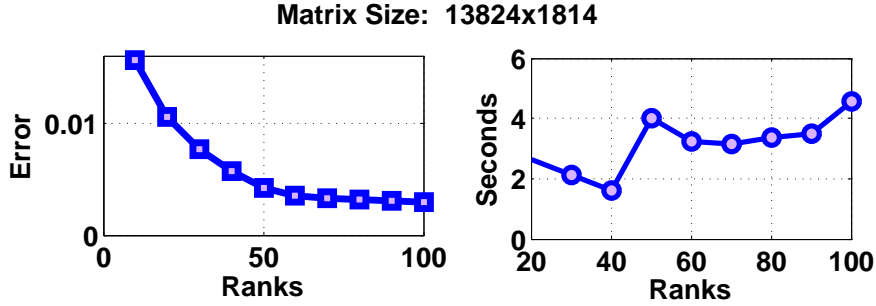


Fig. 2.5 Left: Singular values $\sigma_{k+1} = \|\mathbf{C} - \mathbf{C}_k\|^2$. Right: CPU time to compute the rank- k approximation \mathbf{C}_k .

$$\mathbf{A}_k := \sum_{y=1}^k \sigma_y \mathbf{u}_y \mathbf{v}_y^T = \arg \min_{\text{rank}(\hat{\mathbf{A}}) \leq k} \|\mathbf{A} - \hat{\mathbf{A}}\|_F^2 \quad (2.3)$$

where $\|\mathbf{L}\|_F^2$ is the squared Frobenius norm of a matrix which is used in low-rank based problems since it is invariant to rotations and to the rank. The importance of the rank- k -approximation in (2.3) is given by the following theorem:

Theorem 2 (Eckart-Young) Take a matrix **A** with a SVD as in (2.2), and let $k < r := \text{rank}(\mathbf{A})$. Let \mathbf{A}_k be the rank- k approximation in equation (2.3). Then

$$\|\mathbf{A} - \mathbf{A}_k\|^2 = \min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|^2 = \sigma_{k+1}^2 \quad (2.4)$$

For proof, see Appendix A.

For our ultrafast ultrasound data, we compute the Casorati matrix $\mathbf{C}(F)$ and different rank- k -approximations \mathbf{C}_k in equation (2.3). The resulting matrix decomposition is displayed in Table 2.1, in which the original matrix (of size 13824x1814) was reduced to rank 100. Both the error (2.4) and the computational time are plotted against the rank k in Figure 2.5. On the one hand, only an average of 2.08984 seconds was needed to obtain \mathbf{C}_k , and on the other hand, the approximation error for \mathbf{C}_{100} is $1.44e^{-5}$.

As we stated before, another main motivation to promote a low-rank representation \mathbf{C}_k instead of using the full Casorati matrix **C** is to reduce

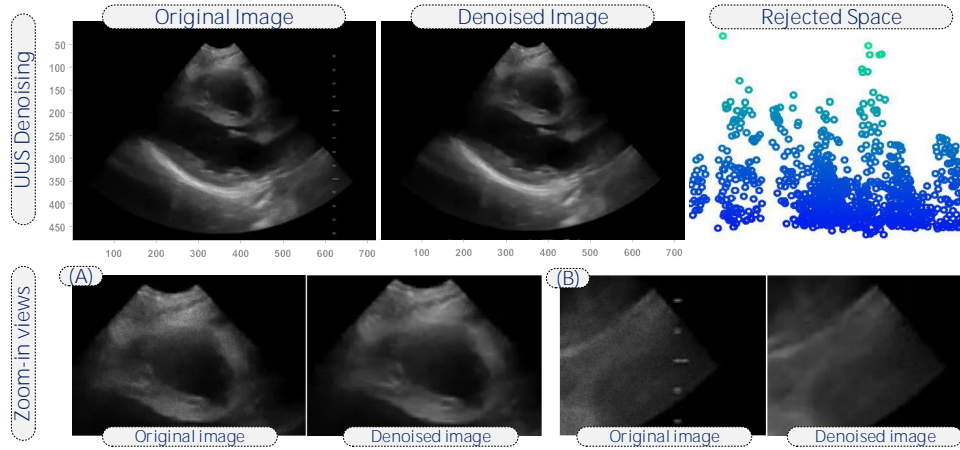


Fig. 2.6 Top row shows an original and denoised frame after applying low-rank process along with the rejected space. Bottom row, A and B, show zoom in views of the same frames in which we can see that both noise and some artifacts were removed.

noise. This is accomplished by eliminating the subspace where the noise relies, which results in retrieving a subspace with only relevant information. Noise, which normally relies on another subspace due to its characteristics, is rejected from the solution (see Figure 2.6). This eliminates artifacts in the subsequent deformation computation.

In the last step, we use the invertibility in equation (2.1) to invert the low-rank representation $\mathbf{C}_k(F)$ back to the denoised video sequence $F_k = ((f_s)_k)_{s=0}^{S-1}$. We will work with that sequence in the following section to extract the mechanical deformation field of the heart.

2.4 Deformation Recovery

Let $F = \{f_s\}_{s=0}^{S-1}$ be the image sequence of S frames, where each image f_s is a function over the bounded domain Ω .

We will find a deformation vector field \mathbf{h} defined in the domain Ω as a minimizer of an energy functional:

$$E(\mathbf{h}) = \sum_{s=0}^{S-2} E_{\text{dsc}}(f_s, f_{s+1}; \mathbf{h}) + E_{\text{reg}}(\mathbf{h}) + E_{\text{tp}}(\mathbf{h}) \quad (2.5)$$

where the three terms used have the following purposes:

E_{dsc}	...	discrepancy measure
E_{reg}	...	regularization term
E_{tp}	...	topology preservation

We will now go through our variational framework and explain each of the terms in the energy functional. We begin with the representation of the deformation field \mathbf{h} , then go on to the discrepancy term E_{dsc} and the regularization term E_{reg} . In Subsection 2.5, we describe the topology preservation term E_{tp} necessary to achieve realistic deformations.

How to represent the deformation field \mathbf{h} ? The deformation model is an essential factor that defines how fast and accurate the approach is. In order to find a compromise between computational cost and accuracy, we will handle the changes over time using a lattice as in the following definition:

Definition 1 *A m-dimensional lattice is the \mathbb{Z} -linear span of a set of k linearly independent vectors in \mathbb{R}^m .*

We will then use a lattice in which its points are characterized by the tensor product of the b-splines [222]. These are widely used in medical applications. The advantage of this lattice deformation model is that it demands low running time, allows multiresolution, has optimal mathematical properties and keeps affine invariance. Using b-splines has the additional advantage of being able to handle complex deformations.

Consider a given position $\mathbf{w} = (w_1, \dots, w_d)$ in \mathbb{R}^d . Let $\{\xi_i(\cdot)\}$ be a basis of spline functions and let $\mathbf{P}_{j_1, \dots, j_d}$ be control points. Then, we express the deformation vector \mathbf{h} at point \mathbf{w} through the model:

$$\mathbf{h}(\mathbf{w}) = \sum_{j_1=0}^n \dots \sum_{j_d=0}^n \overbrace{\mathbf{P}_{j_1, \dots, j_d}}^{\text{control points}} \underbrace{\prod_{k=1}^d \xi_{j_k}(w_k)}_{\text{tensor product}} \tag{2.6}$$

In this work, we use cubic basis splines:

$$\begin{aligned} \xi_0(x) &= (1 - x)^3/6 \\ \xi_1(x) &= (4 - 6x^2 + 3x^3)/6 \\ \xi_2(x) &= (1 + 3x + 3x^2 - 3x^3)/6 \\ \xi_3(x) &= x^3/6 \end{aligned} \tag{2.7}$$

The deformation model for \mathbf{h} in (2.6) is in \mathbb{R}^d and it shows that within our framework, we actually reconstruct the control points $\mathbf{P}_{j_1, \dots, j_d}$ in order to get the deformation \mathbf{h} . – In our application, we will exploit this deformation model for dimension $d = 2$.

We now turn to the discrepancy term E_{dsc} . Since the images are acquired by the same sensor, it is not expected that they have a big intensity variation between them. Therefore, an iconic method is a perfect match for this application. One possible option is to use the Sum of Squared Differences (SSD):

$$\int_{\Omega} (f_0(\mathbf{w} + \mathbf{h}(\mathbf{w})) - f_1)^2 d\mathbf{w} \quad (2.8)$$

SSD offers a low computational cost but it has the disadvantage of not dealing well with outliers.

For the actual expression for E_{dsc} in (2.5), we use:

$$E_{\text{dsc}}(f_0, f_1; \mathbf{h}) = \int_{\Omega} \rho(f_0(\mathbf{w} + \mathbf{h}(\mathbf{w})) - f_1) d\mathbf{w} \quad (2.9)$$

Here, the function ρ is a maximum likelihood type estimator motivated by robust statistics:

Definition 2 *An M-estimator is a symmetric and positive definite function ρ with a unique minimum at zero.*

The M-estimator substitutes the minimization of $\sum_i \mathbf{r}_i^2$, where \mathbf{r} is the residual error, with $\sum_i \rho(\mathbf{r}_i)$, in order to deal with the effect of outliers. This increases the robustness and accuracy of the result.

In this work, we use the Turkey estimator for ρ :

$$\rho(x) = \begin{cases} \frac{c^2}{6} \left[1 - \left(1 - \left(\frac{x}{c} \right)^2 \right)^3 \right] & \text{if } |x| \leq c \\ \frac{c^2}{6} & \text{if } |x| > c \end{cases} \quad (2.10)$$

where c is a tuning parameter. The discrepancy term with the Turkey estimator in (2.10) is known to offer a hard rejection of outliers [209].

For the regularization term E_{reg} , we use the Tikhonov method to impose stability to the energy functional in the sense of Hadamard [80]. Let $\gamma \in \mathbb{R}^+$ be the regularization parameter. Then we use for $\mathbf{h} = (h_1, \dots, h_d)$ the term:

$$E_{\text{reg}}(\mathbf{h}) = \gamma \sum_{l=1}^d \int_{\Omega} \|\nabla h_l(\mathbf{w})\|^2 d\mathbf{w} \quad (2.11)$$

as the regularizer in our variational framework (2.5).

2.5 Topology Preservation

A common drawback of most of the solutions working with complex deformations is that topology-preserving is not guaranteed, which leads to unrealistic deformations becoming a source of error. An example of such deformations is shown in Figure 2.7 in which topology is not preserved. Particularly, in medical applications this issue is of huge importance in order to maintain the anatomical structures. In order to preserve the topology, we will require that the deformation is a diffeomorphism according to the following definitions:

Definition 3 *A manifold \mathbf{G} , according to Boothby [36], of dimension d is a topological space with the following properties:*

- \mathbf{G} is Hausdorff
- \mathbf{G} is locally Euclidean of dimension d , and
- \mathbf{G} has a countable basis of open sets

Taking previous definition, now we can set the diffeomorphic definition.

Definition 4 *Given the manifolds L and M , a map $f : L \rightarrow M$ is called diffeomorphic if f is a bijection, f carries L homeomorphically into M , and f and f^{-1} are differentiable.*

Based on the previous definitions, we can set Dacorogna's theorem which was the first work to discuss the Jacobian determinant equation.

Theorem 3 (Dacorogna) *Let $k \geq 0$ be an integer, $0 < \alpha < 1$, and Ω has a $C^{k+3,\alpha}$ boundary $\partial\Omega$ ($C^{k,\alpha}$ denoting the usual Hölder spaces). Let $f, g \in C^{k,\alpha}(\bar{\Omega})$ with $f, g > 0$ in $\bar{\Omega}$. Then there exist a diffeomorphism φ with $\varphi, \varphi^{-1} \in C^{k+1,\alpha}(\bar{\Omega}, \mathbb{R}^n)$ and satisfying*

$$\begin{cases} g(\varphi(x)) \det \nabla \varphi(x) = \lambda f(x), & x \in \Omega \\ \varphi(x) = x, & x \in \partial\Omega \end{cases} \quad (2.12)$$

where $\lambda = \int g / \int f$.

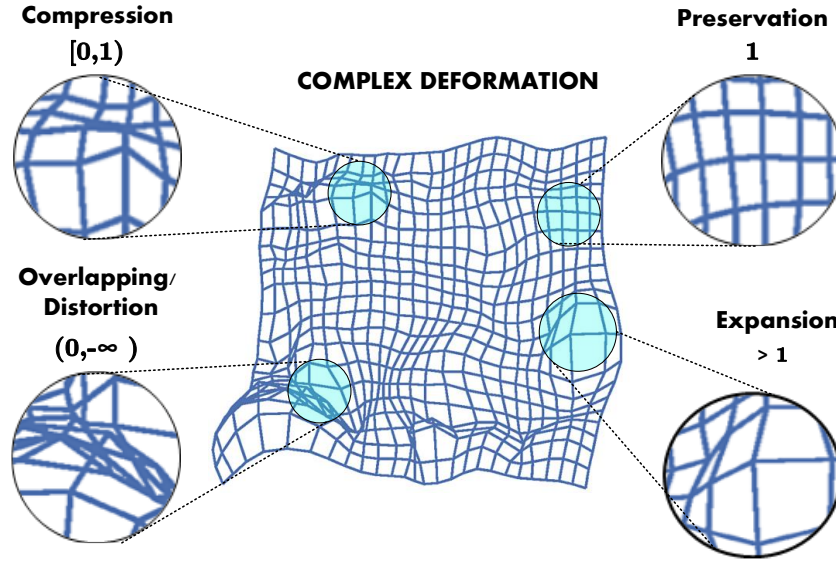


Fig. 2.7 When topology-preserving is not enforced, unrealistic transformations can appear in the result. A way to ensure topology preservation is by checking the Jacobian determinant $|J|$. When $|J|$ is equal to 1 then the volume is preserved. Small positive or large positive numbers of $|J|$ result in contractions or expansions. But having $|J| \in (-\infty, 0)$ can result in distortions, overlapping, and creation of new structures.

As an application of Theorem 3, we can achieve topology-preserving diffeomorphism. Assume that Ω is defined as in Theorem 3 and $\psi_0 \in Dif f^{k+1,\alpha}$, then there exists $\psi \in Dif f^{k+1,\alpha}$ such that:

$$\begin{cases} \det \nabla \psi \equiv 1 \text{ in } \Omega \\ \psi = \psi_0 \text{ on } \partial \Omega \end{cases} \quad (2.13)$$

As stated in Theorem 3 and Eq. 2.13, to achieve a diffeomorphic deformation, the Jacobian determinant can be used. This makes sense since it allows measuring the changes in the area/volume produced by the deformation at each patch.

Let $|J_{\mathbf{h}}(\mathbf{w})|$ be the Jacobian determinant of the deformation $\mathbf{h} = (h_x, h_y)$ in \mathbb{R}^2 . The Jacobian determinant is described as:

$$|J_{\mathbf{h}}(\mathbf{w})| = \det \begin{pmatrix} \frac{\partial h_x(\mathbf{w})}{\partial x} & \frac{\partial h_x(\mathbf{w})}{\partial y} \\ \frac{\partial h_y(\mathbf{w})}{\partial x} & \frac{\partial h_y(\mathbf{w})}{\partial y} \end{pmatrix}. \quad (2.14)$$

The characteristics of the deformation \mathbf{h} encoded in the Jacobian are shown in Table 2.2.

An illustration of these behaviors can be seen in Figure 2.7.

Condition	Local type of deformation
$ J_{\mathbf{h}}(\mathbf{w}) \leq 0$	topology is destroyed, overlapping, distortion and penetration of boundaries may occur
$ J_{\mathbf{h}}(\mathbf{w}) > 0$	diffeomorphism, topology preservation
$0 < J_{\mathbf{h}}(\mathbf{w}) < 1$	contraction
$ J_{\mathbf{h}}(\mathbf{w}) = 1$	volume preservation
$1 < J_{\mathbf{h}}(\mathbf{w}) $	expansion

Table 2.2 Jacobian determinant conditions for topology preservation

From our deformation model in (2.6), the partial derivatives of $\frac{\partial h_x(\mathbf{w})}{\partial x}, \dots, \frac{\partial h_y(\mathbf{w})}{\partial y}$ can be easily evaluated, as they come from the tensor product of independent functions. Using Equation (2.7), we have for the derivatives of the cubic basis splines:

$$\begin{aligned}
 \xi'_0(x) &= (1 - x)^2/2 \\
 \xi'_1(x) &= (3x^2 - 4x)/2 \\
 \xi'_2(x) &= (-3x^2 + 2x + 1)/2 \\
 \xi'_3(x) &= x^2/2
 \end{aligned} \tag{2.15}$$

Then, the determinant in (2.14) can be straightforwardly evaluated as a function of the lattice points $\mathbf{P}_{j_1, \dots, j_d}$ characterizing the deformation \mathbf{h} .

Now that we have stated how to compute the Jacobian determinant, we turn to formulate our penalization term E_{tp} for the energy functional (2.5). We use a weak constraint, but do not penalize values lying near 1:

$$\begin{aligned}
 E_{tp}(\mathbf{h}) &= \int_{\Omega} \delta_{\mathbf{h}}(\mathbf{w}) d\mathbf{w}, \text{ with} \\
 \delta_{\mathbf{h}}(\mathbf{w}) &:= \begin{cases} e^{-|J_{\mathbf{h}}(\mathbf{w})|} + \varphi \sqrt{|J_{\mathbf{h}}(\mathbf{w})|^2} & \text{if } ||J_{\mathbf{h}}(\mathbf{w})| - 1| \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.16}
 \end{aligned}$$

Here, $\varphi \in \mathbb{R}^+$ offers a balance in our penalization, and $\tau \in \mathbb{R}^+$ is the margin of acceptance for values close to one.

The first term

$$e^{-|J_{\mathbf{h}}(\mathbf{w})|}$$

heavily penalizes negative values of the deformation and thus it prevents the field \mathbf{h} from having distortions or penetration of boundaries. The term

$$\varphi \sqrt{|J_{\mathbf{h}}(\mathbf{w})|^2}$$

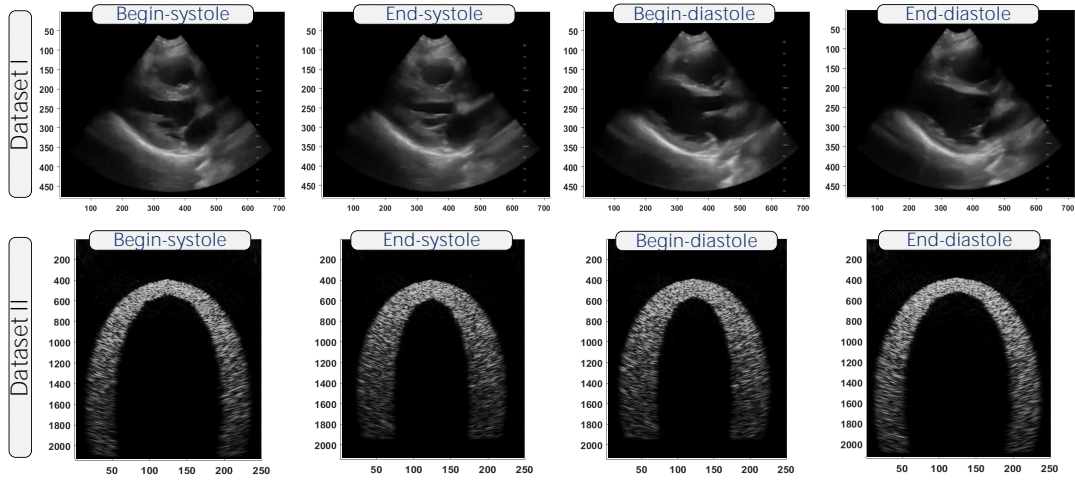


Fig. 2.8 Sample frames of the raw data extracted from the two datasets used for evaluating our approach.

with the parameter φ controls the magnitude of the expansions and contractions.

Unlike most of the Jacobian determinant constrains, for example $\log(|J|)$ (see [11]) and $e^{(|J|)}$ (as in [115]), we do not only guarantee the positivity of the Jacobian determinant but also enforce its value to stay near one. Moreover, we penalize big expansions in order to achieve more realistic deformations.

There are different options for solving the L^2 -regularized class. Traditional methods include Gradient Descent, Newton’s method, Nonlinear Conjugate Gradient, and Evolutionary Optimization Algorithms. However, they can get stuck at local minima, might need an infinite number of iterations to converge or have a slow rate of convergence. A better alternative is the well-known Levenberg-Marquardt (LM) which offers better results with less computational time. For these reasons, in this work we use LM method to minimize our energy functional.

2.6 Experimental Results

This section describes in detail the experiments that we conducted to validate our proposal.

2.6.1 Subjects and acquisition

We used two ultrafast ultrasound datasets (see Fig. 2.8) to evaluate our proposed approach.

The first is a realistic dataset from one patient. During the acquisition, the patient was placed in the supine position or left lateral decubitus. Then, the probe is placed on the left parasternal line at the fourth intercostal space with the marker pointing toward the right shoulder of the patient. The images were thus taken from the parasternal long axis view. This view is useful for global assessment of the motion of the heart’s wall and the function of different areas including the right and left ventricle, the mitral and aortic valves and the interventricular septum. This dataset is composed of 1814 frames with size of 720x480 and a scaled version of size 144x96. This data was acquired with an UUS device with a (fc) transducer with a bandwidth of 6MHz and 192 elements using coherent compounding of plane waves. The output is a 2D plane wave image sequence of the long axis view of a healthy heart.

The second dataset (see [156] for details) is a simulated ultrafast ultrasound sequence that has a realistic cardiac deformation field and describes the mechanics of a healthy left ventricle (see [37] for the description of the mechanics). This data was generated using (fc) transducer with a frame rate of 5000 Hz (single transmits) and effective frame rate of 500 Hz. The output of this simulated data is a set of 2D apical imaging planes. This view is helpful to study the left ventricle and the mitral inflow. The sequence is composed of 399 frames with size of 2143x250. This simulated data is very helpful to assess the performance of our approach since it includes a ground truth of the displacements which can be compared against our estimation.

All the measurements and reconstructions in this section are taken from these two ultrafast ultrasound cardiac sequences. All test and comparison were run under the same condition on a CPU-based implementation. We used an Intel(R) Core i7- 6700 CPU at 3.40GHz-32GB RAM, and a Nvidia GeForce GT 610.

2.6.2 Validation scheme

We divided our validation scheme into two parts. The first part makes use of the realistic dataset and relies on the following measurements to evaluate the performance of our topology-preserving technique:

- Comparison between our proposed topology regularizer and two from the literature: Table 2.3;
- Assessment of using low-rank as a preprocessing step: Fig. 2.9;
- Inspection of the displacement field: Figure 2.10 (A)/(B);

- Numerical results offered by the Jacobian determinant: Figure 2.10 (A.2)/(B.2); Table 2.4;
- Careful comparison of the residual error for both the low-rank tool and the topology preservation tool and study of their synergy: Table 2.4;

In the second part, we use a simulated dataset with a provided ground truth in which we perform the following evaluations:

- Numerical visualization and comparison of the displacement field: Figures 2.10 and 2.12;
- Numerical visualization of mean accumulated displacement of the seven segments of the left ventricle Figure 2.13;
- Nonparametric statistical analysis between the real and the estimated displacements;
- Illustration of the computed strain as a reasonable clinical measure: Figure 2.14.

2.6.3 Results

In order to prove the benefits of using low-rank (SVD) as a preprocessing step, we compared it against two common preprocessing techniques: Gaussian smoothing (kernel 5×5 and $\sigma = 0.7$) and Wavelets (Biorthogonal Spline Wavelet, 4 levels). We carried out the comparison of the three preprocessing techniques using our topology regularizer and two more from the state of the art [186, 86].

According to the results (Table 2.3 and Fig. 2.9), we found that low-rank was able to find the best minima in our case study in a computationally efficient manner. The results showed that Wavelet was able to find an acceptable minima but, it needed an average of 30 iterations per frame to converge compared to the 16 needed by low-rank (see plots in Fig. 2.9). Gaussian smoothing on the other hand performed the worst in terms of minima and average iterations per frame.

Overall, out of the three preprocessing techniques, low-rank offered a good tradeoff between accuracy and computational time since it requires less iterations per frame. This is further reflected in the overall CPU time as illustrated in the box-plot at right side of Fig. 2.9 where low-rank only needed an average of 2.3217 seconds of computational time while Gaussian and Wavelet both required more than 11 seconds in average. Moreover, it performed the best across the different regularizer giving the best minima each time. In general, our topology regularizer approach was the best compared to the regularizers offered by Rohlfing [186]

Table 2.3 Performance comparison between our proposed approach and other state of the art approaches

Preprocessing	Topology Regularizer Our Approach	Minimum	Topology Regularizer Rohifing et al. [186]	Minimum	Topology Regularizer Heyde et al. [86]	Minimum
Low-Rank (SVD)	$E_{tp}(\mathbf{h}) = \int_{\Omega} \delta_h(\mathbf{w}) dw$ with $\delta_h(\mathbf{w})$ as in Eq. [14]	$2.0357e^{-12}$	$E_{tp}(\mathbf{h}) = \int_{\Omega} \log(J_h(\mathbf{w})) dw$	$4.0028e^{-6}$	$E_{tp}(\mathbf{h}) = \int_{\Omega} J_h(\mathbf{w}) - 1 ^2 dw$	$3.2012e^{-7}$
Wavelet		$6.0596e^{-6}$		$9.6984e^{-4}$		$1.6443e^{-5}$
Gaussian Smoothing		$1.7417e^{-3}$		$7.1822e^{-2}$		$3.3613e^{-2}$

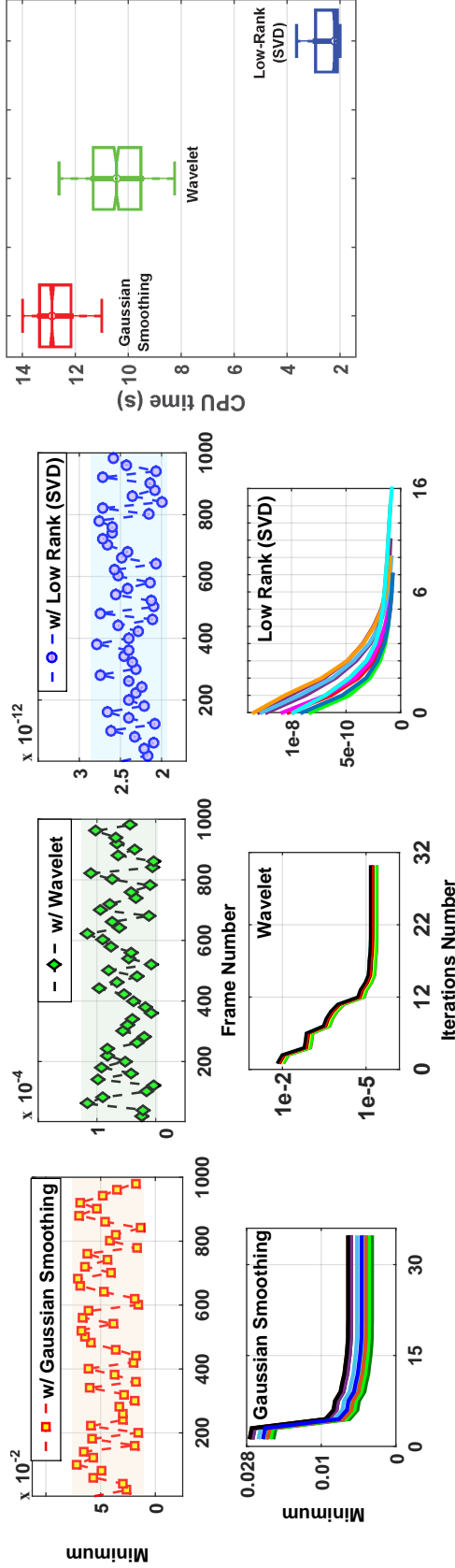


Fig. 2.9 The first row shows the convergence history of the complete sequence (1000 frames) using three different pre-processing techniques, while the second row shows the number of iterations it took to find the minimum for few of those frames. Box-plot at the right side shows the CPU time comparison of the different techniques.

and Heyde [86]. It has the advantage of enforcing the value to be close to one and penalizing very strong expansions and contractions.

The performance of our proposed topology-preserving technique is illustrated in Fig. 2.10 with and without topology preservation. First we show the performance without topology preservation ($E_{tp} = 0$ in the energy functional (2.5)) where the resulting displacement fields are shown in (A) and the corresponding Jacobian determinant are shown in (A.2). The example frames in the columns show critical details of the respective displacements fields and Jacobian. Row (A.1) shows zoom-in parts where different violations of the topology occur in part (A), such as overlapping, boundaries penetration, and mesh elements distortion.

The plots of the Jacobian clearly show huge variations in the value of the determinant which results in an unstable representation of the anatomical structure. Not only that, but in some parts, the Jacobian determinant presented big values (greater than 3), which indicates that some transformations produced very big expansions, while in others the Jacobian determinant gave negative values, which indicates that new structures were formed. These violations create new structures and result in an unrealistic representation of the complex deformation of the heart’s motion. This is not acceptable particularly in medical applications where preservation of the anatomical structure is remarkably needed.

We then ran the same tests after applying the proposed topology preserving approach (E_{tp} in (2.16) with $\varphi = 5 \cdot 10^{-3}$) and show the results in part B. Looking at the zoom-in parts in B.1, and comparing it with part A.1, we can verify that our approach allows controlling expansions and contractions, maintaining region convexity, and avoiding foldings. The corresponding plots of the Jacobian determinant are shown in part B.2. In comparison to the Jacobians without topology preservation, we can see stabilities on the values as they do not suffer large variations (mostly stay on 1) with guaranteed positivity. These are realistic values, as over two pairs of time-frames, the volume should be approximately preserved. From these illustrations, one can therefore conclude that the approach was successful in avoiding topology violation even with complex deformations of the cardiac motion.

For a more detailed quantitative analysis, we evaluated the global performance of our approach by comparing its performance with low-rank and full-rank data using 600 frames of the sequence (see Figure 2.8 dataset I). Unlike works where speckles were useful as indicators for tracking the heart’s motion, in the framework we proposed, results from Table 2.4 show that promoting low-rank offered a

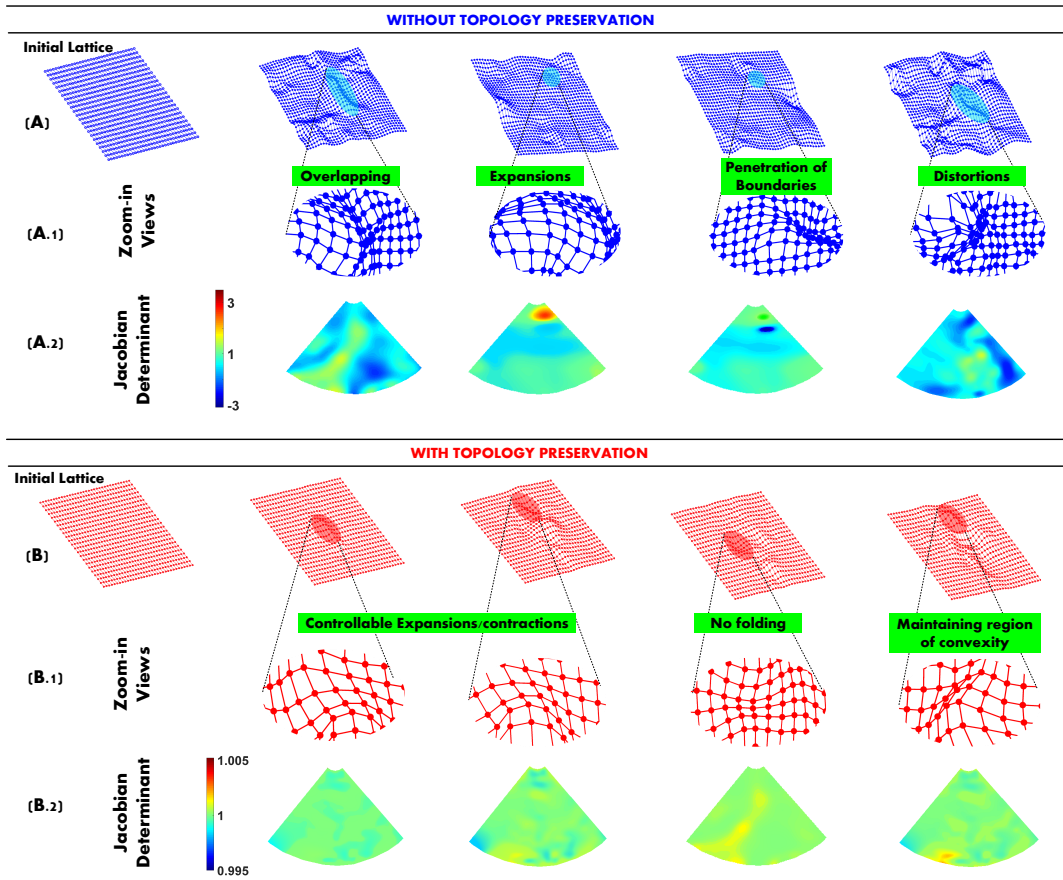


Fig. 2.10 (A) Resulted transformations, during complex deformations, without applying topology preservation. Highlighted areas denote structure violations that are more clearly displayed in the zoom-in views (A.1). The resulted Jacobian determinant are shown in (A.2). Resulted transformations after applying topology preservation are shown in (B) and can be compared with (A), in which (B.1) and (B.2) show that they keep the mesh structures with most of the Jacobian determinant staying at 1.

positive effect on the solution as it significantly reduced the computational time and allowed faster convergence of the energy functional (less iterations per frame).

In Table 2.4, Exp. 1 and Exp. 5 show that without topology preservation in the functional (2.5), the residuum was about 0.1 with and without the low-rank constraint. After applying low-rank, the Jacobian determinant had higher values but still suffered from heavy distortions. With topology preservation in Exps. 2-4, and 6-8, reasonable values of the Jacobian determinant were obtained, avoiding any penetration of boundaries. Thus, topology preservation is a necessary tool to get realistic deformation results for cardiac motion estimation.

Table 2.4 Performance analysis: low-rank vs. full-rank data and their reaction to different degrees of topology preservation (see text for discussion).

Exp. Data	Energy functional in (2.5): specifics of topology preservation	CPU time (seconds) per frame	Minimum	Discrepancy (mean) (2.8)	Average [min max] of $ J_h(w) $
1) Full-Rank	$E_{tp} = 0$	12.0830	0.1069		[-2.8896 3.6884]
2) Full-Rank	(2.16) with $\varphi = 0$	12.0672	$3.9945e^{-3}$	0.0016	[0.0617 1.0096]
3) Full-Rank	(2.16) with $\varphi = 10^{-2}$	11.3953	$6.0155e^{-3}$		[0.9034 1.0023]
4) Full-Rank	(2.16) with $\varphi = 5 \cdot 10^{-3}$	11.0304	$4.5563e^{-4}$		[0.9546 1.0059]
5) Low-Rank	$E_{tp} = 0$	3.8771	0.1191		
6) Low-Rank	(2.16) with $\varphi = 0$	3.4895	$2.1994e^{-09}$	7.9505e-06	[0.0213 1.0031]
7) Low-Rank	(2.16) with $\varphi = 10^{-2}$	3.0214	$3.2975e^{-11}$		[0.9184 1.0028]
8) Low-Rank	(2.16) with $\varphi = 5 \cdot 10^{-3}$	2.32170	$2.0357e^{-12}$		[0.9950 1.0100]

The primary role of the low-rank representation seems to be the radical decrease in computational time, as seen from a comparison of Exps. 1-4 and Exps. 5-8, where the computational time was decreased by about 75 %.

The residual errors in Exps. 2-4 in the topology-preserved full-rank case were in the order of magnitude 10^{-3} resp. 10^{-4} . Contrary to that, topology preservation in the low-rank case in Exps. 6-8 yielded minima in the order of magnitude 10^{-9} to 10^{-12} . Moreover, the discrepancy error showed that low-rank achieved an order of magnitude 10^{-6} in comparison with 10^{-3} given by the full-rank case. In practice, topology preservation and low-rank constraint act synergistically together to get a more realistic deformation field in less time.

As a second part of our validation scheme and for an extended evaluation of our proposal, we illustrate the axial and lateral accumulated displacement of both the ground truth and our estimation of a single heart cycle (left side of Fig. 2.11). It is clear by visual inspection of the colored bar that our estimation is very close to the ground truth. In order to support this statement, we computed the Root-Mean Square error (RMSE) for the axial and lateral displacement and plotted the results in Fig. 2.12 where we can see RMSE values less than 1mm for the axial direction and less than 1.2mm for the lateral direction. The plots also show a concentration of values much lower than 1mm in both displacement directions.

To complement the analysis of the estimated displacement, in Fig. 2.13 we offer an analysis of the seven segments of the left ventricle during one cycle of the heart. Since the inferior and anterior parts are symmetric, we can evaluate their behavior for the axial and lateral displacements. As expected, axial displacements (blue circle) reported positive values acting in a similar way in both sides while lateral displacements (red squared) showed similar behavior with contrary signs since they go into opposite directions during heart motion.

We used the nonparametric Wilcoxon signed rank sum test to answer whether there is a statistical significant difference between the real values (ground truth) and the estimated values. We found that the null hypothesis was not rejected with $p < 0.05$ significance level. This lead us to conclude that we obtained a good estimation of the displacement.

To further support the results obtained in Fig. 2.10 – (A.1 and A.2) and Table 3, the right side of Fig. 2.11 shows the corresponding Jacobian determinant of the illustrated frames where we can see that most of the values met the desired criteria (stay at 1) for achieving topology preservation, which proofs the efficiency of our proposed term.

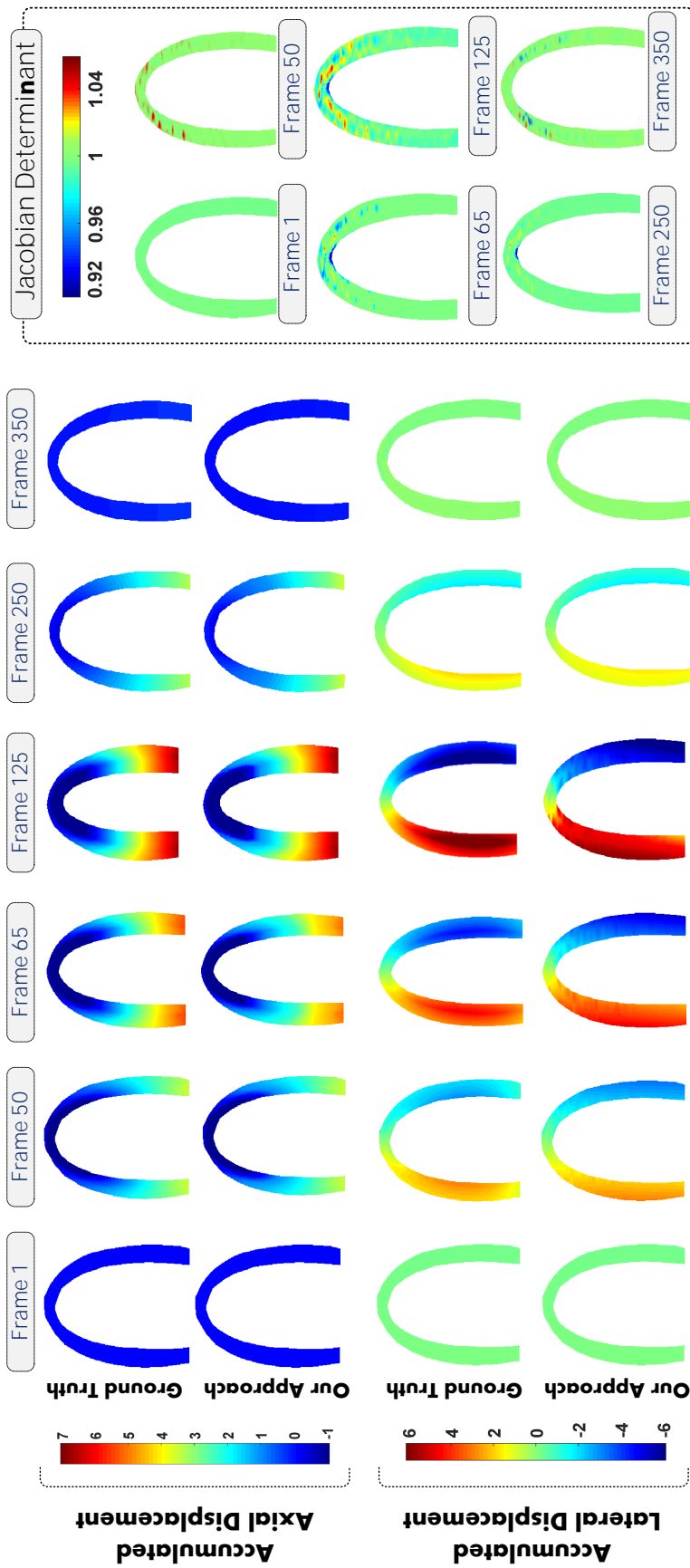


Fig. 2.11 (From right to left) Accumulated displacement for the apical view of the left ventricle. Few samples of the approximated axial and lateral displacements (top and bottom) are compared against the ground truth. Left side shows the Jacobian determinant of the same sample frames which reflects preservation of the anatomy.

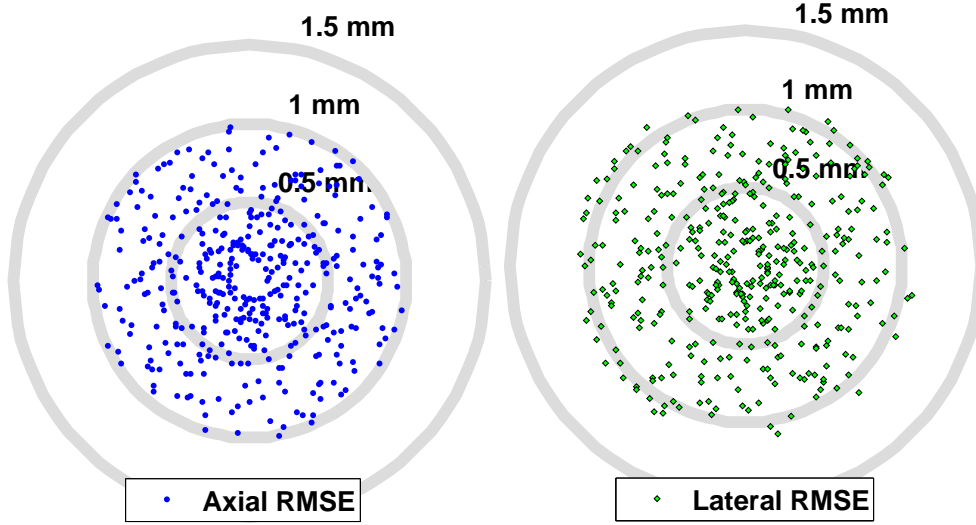


Fig. 2.12 Numerical comparison (in mm) between the real and estimated displacement values using Root-Mean-Square Error (RMSE).

Finally, we provide *strain images* as these are often used clinically. It is well-known that the strain can be calculated in terms of the components of the displacement vector field $\mathbf{h}(\mathbf{w})$:

$$\varepsilon(\mathbf{w}) = \begin{pmatrix} \varepsilon_{xx}(\mathbf{w}) & \varepsilon_{xy}(\mathbf{w}) \\ \varepsilon_{yx}(\mathbf{w}) & \varepsilon_{yy}(\mathbf{w}) \end{pmatrix} = \frac{1}{2}(C - \mathbf{I}) = \frac{1}{2}(\mathbf{F}^T \mathbf{F} - \mathbf{I}),$$

where C is the Green deformation tensor, $\mathbf{F} = \nabla \mathbf{h}(\mathbf{w})$ is the displacement gradient, and \mathbf{I} is the identity.

Strain is useful to evaluate the heart muscle and to identify subtle changes in heart's function [1]. Moreover, it allows representing the percentage change in dimension from a resting state to a stressed state (after applying a force). Fig. 2.14 shows the radial and longitudinal strains related to the left ventricle. The resulted plots can be evaluated according to the sign of the strain in which negative values indicate shortening and positive values denote stretching. According to the results at the upper part of the figure, radial strain reported a stretching behavior while longitudinal strain a shortening behavior. For illustration purposes, the radial strain of few frames from the sequence are displayed at the lower part of the same figure. The strain profiles and displacements of the left ventricle exhibit distinct features and clinically meaningful motion patterns.

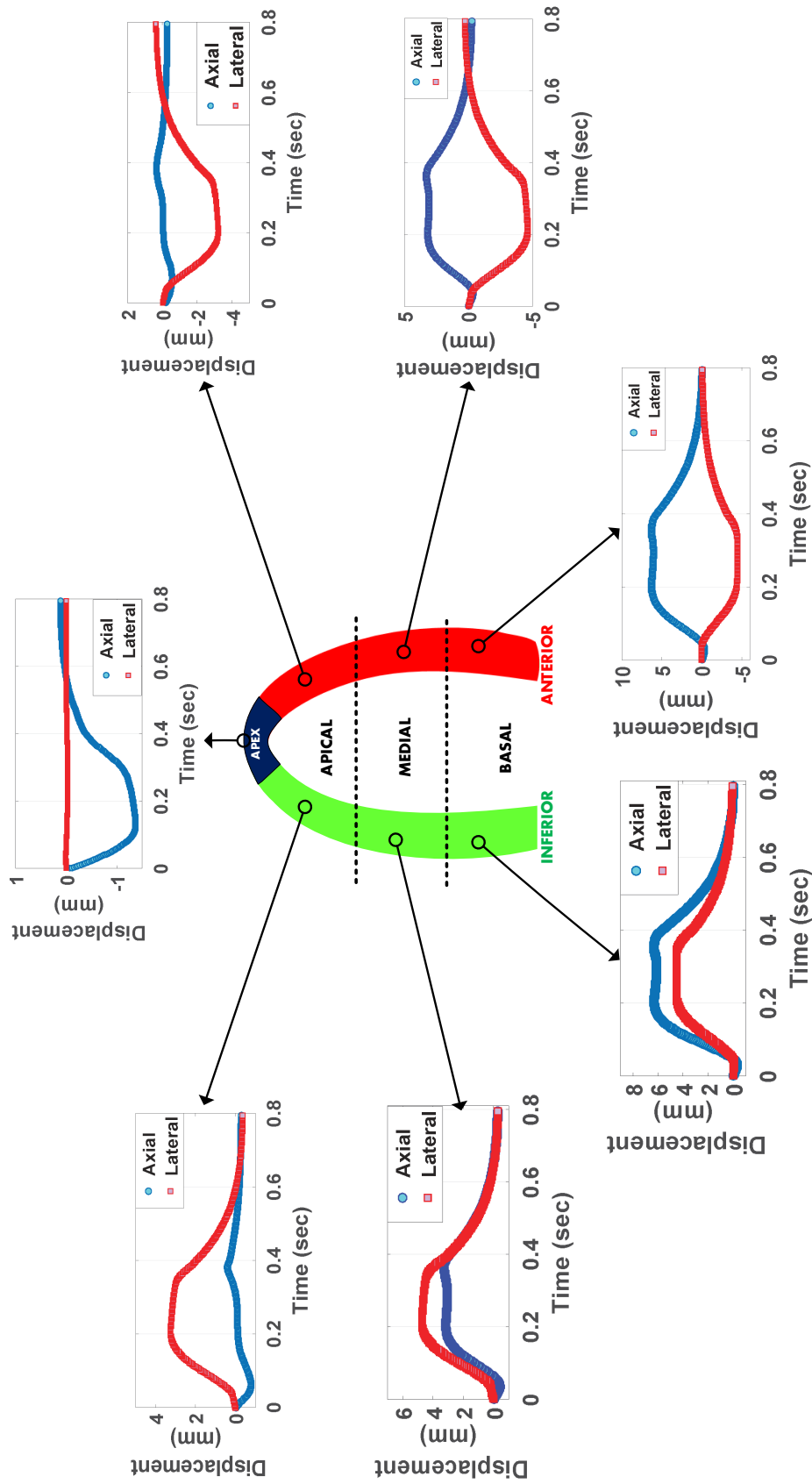


Fig. 2.13 Mean accumulated displacement of the seven segments of the left ventricle. Blue circles make reference to the axial displacement while red squares refer to the lateral displacement.

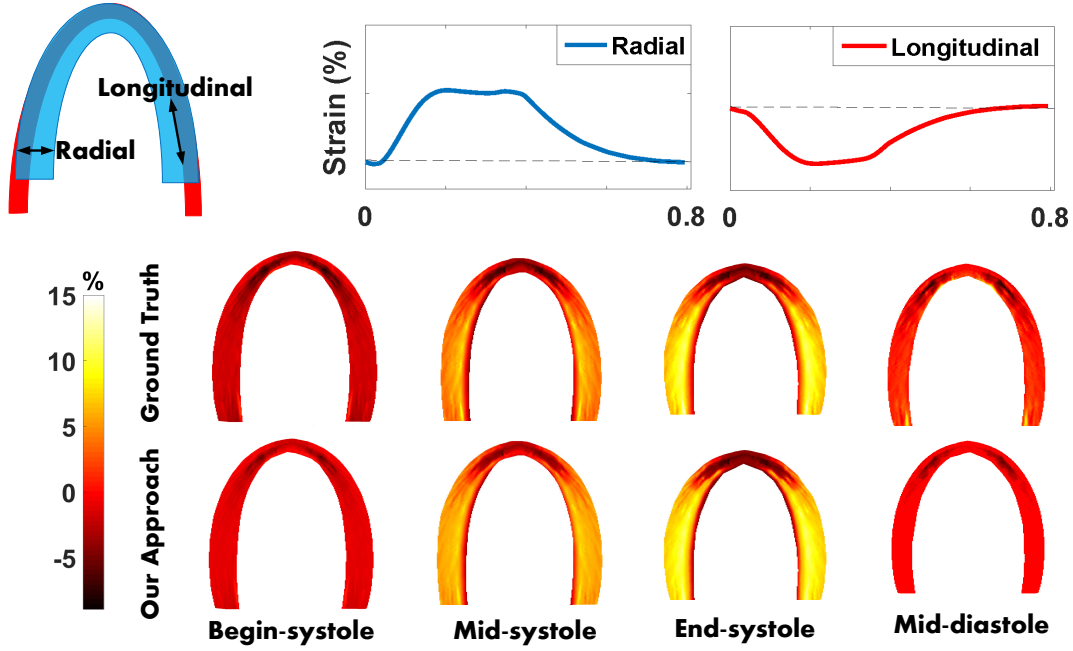


Fig. 2.14 (Top) Radial and longitudinal strain profiles of the left ventricle. These profiles are evaluated by their sign where negative values reflect shortening and positive ones reflect stretching. (Bottom) Few frames of the cardiac cycle showing the radial strain.

To further enhance our proposal, we optimized our variational framework [20] (Eq. 2.5) as follows. We first changed the M-estimator, ρ , from Eq. 2.9 for the Huber's M-estimator in which c is a positive tuning constant given by:

$$\rho_{huber}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq c \\ c|x| - \frac{1}{2}c^2 & \text{otherwise} \end{cases} \quad (2.17)$$

Moreover, as regularization term, E_{reg} , we used the curvature method. We chose this regularizer since it penalizes oscillations and contains harmonic functions (i.e. affine linear transformations). Let $\gamma \in \mathbb{R}^+$ be the regularization parameter. Then we use for $\mathbf{h} = (h_1, \dots, h_d)$ the following term:

$$E_{reg}(\mathbf{h}) = \gamma \sum_{l=1}^d \int_{\Omega} (\Delta h_l(\mathbf{w}))^2 d\mathbf{w} \quad (2.18)$$

The last change is the topology preservation term, E_{tp} . We propose a new term to guarantee the preservation of the anatomical structure of the tissue. It is defined as:

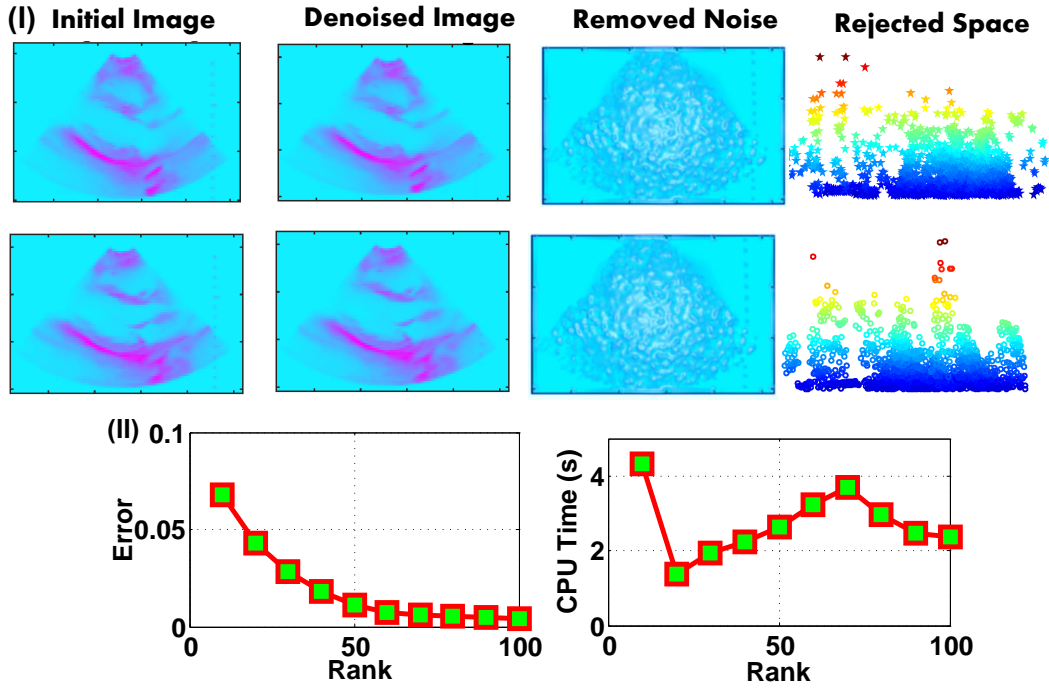


Fig. 2.15 Noise reduction achieved by low-rank representation. Part (I) at first column shows two noisy input frames while the next three columns show the denoised sequences, the removed noise, and the space where the eliminated noise lies. Part (II) shows the error and computational time of the rank- k approximation.

$$E_{\text{tp}}(\mathbf{h}) = \int_{\Omega} \delta_{\mathbf{h}}(\mathbf{w}) d\mathbf{w}, \text{ with} \quad (2.19)$$

$$\delta_{\mathbf{h}}(\mathbf{w}) := \begin{cases} \frac{1}{2} \frac{\pi - \arctan(|J_{\mathbf{h}}(\mathbf{w})|)}{\pi} + \varphi \sqrt{|J_{\mathbf{h}}(\mathbf{w})|^2} & \text{if } (\star) \\ 0 & \text{otherwise} \end{cases}$$

$$(\star) \quad | |J_{\mathbf{h}}(\mathbf{w})| - 1 | \geq \tau$$

where $\varphi \in \mathbb{R}^+$ offers a balance in our penalization, and $\tau \in \mathbb{R}^+$ is the margin of acceptance for values close to one.

We ran our variational framework taking Eqs. 2.17-2.19 in Eq. 2.5 using the Dataset I (as described in subsection 2.6.1). The results are explained next.

We first obtain the low-rank representation of the data. An important source of disturbance is noise, which affects the performance of the motion estimation of the heart. In part (I) of Fig. 2.15, we can see the result of doing noise reduction. The first column shows a couple of frames of the original sequence while the second one displays the frames after promoting low-rank representation. The

third column illustrates the removed noise and its corresponding subspace. The resulting matrix decomposition is displayed in Table 2.5 in which the original matrix of 13824x1814 was reduced to rank 100. The error and the CPU time against the rank can be seen in part (II) of Fig. 2.15. The error for \mathbf{C}_{100} was $2.3243e^{-4}$ and the average time to obtain \mathbf{C}_k was 2.3541 seconds. By promoting low-rank, we keep the subspace where relevant information relies and thus, artifacts can be eliminated.

It is well-known that evaluation of the deformation is complicated due to the lack of ground truth. Thereby, we evaluate our solution based on the following: 1) numerical results of the Jacobian determinant (Table 2.5 and Fig. 2.16 - (A) and (B)), 2) comparison of the residual error for low-rank and/or topology preservation (Table I), and 3) computation of the strain as a clinical measure (Fig. 2.16 - (C)).

Fig. 2.16 part (A) shows the deformation vector fields without topology preservation $E_{tp} = 0$. The color bars correspond to the values of the Jacobian determinant and we can see that negative or big values results on mesh distortions. It is also clear the fluctuation of the Jacobian determinant which result in unrealistic representation of the anatomical structure. We ran the same test but after including our topology preservation term. The results are illustrated in part (B), in which the Jacobian determinant values reflect stability, close to 1, with guaranteed positivity. Finally, part (C) shows both the displacement field and the strain magnitude on three areas of the heart: left ventricle posterior wall, right ventricle and left atrium. Strain helps to evaluate the heart muscle and identify subtle changes in heart function.

Previous results are supported by a set of numerical experiments reported in Table 2.5 where we explored different options for our variational framework. We can see that including topology preservation yields a better minima in the order of magnitude 10^{-3} to 10^{-15} . Another thing to note from experiments 4-5 and 9-10 is that promoting low-rank decreases the CPU time dramatically (about 77% in comparison with full-rank in experiments 1-3 and 6-8). Also, we compared the performance of different M-estimators, specifically Turkey (from our previous work [23]) against Huber estimator. Experiments 1-2 and 6-7 show that Huber outperformed Turkey and was able to find better minima. Moreover, from the results given by the experiments 9 and 10, we can conclude that synergy between low-rank and topology preservation is promising as it reduces the computational time and finds a better minima while ensuring a realistic representation of the heart anatomy.

Table 2.5 Performance analysis of full vs low rank for different cases of our variational framework

Data Representation	Casorati Matrix's Decomposition		Experiments for the Energy Functional (4)	CPU Time per frame (s)	Minimum	[min max] of $ J_h(\mathbf{w}) $
FULL RANK	Full Matrix \mathbf{C}		1) $\rho_{\text{unkey}}, E_{\text{disc}} = 0$ and $E_{\text{tp}} = 0$	13.6382	0.7810	[-3.1793 3.1029]
	\mathbf{U}	\mathbf{S}	2) $\rho_{\text{huber}}, E_{\text{disc}} = 0$ and $E_{\text{tp}} = 0$	13.4536	0.5932	[-3.6421 3.0564]
	13824x13824	1814x1814	3) $E_{\text{tp}} = 0$	12.6571	0.3419	[-2.7954 3.5973]
			4) (2.16) with $\varphi = 0$	12.5349	$4.1654e^{-3}$	[0.6420 1.1000]
			5) (2.16) with $\varphi = 5 \cdot 10^{-3}$	12.6259	$6.7468e^{-4}$	[0.8063 1.0089]
LOW RANK	Rank 100-approximation \mathbf{C}_{100}		6) $\rho_{\text{unkey}}, E_{\text{disc}} = 0$ and $E_{\text{tp}} = 0$	4.2340	0.5004	[-3.3049 3.2390]
	\mathbf{U}	\mathbf{S}	7) $\rho_{\text{huber}}, E_{\text{disc}} = 0$ and $E_{\text{tp}} = 0$	4.1948	0.3715	[-3.4915 3.2491]
	13824X100	100X100	8) $E_{\text{tp}} = 0$	3.7489	0.1519	[-3.1864 2.9543]
			9) (2.16) with $\varphi = 0$	3.7871	$7.1137e^{-12}$	[0.7634 1.0015]
			10) (2.16) with $\varphi = 5 \cdot 10^{-3}$	3.0597	$5.1597e^{-15}$	[0.9770 1.0038]

2.7 Conclusions and Future Work

In this Chapter, we presented a new approach to estimate cardiac motion using ultrafast ultrasound data. In a variational framework, we combined a penalizer for topology preservation with a low-rank data representation. Together with the better temporal resolution of ultrafast ultrasound, our proposed approach overcame challenges of non-rigid registration, including noise and complex heart motion and inaccurate results exhibiting distortions. While keeping the computational time relatively low, a realistic and clinically meaningful displacement field was produced, with the diffeomorphic features and preserved structures.

In our variational framework, the displacement was represented by a lattice with splines, and a maximum likelihood estimator was used in the discrepancy term to provide robustness against outliers. The regularizer for the topology has two features: eliminating radically negative values and carefully controlling the volume expansion and compression. We validated the accuracy of our approach and showed that it offers a RMSE less than 1 mm in comparison to the ground truth.

While this variational framework already gives good results and is strong individually, the CPU time and artifacts consumed in the ultrafast ultrasound sequence motivated us to promote a low-rank data representation, as it has proved to be useful in other areas of imaging and computer vision. We represented the data in a single Casorati matrix and used the dominant singular values to compute the deformation. Apart from removing the noise in the ultrasound data, this technique greatly reduced the computational time and produced, together with the topology penalization term, significantly less discrepancy in the results.

While we wanted to show the potentials of combining ultrafast ultrasound with low-rank techniques and topology preservation, from a technical point of view, the objective of this work is to have a first study as a proof of concept and to open a new line of research for further clinical investigation. In this work, we evaluated the technique using both simulated and realistic datasets. Future work will include more extensive evaluation with more subjects to examine the clinical potentials of the approach. Moreover, the technique promises to be useful for analyzing organs experiencing complex motion other than the heart, as for example the movement of lungs in respiration.

*“All sorts of things can happen when you’re open to
new ideas and playing around with things”*

Stephanie Kwolek

3

Towards Retrieving Force Feedback in Robotic-Assisted Surgery: A Supervised Neuro-Recurrent-Vision Approach

Robotic-Assisted Minimally Invasive Surgery (RAMIS) emerged from the need to address some deficiencies associated with traditional Minimally Invasive Surgery (MIS) and open procedures [233]. The revolutionary technologies utilized by RAMIS systems provide motion scaling and tremor filtering which stabilize the instruments and improve surgical precision [233, 198]. Furthermore, the added degrees of freedom in the tool tip enhances surgeons’ dexterity and results in better clinical outcomes [198]. The small incisions used in RAMIS allow reducing the amount of blood loss during surgery, minimizing trauma, and improving cosmetic results. Patients who undergo RAMIS experience less post-operative pain, faster recovery, and lower mortality and morbidity events [113, 208].

Despite all the benefits offered by RAMIS, current commercially available systems suffer from one major limitation which is the lack of force feedback [208, 54]. This feature is of huge importance since it increases surgeon-patient transparency [169] and allows more natural interaction with delicate tissues, as in the case of the heart (Fig. 3.1). Without force feedback information,

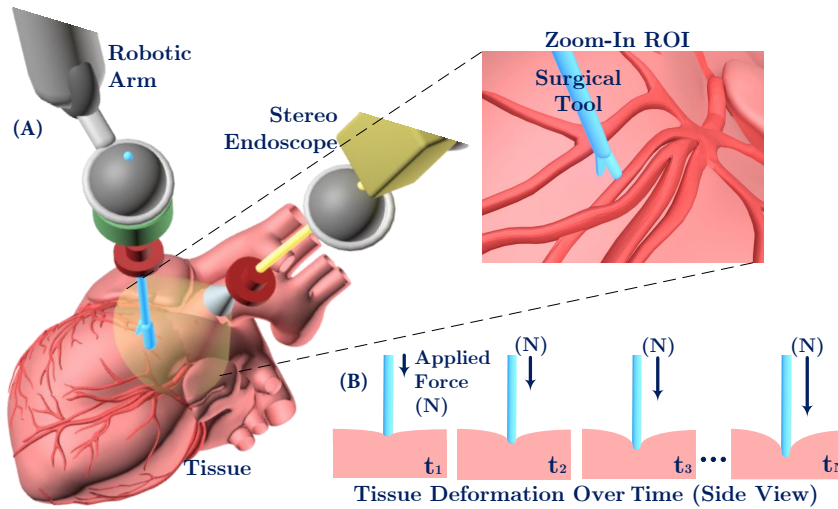


Fig. 3.1 (A) shows tool-tissue interaction during Robotic-Assisted Surgery which lacks force feedback that informs the surgeon about the amount of applied force. (B) shows the observable displacements after applying a force, which we obtain using a sensorless approach that relies, in part, on computing the 3D shape of the tissue over time.

surgeons have no means of knowing how much force is applied to the tissue, which could complicate the surgical task, increase its completion time and, what is worst, result in irreversible injuries [223, 167]. Furthermore, dealing with the absence of this primary sense of touch creates a high mental workload for surgeons and might be a hazardous source of distraction [121]. For these reasons, numerous researchers have dedicated significant efforts to address the problem of force feedback. However, up to date it is still considered an open problem [29].

In the search for solutions for the lack of force feedback, some researchers have focused their efforts toward developing force sensing devices (FSDs) [238, 221, 66]. These devices can be placed either inside or outside the patient’s body. When placed outside, the devices are attached to the robot or its instruments and offer indirect sensing. With this option, the devices measure not only the instrument-tissue interaction forces but also irrelevant force data given by the external/internal surgical environment. Removal of these undesirable measurements is not possible due to hysteresis and because they greatly depend on ambiguous starting conditions [162].

Alternatively, FSDs can offer direct sensing if they are placed close or on the tip of the instrument inside the patient’s body. However, the internal location of the sensor introduces numerous problems, including: biocompatibility and sterilization constrains; long-term stability; adaption to surgical tool; size and

high cost [83, 205]. All these limitations put severe restrictions to the adoption of FSDs in real surgical environments. An alternative to the use of FSDs is to compute the interaction forces by the observable deformation of the tissue in what is called Vision-based force estimation (see Subsection 1.1 for details). Whether it is FSD or VBFE, an important factor after having the force information is how to transmit it to the surgeon. The direct way is to use haptic devices but they suffer from several limitations including stability, number of degree of freedom, cost and space, which make their use complicated. An alternative way is to use Force Sensory Substitution (FSS). When FSS is used physical properties of the environment are sensed using an alternate sensing modality. The potential benefits of FSS for force feedback in teleoperation tasks were first explored by Massimino and Sheridan [133].

When force sensory substitution is used, force feedback can be transmitted to the surgeon through other sensory modalities such as vision, audio or tactile. For further explanation about sensory substitution refer to Chapter 4. Particularly in this work, we use vision modality, which is considering a promising sensory substitution suitable for clinical adoption [162]. With this alternative, surgeons perceive force information via visual cues of tool-tissue interaction. Various studies have investigated the feasibility of visual feedback on conveying force information for surgeons while performing delicate tasks. Investigation results show improved performance among novice surgeons and decreased inconsistencies [? 162]. Out of the different FSS modalities, in this work we chose visual feedback as it has proven to offer more advantages over other alternatives (clinical evaluation is presented in Chapter 4). Moreover, the use of visual information has been proven to be very reliable for force estimation as all RAMIS settings include a videoscopic view of the operation. Thereby, in order to avoid using force sensors, we can employ the available visual information of the tool-tissue interaction and relate it to the applied force.

In this chapter, we describe a novel approach to estimate the applied forces during RAMIS interventions that is illustrated in Fig. 3.2. Since all RAMIS settings include a videoscopic view of the operation, we can employ the available visual information of the tool-tissue interaction and relate it directly to the applied force. From the conservation principles of continuum mechanics, it is clear that the change in shape of an elastic object is directly proportional to the force applied. Following this principle, we propose a novel approach that is based on a variational framework that allows computing the observable deformation after a force is applied. Then, this information is used in a learning system that

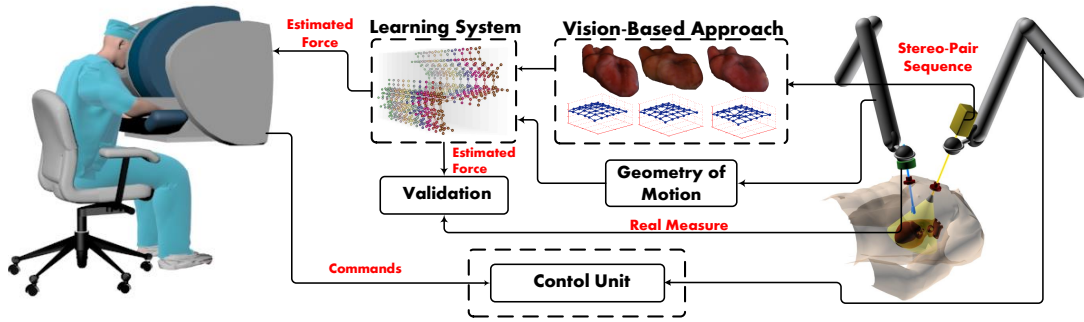


Fig. 3.2 Flowchart of our approach for estimating applied forces in robotic surgical systems. We first propose a visual approach to compute the deformation structure over time. Then, the available information is used as input to an artificial neural network which accurately estimates the applied force.

finds the nonlinear relationship between the given data and use it to estimate the applied force. In particular, our contributions are:

- A new energy functional to compute the 3D tissue deformation. We prove numerically that it offers a better minima with a low computational cost.
- We propose the use of a powerful supervised learning system that allows finding the optimal nonlinear relationship between the given data and the applied force. We demonstrate the adaptability across subjects and the stability of our solution during long periods of time based on in-vivo and ex-vivo datasets.

Our proposed force estimation solution avoids the drawbacks usually associated with force sensing devices, such as biocompatibility and integration issues. We evaluate our approach on phantom and realistic tissues in which we report an average root-mean square error of 0.02 N.

The remainder of this chapter is organized as follows. A revision of the literature related to vision-based force estimation is presented in Section 3.1. In Section 3.2 we describe our energy functional to compute the 3D deformable shape recovery, whereas in Section 3.3 we describe the deep architecture used to learn the relationship between the extracted visual-geometric information and the applied force, and to find accurate mapping between the two. In Section 3.4, we numerically evaluate our approach on phantom and realistic tissues. Finally, Section 3.5 gives a global conclusion and future works.

3.1 Vision-based Force Estimation

Vision-based force estimation can incorporate explicit knowledge of the mechanical properties of the tissues. However, this requires both complex calculation and adaptation to each tissue. To avoid these drawbacks, knowledge about the tissue properties can be learned implicitly from the data itself. This makes the system more suitable for real-time solutions since the learning can be optimized for faster computation.

The viability of using visual information to estimate the applied forces has been demonstrated in different scenarios. In 2D, Greminger et al. in [76] used Dirichlet to Neumann map to estimate the force distribution applied to a deformable object for microassembly and biomanipulation. Authors measured the displacement field of the contour of the object and then used a template matching based on linear elasticity equations. Similarly, authors in [105] modeled the deformation by introducing contour information of the object, together with its mechanical properties, into the boundary element method. Then, deformation data was used to compute the applied forces by means of a capacitance matrix. The disadvantage of this proposal is the need of a prior knowledge of the object's material properties.

The concept of virtual template for computing the deformation of the object, using monocular images, was presented in [158]. In that work, authors assumed that the surface of the object is a smooth function with local deformation. Then, they used a strain-stress relation together with the penetration depth to estimate the force. Authors in [106] applied a mesh-based model to characterize the deformation based on stereo-endoscopic images. Afterwards, a spring-damper system was used to compute interaction forces. Authors in [7] attempted to improve the realism of visual and haptic feedback in a cell injection system by using a 3D nonlinear mass-spring-damper model. The model parameters were identified using offline Finite Element Method (FEM) simulations and the biomembrane geometry deformation was reconstructed using snakes based visual tracking. However, as shown in [104], the use of mass-spring models offers limited accuracy, and the FEM-based parameters computation requires additional modeling efforts.

More recently, some researchers have investigated the use of soft computing to improve the accuracy of the force estimation. Authors in [75] computed the applied force using a 2-layers feedforward network incorporated into a deformable template matching algorithm. The deformable template was an iterative computation of the object's edge, using Canny's method. Karimirad

et al. in [103] used a feedforward Artificial Neural Network (ANN) to estimate the force applied to cells during micromanipulation. The neural network was trained on geometric features of the cells, including deformation, orientation, and size. These features were extracted using various image processing techniques under different known force conditions. Two different hybrid intelligent systems were proposed in [145] to model the tool-tissue force in laparoscopic surgery: an adaptive coevolutionary fuzzy inference system and an adaptive neuro-fuzzy inference system. Both systems were trained on three different geometric features extracted from a 2D simulated deformable model: angle and depth of maximum deformation and width of displacement constraint. Nonetheless, experiments in both works, [103] and [145], were only conducted in 2D.

We have also contributed in the interaction forces recovery. In our work presented in [12], an energy minimization strategy was applied to compute a deformation structure from the acquired stereo image sequences. The deformation structure, along with geometric data from the robotic manipulator, was used as an input to a Recurrent Neural Network (RNN) which was trained using the adapted Levenberg-Marquardt method. A modification of the RNN architecture was presented in [16], in which three types of feedback were defined: local, global and no feedback. With the aim of increasing previous system accuracy, in a recent work [17] we used a Long-Short Term Memory RNN (LSTM-RNN) architecture. This LSTM-RNN allowed preserving information for a longer period of time, which enforced constant error flow.

3.2 3D Deformable Shape Recovery

The first part of our solution for estimating the applied force is the computation of the deformation structure as shown in Fig. 3.2. In this work, 3D shape recovery is accomplished by minimizing an energy functional reformulated using the l_2 -regularized optimization class. Moreover, in order to reduce the computational time, we parametrize the changes produced on the tissue surface using a set of linearly independent vectors. In the remainder of this section, we present a formulation that allows recovering the deformation produced when a force is applied on the tissue surface over time.

Let us assume that $\mathbf{I}_l^t : \Omega_{\mathbf{I}_l} \rightarrow \mathbb{R}^2$ and $\mathbf{I}_r^t : \Omega_{\mathbf{I}_r} \rightarrow \mathbb{R}^2$ are the left and right image views from a stereo pair image acquired at each instant time t , where $\Omega_{\mathbf{I}_l}$ and $\Omega_{\mathbf{I}_r}$ are their corresponding domains. Since during a medical procedure

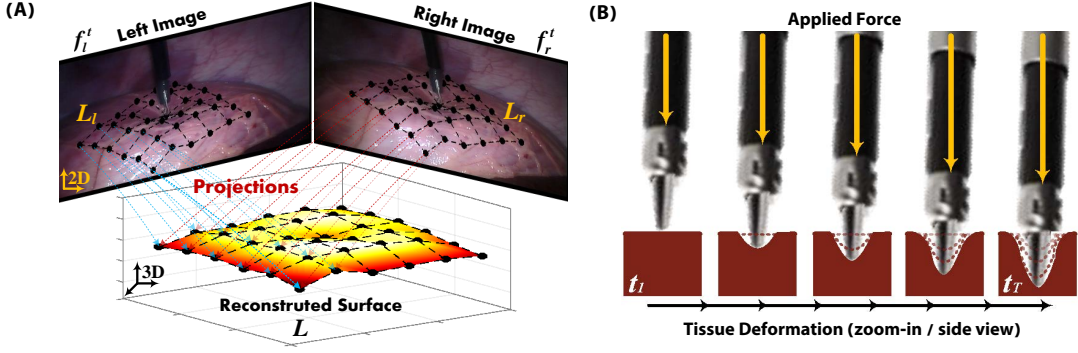


Fig. 3.3 (a) The 3-dimensional tissue surface is reconstructed from the projections of homologue points on the left and right lattices. (b) Illustration of how tissue deformation is directly proportional to the applied force.

the surgeon is interested only in the region to be repaired, for example a vein, computational cost can be reduced by defining a region of interest (ROI).

We handle the specified ROI using a 3D lattice. Let $\mathcal{L}_l : \Omega'_{I_l} \subseteq \Omega_{I_l}$ and $\mathcal{L}_r : \Omega'_{I_r} \subseteq \Omega_{I_r}$ be 2D lattices defined at each image view respectively. Then, the 3D lattice is computed from the projections of the corresponding lattice points on $\hat{I}_l \subseteq I_l$ and $\hat{I}_r \subseteq I_r$ (see Fig. 3.3). Let \mathbf{P} be the result of such correspondences and v be the number of lattice points, $P_v \in \mathbf{P}$ where $P_v = (y_1, \dots, y_m) \in \mathbb{R}^m$. Considering that initially lattice points are evenly spaced, then the changes produced on the tissue surface, over time, are computed by minimizing the total energy, \mathbf{E}_t , such that the optimal \mathbf{P} can be found using the following equation:

$$\begin{aligned} \mathbf{E}_t(\mathbf{P}) = & \mathbf{E}_\Phi(\hat{I}_l^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}), \hat{I}_r^t(\mathbf{x})) + \\ & \gamma \mathbf{E}_\Psi(\Gamma(\mathbf{x}; \mathbf{P})) + \mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P}) \end{aligned} \quad (3.1)$$

where \mathbf{E}_Φ is the discrepancy measure term, \mathbf{E}_Ψ denotes the penalization term used to obtain a plausible transformation, $\gamma \in \mathbb{R}^+$ is the parameter that controls the quality of the data fit, \mathbf{E}_Λ gives a constraint to preserve shape, \mathbf{x} is a vector containing the coordinates, and Γ is the deformation model.

The deformation model is an essential factor that determines how fast and accurate the approach is. In order to find a compromise between computational cost and accuracy, we characterize the lattice points using the tensor product of b-splines as they demand low running time, allow multiresolution, have optimal mathematical properties and keep affine invariance [222]. The mapping of the changes over the reconstructed lattice Γ at position \mathbf{x} is given as follows:

Definition 5 Let $\mathbf{P}_{lmn} \in \mathbb{R}^3$ denote the displacement of a control point with $z := y_1 y_2 y_3$ number of points. Then,

$$\Gamma(\mathbf{x}; \mathbf{P}) = \sum_{l=1}^{y^1} \sum_{m=1}^{y^2} \sum_{n=1}^{y^3} \mathbf{P}_{lmn} \prod_{k=1}^K \xi_{k,c}(x_k) \text{ for } k = 1, \dots, 3 \quad (3.2)$$

where $\xi_{.,c}$ are the cubic basis splines function expressed as:

$$\begin{aligned} \xi_{k,0}(x) &= (1-x)^3/6 & \xi_{k,1}(x) &= (4+3x^3-6x^2)/6 \\ \xi_{k,2}(x) &= (1-3x^3-3x^2+3x)/6 & \xi_{k,3}(x) &= x^3/6 \end{aligned} \quad (3.3)$$

We now turn to reformulate the energy functional defined in Eq. 3.2. The discrepancy term, \mathbf{E}_Φ , is computed using the sum of squared differences method. This was selected because it has a low computational cost and offers an optimal result when images are acquired with the same sensor, as it is in our case.

Moreover, since the 3D deformation recovery is an ill-posed problem, as defined in Definition 6, it is necessary to have a penalization term to restrict the solution space and impose stability to the energy functional.

Definition 6 *Consider the problem $Bf = g$ where $B \in \mathfrak{L}(J, V)$ and J, V –Hilbert Spaces. Then the problem is well-posed, in the sense of Hadamard [80], if:*

- $g \in V$ has solution $f_* \in J$ i.e. g is in the range of B
- the solution of $Bu = g$ is unique
- f_* depends continuously on the data

Thus, to obtain a well-posed problem we rewrite the penalization term \mathbf{E}_Ψ using Tikhonov regularizer. The third term of the functional is given by a soft constrain for volume-preserving mappings. Taking previous statements, Eq. 3.2 results in:

$$\begin{aligned} \hat{\mathbf{E}}_t(\mathbf{P}) &= \underbrace{\|\mathbf{E}_\Phi\|_{L^2}^2}_{\text{discrepancy}} + \underbrace{\gamma\|\mathbf{E}_\Psi\|_{L^2}^2}_{\text{penalization}} + \underbrace{\|\mathbf{E}_\Lambda\|_{L^2}^2}_{\text{constraint}} \\ &= \frac{1}{S} \left(\int_{\mathbf{x} \in \Omega'} \|\hat{\mathbf{I}}_l^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}) - \hat{\mathbf{I}}_r^t(\mathbf{x})\|^2 d\mathbf{x} \right. \\ &\quad \left. + \gamma \sum_{i=1}^d \int_{\mathbf{x} \in \Omega'} \|\nabla \Gamma_i(\mathbf{x}; \mathbf{P})\|^2 d\mathbf{x} + \int_{\mathbf{x} \in \Omega'} \|\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P})\|^2 d\mathbf{x} \right) \end{aligned} \quad (3.4)$$

where S is the number of overlapping pixels. In search of practicality and efficiency, we use a discretize-then-optimize process. Strictly speaking, after defining the continuous optimal energy functional, as defined in Eq. 3.4, we transform it into a standard optimization problem by discretizing it resulting in:

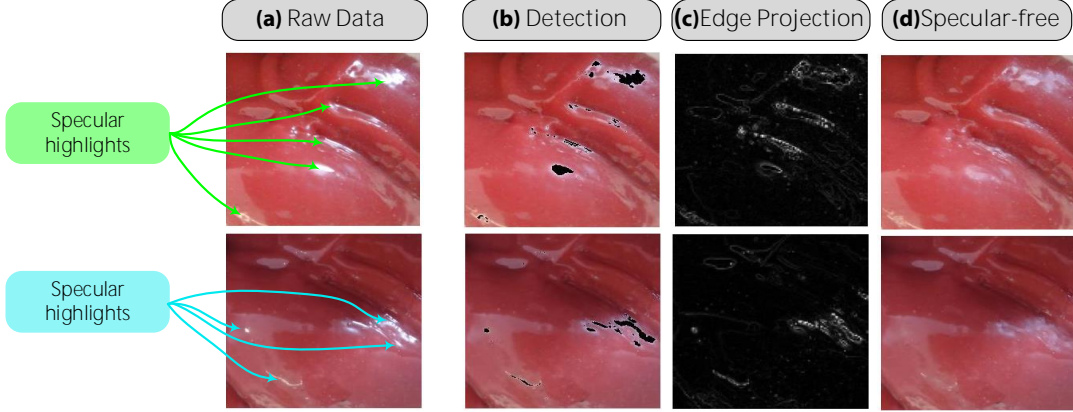


Fig. 3.4 Specular highlights cause major tracking disturbance. We deal with this issue using a real-time detection and inpainting approach that accurately recovers a specular-free image.

$$\hat{\mathbf{E}}_t(\mathbf{P}) = \frac{1}{S} \left(\sum_{\mathbf{x} \in \Omega'} \|\hat{\mathbf{I}}_t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}) - \hat{\mathbf{I}}_r^t(\mathbf{x})\|^2 + \gamma \sum_{i=1}^d \sum_{\mathbf{x} \in \Omega'} \|\nabla \Gamma_i(\mathbf{x}; \mathbf{P})\|^2 + \sum_{\mathbf{x} \in \Omega'} \|\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P})\|^2 \right) \quad (3.5)$$

where the soft constraint $\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P}) = \det(\nabla \Gamma(\mathbf{x}; \mathbf{P}))$.

3.2.1 Robust 3D Shape Recovery

During the process of recovering the temporal 3D deformable structure, different factors can affect the performance of the visual approach.

One source of error that might affect the reconstruction precision is the specular highlight regions (see Fig. 3.4-(a)) that appear on the surface of the heart. These bright spots appear on surfaces with high reflectivity and occlude the underlying visual information causing uncertainty in the tracked Region of Interest (ROI). To eliminate this artifact, we carried out two steps:

1. Detection of the specular highlights – using a hybrid detection technique based on saturation and intensity color attributes since the local coincidence of the intense brightness and unsaturated color characterize these kind of artifacts (Fig. 3.4-(b)). Then, we applied a refinement of the detection process based on the computation of the local singularities based on the Wavelet Transform Modulus Maxima (WTMM) (Fig. 3.4-(c)).

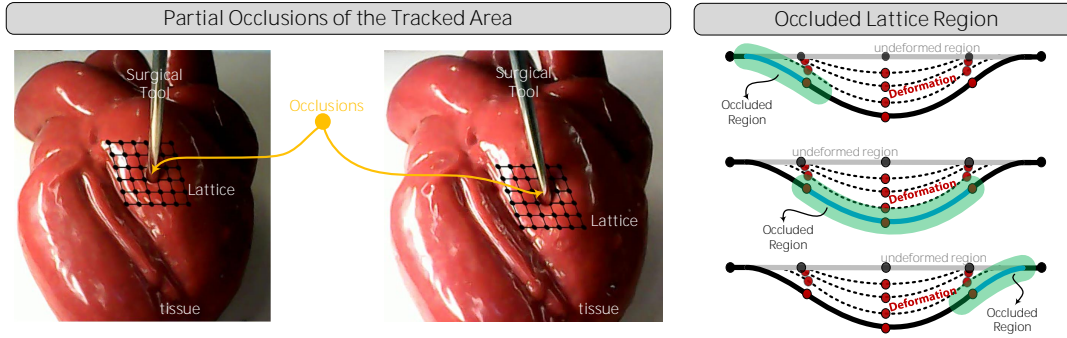


Fig. 3.5 Surgical tools can partially occlude the tracked region of interest which affects the 3D shape recovery over time (Left side). Right side shows a side view of occluded lattice regions from different views.

2. Inpainting – we reconstructed the damaged regions using a dynamic search based approach to smoothly propagate pixel information from the surrounding areas (Fig. 3.4-(d)). We also optimized the process to perform in real time. Details of these two steps can be seen in our collaboration work published in [4].

Another potential source of error when tracking the surface deformation is the partial occlusion of the tracked region of interest (ROI). Occlusion makes the tracking process more challenging and can cause tracking failure as the algorithm will not have enough information about the occluded part of the surface. In RAMIS settings, the tracked ROI may be partially occluded for a short period of time, by a surgical tool or blood, which might hide useful information about the surface and affect the tracking precision (see Fig. 3.5). This source of error needs to be eliminated as precision is an essential factor in medical applications.

As it was shown by Turkey in 1920 [ref], the L_2 norm is very sensitive to small deviations leading to affect the residuals. To increase robustness and deal with the aforementioned problem, we took Eq. 3.5 and included a maximum likelihood estimator called Huber’s M-estimator as stated in the following definition:

Definition 7 Let \mathbf{r} be the residuals and \mathbf{z} the L_2 estimator in the form:

$$\sum_{i=1}^n \mathbf{r}_i(\mathbf{z})^2 = \min \quad (3.6)$$

then the Huber’s maximum likelihood estimator ρ with a positive tuning constant c is given by:

$$\rho_{huber}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq c \\ c|x| - \frac{1}{2}c^2 & \text{otherwise} \end{cases} \quad (3.7)$$

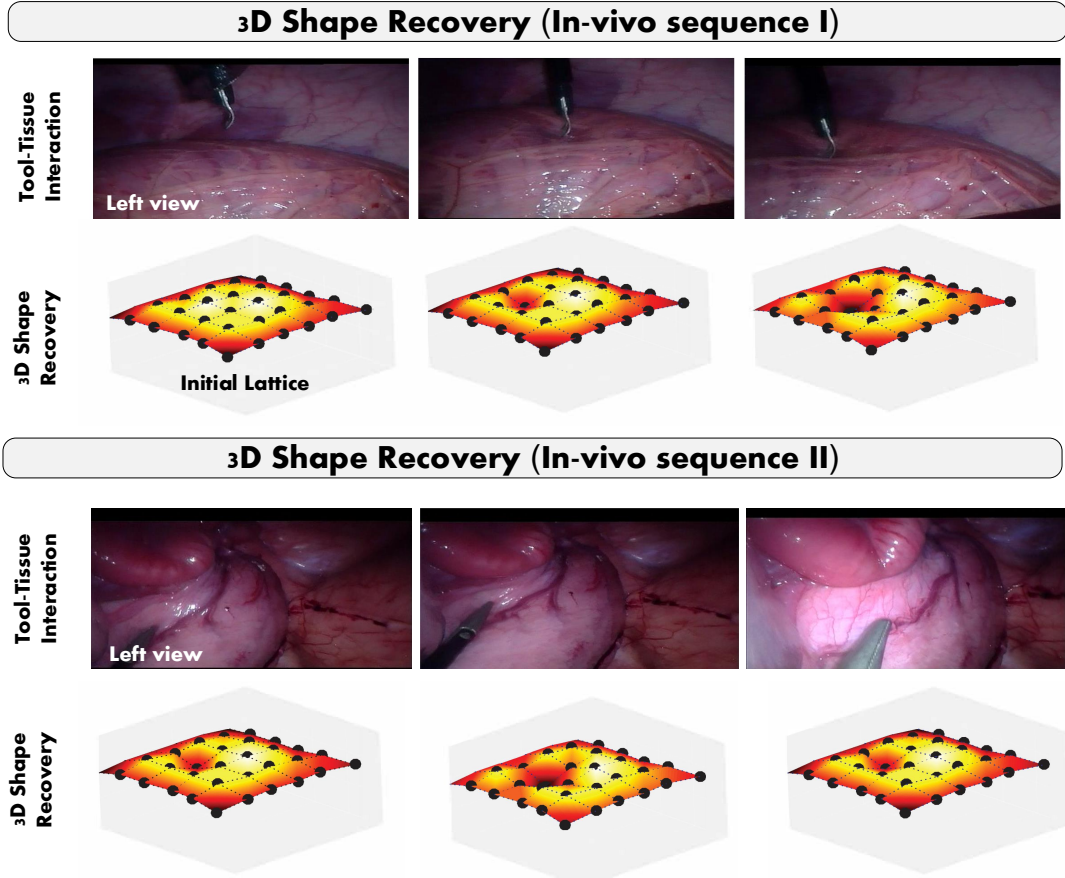


Fig. 3.6 3D deformation structure of the tissue, obtained by our vision approach, plotted at different time instants.

and substitutes Eq. 3.6 in the form:

$$\sum_{i=1}^n \rho(\mathbf{r}_i(\mathbf{z})) \quad (3.8)$$

Rewriting Eq. 3.5 using Definition 7, the new total energy, $\check{\mathbf{E}}_t$, is expressed as:

$$\begin{aligned} \check{\mathbf{E}}_t(\mathbf{P}) = \frac{1}{S} & \left(\sum_{\mathbf{x} \in \Omega'} \rho(\hat{\Gamma}_l^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}) - \hat{\Gamma}_r^t(\mathbf{x})) \right. \\ & \left. + \gamma \sum_{i=1}^d \sum_{\mathbf{x} \in \Omega'} (\Gamma_i(\mathbf{x}; \mathbf{P}))^2 + \sum_{\mathbf{x} \in \Omega'} (\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P}))^2 \right) \end{aligned} \quad (3.9)$$

Once the total energy, $\check{\mathbf{E}}_t$, is defined, we turn to finding the optimal value. To do that, we use the Levenberg-Marquardt (LM) method [122, 132]. LM combines the stability of the gradient descent and the fast convergence of the Gauss-Newton. LM makes use of a damping parameter, δ , in order to switch

between the gradient descent and the Gauss-Newton. When δ is small, it acts as Gauss-Newton with the difference that it uses a trust-region with radius Δ_h instead of a line search. While when δ is large, it performs as gradient descent. The search direction, d_h , at iteration h is computed as follows:

$$\begin{aligned}(\mathbf{J}_h^\top \mathbf{J}_h + \delta \mathbf{I})d_h &= -\mathbf{J}_h^\top \mathbf{r}_h \quad \delta > 0 \\ d_h(\delta) &= -(\mathbf{J}_h^\top \mathbf{J}_h + \delta \mathbf{I})^{-1} \mathbf{J}_h^\top \mathbf{r}_h\end{aligned}\tag{3.10}$$

where \mathbf{J} is the Jacobian, \mathbf{r} is the residual vector, and \mathbf{I} is the identity matrix.

In order to illustrate the 3D deformation recovery mapping, in Fig. 3.6 we show the deformation structure, bounded by our defined lattice, recovered using our proposed visual approach from two in-vivo datasets. The tissues experience deformation from applying force over time and darker shades represent intense deformation at contact point. The plots clearly show pleasant visual results of the deformation field even during changes of illumination or complex deformation. Detailed numerical results can be found in Section 3.4.

3.3 Retrieving Force Feedback

The force estimation strategy proposed is part of the robotic surgical system shown in Fig. 3.2. In a general RAMIS setting, a surgeon controls the robotic manipulator through a teleoperation control unit that scales and transforms the given commands into relative motion. A stereo pair camera is used to track this motion and feed the image sequences to the vision-based module, which is the first part of our estimation strategy. This module uses the acquired visual information to retrieve the deformation observed on the tissue surface after applying a force. The structure deformation information, along with the geometry of motion given by the robotic manipulator, are given as input to the second module, which is the neural approach. In this module, we use recurrent learning to analyze the given information and map it into an accurate force estimation. As a final step, the estimated force is validated against the real force measurement given by a robotic sensor attached to the surgical tool for training and validation purposes.

3.3.1 Force Estimation: Supervised Recurrent Learning

In this subsection, we describe our strategy to estimate the applied forces. In particular, we use Artificial Neural Networks to find the relationship between the input data (visual and geometry information) and the applied force.

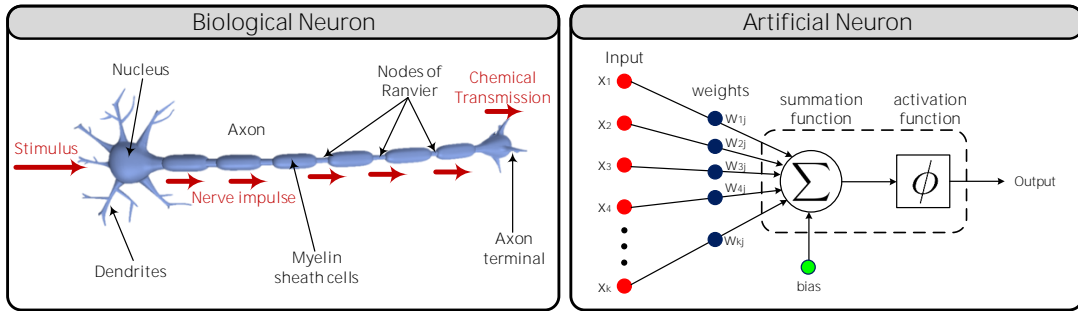


Fig. 3.7 Left-side shows the structure of a biological neuron of the human brain while the right-side shows an artificial neuron that imitates the functioning of the biological one.

From the biological point of view and roughly speaking, a neural network is a collection of neurons in which each neuron is composed of a cell body and a set of dendrites that transmit information through the axon using electrical and chemical signals (see Fig. 3.7 left side). While this definition is enough for our purpose, details about the biological meaning can be found for example in [8, 28]. ANNs were inspired by this biological view in order to capture the human brain function. The first mathematical model was proposed in 1943 by McCulloch and Pitts in [134] where they described the neural events by means of propositional logic. Later on, these mathematical models were extended to computational models (for instance see [93, 42, 191]) capable of solving large number of problems.

An ANN is composed of interconnected neurons which activations describe a well-defined path. The functioning of an ANN starts by having a set of inputs which are multiplied by weights in order to obtain the connection strength between them. These values go to a summation function and then the values are modified based on an activation function. The activation goes through all available neurons until reaching the output neuron (see Fig. 3.7 right side). A formal definition of an artificial neuron is given next:

Definition 8 *An artificial neuron is a function of two vector variables, the weights (w) and the input set $X \subset \mathbb{R}^k$ expressed as $\Phi(w, x) = \phi(\langle w, x \rangle) \in \mathbb{R}$, where $\langle \cdot, \cdot \rangle$ is the dot product and ϕ is the activation function.*

Based on Definition 8, we can now formally define an Artificial Neural Network as:

Definition 9 *An artificial Neural Network (ANN) is a triple (N, V, w) where N is a set of neurons which connections between the i – th and j – th neurons are*

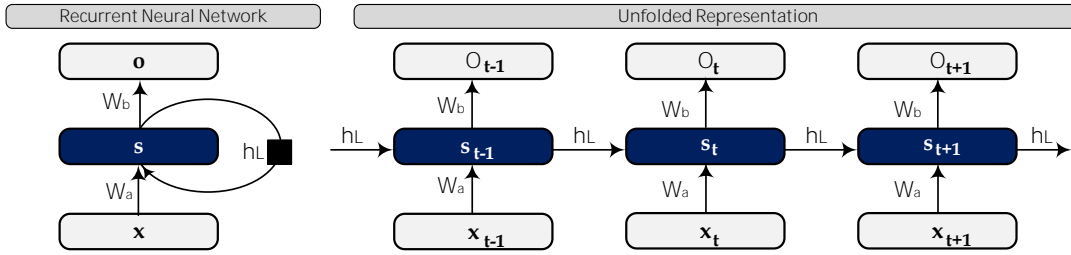


Fig. 3.8 Left side a simple recurrent neural network while right side shows its unfolded version through time.

given by the set $V = \{(i, j) | i, j \in \mathbb{N}\}$. The synaptic coupling between the i -th and j -th neurons is given by $w_{i,j}$.

When using common feedforward neural networks, the inherent assumption is that the inputs/outputs are independent between them. However, there are a lot of applications in which it is convenient to make use of the temporal information as in the case of speech and handwriting recognition, weather forecasting, music synthesis, financial prediction and the estimation of the applied forces, to name a few. In these cases, a more advantageous class of ANNs is Recurrent Neural Networks (RNNs). A RNN allows introducing memory by having feedback connections at their units, which enables dynamic temporal processing instead of a hierarchical one (for illustration purposes and a better understanding refer to Fig. 3.8). Moreover, RNNs can be seen as deep neural networks (DNN) when folded out in time with indefinitely layers [85]. They have demonstrated to exhibit different advantages including handling noise-contaminated data and creating complex input-output relationships.

One of the most impressive characteristics of the human brain is its ability to learn. Based on this, a natural question that arise is – how do ANNs learn? In machine learning, three major learning paradigms for ANNs can be distinguished:

- Supervised Learning – learning algorithms within this category rely on input-ouput pairs provided during the training process to produce an inferred function. This function will serve to map new coming data (i.e. unseen instances). Updating the network is based on an error function between the target and actual output.
- Unsupervised Learning – unlike supervised learning, this kind of algorithms receive a set of unlabeled inputs and have as an objective finding a function such that hidden patterns can be found.

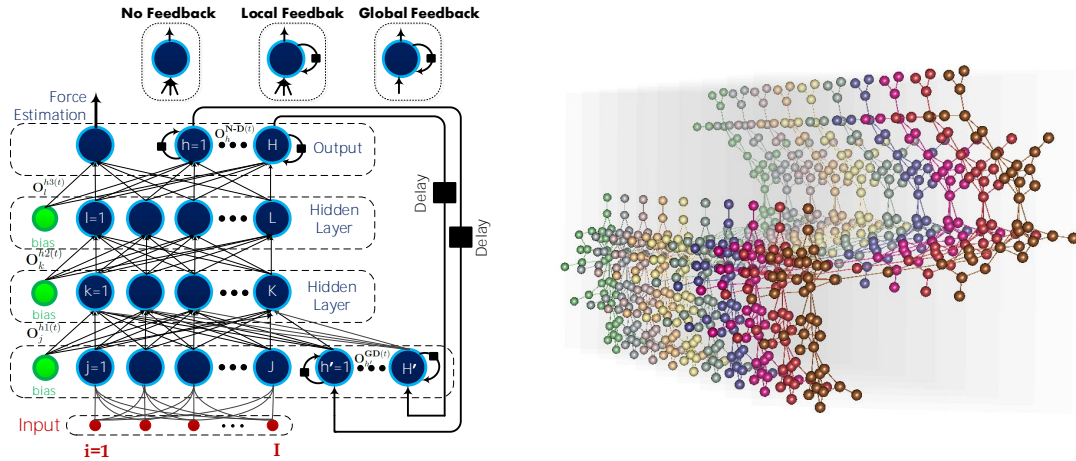


Fig. 3.9 Estimation of the applied forces is achieved by means of a RNN in which three types of output units can be identified (zoom in the upper row). Those units with delayed feedback save past information that helps to increase accuracy. Additionally, at the right side a visualization of the network over time is displayed.

- Reinforcement Learning – it is close related to supervised learning, however, it differs in the fact that instead of providing a target, it gives a reward based on the system actions and how well it performs (i.e. online performance).

In this chapter, the supervised learning paradigm is used in the RNNs. The formal definition is stated next.

Definition 10 *Given a set of \mathfrak{N} training samples in the form of input-output pairs $\{(x_1, y_1), \dots, (x_{\mathfrak{N}}, y_{\mathfrak{N}})\}$, where x is a feature vector and y is its corresponding target value, supervised learning finds a function $f: X \rightarrow Y$ that maps the input space, X , to the output space, Y , and works well on unseen inputs x .*

2-type feedback RNN: A first Approach

As a first solution, in [16] we propose the use of RNNs for estimating the interaction forces. In this work, two main types of feedback are used. The former is a local feedback that creates a loop to a unit itself whereas the later is a global feedback that goes from the output to the input of the network. According to these feedback, three types of outputs are defined: no feedback \mathbf{O}^{N-D} , local feedback \mathbf{O}^{LD} and global feedback \mathbf{O}^{GD} . Our architecture can be seen in Fig. 3.9. Let ξ and \mathbf{b} be the input vector with I inputs and the bias respectively. Moreover, consider J, K, L and H as the number of units at each layer and w the weights.

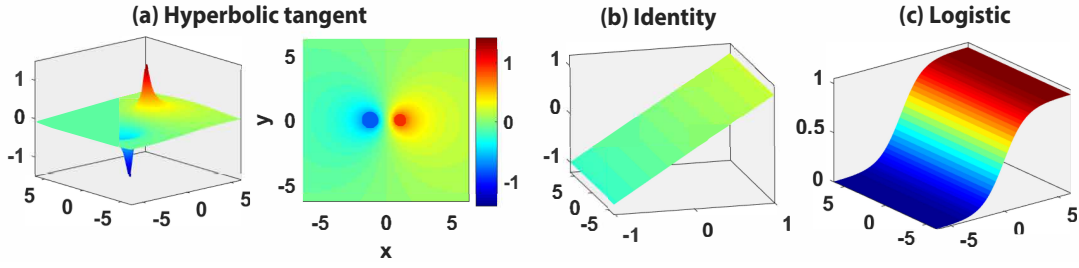


Fig. 3.10 Three dimensional illustration of the activation functions used in the architecture.

Activation Function	Equation	Derivative	Range	Order of Continuity
Identity	$\phi(x) = x$	$\phi'(x) = 1$	$(-\infty, \infty)$	C^∞
Hyperbolic tangent sigmoid	$\phi(x) = \frac{1}{1+e^{-2x}} - 1$	$\phi'(x) = 1 - f(x)^2$	$(-1, 1)$	
Logistic	$\phi(x) = \frac{1}{1+e^{-x}}$	$\phi'(x) = f(x)(1 - f(x)^2)$	$(0, 1)$	

Table 3.1 Summary of Activation Functions used in this Chapter

Thus, following previous notation and from Fig. 3.9, the units outputs are given by:

$$\begin{aligned}
 \mathbf{O}_j^{h1(t)} &= \phi\left(\sum_{i=1}^I w_{ij}\xi_i^{(t)} + \mathbf{b}_j\right) \text{ for } j = 1, 2, \dots, J \\
 \mathbf{O}_k^{h2(t)} &= \phi\left(\sum_{j=1}^J w_{jk}\mathbf{O}_j^{h1(t)} + \mathbf{b}_k\right) \text{ for } k = 1, 2, \dots, K \\
 \mathbf{O}_l^{h3(t)} &= \phi\left(\sum_{k=1}^K w_{kl}\mathbf{O}_k^{h2(t)} + \mathbf{b}_l\right) \text{ for } l = 1, 2, \dots, L \\
 \mathbf{O}_h^{\mathbf{N-D}(t)} &= \phi\left(\sum_{l=1}^L w_{lh}\mathbf{O}_l^{h3(t)} + w_{uh}\mathbf{O}_l^{\mathbf{N-D}(t-1)} + \mathbf{b}_h\right) \text{ for } h = 1, 2, \dots, H \\
 \mathbf{O}_{h'}^{\mathbf{GD}(t)} &= \phi\left(\mathbf{O}_h^{\mathbf{N-D}(t)} + w_{vh'}\mathbf{O}_{h'}^{\mathbf{GD}(t-1)}\right)
 \end{aligned} \tag{3.11}$$

where ϕ represents the activation function. The selection of the activation functions highly affects the performance of the RNNs. In Eqs. 3.11 two activation functions were used: the identity function, used in the output layer, and the hyperbolic tangent sigmoid, used in the remaining layers. This combination allows taking advantage of the multilayer configuration helping to improve in some way the performance of the architecture. Details and illustrations of these functions can be seen in Table 3.1 and Fig. 3.10.

Algorithm 1: BPTT Training Algorithm

```

Data:  $\{(x_i, y_i), \dots, (x_{\mathfrak{N}}, y_{\mathfrak{N}})\}$ 
1 Initialize weights  $w$ ;
2 Unfolding process for  $k$  instances;
   //  $\mathfrak{N}$  is the length of the training sequence
3 for  $t$  from 0 to  $\mathfrak{N} - 1$  do
   // forward-propagate the inputs over the unfolded network
4    $\hat{y}$ =forward_propagation(X);
   // error= target-prediction
5   calculate the error using  $\varepsilon = (y[t + k] - \hat{y})^2 / (2)$  ;
   // Back-propagate the error
6   backward_propagation( $\varepsilon$ ) ;
7   Update weights;
8   Average the weights at each instance  $k$ ;

```

When a RNN based architecture is trained, it is not possible to use the well-known backpropagation [230], which calculates the gradient of a loss function with respect to the weights available in the network, $\Delta w = -\eta \frac{\partial \varepsilon}{\partial w}$, since it assumes that there are free loops in the network connections. Thus, there are different options for training RNNs, one popular option relies on algorithms that are based on the gradient. An example of such algorithms is the Real-Time Recurrent Learning (RTRL) algorithm [232] which is computationally expensive. Another well-known and commonly used algorithm is the called Backpropagation Through Time (BPTT) [229]. BPTT is widely used for training RNNs since it is computationally efficient [77]. For this reason, we used it for training our architecture.

The main idea behind the BPTT is to unfold the network in order to capture longer history information. This unfolding process for k instances, which can be seen in Fig. 3.9, is achieved by duplicating the recurrent weights and redirecting them at the network. This process can be seen in Algorithm 1 which is used for training the architecture presented in Fig. 3.9 .

A Long-Short Term Memory Approach

In the previous subsection, we proved the feasibility of using RNNs to estimate the interaction forces. Nevertheless, the vanishing gradients problem, where error-signals exhibit exponential decay as they are back-propagated through time, has a direct impact on the performance of RNNs [31]. In order to mitigate this problem, in [18] we proposed the use of a Long-Short Term Memory (LSTM)

based architecture to overcome this problem and improve the accuracy of force estimation in RAMIS. Moreover, we proof the stability of our solution during long periods and we offer a comparison against our previous proposal based on a RNN [16].

The LSTM is a gradient-based method that was first introduced in 1997 by Hochreiter and Schmidhuber in [88]. LSTM is specially designed to store and retrieve information over long periods of time and enforce constant error flow by using specialized units, called cells. An LSTM layer has one or more recurrently connected memory cells composed of a central unit and specialized input, output and forget gates. The input and output gates are multiplicative units that protect the memory content from perturbations. On the other hand, the forget gates release irrelevant information by resetting the memory cell when the information stored there is not useful anymore [88]. These three gates have access to the central unit through peephole connections.

Our architecture, as illustrated in Fig. 3.11, is composed of two types of hidden layers: with basic units (Layer 1) and memory cells (Layer 2). Following the notation presented in Fig. 3.11-(A), let ξ be the input vector with I inputs, L the number of units, and K the number of cells with C memory cells in each block. Let w and b denote the weights and the bias respectively, and ϕ the activation function, which in this case is the log-sigmoid function (a.k.a logistic function). Then, the outputs, \mathbf{O}_l^t , are computed as follows:

$$\mathbf{O}_l^t = \phi(\sum_{i=1}^I w_{il}\xi_i^t + b_l) \text{ for } l = 1, \dots, L \quad (3.12)$$

Each output of Layer 2, \mathbf{O}_k^t , is defined by the relation of a set of units, as depicted in Fig. 3.11-(B). Consider \tilde{h} , φ and \mathfrak{S} as the input, output and forget gates, their corresponding outputs are defined as:

$$\begin{aligned} \mathbf{O}_{\tilde{h}}^t &= \phi(\sum_{l=1}^L w_{lh}\mathbf{O}_l^t + \sum_{k=1}^K w_{k\tilde{h}}\mathbf{O}_k^{t-1} + \sum_{c=1}^C w_{c\tilde{h}}\mathbf{S}_c^{t-1}) \\ \mathbf{O}_{\varphi}^t &= \phi(\sum_{l=1}^L w_{l\varphi}\mathbf{O}_l^t + \sum_{k=1}^K w_{k\varphi}\mathbf{O}_k^{t-1} + \sum_{c=1}^C w_{c\varphi}\mathbf{S}_c^t) \\ \mathbf{O}_{\mathfrak{S}}^t &= \phi(\sum_{l=1}^L w_{l\mathfrak{S}}\mathbf{O}_l^t + \sum_{k=1}^K w_{k\mathfrak{S}}\mathbf{O}_k^{t-1} + \sum_{c=1}^C w_{c\mathfrak{S}}\mathbf{S}_c^{t-1}) \end{aligned} \quad (3.13)$$

Continuing with the notation presented in Fig. 3.11-(B), the output of the unit \mathbf{O}_u^t and the memory cell state \mathbf{S}_c^t are obtained as follows:

$$\begin{aligned} \mathbf{O}_u^t &= \phi(\sum_{l=1}^L w_{lh}\mathbf{O}_l^t + \sum_{k=1}^K w_{ku}\mathbf{O}_k^{t-1}) \\ \mathbf{S}_c^t &= \mathbf{O}_{\mathfrak{S}}^t\mathbf{S}_c^{t-1} + \mathbf{O}_{\tilde{h}}^t\mathbf{O}_u^t \end{aligned} \quad (3.14)$$

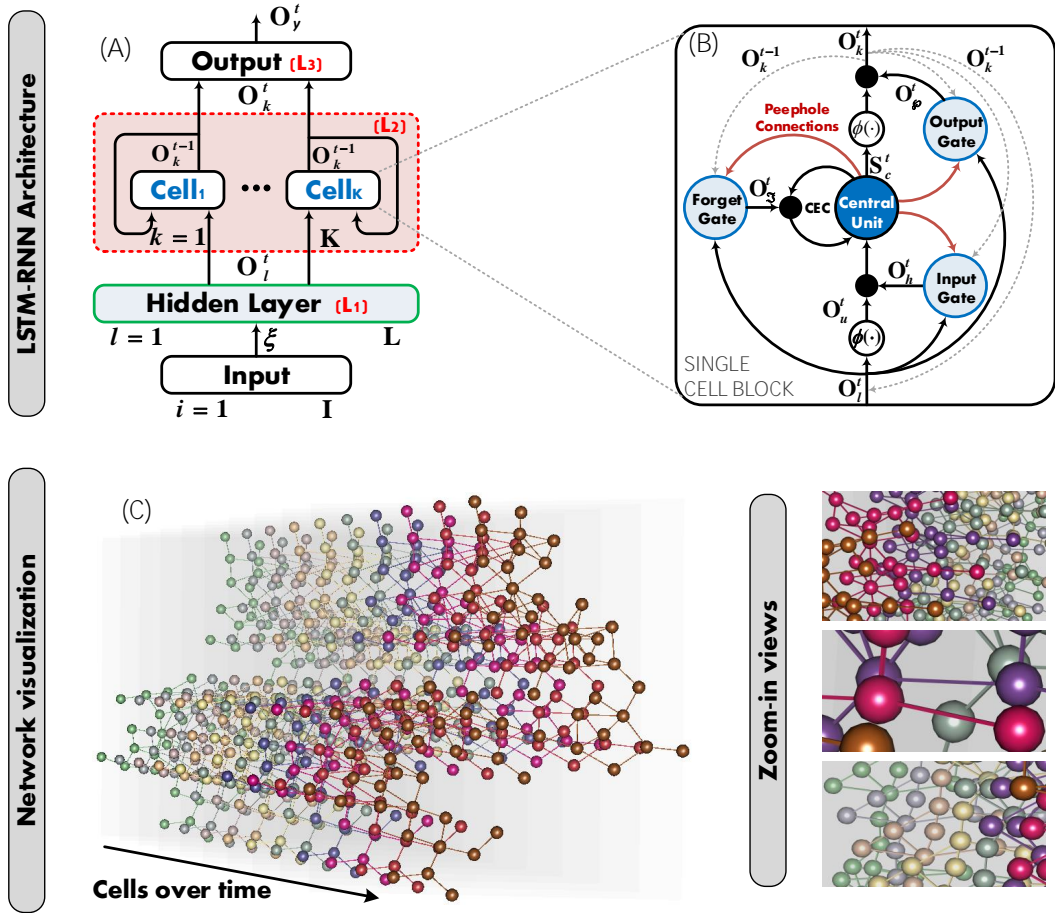


Fig. 3.11 In order to estimate the applied force, we use an architecture based on LSTM-RNN (part A) which combines basic units with cells. Part B shows a single cell block in detail and shows that each of the cells is composed of a set of units that enforce constant error flow which helps stabilizing force estimation over time. Additionally, part C shows an illustration of the hidden layer with 10 cells over time.

using Eqs. (3.13) and (3.14), we can describe the output of each cell as:

$$\mathbf{O}_k^t = \mathbf{O}_\varphi^t \phi(\mathbf{S}_c^t) \text{ for } k = 1, \dots, K \quad (3.15)$$

Finally, the output of the network, \mathbf{O}_y^t , is given by:

$$\mathbf{O}_y^t = \phi\left(\sum_{k=1}^K w_{ky} \mathbf{O}_k^t + b_y\right) \text{ for } y = 1, 2, 3 \quad (3.16)$$

where the output units are the force in X, Y and Z directions. Using this LSTM-RNN based architecture, the vanishing gradient problem is solved. Thus, a gradient-based algorithm can be used. In this work, we apply Backpropagation

Through Time (BPTT), which unfolds the network over time by replicating the network and sharing the weights.

Observation 1 *Notice that the learning process is carried out offline since its goal is to find the optimal parameters. Once the adjusting parameters (weights) are found, they are used in our system in real-time.*

Taking into account observation 1, the training process for the deep network (Fig. 3.11-(C)) is performed once to learn the optimal parameters and the mapping function from the input to the output space. Once the network is trained, it can be used for estimating the applied forces in real time without the need of carrying out the training process for each subject. This is based on the fact that, according to [97, 71], generic material properties of the human heart tissue can be modeled and, in consequence, these properties can be learned using the LSTM-RNN architecture and then generalize the model across subjects. The proposed approach can be generalized to handle other tissues, requiring only a single training run in order to obtain the mapping function for the new tissue. The user can then select the desired function during the real time procedure or can even choose to train a function that can handle different tissue types.

3.4 Experimental Results

This section describes in detail the experimentations that we conducted to validate the accuracy of the proposed solution. We start by describing the datasets and explaining the tasks conditions used while acquiring them. Next part is devoted to explaining the measurements and tools we used to evaluate our solution and finally we present detailed results and discussions of the evaluation.

3.4.1 Data Description

To evaluate our proposal, we used both in-vivo and ex-vivo datasets (see Fig. 3.13).

The in-vivo dataset [143] is from a porcine and exhibits tissue deformation, due to tool interaction, and was used to evaluate our deformation approach. This sequence is composed of stereo-pair images of size 720x288 recorded during a period of 450 sec. During the text, we refer to this in-vivo sequence as *Dataset I* (see top part of Fig. 3.13).

We acquired the ex-vivo datasets using the experimental setup showed in Fig. 3.12. It is composed of a stereo camera, a set of robot manipulators (Stäubli

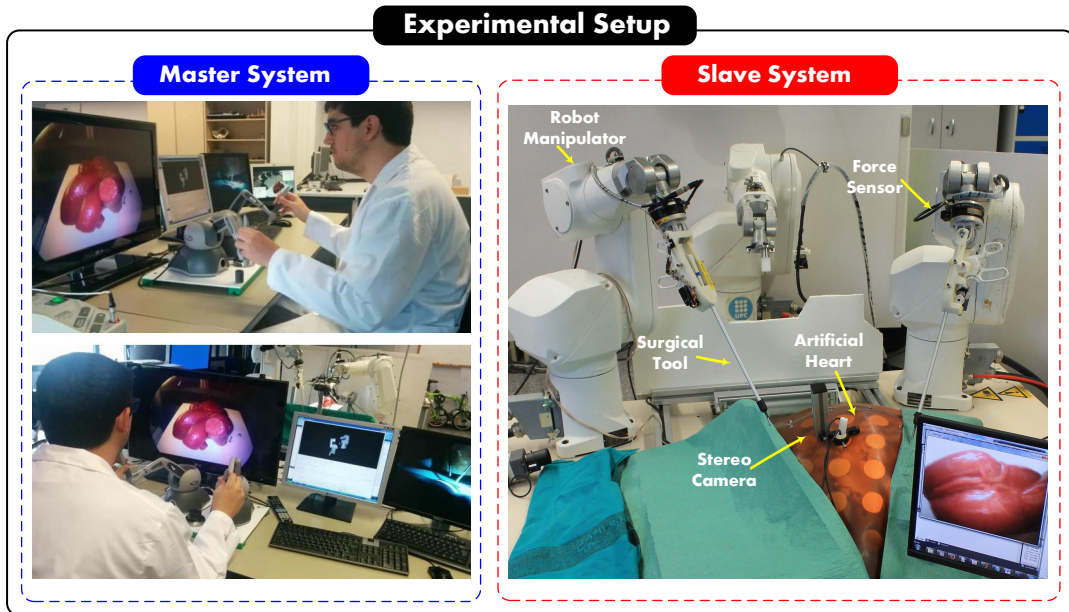


Fig. 3.12 The realistic surgical setting, with typical RAMIS surgical setup, used to obtain the two ex-vivo datasets. The force sensor is used to obtain the ground truth to validate our estimation.

RX60B), and an ATI Gamma SI-32-2 force sensor which we used to acquire a ground truth for the applied force in order to compare it against our estimation. We obtained two stereo-pair image sequences of size 640×480 recorded during 2100 sec. In the remainder of this section we refer to these sequence as *Dataset II* and *Dataset III* (see bottom part of Fig. 3.13).

As for the ex-vivo datasets, we used two artificial hearts made of ECOFLEX 0030, which has mechanical properties similar to those of human tissues, to imitate variations between two different subjects. From a technical point of view ECOFLEX material allows comparing our approach with other research since it is widely used and considered a standard material for experimentation in clinical environments (e.g. [184, 135, 168]). Moreover, ECOFLEX facilitates the continuous experimentations avoiding at the same time hygienic issues.

It is noteworthy that during the acquisition of our ex-vivo datasets we did not take the dynamics of the heart into consideration. However, it will be included in a future work.

3.4.2 Tasks Description

The datasets described in Subsection 3.1 were acquired by doing general inspection through palpation over the tissue and region of interest while varying three main

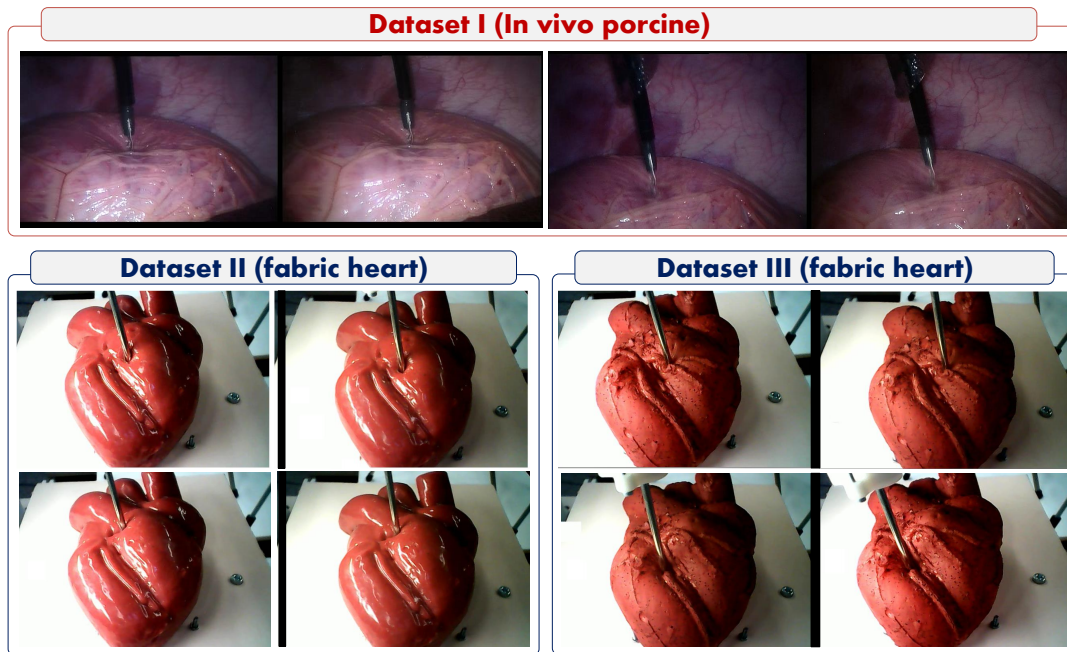


Fig. 3.13 Raw data of the three different datasets used to evaluate our proposal (one in-vivo and two ex-vivo).

factors over time: position, orientation, and illumination. An illustration of palpation actions can be seen in Fig. 3.14.

General palpation is necessary during different clinical activities such as tumor detection, tissue cutting, and needle-based procedures; it is an actuation very representative for this study. Palpation is relevant for RAMIS since during procedures, surgeons perform different tasks part of which requires avoiding penetration of the tissue and control the applied force.

3.4.3 Evaluation Scheme

Our evaluation scheme is divided into two parts. The first part uses an in-vivo dataset to evaluate the following:

- Inspection of the displacement field: Fig. 3.15-(A);
- Careful comparison of the residual error of our deformation approach: Table 3.2;
- Numerical analysis of our visual approach with different degrees of penalization : Table 3.2;

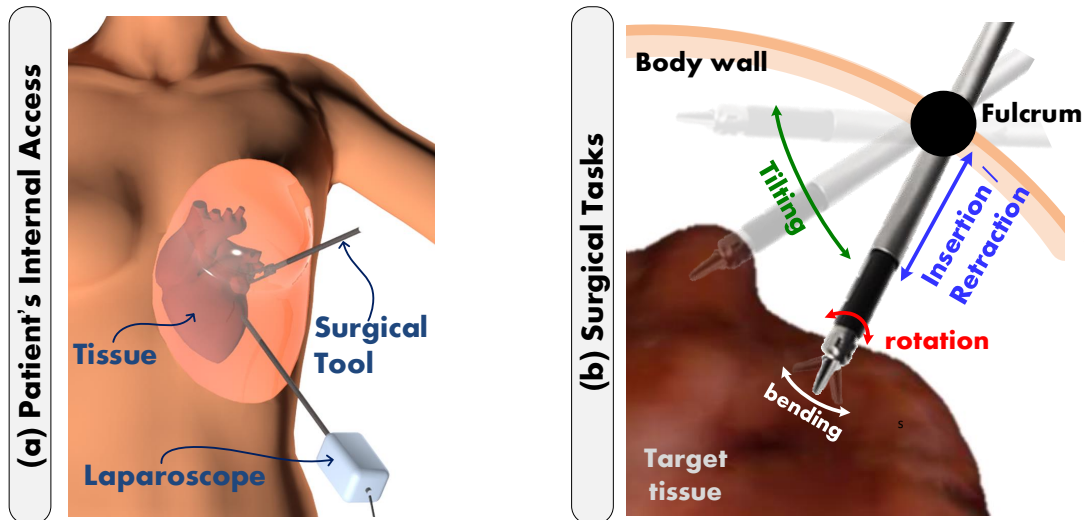


Fig. 3.14 (a) Typical way to access the patient during a RAMIS. (b) Illustration of the palpation and exploration surgical tasks used to test the efficiency of our solution.

- Visual inspection of the 3D shape recovery including change of illumination and complex deformation cases using ex-vivo and in-vivo datasets: Fig. 3.16.

In the second part, we used our two ex-vivo datasets with a provided ground truth and performed the following evaluations:

- Visual examination of the displacement field: Fig. 3.15-(B);
- Convergence of our energy functional (Eq. 3.4): Fig. 3.17;
- Comparison between the estimated and real displacement at contact point: Fig. 3.15-(C);
- Associated strength between the real and estimated force: Fig. 3.18;
- Statistical analysis of adaptability of our force estimation strategy: Table 3.3.
- Comparison between the real and estimated forces in the (X, Y, Z) directions: Fig. 3.19;
- Stability over long periods of time of our proposal for estimating the interaction forces: Fig. 3.20;
- Inference and analysis of the temporal visual uncertainty using Fuzzy theory: Figs. 3.21 and 3.22 and Table 3.5.

Table 3.2 Residual Error evaluation of our deformation approach

Exps.	Energy Functional (Eq. 3.4)	Minimum
1	\mathbf{E}_Φ without ρ_{huber} , $\mathbf{E}_\Psi = 0$, and $\mathbf{E}_\Lambda = 0$	0.2657
2	\mathbf{E}_Φ with ρ_{huber} , $\mathbf{E}_\Psi = 0$ and $\mathbf{E}_\Lambda = 0$	0.1348
3	\mathbf{E}_Φ with ρ_{huber} , \mathbf{E}_Ψ and $\mathbf{E}_\Lambda = 0$	$1.7896e^{-03}$
4	\mathbf{E}_Φ with ρ_{huber} , \mathbf{E}_Ψ and \mathbf{E}_Λ	$3.1584e^{-05}$

3.4.4 Results and Discussion

In order to prove the benefits of our proposal, in this subsection, we offer a detailed evaluation of both our visual-based and force estimation approaches.

Visual-based Approach

We evaluated the performance of our visual-based approach using Datasets I and II (see Fig. 3.13). First, Fig. 3.15-(A)/(B) show tissue deformation that results from applying force and illustrates the recovered 3D deformation structure, bounded by our defined grid, over some time instants where darker shades of red represent more intense deformation at contact point. The plots clearly show pleasant visual results of the deformation field with both in-vivo and ex-vivo data.

We then took the results from the in-vivo data (Dataset I), Fig. 3.15-(A), and offer a quantitative analysis of our energy functional (Eq. 3.4). The results are reported in Table 3.2 in which experiments Exp. 1 and Exp. 2 show that without penalization given by the M-estimator and the two regularizers (refer to Eqs. 3.7 and 3.9), the residuum was about 0.1348 and 0.2657 respectively.

Comparing that to Exp. 3, we can see that including Tikhonov regularizer resulted in a minima in the order of magnitude 10^{-3} ; while adding a volume preserving term, as in Exp. 4, clearly offered the best minima in the order of magnitude 10^{-5} . With this, we conclude that the combination between the ρ_{huber} with the two regularizers offered a significant difference in order of magnitude.

Moreover, by acquiring the geometry of motion from the robotic manipulator, we were able to have a ground truth reference at least for the contact point between the tool and the tissue. So in order to evaluate the accuracy of our computed deformation, we compared the displacement value at contact point in X, Y, and Z directions against the reference measurements given by the geometry of motion. The plot at Fig. 3.15-(C) shows that comparison and the zoom-in

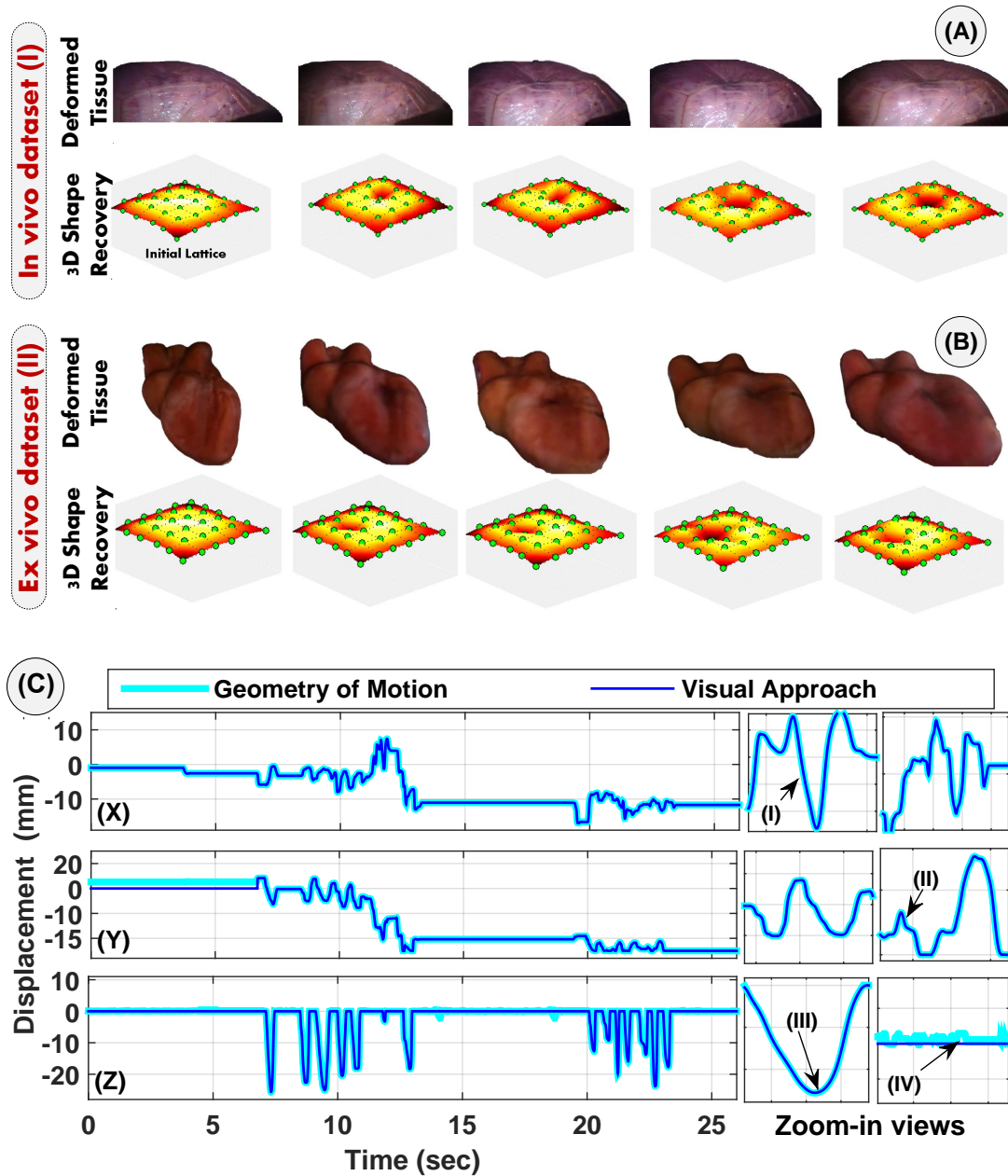


Fig. 3.15 Tissue deformation that result from applying a force at different time instants is illustrated in parts (A) and (B) along with the recovered 3D deformable structure using our proposed visual approach. Finally, plots at part (C) show a comparison between the computed displacement (at contact point) in X,Y,Z directions against the reference measurements given by the geometry of motion of the robot from dataset II. The zoom-in views demonstrate the high estimation accuracy of our approach even during complex deformation as it can capture small (I-II) and large displacements (III). It also eliminates the noise in the geometry of motion as shown in (IV).

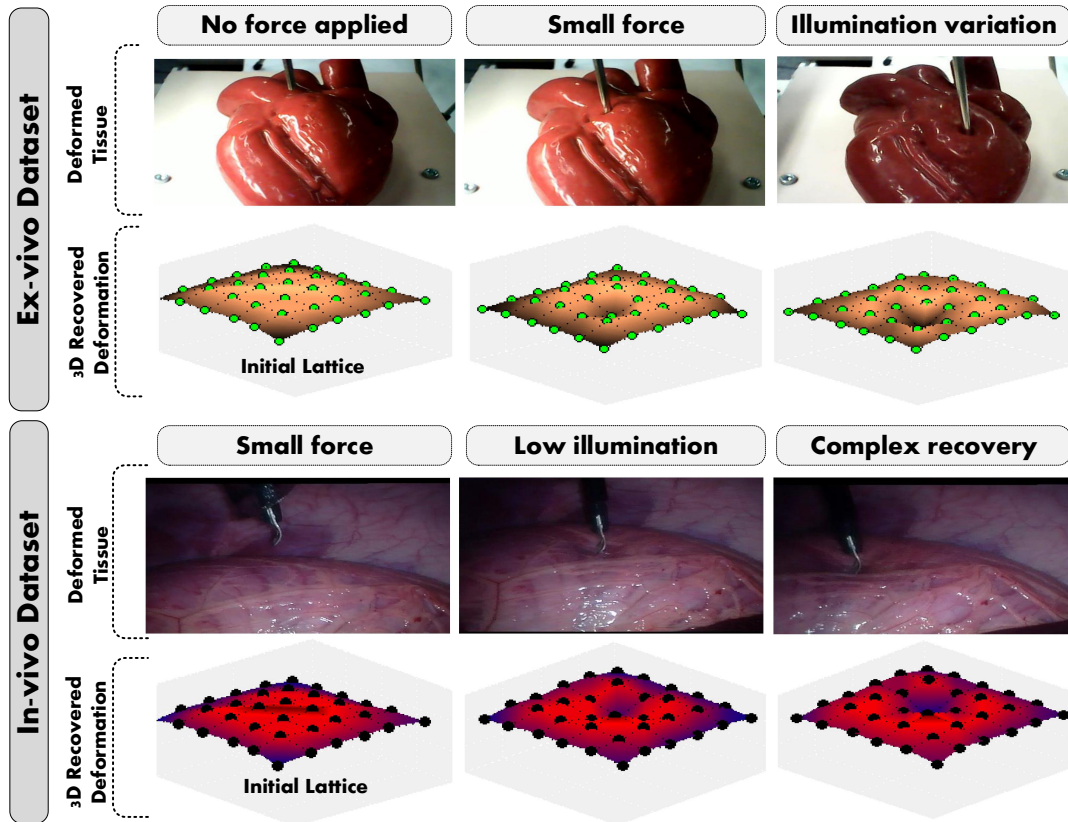


Fig. 3.16 Illustration of tissue deformation that result from applying force at different time instants along with the 3D deformable structure recovered using our proposed visual approach. Our proposal was tested under different variation of illumination, occlusions and complex deformation.

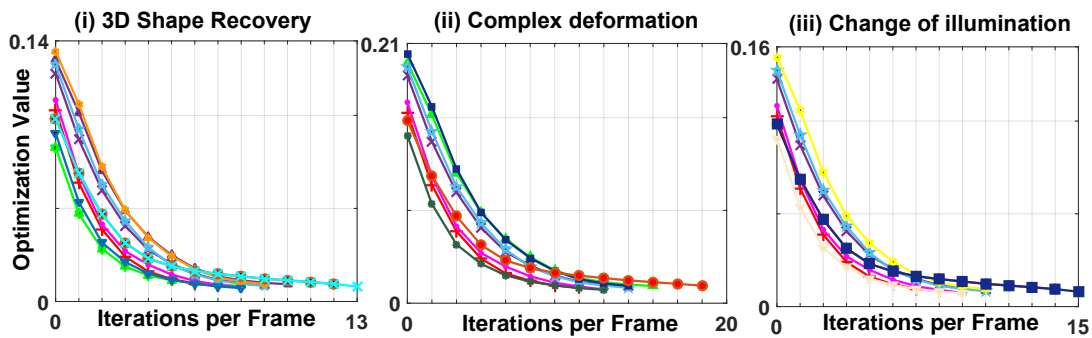


Fig. 3.17 Optimization plots resulted from our energy functional for different cases in which retrieving the 3D shape is challenging including complex deformations and change of illumination.

views, together with a root-mean-square error (RMSE) smaller than 1mm in all directions, demonstrate the accuracy of our computed measurements even during complex deformation. Furthermore, our visual approach was even able to

deal with the mechanical issues that usually exist in the geometry of motion and eliminated the noise as shown in view (IV).

For further support, in Fig. 3.16 we show cases where recovering the 3D deformation is complicated including change of illumination, occlusions and specular highlights. Apart from offering visual results, we also analyzed the convergence of our energy functional. The plot at Fig. 3.17 shows that the minimization of our functional, on different frames and for the different cases shown in Fig. 3.16, needed less than 25 iterations to get the minima. For this reason, we limited the number of iterations according to Observation 1.

This supports the good accuracy and fast convergence of our proposed visual-based approach.

Force Estimation Approach

The ultimate goal of this work is to estimate the applied force in RAMIS scenarios accurately over time. Therefore, we conducted large number of tests to validate our complete neuro-recurrent-vision solutions presented in section 3.3.1 against the ground truth of the force provided with the ex-vivo datasets.

To validate the accuracy of our solutions, we first tested the associated strength between the estimated (using dataset II) and actual forces (ground truth). The regression plots are shown in Fig. 3.18. Top part illustrates the resulted regression of our model [16] presented in Eq. 3.11 while bottom part the RNNLSTM model [17, 18] described in Eqs. 3.12- 3.16. The plots show a strong correlation between the two measures. The red dashed lines in the plots show the ideal solution while the straight black lines are the best linear regression fit between the target and the output. The tight relationship is clear in both training and test datasets as they reported R-values of 0.98 and 0.96 and, 0.99 and 0.98 respectively.

It is worth mentioning that our force estimation model was trained once using dataset II, but then the optimal parameters were tested on the two artificial hearts that have slightly different mechanical properties (Dataset II and Dataset III). The results show that our model adapts well to the two datasets and prove that it was able to handle small variations across subjects.

In order to support the previous statement, we ran a statistical analysis on the force estimated on both datasets II and III and the corresponding ground truth. More specifically, we used the nonparametric Wilcoxon rank sum test to answer the question of whether there is a statistical significant difference between the two estimated forces. The results at Table 3.3 show that the null

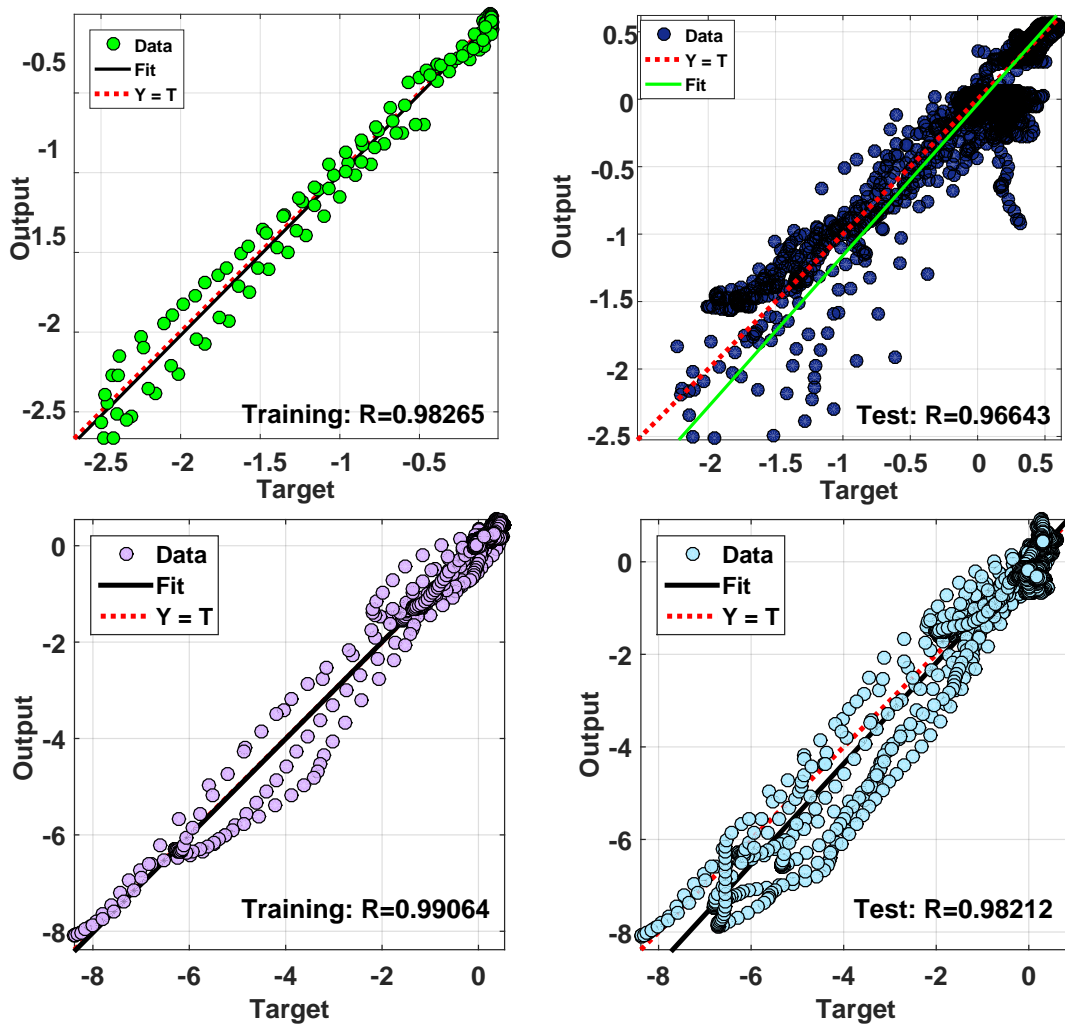


Fig. 3.18 These linear regression plots show the associated strength between the real (target) and estimated force (output) measurements of both training and test datasets. In both sets, the points fit a line showing a tight relationship between the measurements and demonstrating the accuracy of the force estimation.

hypothesis was not rejected at $p < 0.05$ of significance level. This, together with the big p-value for the three directions (X,Y,Z), led us to conclude that there is no significant difference between the two groups which support our proposal in the sense of adaptability to different subjects.

The accuracy of our solution is further validated by the top plots shown in Fig. 3.19 in which we compared the estimated force (dataset II, test data) against the real one in the (X,Y,Z) directions. The results show that the measurements are very close to each other. We also show the RMSE results at the bottom plots in which the error remained less than 0.03N with a concentration of values much lower than 0.02 N in all directions.

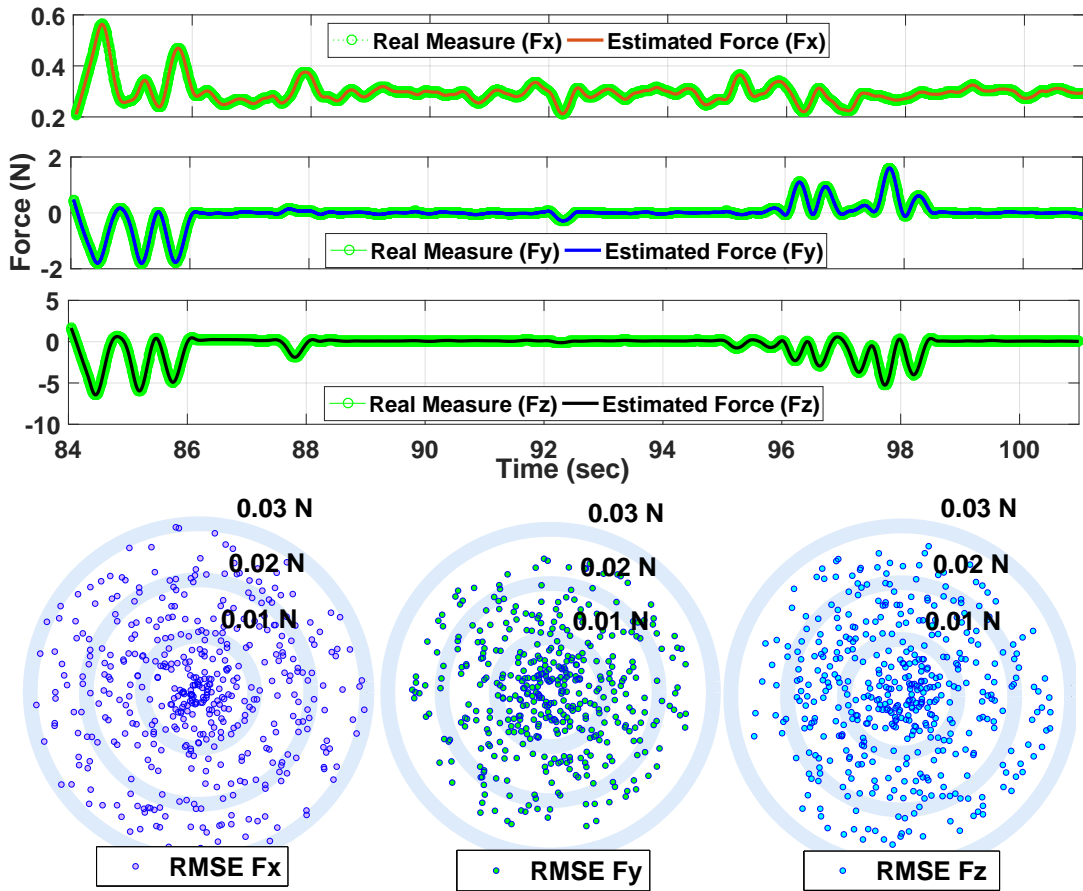


Fig. 3.19 Plots in top part show the real force measures, in X,Y and Z directions, and those estimated by our approach. Bottom plots illustrate the RMSE results in all directions.

Table 3.3 Statistical nonparametric analysis of our proposal to estimate the applied forces. It takes into consideration the ex-vivo datasets and the real measure.

Input Values	Direction	p-value	Null Hypothesis
<i>Real and Dataset II</i> <i>(ex-vivo data)</i>	x	0.8105	h=0
	y	0.8026	
	z	0.7598	
<i>Real and Dataset III</i> <i>(ex-vivo data)</i>	x	0.8654	h=0
	y	0.8287	
	z	0.8045	

Furthermore, we demonstrate the stability of our solution over time by inspecting the results of both datasets at different time intervals (see Fig. 3.20). The zoom-in views show comparison of the force measures as given by the force

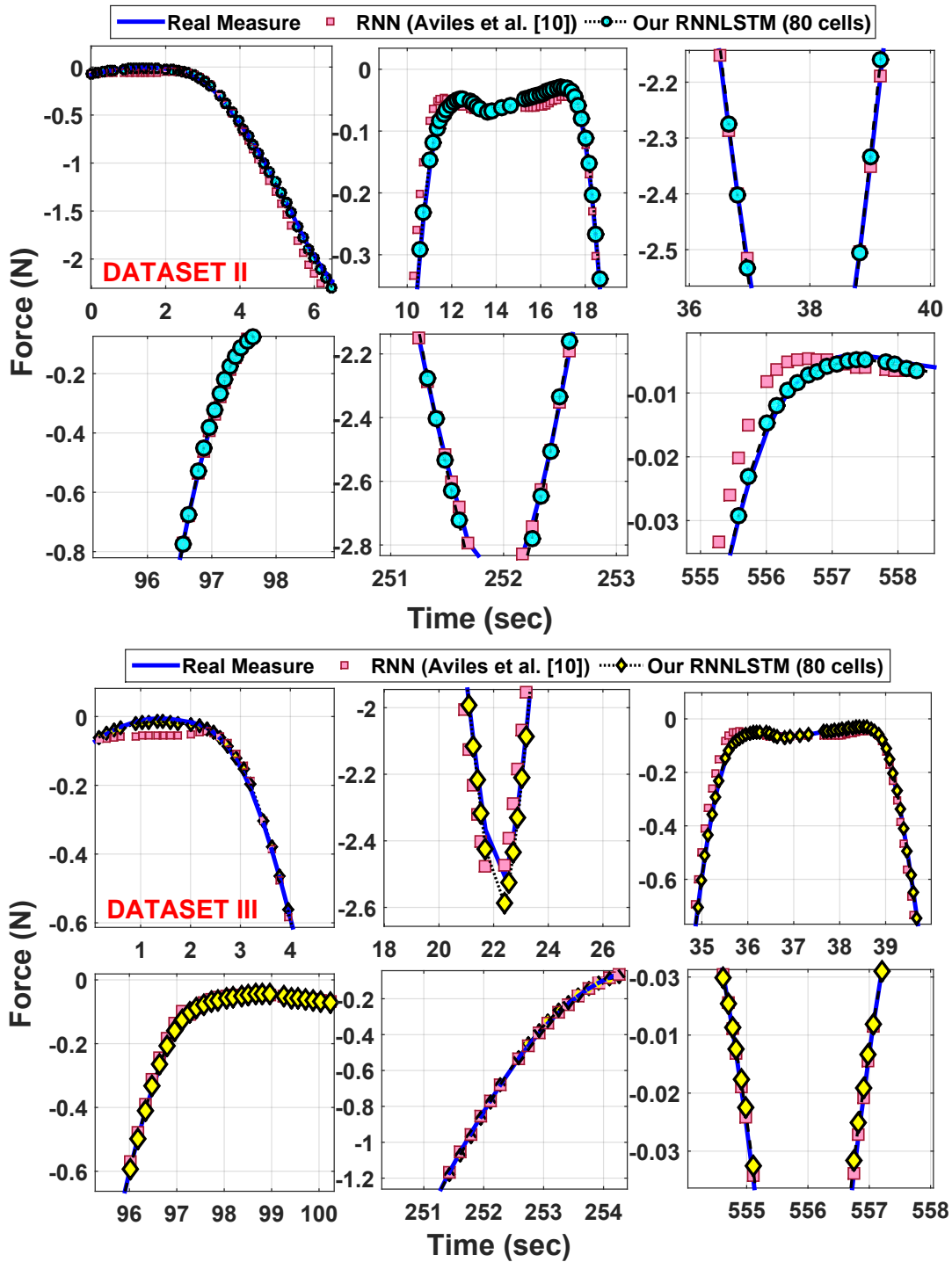


Fig. 3.20 Stability criteria is shown in these plots using the ex-vivo datasets where the estimated and real force measures are plotted at different time intervals of a longer period of time.

sensor, our previous RNN solution [16], and our newly proposed LSTM-RNN solution. As we can see, the added cells with the LSTM architecture improve

the accuracy of the results and bring the force measure closer to the actual one. Furthermore, the results we obtained from the LSTM-RNN tend to be more stable and with no error decay, during long periods of time as clearly visible in the last time interval. Based on the results obtained in our evaluation scheme, we report an average RMSE of 0.02 N for all our experiments.

Deep-Neuro-Fuzzy Approach for Visual Uncertainty

Uncertainty is inherent to computer vision and it occurs at low and high levels. For example, at low level during the acquisition due to the sensor, and at high levels during processes such as tracking, segmentation etc. Therefore, for a computer vision system to be robust, it has to have at its disposal the machinery allowing vagueness representation [204]. With the problem at hand, our starting hypothesis is that the force estimation can improve by dealing with visual uncertainty during the 3D shape recovery process. Uncertainty is reflected mainly in two stages:

1. During the acquisition process using the endoscopic camera
2. When part of the target is occluded by the surgical instrument, which entails lack of knowledge about a portion of tissue

In order to deal with uncertainty during the estimation of the applied forces, in [15, 19] we proposed a novel approach that we called Voting-Adaptive Neuro-Fuzzy Inference System (V-ANFIS) which is divided into two main steps. The former is a voting process that allows decreasing the neighboring points error (of the lattice) using combinatorics and agreement processes. The latter is a prediction step based on the ANFIS [99] properties, in which the main idea is to estimate the lost information, when an occlusion occurs at contact point, using patterns of the available data. Also, to help to the maximum likelihood estimator to increase robustness.

The voting process is explained in Algorithm 2 in which the following definitions and conventions are given for clarification.

- Control Points (**P**): A set of points, that compose the lattice, uniformly spaced.
- Neighboring Point (**NP_i**): Point that has direct connection to the contact point between the tissue surface and surgical tool.

Algorithm 2: Voting process of our V-ANIFS approach

Data: Set of control points \mathbf{P}

- 1 Initialization;
- 2 $i \leftarrow$ Number of neighboring points;
- 3 $NP_i \leftarrow$ Neighboring points;
- 4 ${}_iC_2 \leftarrow$ Combinations matrix;
- 5 **while** *stereo-pair image available* **do**
- 6 $NP' \leftarrow$ current Z-displacement value for all NP_i ;
- 7 **if** *all values in NP' are equal* **then**
- 8 Preserve actual values;
- 9 **else**
- 10 **forall** *pair-elements in C_2* **do**
- 11 **if** \exists *one pair-agreement* **then**
- 12 $NP' \leftarrow$ agreement value;
- 13 **else if** $\exists >$ *one pair-agreement or none* **then**
- 14 $NP_{min} \leftarrow$ min value in NP' ;
- 15 $NP_{max} \leftarrow$ max value in NP' ;
- 16 Find minimum difference (NP_{min}, NP_{max}) in NP' ;
- 17 $NP' \leftarrow$ value with minimum difference;

- Combination Matrix (**C**): Allows carrying out a pair-search, identifying the points that need correction.
- Contact Point (**CP**): Point generated by the contact of the tissue surface and the surgical tool.

Definition 11 *Let i be a positive integer denoting the number of neighboring points, $B_c = \binom{i}{b}$ the binomial coefficient and \mathbf{C} the combinations matrix, having b columns and B_c rows. Then, a pair-element is a combination in \mathbf{C} .*

Definition 12 (*Pair-agreement*). *Given a pair-element in \mathbf{C} , we say that it fulfills the pair-agreement condition if and only if at current time both points have the same value.*

Once the voting process is performed, we use the capability of ANFIS as universal approximation (see proof in Appendix A) in order to guarantee more reliable information at each time instant. The ANFIS architecture used is described in Table 3.4

We used Dataset II to test and validate the performance of our proposed system combing Fuzzy Theory and a Deep Network (RNNLSTM). We started by

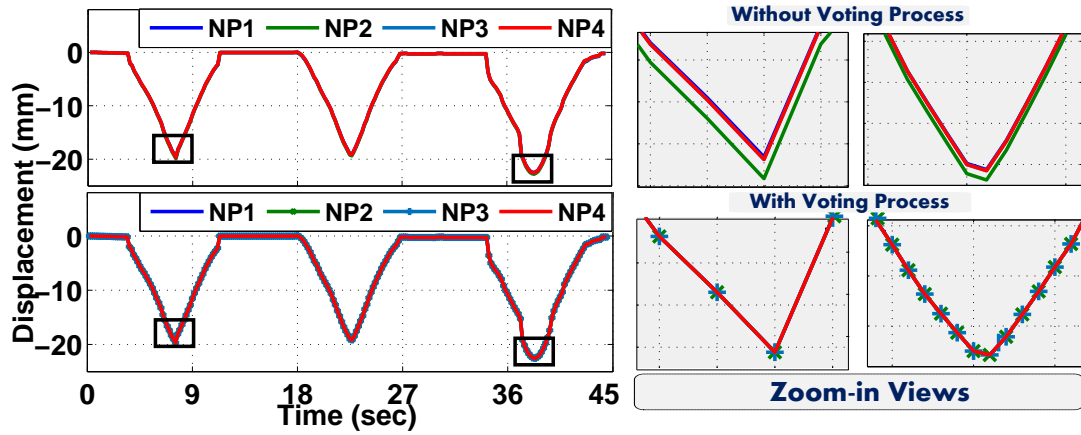


Fig. 3.21 Retrieved displacements of the four immediate neighboring points are plotted first without voting process correction (top) then with voting process (bottom).

Table 3.4 The description of the architecture used in our V-ANFIS approach

Layer's Output	Description
$y_{ij}^{L1} = \mu S_j^i(x_i)$	Fuzzification layer
$y_j^{L2} = \prod_{i=1}^{N_2} y_{ij}^{L1}$	Belief strength of the rules
$y_j^{L3} = \frac{y_j^{L2}}{\sum_{n_2=1}^{N_2} y_{n_2}^{L2}}$	Normalization of the firing strengths
$y_j^{L4} = y_j^{L3} f_j(x_i)$	Consequent parameters
$y^{L5} = \sum_{j=1}^{N_2} y_j^{L4}$	Output

evaluating the proposed *voting process*. Fig. 3.21 shows the displacement values at neighboring points with and without applying the voting process. Zoom-in view shows the significant improvement after applying the voting step, which brings the displacements of all neighboring points into a complete agreement.

Fig. 3.22, left side, shows a plot comparing the recorded displacement values during complex recovery (i.e. occlusions). We can see that applying V-ANFIS brings the estimated values closer to the real-geometric measure. This assertion can be better appreciated in the four zoomed-in plots at right side of Fig. 3.22. For further support, the RMSE between the real measure and the estimated one was less than 1mm in average.

Besides improving the displacement estimation and having in mind that our ultimate aim is to improve the estimated force and bring it closer to the target values, we have checked the effect of our V-ANFIS in the applied force estimation (RNNLSTM approach [17, 18]). We compared our approach against state of

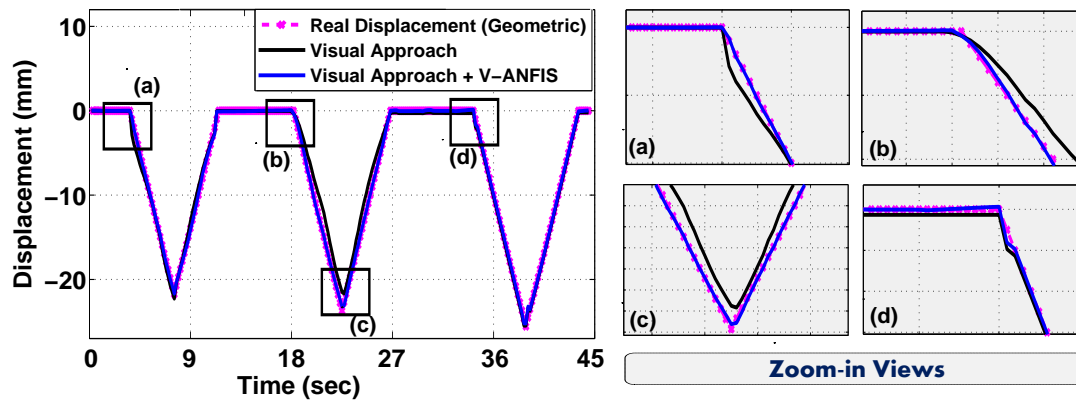


Fig. 3.22 (From left to right) Comparison of displacements, at contact point, between real-geometric measure and visual approach with and without V-ANFIS. Zoom-In displacements are also shown in order to observe the improvement when V-ANFIS is applied.

Table 3.5 Performance analysis of existing and proposed approaches.

Method	RMS Error(N)	% Improvement
Yip et al. [238]	0.13	84.15
Aviles et al. [15]	0.315	34.60
Faragasso et al. [66]	0.1355	84.79
Noohi et al. [158]	0.07	70.54
Our Approach [19]	0.0206	—

the art studies on force estimation with similar settings. Table 3.5 shows the RMSE of our approach and reported by the studies under comparison and our solution outperforms them with the least RMSE of 0.0206. Moreover, it offers a percentage of improvement over those studies that ranges from about 35% to 85%. These observations prove the feasibility of using DN together with Fuzzy Theory to estimate applied forces in RAMIS settings.

3.5 Conclusions and Future Work

In RAMIS, surgeons perform delicate procedures remotely through robotic manipulators without directly interacting with the patients. As a result, they lack force feedback that informs them about how much force the surgical tool is applying to the tissue. While force sensing devices are able to provide that information, their size and cost, along with biocompatibility concerns, prevent them from being fully integrated into the surgical environment. The approach presented in this work offers a feasible alternative that overcomes on the one side,

the limitations of integrating sensors on the surgical tools and on the other side, provide an alternative way to transmit the estimated force to the surgeon. The proposed approach combines a computationally efficient visual shape recovery approach with an accurate recurrent learning based force estimation model.

By minimizing an optimized energy functional, we were able to recover the 3D deformable structure of the region of interest over time. We ensured the robustness of our shape recovery approach by handling sources of errors and outliers that exist in real surgical environments such as occlusions and uncertainties. Furthermore, we utilized the learning power of deep network by using a RNNLSTM architecture to relate the extracted visual-geometric information to an accurate force estimation. We obtained a trade-off between computational time and accuracy of our deep network by reducing the complexity of the input space and only considering features with high correlation to force.

The experimental results presented in Section 3.4 verified that vision-based techniques combined with supervised learning provide a feasible and accurate estimate of the applied forces without using force sensors. Experiments included various datasets, in-vivo and ex-vivo, and the computed and estimated results were validated against the ground truth obtained from the robotic manipulator and the force sensor. The advantages of this approach include robustness, accuracy, and stability over long periods of time. This methodology would allow surgeons performing RAMIS to have force feedback and would increase the transparency of interaction with the patient without using force sensors.

Moreover, our solution promises to be useful in robotic-assisted surgery as well as in different situations in which knowing the applied force make a difference in the results, including: detection and prevention of diseases or abnormal behavior (e.g. [124]), needle-based procedures (e.g. [57]), microsurgery (e.g. [225]) and knot tying (e.g. [108]). Thus, this approach can be extrapolated in the above-mentioned situations avoiding in this way the space restrictions, biocompatibility issues and cost of designing a new miniaturized force sensor.

As mentioned earlier, the goal of this chapter is to prove the feasibility of combining visual information with deep network to estimate the applied forces during robotic-assisted surgeries. However, when we talk about haptics technology in RAMIS settings, we have to consider two questions. One is how to acquire the significant information, and the other is how to transmit that information to the surgeon. While the first question was tackled in this Chapter, the second one is part of Chapter 4 in which we conducted an experimental study to provide an efficient way to display the estimated force to the surgeon.

*“The world is full of magic things, patiently waiting
for our senses to grow sharper.”*

William B. Yeats

4

From Motion Estimation to Clinical Evaluation: A Perception Experimental Study

Technological advancements are revolutionizing the field of medicine by creating and integrating robotic devices in multiple clinical scenarios such as diagnostic and surgery. In particular, Robotic-Assisted Surgical Systems (RASS) have taken the advantage of the recent technological developments allowing performing Robotic-Assisted Minimally Invasive Surgeries (RAMIS). When using a RASS, both patients and surgeons benefit as it offers better ergonomics for surgeons by helping them regain dexterity and extending their surgical capabilities offering optimal hand-eye alignment, motion scaling and tremor filtering [208, 113]. For patients, RASS offers all the inherent benefits of minimal intra-operative invasiveness [233] (e.g. less trauma and recovery time).

A RASS attempts to reproduce the surgeon’s motion in a master/slave teleoperated setting. Fig. 4.1 shows the architecture of one such system. At the master side, the operating surgeon is immersed into a three dimensional environment in which additional useful information can be added to improve the transparency in the teleoperated system [169]. Nonetheless, the physical

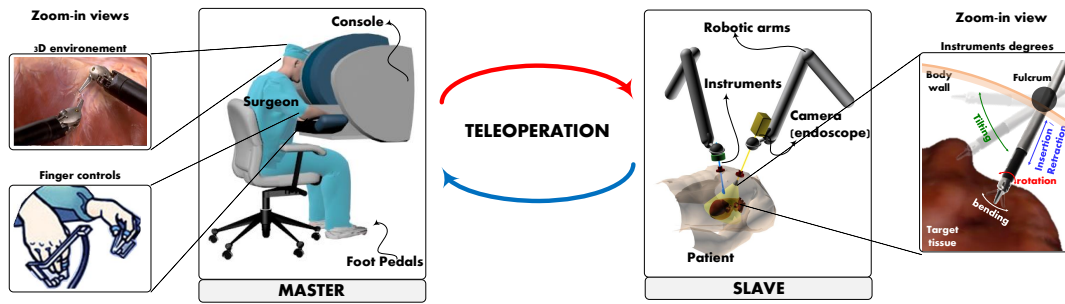


Fig. 4.1 A typical teleoperated robotic surgical system using a master-slave configuration. At the master side, surgeon is provided with a 3D patient view and is able to perform the procedure using finger controls and foot pedals. All surgeon’s actions are reproduced by the slave which holds the surgical instruments.

separation between the operating surgeon and the instruments in the operating field leads to a complete deprivation of force feedback during the surgery. This means that the surgeon would have no information about the force applied to the tissue which is considered of great importance for a better surgical performance [136].

This loss of direct interaction has proven to be a major limitation in currently available surgical systems since humans rely on haptics as a main sensory input during object manipulation tasks [64]. Without force feedback, the operating surgeon has to interpret the force load from indirect cues which produces a high mental workload and complicates the task at hand. Any misinterpretation from the surgeon side could lead to irreversible damage such as torn tissues or broken sutures [223]. This limitation is reputed to be one of the causes that restricts further spread of medical robotics [64].

Given that it is still an open problem in surgical robotics, different researchers have attempted to acquire the force information using direct sensing devices. However, the miniaturization constraints on the design of the device, together with the list of medical regulations and restrictions including sterilization, biocompatibility, stability, and robustness [83, 205], is why direct force sensing devices have not yet been integrated into current robotic surgical systems. An alternative group of force measuring solutions emerged to overcome these limitations by estimating the interaction forces using visual information. The idea behind what is called Vision-Based Force Estimation (VBFE) comes from the conservation principles of continuum mechanics in which it is clear that the change in shape of an elastic object is directly proportional to the force applied. These kinds of solutions depend mainly on visual information, such as

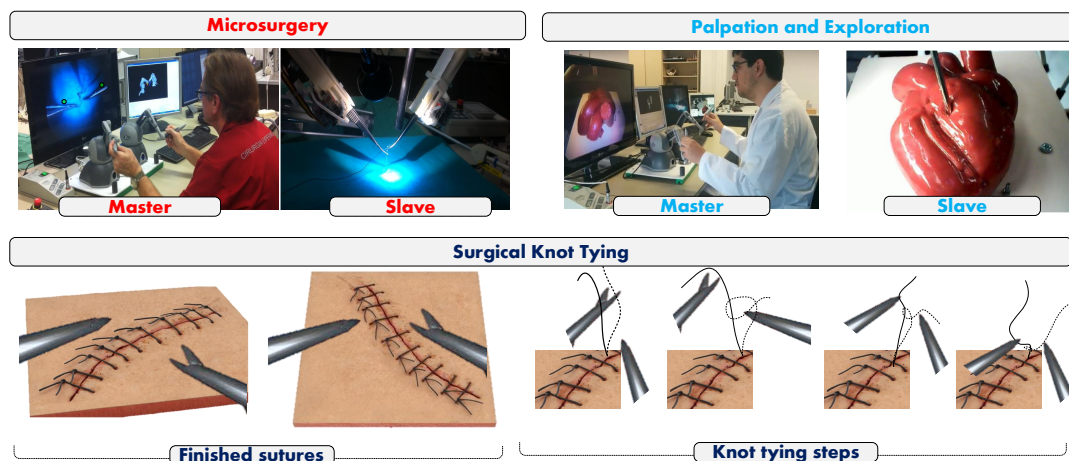


Fig. 4.2 Examples of surgical tasks where knowing the applied force is relevant and helps decreasing the procedure completion time and avoiding injuries.

the deformation of tissue under load, to estimate the applied forces. Different authors have proved the benefits of VBFE such as in [76, 103, 158, 17].

Whether direct or estimated forces are available, a natural question arises, that is, *how to provide this information to the surgeon?* Studies show that force feedback information enables surgeons to have a better control and precision with tissue manipulation [68, 98]. Moreover, force feedback is specially relevant in the performance of many surgical tasks. For example, in many situations, surgeons perform exploration and palpation tasks in order to identify abnormal or cancerous tissue regions. Force feedback is useful in these situations as it enables surgeons to sense tissue mechanical properties and identify specific tissue features that are hard to be identified visually. Other surgical tasks involve tissue manipulation, such as dissection and suturing in which having force feedback is important to prevent puncturing the tissue or breaking sutures due to the application of large forces. Fig.4.2 shows an illustration of such common surgical tasks.

A straightforward solution for having force feedback would be to transmit the force information to the surgeon's hands using a haptic master device. Nonetheless, there are many concerns associated with this option including cost, stability of the controller, degrees of freedom, and space limitations [92, 161, 64]. Moreover, when the gains size is too large, it can result in fatigue for the surgeon and consequently, affects his/her performance [130].

An attractive alternative to direct force feedback to the surgeons hands is what is called sensory substitution, which was first introduced by Bach-y-Rita in [26], in which one sense, in this case touch, is replaced by another sensory

modality, vision or auditory for example, to convey the lost information indirectly. This option is inspired by theories of perception and mechanisms of brain plasticity, in which different studies have demonstrated that the brain is a complex machine that is able to restore functions using input from other stimuli or sensory modalities [25, 148]. This was the start of different studies and developments in terms of sensory substitution in teleoperation settings [24, 180, 182, 146].

4.1 Sensory Substitution in Teleoperation

The term *sensory substitution* refers to the ability of the central nervous system to learn a new mode of perception and has been successfully used for many years now to develop sensory aids for people with full or partial impairment in one or more of their sensory systems [120]. In engineering, this term has come to take on a much more general definition than was originally described by Bach-y-Rita [24] and now means transcoding information from one sensory modality to present it to a different sensory modality, and this is the way we will use the term in this manuscript. Since direct force feedback has not yet been integrated into current commercial surgical robotic systems, many research works have investigated the substitution of the true sensing modality to convey to the surgeon a representation of the forces applied by the robotic tele-manipulators. This might offer a significantly more practical solution in RASS settings as it can be easily integrated into existing consoles, is less expensive to implement, and is more stable, manageable and controllable than direct force feedback [161]. Furthermore, sensory substitution can be very effective in training surgeons to use RASS and can compensate for the lack of haptic feedback in the robotic system.

Several studies have presented and evaluated different sensory substitution options to transmit forces and tissue properties information. The most commonly used sensory modalities for feedback in this context can be classified into two groups: (i) monomodality including tactile, auditory or vision and (ii) multimodality which refers to the combination of two or more sensory modalities.

Single Sensory Modality (Monomodality)

Starting with tactile feedback, early investigations noted that fingertips contain sensitive sensory receptors that allocate large areas in the sensory cortex for information processing, which makes vibrotactile modality a good option for successfully presenting force feedback information [133]. The potential benefits

of vibrotactile sensory substitution for force feedback were first explored in the work of Massimino and Sheridan, in which they tested the use of tactile and auditory senses to convey forces in teleoperation tasks. In that work, force was scaled to a vibration stimulus presented to the index finger and thumb, and the subjects were required to react as quickly as possible once they recognized the presence of a contact force. According to the experimental results, the operators performed better than when working with no force feedback.

Researchers in a different study designed a simulated tissue probing task to measure the effect of vibrotactile feedback on surgeon's performance. Particularly, they measured its effect on three main aspects: control of force application, tissue material differentiation and task completion time [196]. According to the results, having vibrotactile feedback allowed subjects to perform better, reducing the depth error and maximum force applied, and achieving more consistency compared to when no vibrotactile feedback was available. Similar results were reported in a more recent study in which authors tested the value of adding vibration feedback to the surgical setup during robotic surgery. The study illustrated that vibration feedback increases the level of awareness about tool contacts and demonstrated the users' strong preference for this technology [109]. Despite these capabilities, vibrotactile feedback is limited in the amount of information it provides as it is difficult to convey both force direction and magnitude at the same time with vibration. Other drawbacks of vibrotactile feedback start to appear when it is used for long periods of time, as they become uncomfortable for the surgeon and the skin starts to lose its sensitivity to the vibration stimuli [32, 160, 109].

Another form of feedback is audio modality, which has been shown to improve task performance in many teleoperation settings. Authors in [107] studied the effect of sensory substitution on suture-manipulation forces and evaluated four feedback scenarios. One of them was audio feedback, in which a single tone was provided to the operating surgeon when the applied force reached a specified ideal value. Even though the audio cues did not differentiate forces applied by the left or right hand instrument, they still improved the consistency of the robotically applied forces. However, surgeons who participated in that study preferred having a continuous real-time feedback over a discrete single-event information.

This modality was examined in a different study in which authors presented force feedback as an auditory signal to both ears with the tone loudness being proportional to the magnitude of the force [133]. The experimental results in

that study revealed that the reaction speed for recognizing the presence of a contact force was quickest with the auditory feedback compared to vibrotactile and traditional force feedback. Even though some studies showed that continuous frequency modulated audio feedback was easier to interpret by surgeons, there were still concerns about such continual auditory signals being disruptive and confusing in the operating room as it is already noisy with different sounds coming from medical instruments and verbal communication [223]. Additionally, continuous sounds during long procedures can be a source of discomfort and annoyance to the surgeon and might distract communication between assistants and the surgeon [176].

Early investigations showed the feasibility of vision modality, sight-to-touch, for sensory substitution during delicate surgical tasks. In the work of Bethea et al. [34], a subject study was presented in which participating surgeons were instructed to perform a robot assisted knot tying task with and without the aid of a color bar as sensory substitution. In that experiment, the visual color bar scale was used to convey the mean tension applied to the suture and it progressed as the tension increased. After statistically analyzing the results, the authors found evidence that visual sensory substitution allowed surgeons to have more consistent, precise, and greater control over the tension applied to the fine suture material without breaking it.

Visual feedback was also compared against other sensory substitution alternatives in [107] where authors presented a visual feedback in the form of two bars, one for each hand, in the upper right corner of the display with the height and color of the bars changing according to the measured force. Out of the different sensory substitution options, visual feedback appeared to enhance most the consistency of applied forces and was superior to the other alternatives. A real-time visual force feedback graphic overlay was presented in [181] during the performance of delicate repetitious robotic manipulation of fine sutures. The graphic overlay in that experiment consisted of two semi-transparent circles superimposed over the corresponding moving instrument tips, which changed color in relation to the force magnitude. Subjects reported preference on the use of visual feedback as it helped them avoid applying excessive forces and gave them more control over the task.

Multiple Sensory Modality (Multimodality)

Apart from the use of single sensory modality, multimodal feedback has also been reported in the literature. In [41], authors carried out a meta-analysis to compare

the effects of visual-auditory and visual-tactile feedback against the use of visual feedback alone. They reported that multimodal feedback helped them to improve reaction time, but it was not effective in decreasing error rates. The influence of multimodal feedback was also explored in [55] where authors suggested that using a combination of modalities can improve realism between the user and the environment leading to a better performance. In a more recent work [212], authors performed a study of all possible combinations between visual, auditory and tactile feedback. They stated that there is no statistical significant difference between them. However, using visual modality along with another modality was preferred by the users.

The improved performance of multimodal feedback was further supported by the Wickens multiple resource model [231] that stated that human task performance improves by increasing the number of sensory resources. However, in those studies, authors did not take into account tasks that require long periods of time, the constraints of real surgical tasks and the limitations of the human cognitive system. The use of multimodal feedback can be affected in real clinical environments. Firstly, it can be affected by the attentional capacity of humans since we can keep only a limited amount of information reaching our senses [185, 112]. Secondly, the selective attention can cause loss of information which leads to more error during the procedure [58, 38].

4.2 Aim of this Work

In this work, out of all the aforementioned modalities, we use visual feedback due to the following reasons:

- Surgeons who operate the robotic systems primarily rely on their visual system to view and control the remote task via the console monitor. This makes the use of vision modality one of the most promising sensory substitution options for clinical adoption as it does not put any extra burden on the operator [161].
- Humans are able to acquire more information through vision than through all other senses combined [228]. The human visual system allows a bandwidth of 10^6 (bit/s) while tactile (skin) 10^2 and auditory (human ear) 10^4 [110].
- Vision modality allows continuously transmitting spatiotemporal information of the environment, during long periods of time, without

producing interference as it happens with auditory modality, or losing skin sensitivity as when tactile modality is used [176, 32, 160, 109].

With visual feedback, visualizations representing the information should be integrated in the three dimensional environment displayed to the surgeon. The correct visualization of this information is essential to avoid workload on the surgeon which could cause fatigue, tissue damage or increase the procedure time. Although vision modality is a feasible and comprehensible option, there still exists perceptual and cognitive burden of transcoding visual information into the force domain. To deal with this burden and to get a good representation of the information that can be quickly interpreted, one can come with natural questions such as:

Do all users interpret the different visual representations in the same way? How do users perceive these visualizations? Are they understood correctly? Leading us to formulate a particular question: – *How to display the interaction forces in an efficient way that does not disturb the surgeon’s perception flow during a surgery?*

In this work and throughout a perceptual user study, we offer an extensive discussion of the aforementioned questions with the ***aim to report our findings and recommendations on the best options to display the force information in an efficient way based on the surgeons’ preferences***. We also prove the feasibility and potentials of using vision modality in Surgical Systems. To our best knowledge, there are no works that address this issue or analyze how to efficiently represent the information in RASS.

In the remainder of this paper, our perceptual experimental study is structured as follows: Section 4.3 describes all relevant details about how our user study is carried out. In Section 4.4 we report our findings using statistics and graphical methods, and describe the results from a perceptual and cognitive point of view. Finally, we present the conclusions of the work in Section 4.5.

4.3 Perceptual Study

This section describes in details aspects that are particularly relevant for the scope of this study. Particularly, we describe three relevant aspects: the participants’ characteristics, the evaluated visualizations, and the experimental and evaluation procedure.

4.3.1 Subjects Description

Twenty eight surgeons, on a voluntary basis, participated in the study. The participants came from four specialties: Obstetrics and gynecology (OB/GYN), Neurosurgery (NS), Pediatric surgery (PDS) and Cardiovascular surgery (CS) from the Josep Trueta University Hospital, the Vall d’Hebron University Hospital and the Sant Joan de Deu Hospital in Spain. This population was divided into two main groups: experts and novices.

What defines participants to be experts or novices? This question has been a central point in psychology since we rely on experts to make decisions that affect our environment almost everyday. Examples of works that address this question can be seen in [199, 67, 49]. Distinguishing experts from novices depends heavily on individual psychological differences and behavioral characteristics and varies according to the area of study since expertise is domain-specific [199].

In the medical domain, this question is a thought-provoking topic and is part of vast psychological research. Nonetheless, distinction between experts and novices can be determined based on the number of practice hours in which surgeons can improve skills such as reduction of task completion time, movement accuracy, and identifying and solving errors [90, 101]. Based on this, we define the two groups as:

- Experts: surgeons who perform more than 20 robotic-assisted surgeries, minimally invasive procedures and non-invasive procedures, each, per month.
- Novices: surgeons who perform more than 20 minimally invasive procedures and non-invasive procedures, each, per month, but no robotic-assisted surgeries.

Defining experts by the number of surgeries is not trivial and also depends on the speciality. In this work, we selected our threshold of 20 surgeries based on works such as [172, 206]. With this criteria, we worked with 19 novices and 9 experts. All the analyses of the remaining sections is taken from these two subgroups.

4.3.2 Visualizations Description

Information is essential to understand our environment and its proper visualization determines our level of interpretation. Particularly in human-machine interaction, visual displays offer the highest bandwidth channel

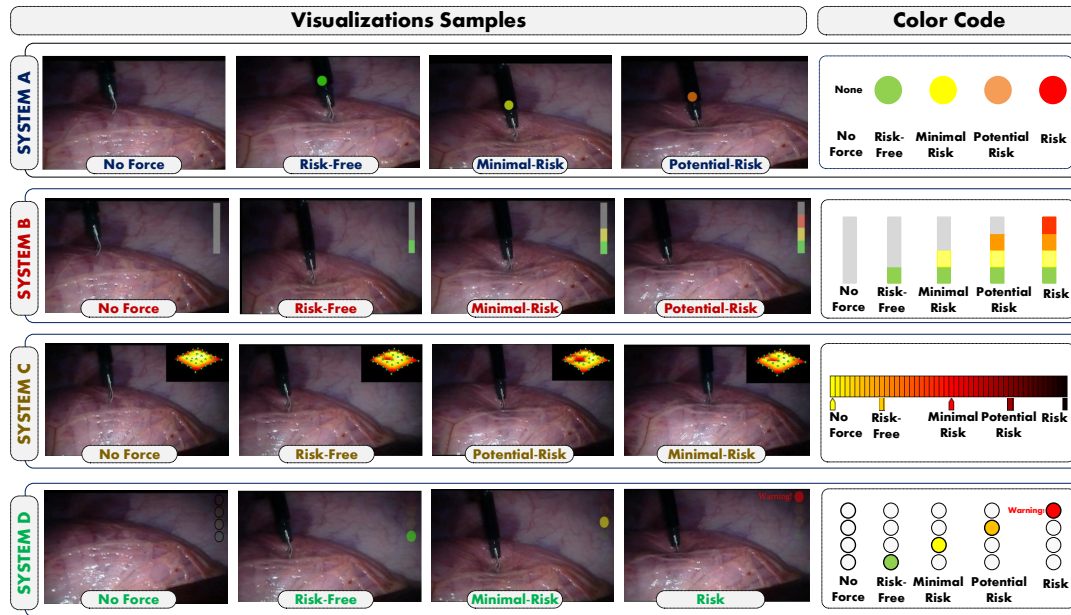


Fig. 4.3 (From left to right) The four visualizations used to carry out our experiments at different time instants. The color-coding used to indicate the level of risk according to the magnitude of the applied force.

since our visual system is capable of acquiring more information than all other senses. Finding the most adequate representation of the information is crucial to enable participants to quickly interpret what is happening in the environment and to make proper decisions. Our take on visualization design along with tasks and data descriptions are presented next.

Data and Tasks Description

We use an in-vivo porcine dataset from the Hamlyn Center Laparoscopic / Endoscopic Video Library [143]. This video sequence is composed of stereo-pair images of size 720x288 recorded during 450 sec showing the tissue deformation produced by the tool-tissue interaction. The dataset was acquired while doing palpation on the tissue and varying different factors such as illumination, position and tool orientation. This task is clinically relevant since it is commonly used to identify tumors, cut tissues and avoid tissue penetration.

As in any Robotic-assisted surgical system, which inherently lose all patient-surgeon interaction forces, participating surgeons were provided with this internal view of the surgical region of interest.

Estimating the Interaction Forces

We computed the interaction forces using our approach presented in Chapter 3 (see also [17, 18]), in which we proposed an energy functional based on L^2 in which the minimization of the residual error was changed by a maximum likelihood type estimator. Once the deformation is computed, we use a learning system to find the nonlinear relationship between deformation and force to assign the value to a particular color label. With this option, a set of target values are necessary during the training stage. However, there are some cases in which having target values is complicated. To accommodate with this problem, in [21] we proposed using a generative model in which the main idea is to model in a parametric form of several mixture elements the computed tissue deformation, resulting in an assignment of the deformation to a perceived force.

Assume having the deformation mapping as presented in Eq. 3.9 Chapter 3, R observations $\{x_1, \dots, x_R\}$ and K -groups, the process of identifying where each observation belongs is given by modeling them in a parametric form of several mixture components and then assigning them to each indicator based on its posterior probability. To do this, and on the basis of a training set, consider the set of points \mathbf{P} at a given position w as \mathbf{x} in a form $L \times D$, then, we can model the k -component by maximizing the likelihood function given by:

$$\ln p(\mathbf{x}|\theta) = \sum_{l=1}^L \ln \left\{ \sum_{k=1}^K \pi_k g(x_l|\theta_k) \right\} \quad (4.1)$$

where $\theta = \{\pi, \mu, \Sigma\}$ such that μ and Σ are the mean and the covariance matrices respectively and π is the mixture coefficients satisfying $\sum_{k=1}^K \pi_k = 1$. Moreover, let g be a the D -dimensional multivariate gaussian density function expressed as:

$$g(x_l|\theta_k) = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{-\frac{1}{2}}} e^{\frac{1}{2}(x-\mu_k)^\top \Sigma_k^{-1} (x-\mu_k)} \quad (4.2)$$

From Eqs. 4.1-4.2, the objective is to find the set of K parameters $\theta = \{\pi_1, \mu_1, \Sigma_1, \dots, \pi_K, \mu_K, \Sigma_K\}$ such that we can have an assignment of each deformation mapping. To find Eq. 4.1, we use the Expectation-maximization algorithm [53]. The advantages of using this generative model to assign a given deformation to a perceived force include the computational tractability as well as handling uncertainty. Moreover, since it is an unsupervised approach, having pre-labelled data is not required.

Thus, estimation of the applied forces can be computed by either [18] or [21] depending on the data at hand and the application. The top of Fig 4.4 shows

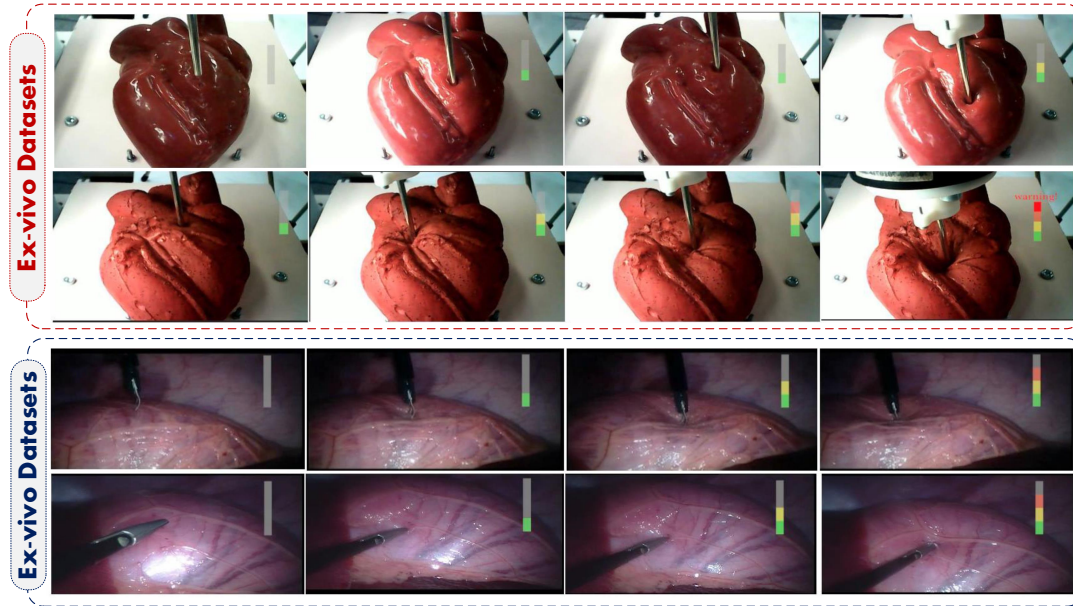


Fig. 4.4 Sample frames from four datasets with the force feedback visual cue displayed.

sample frames with force estimated using the approach described in [18] while the bottom shows the results of the approach in [21]. This information is the one displayed to the users.

Visualizations Design

For displaying the force information, we designed four different visualizations labelled as: System A, System B, System C and System D. The visualizations are shown in Fig. 4.3 and described in the following points:

- **System A.** This visualization provides force feedback information by the means of a dynamic circle that tracks the tool tip. The circle alternates between four different color indicators corresponding to the force magnitude. When no force is applied, no circle is embedded in the environment.
- **System B.** The force is represented by a dynamic bar at the top-right corner of the display. The bar alternates between five states representing the intensity of the applied force. In contrast with System A, this visualization presents stacked states, that is, past states remain displayed during the current state.
- **System C.** A heat map is shown at the top-right corner of the view. In this visualization, force is denoted by the level of deformation that the

Table 4.1 Color-coding used at each visualization to indicate the level of risk depending on the force applied

Indicator	Coding-color	Meaning
Circle (System A) Bar (System B) Traffic light (System D)	Green	Risk-free – represents minimal interaction forces
	Yellow	Minimal Risk – symbolizes a safe amount of force
	Orange	Potential Risk – refers to a potential tissue damage
	Red	Risk – warning the physician of a tissue damage
Heat map (System C)	Yellow	No force – There are no interaction forces
	Red	Minimal Risk – represents a safe amount of force
	Black	Risk – denotes tissue damage

tissue undergoes. The level of risk is represented by the color intensity where darker shades correspond to large forces (risk).

- **System D.** For this option, a traffic light type visualization is displayed at the top-right corner of the environment. It alternates between four color indicators illustrating the magnitude of the applied force. The traffic light also shows a void state (colorless) which indicates no force.

The selection of appropriate colors is an important factor to transfer the information adequately. We followed a color-coding based on the perceptual phenomena related to colors. Based on the functional and sensory-social meaning of colors [201, 96], we used red for example to convey warning messages and green to indicate a small magnitude of force. Details of the color-coding we used in relation to the amount of applied force can be seen in Table 4.1.

4.3.3 Experimental Procedure

After explaining the problem to the participating surgeons and giving the required instructions, they were provided with the four visualizations each with a corresponding computer-based questionnaire. The instructions given were as follows: We will display four different visualizations embedded in the robotic surgical system environment and labeled as System A, System B, System C and System D. Please interact with each visualization mode and respond to the corresponding questionnaire. At any time, you can return and interact again with any system and change any response in the questionnaire. The interaction entailed users watching the prerecorded video demonstrating tool-tissue interaction, with one of the four visualization systems being overlaid on the video to convey information about the force applied in the video.

Table 4.2 Questionnaire used to evaluate each visualization option based on five human factor criteria

#	Statements
(i) Perceived Usefulness	
1.	Using the system during surgery would enable me to accomplish surgical tasks more quickly
2.	Using the system would improve the surgery performance
3.	Using the system would enhance my effectiveness during the surgery
4.	Using the system would make it easier to carry out the surgery
5.	It gives me more control over the surgical task.
6.	I would like to use this system during my surgeries
(ii) Learnability	
7.	It is easy to understand the meaning of the visualization
8.	It is easy to understand without instructions
9.	It is easy to interpret the meaning of the color coding
10.	I found it easy to adapt to the visualization
11.	The system is designed for all levels of users
12.	I quickly became skillful with the system
(iii) Perceptual Limitation	
13.	The visualization tool is distracting
14.	The visualization tool is logical
15.	The visualization tool has a useful location
16.	The provided colors are easily distinguished
17.	I found the system unnecessarily complex
(iv) Consistency	
18.	Is the assignment of color codes appropriate?
19.	The display format is consistent
20.	The display orientation is consistent
21.	The data display is consistent with user conventions?
(v) Satisfaction	
22.	Overall, I am satisfied with this system
23.	Overall, the system is pleasant to use
24.	Overall, the system works the way I want it to work

All questions had a rating scale that goes from Strongly disagree to Strongly agree. Moreover, two open-ended questions were asked: List the most negative aspect(s) and List the most positive aspect(s).

Visualizations Evaluation

It is well-known that questionnaires are a powerful tool to assess the usability and reliability of human-machine interfaces [187, 173]. For our study, we designed a questionnaire as a five-point Likert rating scale in which participants were asked to indicate the level of agreement with the given statements, ranging from Strongly disagree to Strongly agree. The questionnaire was composed of twenty-four questions, shown in Table 4.2, that evaluates *five human factors* that are relevant in the context of human-machine interfaces, which are:

1. Perceived Usefulness – Refers to the extent to which each participant believes that using each one of these systems will improve his/her surgical performance.

2. Learnability – Denotes how easy it is to accomplish basic tasks and interpret outputs of a system.
3. Perceptual Limitation – Indicates the degree to which participants respond to changes in each one of these systems using their sensory system.
4. Consistency – The extent to which a participant agrees with the stimulus-response compatibility. That is, the input-output of the system exhibits logical accordance.
5. Satisfaction – Refers to the participants' level of comfort and acceptability of a given system.

It is also worth mentioning that our questionnaire contained two additional sections. The first was shown at the beginning of the session and collected information about the expertise of the surgeon. The second was presented at the end of each questionnaire in the form of an open-ended question that asked surgeons to list the most positive and negative aspects of each visualization.

4.4 Experimental Results

This section presents the analysis and results obtained from the data collected in the study. We have divided our evaluation into two main parts. The first is based on the use of statistical and graphical methods to extract participants preferences, while the second analyzes the results from the perceptual and cognitive point of view.

4.4.1 Evaluation Scheme

The aim of this study is to report findings and answer the questions presented in subsection 4.1. To achieve this, we used the following evaluation scheme:

1. We divided the collected data into two subgroups, experts and novices, to obtain the following:
 - Analysis of the results based on the five different human factors: Fig. 4.5
 - Non-parametric analysis of each human factor: Table 4.3
 - False Discovery Rate adjustment using the Benjamini-Yekutieli (BY) procedure: Table 4.3

- Overall analysis of data by subgroups (i.e. experts vs novices): Fig. 4.6
2. We used the entire population of experts and novices combined, on a basis of a statistical justification, to obtain the following:
 - Evaluation of the combined results based on the five human factors: Fig. 4.7
 - Overall analysis of the data system by system: left side Fig. 4.8
 - Finding relationships in the population that were undiscovered (Pairwise comparison): right side Fig. 4.8
 3. Finally, we reported an analysis based on different perceptual and cognitive principles.
 - Evaluation of the principles based on attention, mental models, perception, and memory: Fig. 4.10

4.4.2 Analysis and Results

Statistical and Graphical Analysis

We started our analysis by evaluating the preference degree over the four visualizations per human factor of experts vs. novices. As Fig. 4.5-(a) evidences, experts expressed the strongest preference for System A (70%) to improve their performance during surgeries, while novices expressed preference for System D (71%). Results also show a noticeable difference in opinion between experts and novices (22%) for System B. Another clear finding was the rejection of System C by both groups with an average of 37%.

The ease of completing tasks and interpreting outputs using each system were assessed by the learnability factor which results are reported in Fig. 4.5-(b). Although the plot shows a very slight difference, smaller than 10%, in opinion between expert and novices for all visualizations, the results revealed an inclination for system A by experts (76%) and System D by novices (83%). Similar to the previous factor, both groups showed a clear dislike for System C in terms of learnability.

The extent to which the users responded to the visual changes in a given system using their sensory modalities is illustrated in Fig. 4.5-(c). The plot shows a strong preference for System D by experts (72%) while novices preferred

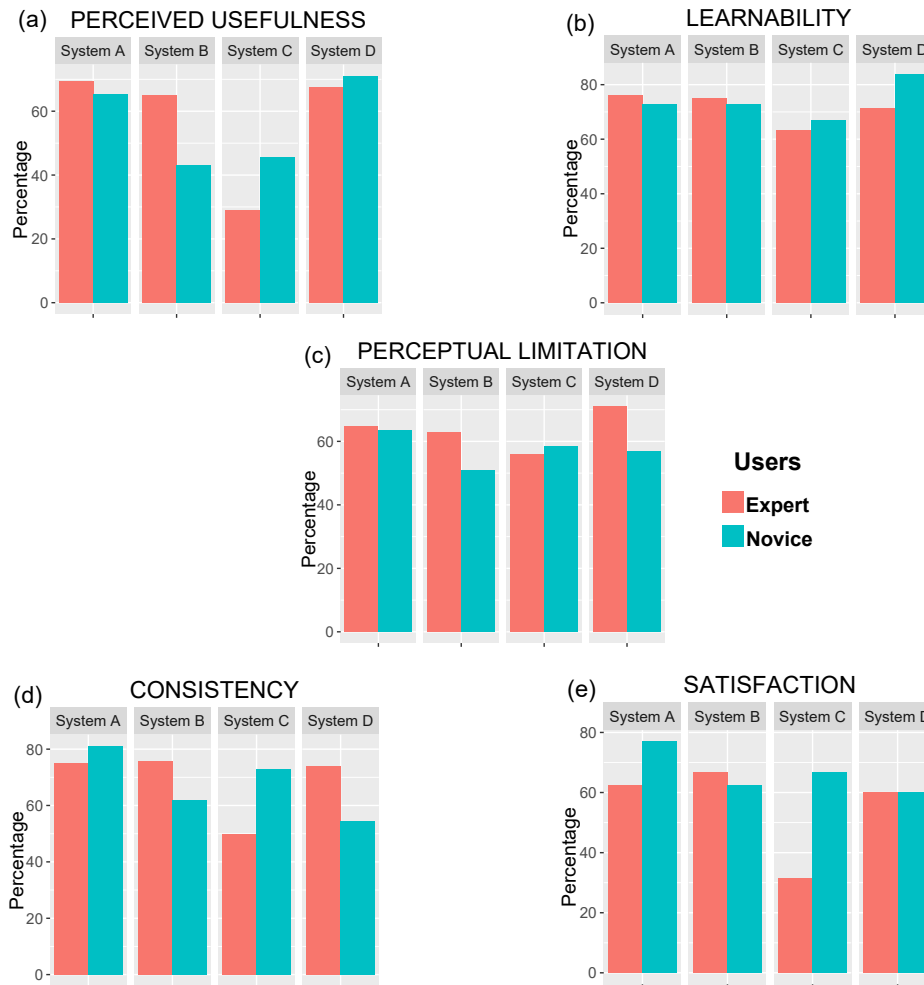


Fig. 4.5 Expert and Novices preference level per human factor. Each plot shows the percentage of acceptance of each system.

Table 4.3 Statistical nonparametric analysis of the results obtained from the experts and novices preferences. Left side shows the p -values while right side shows the adjusted ones.

<i>Experts vs Novices</i> – (ρ -value / adjusted ρ -value)				
<i>Evaluated Factors</i>	<i>System A</i>	<i>System B</i>	<i>System C</i>	<i>System D</i>
(a) Perceived Usefulness	0.5283 / 1.00	0.0279 / 0.7840	0.1291 / 1.00	0.8478 / 1.00
(b) Learnability	0.0326 / 0.7840	0.0965 / 1.00	0.6309 / 1.00	0.0950 / 1.00
(c) Perceptual Limitation	0.6294 / 1.00	0.3715 / 1.00	0.8220 / 1.00	0.3066 / 1.00
(d) Consistency	0.8894 / 1.00	0.1721 / 1.00	0.3553 / 1.00	0.0826 / 1.00
(e) Satisfaction	0.6998 / 1.00	0.0278 / 0.7840	0.9809 / 1.00	0.4965 / 1.00

System A (65%). A comparison between experts and novices in Systems A, B and C results in no significant disagreement (smaller than 5% in difference).

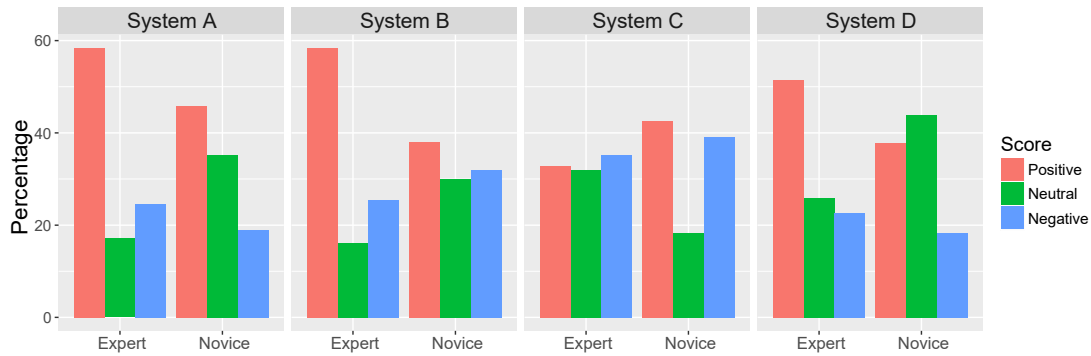


Fig. 4.6 Global view of the obtained results showing experts vs novices responses of each systems.

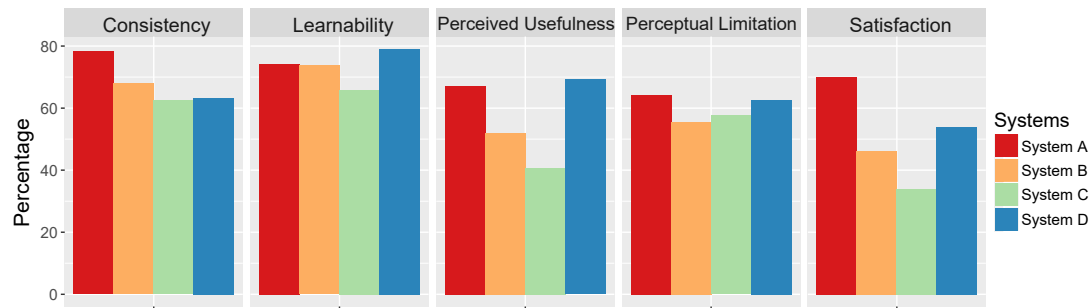


Fig. 4.7 Plots show the percentage of positive responses that each system received from the complete population per human factor.

The logical consistency of each system is assessed in Fig. 4.5-(d). Both experts and novices indicated a strong preference for System A with a percentage greater than 75%. The plot also shows that experts reported System C as the one with the lowest consistency (50%) while novices reported System D (54%) as the one with lowest score. Finally, the level of comfort and acceptability for each visualization can be seen in Fig. 4.5-(e). The plot shows a clear preference for System A by novices (77%) and a noticeable disagreement (4%) between experts and novices in regards to Systems A and B. Both subgroups reported much less satisfaction for System C.

To further support the aforementioned analysis, we used the nonparametric Wilcoxon test to check whether there is statistical significant difference in opinion between experts and novices for each human factor. Table 4.3-(i) shows the resulted p -values in which the null hypothesis was not rejected with $p < 0.05$ significance level for all systems except for the three reported in red. To eliminate the chance factor from the obtained p -values, we adjusted the false discovery rate using the Benjamini-Yekutieli (BY) method and the results are presented in gray at Table 4.3. After BY test, we found that the three results in red turned

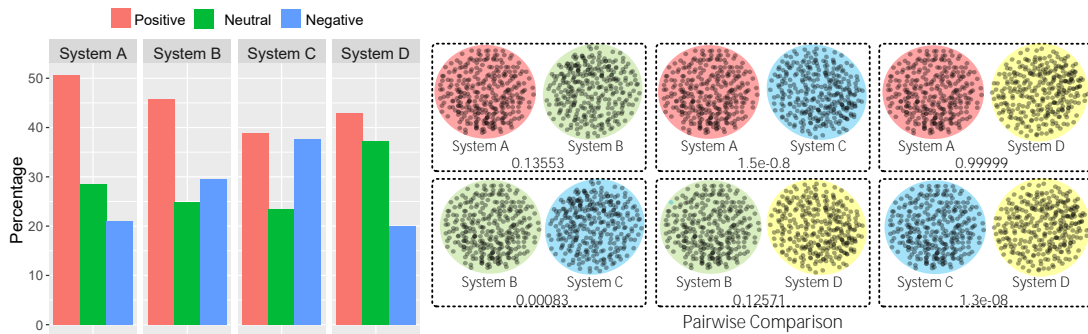


Fig. 4.8 (From left to right) Total responses evaluating each system received from the whole population. Results obtained from a post-hoc test for multiple comparison.

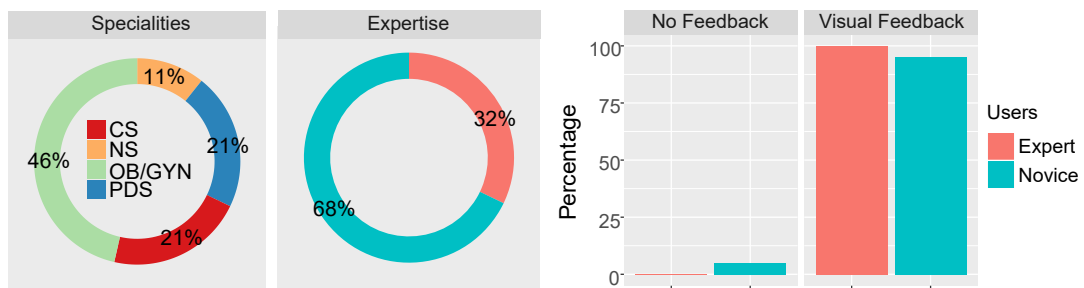


Fig. 4.9 (From left to right) Population in our experiments comes from four specialities which we divided in two subgroups: experts and novices. Distribution of the users preference in which 95% of the novices and 100% of the experts preferred visual feedback over no feedback.

to be not significant (adjusted values in blue color). This led us to conclude that overall, there is no statistical significant difference between expert and novices.

For a better understanding and visualization of the results, we carried out an aggregation process of the data, from the five-points Likert scale, in which we took into account strongly agree and agree as positive while strongly disagree and disagree as negative responses. Fig 4.6 evidences that experts and novices not only expressed a strong preference for System A, but also less confusion (see green bars). Moreover, looking at the blue bars, we found that both groups gave the most negative rating to SystemC.

Based on the previous statistical justification, we combined both subgroups and performed further analysis in the combined population. As Fig. 4.7 shows, System A was the best rated in 3 human factors: perceptual limitation, consistency and satisfaction. Although System D was preferred in terms of perceived usefulness and learnability, we found a slight difference in preference in these two factors, of 2% and 5% compared with System A. In a global inspection,

System C was the worst rated with an average percentage of all factors of $\sim 52\%$ while the most preferred were Systems A and D with averages of $\sim 71\%$ and $\sim 66\%$.

Further analysis is shown at the left-side of Fig. 4.8 with an overview of the surgeons' preference after aggregating the data (in the same way explained before). The red bars show that surgeons reflected a strong preference for System A and a hard rejection of System C (blue bars). It is worth noticing that by inspecting the green bars, surgeons expressed certain level of confusion for each system. An interesting rating was given to System D where while it had good positive rating, it also had a high level of neutral responses and the least negative feedback.

To further investigate the results, we pose the question of whether there is a statistical significant difference in preference between the four visualizations. To answer this, we computed the nonparametric Friedman test to detect differences across multiple tests and the results indicated statistically significant difference, $\chi^2(3) = 68.009$, p -value = $1.139e-14$. We then performed Nemenyi post-hoc test for multiple samples comparison and according to the results shown at the right-side of Fig. 4.8, we can see that system C highly differs from Systems A, B and D with p -values of $p < 0.05$, $1.5e-8$, 0.00083 and $1.3e-8$ respectively. Other comparisons are not significant ($p > 0.05$).

Apart from all the results reported so far, one of the most important findings is related to the significance of having a visual force feedback compared to having no feedback at all. The plots at the left side of Fig. 4.9 show a visual illustration of the surgeon population used in this study classified based on their specialties and then based on their level of expertise. Out of the entire population, only 5% of the novices reported that they prefer not having a visual feedback as illustrated in the bar chart at the right side of Fig. 4.9. Nonetheless, the majority, composed of 100% of the experts and 95% of the novices, actually preferred the option of having a visual feedback of the force information. After interacting with the different systems, they reported that the visual cues helped them to be more aware of the interactions taking place in the remote location and increased the level of transparency between them and the patient.

Perceptual and Cognitive Analysis

From a statistical point of view, it was clear that the overall population disliked System C and preferred System A. There was an intermediate and statistically indistinguishable level of acceptance for Systems B and D.




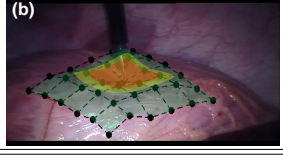

	Visualizations	Advantages	Disadvantages
SYSTEM A		<ul style="list-style-type: none"> ✓ Least eye movement ✓ Color-coding ✓ Predictive Aiding ✓ Less memory charge 	<ul style="list-style-type: none"> ✗ Non-gradual changes ✗ No close mental model
SYSTEM B		<ul style="list-style-type: none"> ✓ Redundancy gain ✓ Color-coding ✓ Predictive Aiding ✓ Less memory charge ✓ Familiar Representation 	<ul style="list-style-type: none"> ✗ Constant eye movement ✗ Position
SYSTEM C	<div style="display: flex; flex-direction: column;"> <div style="margin-bottom: 5px;">(a) </div> <div>(b) </div> </div>	<ul style="list-style-type: none"> ✓ Least eye movement (b) ✓ Spatial information (b) ✓ Color-coding (a), (b) 	<ul style="list-style-type: none"> ✗ Overlapping information (b) ✗ Loss of information (b) ✗ Position (a) ✗ Constant eye movement ✗ High memory charge ✗ Complex to understand ✗ Constant eye movement (a)
SYSTEM D		<ul style="list-style-type: none"> ✓ Color-coding ✓ Familiar representation ✓ Redundancy gain ✓ Less memory charge 	<ul style="list-style-type: none"> ✗ Position ✗ Constant eye movement

Fig. 4.10 The advantages and disadvantages of each visualization system as reported by the users.

While past work has dealt with the lack of force feedback for RASS, most of these studies have not fully taken into account the *end user*. The challenge of developing new visualization techniques while keeping in mind the end user has been pointed out in different works such as [228, 39, 61]. In Fig. 4.10, we summarize the advantages and disadvantages of each visualization based on the surgeons' feedback.

Fig. 4.10 shows that Systems B, C-(a) and D present the force-feedback information at the top-right corner of the display. This placement requires users to make frequent eye movements back and forth between the visualized tissue and the feedback display. Consequently, this places additional demands on the perceptual and cognitive resources that could impact both user performance and user preference with frequent use. System A minimizes this factor by placing the feedback display closer to the tissue of interest.

Our participants' responses suggested a linkage between display complexity and learnability, such that the system rated most unnecessarily complex (System C) was also rated lowest in learnability. By contrast, Systems A, B and D were rated as less complex, and also were rated higher in acceptability.

Another advantage of Systems B and D is based on the perceptual principle called *redundancy gain* which states that redundantly encoding the same information in more than one way can be beneficial for performance by providing multiple opportunities for the information to be detected and processed. In the case of these two systems, information was not just coded by color but also by position. Although color is a fundamental component for visualization, it is also known that $\sim 10\%$ of the world population have a color vision deficiency. In these cases, having the information redundantly coded using a modality other than color would be essential. Redundancy gain is thought to promote faster learning and understanding of information.

Although surgeons agreed that all the visualizations displayed the information according to what is happening in the surgical environment, they rated System C relatively low in terms of consistency with user conventions and low in the learnability factor. This suggests that the representation used in this system was relatively unfamiliar for the surgeons. This was likely related to the increased ratings of complexity and distraction for System C.

Working memory (or short-term memory) capacity is also important to consider when displaying information. Working memory capacity is limited, such that only a small number of events or items can be maintained in the working memory at once. Surgeons have a variety of factors to monitor simultaneously when they perform a robotic-assisted procedure like controlling the pedals and fulcrum. Thus, it is important to reduce the working memory load wherever possible. Importantly, users reported that Systems A, B, and D were less demanding of memory and rated them relatively high in learnability.

Finally, in order to analyze deeper the rejection of System C, we offered an alternative which is labeled as System C -(b) in Fig. 4.10. Although we improved the color-coding and relocated the force-feedback display to be closer to the tissue of interest, this system was still disliked by the surgeons. We believe that the main reason for the rejection in this case was due to the partial occlusion of the tissue by the semi-transparent force-feedback display.

Based on the previous findings, we can summarize our recommendations for displaying information in robotic surgical systems as follows:

- Avoid overlapping information over the region of interest (e.g. see Fig. 4.10 System C-(b)).
- Place the visual cue as close as possible to the surgical tool if it is of a small size (e.g. see Fig. 4.10 System A).
- If the visual cue is big, such as the visualizations shown in Fig. 4.10 System B and D, avoid placing it on the surgical tool. This is because it could cause distraction and might not fit on the tool at all times when the tool is partially visible (i.e. only part of the tool is in the current field of view).
- Use a color-coding that is compatible with the mental model of the surgeon. A simple but good example is using green-yellow-red scheme.
- Use a simple but efficient geometric shapes (for example see Fig. 4.10 Systems A, B and C).
- Do not overburden the display. You can include text but only in cases where it is needed, such as in a dangerous situation (e.g. see Fig. 4.10 Systems D).
- Offer visual cues that represent the information with more than one cue, such as position and color (e.g. see Fig. 4.10 Systems B and D).

4.5 Conclusion

The absence of force feedback in robotic surgical systems continues to be one of its major limitations and is one of the reasons why surgeons need to go through an extensive training effort to accommodate the indirect interaction. Having interaction forces information is of huge importance since it is directly related to reducing the complexity of the surgical task's execution. Force feedback also increases transparency between the operating surgeon and the patient as it gives the sensation of direct interaction. Although current literature in medical robotics is vast, the topic of designing a proper visual display of force feedback has not been sufficiently discussed yet. This is a very important aspect since having a proper visualization of the force information has direct repercussion on the surgeons' performance, particularly, when it takes into account perceptual and cognitive principles that are compatible with the surgeon.

The main goal of this work was twofold. First, to carefully assess the use of visual cues to transmit the interaction forces information. Second, to offer

recommendations for proper design of visual displays based on the surgeon's preference. To achieve these two goals, we conducted a perceptual user study to demonstrate the potential benefits of using visual feedback, taking into account the opinion and preference of the end users, i.e. the operating surgeons. Out of the entire population, 96% of participating surgeons preferred having visual feedback over none. Going back to the questions we posed in subsection 1.1, we found that in order to present the force information in a way that can be easily interpreted; we have to take into account the surgeon's mental model. Meaning that the design of the visual cues should fit the perceptual and cognitive principles of the end user.

*“The future of surgery is not about blood and guts:
the future of surgery is about bits and bytes.”*

Richard Satava

5

Sliding to Predict: Improving Vision-Based Beating Heart Motion Estimation by Modeling Temporal Interactions

Robotic-Assisted Minimally Invasive Surgery (RAMIS) has been an attractive alternative to traditional and laparoscopic surgeries during the last years since it offers diverse advantages to both surgeons and patients [198, 233]. Particularly in the last decades, RAMIS has allowed performing complex procedures including Off-Pump Coronary Artery Bypass Grafting (OPCABG). This procedure avoids the associated complications of using Cardiopulmonary Bypass (CPB) since the heart is not arrested while performing the surgery. Thus, surgeons have to deal with a dynamic target which compromises their dexterity and the surgery precision. When talking about RAMIS, one can observe that since its introduction, the number of robot-assisted coronary procedures remained stable [174].

To compensate the heart's motion, different authors have proposed solutions based on mechanical stabilization (for example see [27, 239, 189, 72]), in which small devices positioned over the heart surface keep the region to be repaired in

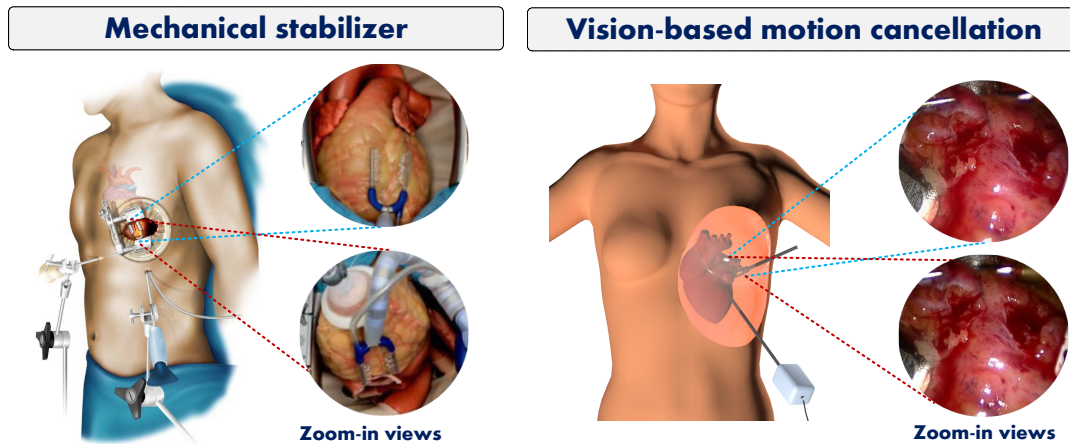


Fig. 5.1 Left side shows a mechanical stabilizer called OctopusTM Nuvo (figure reproduced from [137]) for heart motion cancellation. Compared with these type of devices, the right side illustrates an alternative approach that tracks the area of interest using the endoscopic camera.

a steady state. However, such works presented by Lemma and colleagues [119] reported that there is still a significant residual motion (1.5 – 2.4 mm) after mechanical stabilization. This entails needing manual compensation from the surgeon, which is not possible since the heart’s motion exceeds the human tracking bandwidth [65]. Moreover, these mechanical stabilizers can only be positioned on a small region of the heart’s surface and they can cause irreversible heart damage that affects the cardiac mechanics [63, 125].

To overcome the aforementioned difficulties, the pioneered work of Nakamura [149] reported that motion cancellation is possible by tracking the heart’s dynamics and continuously synchronizing this motion with the robot. This was the start of different studies and developments in terms of cardiac motion compensation using different sensors such as accelerometers [89], laser-scan endoscope system [82] and whisker sensor [30]. However, these kind of sensors present problems such as space restriction, biocompatibility and sterilization constrains, size and cost, long-term stability and the difficulty to adapt them to the surgical system. These issues prevent their adoption in real clinical scenarios [144, 84].

A more practical option is the use of a vision sensor such as the one integrated in the endoscope or a noninvasive real time system such as 3D ultrasound imaging. This direction has been followed by different authors. An image-based motion tracking algorithm was proposed in [116] for retrieving the cardiac surface deformation using a stereo endoscopic system. In that work, authors formulated the tracking problem as a time-varying optimization of a parametric function

which described the disparity map. However, they did not take into account the effect of occlusions from either the surgical tool or the specular highlights, which can severely affect the performance and stability of the tracking algorithm. Later on, Ortmaier et al. presented in [164] a 2D affine matching algorithm using natural landmarks for estimating the heart motion. Authors dealt with occlusions of the region of interest by integrating a prediction scheme based on Takens Theorem and combining electrocardiogram and respiration pressure signals.

The heart motion estimation problem was addressed in terms of displacement and acceleration using calibrated landmarks placed on the heart surface in [195]. Afterwards they proposed tracking small regions of the heart based on texture information and reported tracking failure in some situations, such as a change of illumination. A probabilistic framework for recovering 3D tissue deformation was presented in [126]. The authors used a Markov-Random Field based Bayesian Network to achieve a good representation of the heart's surface. Although they took into account specular highlights using a color based filtering approach, they did not consider occlusion events. In [240], authors proposed a one-degree-of-freedom heart motion compensation using ultrasound imaging data. They compensated the delay caused by the fast motion of the heart by including an Extended Kalman Filter (EKF).

Richa et al. in [183] proposed tracking the heart surface using a thin-plate spline (TPS) deformable model. They included an illumination compensation solution based on finding the element-wise multiplicative lighting variation. Another solution was presented in [35] in which the heart's motion was retrieved using a stochastic physics-based tracking approach. This approach was able to deal with surface occlusions by using the Kalman filter. They tested the approach in a vision system composed of three cameras. Another 3D tracking approach based on a quasi-spherical triangle was introduced in [234]. Authors modeled the heart surface using a triangle with a curving parameter. They handled occlusions by applying an algorithm based on the peak-valley characteristics of motion signals.

In more recent works [235], a tracking scheme for the heart motion using two recursive processes was presented. In the first process, they represented the target region in joint spatial-color space, and in the second, they applied the thin-plate spline model to fit the heart shape around the region of interest. Yang in [236] proposed a motion prediction scheme for tracking the heart motion

during occlusion events. That scheme was based on the dual Kalman filter in which a point of interest was modeled as a dual time-varying Fourier series.

In this chapter, we detail a new approach to compensate the heart motion in a RAMIS setup. Our solution takes advantage of the primary information obtained in a robotic surgical system which is the visual information obtained from the endoscopic camera, avoiding in this way the use of additional devices. We track the heart motion using a variational framework which we formulated in both L^1 and L^2 and we then increase robustness in term of delays and occlusions by adding a prediction stage. While this is an important part of our solution, the main contributions are:

- We propose a diffeomorphic variational framework which is able to deal with the inherent complex deformation of a beating heart. Unlike previous chapters, here we maintain affine linear transformations by means of the curvature penalizer and we extended our topology preserving penalizer to the 3D space which is not trivial due to the optimization process. It also incorporates a preprocessing stage for dealing with specular highlights.
- A key point is our prediction stage which is different from existing approaches that use well-known algorithms from estimation theory such as the Extended Kalman Filter. We slide the given sequential data to formulate a standard supervised learning problem, which is handled via a Conditional Restricted Boltzmann Machine.

The remainder of this chapter is organized as follows. After reviewing the state of the art related to cardiac motion compensation in Section 5.1, we present the challenges to keep in mind when vision-based motion compensation is used. Section 5.2 describes our solution to capture the beating heart motion while Section 5.2.3 details our prediction strategy for cases in which the camera view is occluded or the master-slave communication is compromised. We evaluate our solution in Section 5.3 using phantom and in-vivo datasets. Future work and conclusion are presented in Section 5.4.

5.1 Challenges in Vision-Based Beating Heart Motion Estimation

A suitable and practical solution for compensating the cardiac motion during RAMIS is to track the heart using visual information and then synchronize the

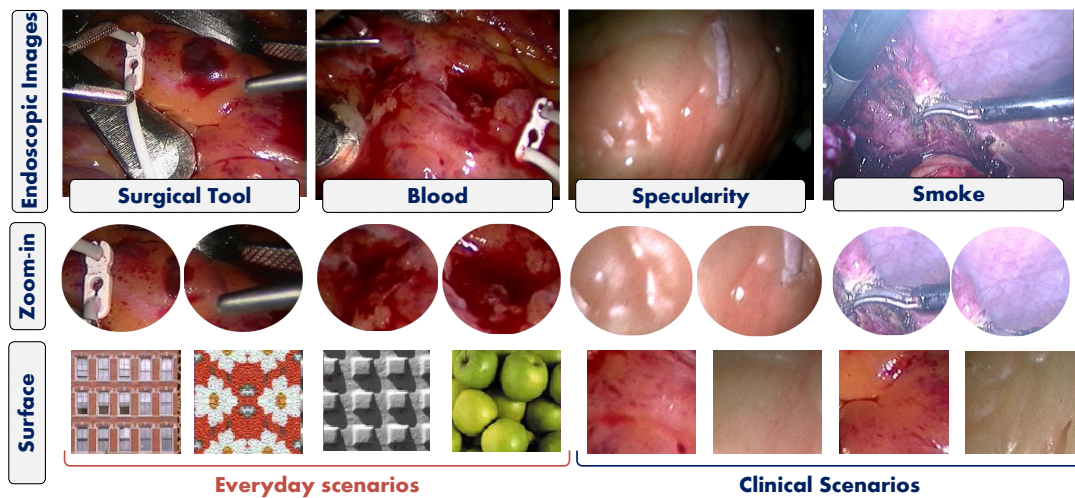


Fig. 5.2 Top row displays samples of typical endoscopic images while middle row highlight.

resulted motion with the robots in what is called Vision-Based Cardiac Motion Compensation. The robustness and accuracy of the visual based algorithms depend heavily on having a consistent surface appearance of the object across the images and it is very crucial to take into account sources of errors that impair their performance. Typical endoscopic images are illustrated in Fig. 5.2 in which the bottom row shows some factors that need to be taken into account in order to have a suitable solution.

A potential source of error is *specular highlights* that appear as bright spots resulting from the light reflection on the organs's glossy surface. These spots potentially occlude the field of view which generates discontinuities in the images and causes loss of texture/color information [9]. Particularly, during heart motion compensation, specularity on the heart's surface cause one of the major tracking disturbances.

A similar challenging factor related to surface occlusion occurs due to the restricted workspace during a RAMIS procedure. The small volume makes it very common that two spatially separated objects interfere with each other causing *partial occlusions*. Particularly the endoscopic camera can be occluded, for a period of time, by surgical tools, blood or surgical smoke. During the laps of time where occlusion occurs, tracking precision is compromised which leads, in some cases, to algorithm failure.

A comparison between everyday scenarios and organ's surfaces is displayed in the bottom row of Fig. 5.2. As evidenced by this comparison, outdoor images are characterized by a global non-homogeneity in contrast to the strong homogeneity

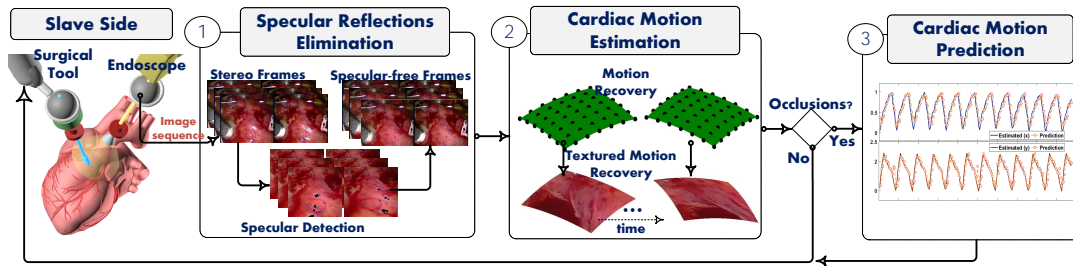


Fig. 5.3 Overview of our proposed approach composed of three main parts. (From left to right) The image sequence acquired from the endoscope at the slave side is passed to the first step which eliminates the specular highlight artifact. The specular-free images then go through our cardiac motion estimation step which recovers a 3D deformable surface of the region of interest. Finally, the last step guarantees information at all time by predicting the motion in cases where there is occlusion.

of the organs’ surface. This is another challenging factor in Vision-based cardiac motion cancellation since the majority of the heart’s surface is homogeneous, thus, it does not have stable features or identifiable textures that the tracking algorithm can use to infer the heart’s motion.

5.2 Towards Cardiac Motion Estimation

When vision-based cardiac motion cancellation is used, it is necessary to estimate the beating heart motion and actively synchronize that motion with the surgical system. In this section, we present our vision-based approach to cancel the cardiac motion. Our approach is composed of three main parts illustrated in Fig. 5.3:

1. Specular Reflection Elimination: this step allows eliminating discontinuities produced in the image domain due to light reflection.
2. Cardiac Motion Estimation: in this step, we model the motion of the beating heart using a variational framework.
3. Cardiac Motion Prediction: in this step, we ensure continues information to our approach by predicting the missing data during occlusions

In this section, we describe our solution and provide more details about the aforementioned parts.

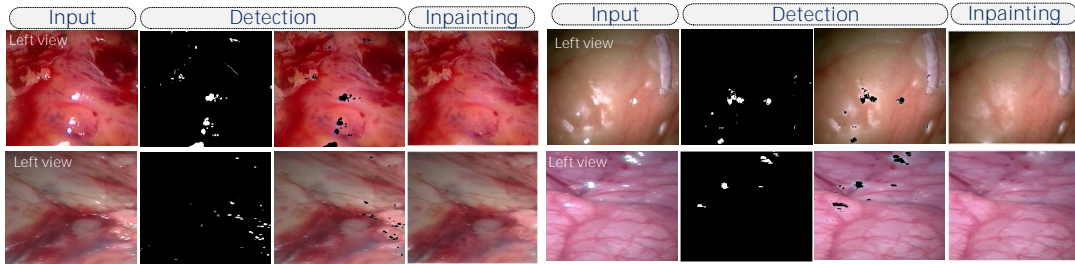


Fig. 5.4 Specular highlights detection and inpainting results of our proposed algorithm on four different datasets. From left to right: input image; detected specular regions and information retrieval via Sobolev inpainting.

5.2.1 Specular Reflection Elimination

Since the performance of the vision based solution are highly dependent on the available visual information, it is very important to handle any source of error that might affect the solution’s robustness and precision.

The visual information available on the robotic surgical setting is mainly compose of videoscopic views of the internal organs which often have glossy surface with strong reflectivity. This results in the appearance of specular highlights on the organ’s surface which are white bright spots that correspond to light reflection. Specular highlights hinder the performance of the vision-based solution as they partially occlude the targeted surface, appear as additional features, generate discontinuities in the images, or cause loss of texture or color information. In this work, we eliminate these artifacts by carrying out a two-steps correction on the input image sequence.

The first step is a segmentation step with the goal of accurately identifying the specular regions. This is accomplished by using a robust threshold that automatically adapts to the color variations in the input image sequence. We coupled this threshold with a gradient-based edge detector to further enhance the segmentation and as a result, the specular highlight regions within the image are accurately isolated.

The second step involves retrieving the missing information in the regions that were occluded by the specular highlights. We used a robust mask-specific Sobolev inpainting approach to recover the missing pixels. The approach is based on minimizing a functional using the projected gradient descendant to achieve smooth and real-time inpainting. More detailed description of both steps can be found in our work [5] and an illustration of our approach is shown in Fig. 5.4.

5.2.2 Cardiac Motion Estimation

Assume a calibrated image sequence $G = \{g_s\}_{s=0}^{S-1}$ composed of S stereo-pair frames where $g_s = \{f_r^s, f_l^s\}$. Let $f_r^s \rightarrow \mathbb{R}^2$ and $f_l^s \rightarrow \mathbb{R}^2$ denote the left and right views of s on its bounded domain Ω . To retrieve the heart's motion, we start defining a lattice in each stereo view according to the next definition:

Definition 13 *A lattice, \mathfrak{L} , is a subgroup in a real vector space V of dimension d that has the form $\mathbb{Z}v_1 + \dots + \mathbb{Z}v_d$*

Consider $\mathfrak{L}_l^s, \mathfrak{L}_r^s \subset \mathbb{R}^2$ as the lattices defined for the left and right views of g_s . We recover the 3D heart surface by computing the projections of the corresponding points from \mathfrak{L}_l^s and \mathfrak{L}_r^s as illustrated in Fig. 5.6, which results in a three dimensional lattice $\mathfrak{L}^s \subset \mathbb{R}^3$ with a set of lattice points \mathbf{B} . In this work, we represent the deformable heart surface by the tensor product of the b-splines ξ_c for $c = 0, 1, 2, 3$. Assume a given position $x \subseteq \mathbb{R}^d$, a d -dimensional lattice point defined as $z := y_1 \dots y_d$ and n degree b-splines. Then, the deformation can be represented as:

$$\begin{aligned} \varphi(x; \mathbf{B}) &= \sum_{j_1=0}^n \dots \sum_{j_d=0}^n \mathbf{B}_{j_1, \dots, j_d} \prod_{k=1}^d \xi_{k,c}(x_k) \\ \xi_{k,0}(x) &= (1-x)^3/6 & \xi_{k,1}(x) &= (4+3x^3-6x^2)/6 \\ \xi_{k,2}(x) &= (1-3x^3-3x^2+3x)/6 & \xi_{k,3}(x) &= x^3/6 \end{aligned} \tag{5.1}$$

The 2D Cardiac Motion Case

Although our goal is to recover the three-dimensional heart motion, we first start by analyzing the repercussion in terms of the number of control points defined in the lattice as in Definition 13. This is important since this application requires a physical time as close to real-time as possible while taking into account getting a good approximation. To do this, we first formulate the problem in 2D to have a better view of this effect. We presented this work in [14, 13].

From the calibrated stereo-pair image sequence defined previously as G for this 2D case and following previous notation, we rewrite our sequence as $\hat{G} = \{f_l^s\}_{s=0}^{S-1}$. So far our energy functionals have been formulated using the L^2 norm, but there are other alternatives such as the L^1 norm which is pointed out to provide, in some cases, a clearer model interpretation [219] or a better performance within a space that contains irrelevant features [154]. Motivated by this, in [14] we formulated the cardiac motion based on the L^1 norm in the form $h(x) \triangleq L(x) + \gamma \|x\|_1$ where L is the loss function with the L1-penalty.

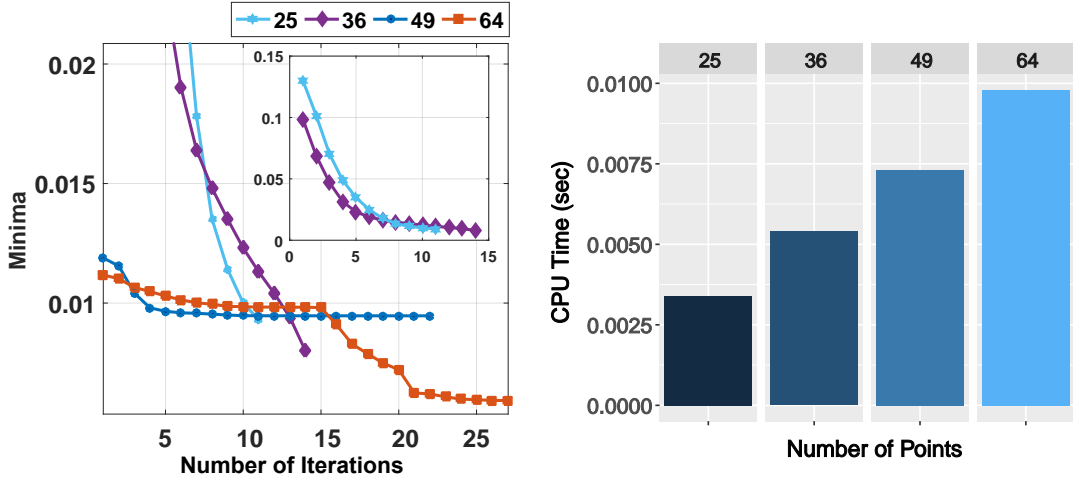


Fig. 5.5 From left to right. Results obtained from our energy function with different number of control points. CPU time reported during the optimization process.

Particularly, we used the Total Variation penalizer [190] which is defined as follows:

Definition 14 For a given image, the Total Variation is defined by duality for $u \in L^1_{loc}(\Omega)$ and is given by:

$$J(u) = \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi \, dx : \phi \in C_c^{\infty}(\Omega; \mathbb{R}^N), |\phi(x)| \leq 1 \forall x \in \Omega \right\} \quad (5.2)$$

using Definition 14 and the sum of absolute differences (i.e. $\|j_g - j_d\|_1$) as loss function then, we can express the changes on the heart surface as:

$$\begin{aligned} E^{2D}(f_0, f_1; \mathbf{B}) &= \int_{\Omega} |f_0(x + \varphi(x; \mathbf{B})) - f_1(x)| \, d\Omega + \gamma \sum_{j=1}^n \int_{\Omega} |\nabla \varphi(x; \mathbf{B})_d| \, d\Omega \\ &\approx \sum_{\Omega} |f_0(x + \varphi(x; \mathbf{B})) - f_1(x)| + \gamma \sum_{j=1}^n \sum_{\Omega} |\varphi(x; \mathbf{B})_d| - \mathbf{f}_{\log} \end{aligned} \quad (5.3)$$

using $n = 1, 2$ and where \mathbf{f}_{\log} is a barrier function used to deal with the non-differentiability of the L^1 norm defined as:

Definition 15 A barrier function is a continuous function $b(x)$ defined over the interior of a feasible set. The particular $b(x)$ that is used in Eq. 5.3 is the logarithmic barrier function \mathbf{f}_{\log} expressed as:

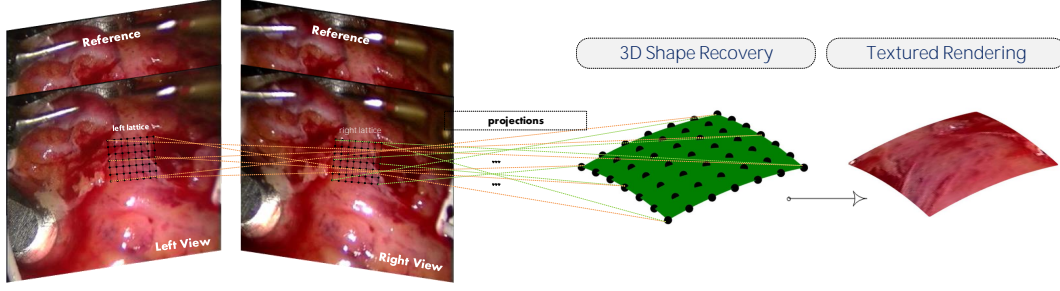


Fig. 5.6 3D Diffeomorphic surface reconstruction from the projection of the lattice points defined at each stereo-pair image.

$$\begin{aligned}
 \mathbf{f}_{\log}(\mathbf{u}) &= -\varphi \sum_{x \in \Omega} \log c_x(u) \quad \left\{ u \in \mathbb{R}^d \mid c_x(u) > 0 \text{ for all } x \in \Omega \right\} \\
 \nabla \mathbf{f}_{\log} &= -\sum_{x \in \Omega} \frac{\varphi}{c_x(u)} \nabla c_x(u) \\
 \nabla^2 \mathbf{f}_{\log} &= \varphi \sum_{x \in \Omega} \left[\frac{1}{c_x(u)} \nabla c_x(u) \nabla c_x(u)^T - \frac{1}{c_x(u)} \nabla^2 c_x(u) \right]
 \end{aligned} \tag{5.4}$$

From the results reported in Fig. 5.5, it is clear that as the better approximation is obtained, the more computational time is demanded. For detailed discussion, refer to the Experimental Results section.

The 3D Cardiac Motion Case

After defining the deformation model in Eq. 5.1, the changes in the heart’s surface deformation over time are computed using an energy functional. The functional is composed of three terms: (i) a data term that allows measuring the discrepancy between the current f_r and f_l , (ii) a regularization term that enforces a plausible transformation and (iii) a topology preservation term which ensures connectivity between the structures created within the lattice.

Particularly, we represent the data term as the Sum of Squared Differences replacing the minimization of the residual error $\sum_i r_i^2$ with $\sum_i \rho(r_i)$, where ρ is the Tukey’s M-estimator (refer to the Appendix) that we used to increase robustness in sense of outliers. The second term is formulated using the curvature method which has the advantage of penalizing oscillation and maintaining affine linear transformations [69].

Definition 16 *A map $f : X \rightarrow Y$ preserves topology if there exist f^{-1} and both f and f^{-1} are smooth.*

For the third term and following Definition 16, we use the topology preservation term that we first proposed in [20], which penalizes the Jacobian determinant to preserve the anatomical structure of organs. Unlike works like [195, 126, 183, 234] where topology preservation is not considered, in this work, we demonstrate the relevance of preserving the heart's anatomical structure specially during complex deformations. Taking these three terms and assuming a given reference \mathbf{R} and a number of m pixels in the overlapped domain Ω_{f_r, f_l} , our energy functional is given by:

$$\hat{\mathbf{E}}_s(\mathbf{B}) = \left(\frac{1}{m}\right) \underbrace{\int_{\Omega} \rho(f_r^s(\varphi(x; \mathbf{B})) - \mathbf{R}(x)) dx + \rho(f_l^s(\varphi(x; \mathbf{B})) - \mathbf{R}(x)) dx}_{\text{data term}} + \underbrace{\sum_{i=1}^d \int_{\Omega} (\Delta\varphi(x; \mathbf{B}))_i^2 dx}_{\text{regularization term}} + \underbrace{\int_{\Omega} \delta_{\varphi}(x; \mathbf{B}) dx}_{\text{topology preservation term}} \quad (5.5)$$

where δ_{φ} is the topology preservation term defined as:

$$\delta_{\varphi}(x; \mathbf{B}) := \begin{cases} \frac{\frac{1}{2}\pi - \arctan(|J_{\mathbf{h}}(x; \mathbf{B})|)}{\pi} + \varphi\sqrt{|J_{\varphi}(x; \mathbf{B})|^2} & \text{if } (\star) \\ 0 & \text{otherwise} \end{cases} \quad (5.6)$$

$$(\star) \quad | |J_{\varphi}(x; \mathbf{B})| - 1 | \geq \tau$$

where $\varphi \in \mathbb{R}^+$ offers a balance in our penalization, and $\tau \in \mathbb{R}^+$ is the margin of acceptance for values close to one. Our penalizer φ was first introduced in Chapter 2 and here we extended it to the three dimensional space, which is nontrivial due to the optimization process.

5.2.3 Sliding to Predict: Improving Cardiac Motion Estimation

The success or failure of any vision-based solution depends heavily on the available visual information given as input. During a RAMIS procedure, a common challenging factor is the presence of *partial occlusions* that results from having two spatially separated objects interfering with each other and occluding the endoscopic camera view. This causes loss of information in the occluded

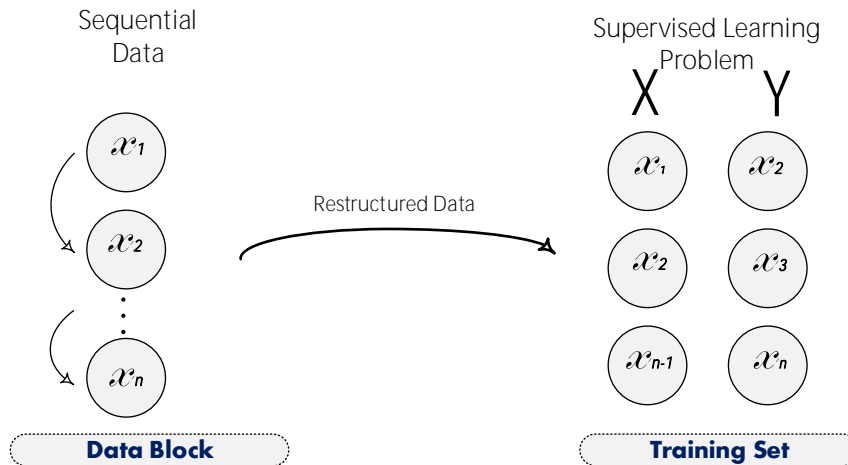


Fig. 5.7 A toy example that illustrates how we restructure our sequential data to be used with a standard supervised learning approach

regions during the laps of time when occlusion occurs, which compromised the tracking precision and leads, in some cases, to algorithm failure.

To deal with this issue, many of the studies in the literature of cardiac motion estimation deal with the lost information by using prediction approaches. The most commonly used algorithms come from the classic estimation theory such as the Extended Kalman Filter (EKF) and the Auto-Regressive eXogenous (ARX) model. The EKF is considered probably the most widely used estimation algorithm particularly because in practice, measurements are nonlinear. In short, the EKF takes the nonlinear model and linearizes the given transformations and then, KF is applied to estimate the next state. The ARX model allows inferring values by relating current state to a finite number of past states and exogenous inputs to the set of states of interest.

Although algorithms from estimation theory have been exhaustively studied and widely applied in diverse applications, in the last decades, machine learning has provided useful tools as another alternative to solve prediction of sequential data. In the standard supervised learning problem, a set of n training samples in the form of input-output pairs $\{(x_i, y_i)\}_{i=1}^n$ is needed to find the function M that maps $X \xrightarrow{M} Y$ and works well on unseen inputs x . Our problem at hand lacks of true observed values Y . In order to use a standard supervised learning approach, we restructure the given sequential data $\{(x_i)\}_{i=1}^n$ in the form $\{(x_i, x_{i-d})\}_{i=2}^{n-1}$ where d is the size of the lag [22]. This process is also known as the Sliding Window Method which allows converting any sequential data into a standard supervised learning [56]. A toy example of this process can be seen in Fig. 5.7.

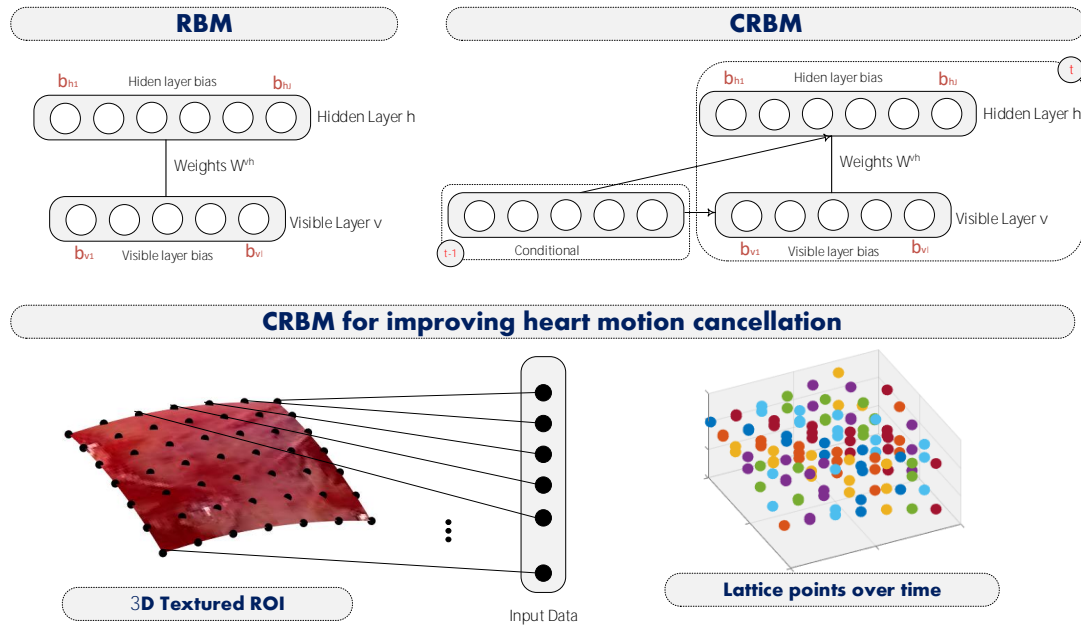


Fig. 5.8 Top part illustrates the architectures of both RBM and CRBM. The left-bottom part shows illustration of how we use the reconstructed heart motion as an input for CRBM while the right side shows the accumulated lattice points over time.

Taking the previous tool, our key idea is to predict the heart's motion within the lattice domain *not just during occlusions events, but as a feedback information for improving the heart motion estimation* restructuring the data to have a standard supervised learning problem. In the literature, one can find works for sequential data using machine learning as a tool for prediction in different applications, using for example, Gaussian processes for human motion prediction [226] or Encoder-Recurrent-Decoder for body pose prediction [70]. Unlike the works reported in the literature related to cardiac motion cancellation that make use of the classic estimation theory, some examples can be found in [240, 35, 236]. In this work we exploit machine learning tools. To our best knowledge, this is the first work that shows the performance of a generative model in the cardiac motion cancellation problem.

Definition 17 *A Restricted Boltzmann Machine (RBM) is a two-layer graphical model that learns a probability distribution of a given set of inputs and can be defined as the energy E where the probability distribution of the visible and hidden units is given in terms of E having:*

$$\begin{aligned}
 E_{RBM}(v, h|W, b^v, b^h) &= -\left(\sum_i \sum_j v_i W_{ij} h_j + \sum_i v_i b_i^v + \sum_j h_j b_j^h\right) \\
 &= -(v^\top W^{vh} h + v^\top b^v + h^\top b^h) \tag{5.7} \\
 P(v, h) &= \frac{1}{Z} \exp(-E_{RBM}(v, h))
 \end{aligned}$$

where W refers to the weights matrix, h and v are the hidden and visible units, b^v and b^h are the unit bias, and Z the normalization factor.

Although RBMs are powerful models, they are not able to capture temporal dependencies from the model data. To cope with this problem, an extension of RBMs called Conditional Restricted Boltzmann Machines (CRBM) [217] have been recently a focus of attention, and in particular, in dealing with motion capture [241, 216, 217]. An important feature of the CRBMs is that, once they are trained, they can build a deep belief network by stacking layers. For illustration purposes refer at the top part of Fig. 5.8.

For improving the cardiac motion, within the lattice domain, we exploit CRBM as a tool to, on the one side, improve the heart motion estimation and, on the other side, predict the motion during occlusion events. Let c be the vector (the conditional) that contains the past information in the form time $t - 1, t - 2, \dots, t - M$ of the lattice (points motion). See illustration at bottom part of Fig. 5.8. The joint probability function, given the hidden and visible layers, the conditional data, and M past elements, is expressed in terms of the energy E_{CRBM} as:

$$\begin{aligned}
 E_{CRBM}(v_t, h_t|c, W, \mathfrak{W}, b^v, b^h) &= E_{RBM}(v, h|W, b^v, b^h) \\
 &\quad - \sum_m \left(\sum_k \sum_i v_{ki,t-m} \mathfrak{W}_{ki,t-m} v_{it} + \sum_k \sum_j v_{kj,t-m} \mathfrak{W}_{kj,t-m} h_{j,t} \right) \tag{5.8} \\
 p(v_t, h_t|c, W, \mathfrak{W}, b^v, b^h) &= \frac{1}{Z} \exp(-E_{CRBM}(v_t, h_t|c, W, \mathfrak{W}, b^v, b^h))
 \end{aligned}$$

For training the CRBM, we used the well-known contrastive divergence algorithm [87]. Details about the architecture, for example number of units, is explained in the experimental results.

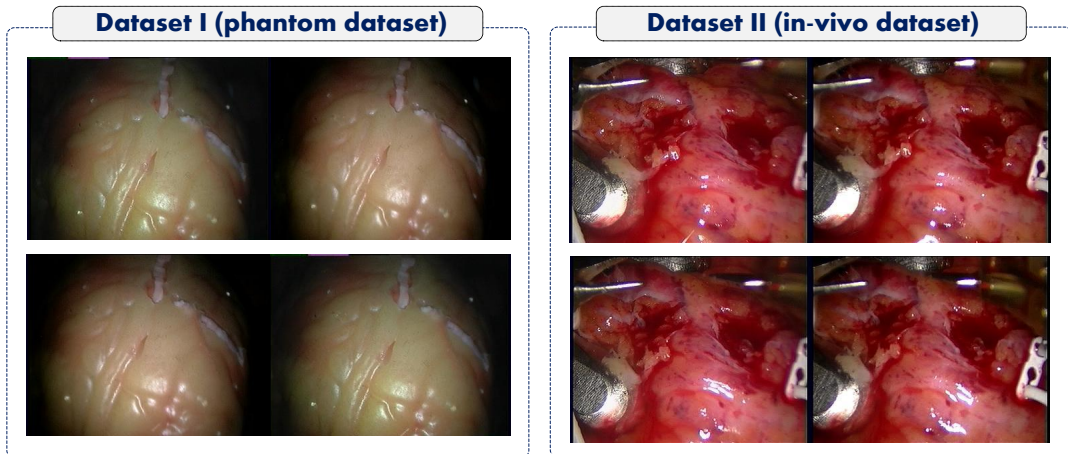


Fig. 5.9 Sample frames of the raw data from the two datasets used in our experimentation

5.3 Experimental Results

This section describes in detail the experimentations that we conducted to validate our proposed solution.

5.3.1 Cardiac Data Description

We used both phantom and in-vivo datasets to evaluate our approach. The phantom dataset [211] is from a silicon heart with cardiac motion. It is composed of stereo-pair images of size 720×288 with 3389 frames. We refer to this phantom dataset as *Dataset I* in the remaining of this section (see left side of Fig. 5.9).

The in-vivo data [210] comes from a robotically assisted Totally Endoscopic Coronary Artery Bypass (TECAB) surgery. It is composed of a stereo image sequence of size 720×288 with 1573 frames. In the remainder of this section, we refer to this sequence as *Dataset II*. See right side of Fig. 5.9.

All the measurements and reconstructions in this section are taken from these sequences. All test were ran on a Python based implementation under an Intel(R) Core i7- 6700 CPU at 3.40GHz-32GB RAM, and a Nvidia GeForce GT 610.

5.3.2 Evaluation Scheme

Using Datasets I and II, we designed the next validation scheme to evaluate our approach:

- Inspection of our specular-free approach: top part of Fig. 5.10

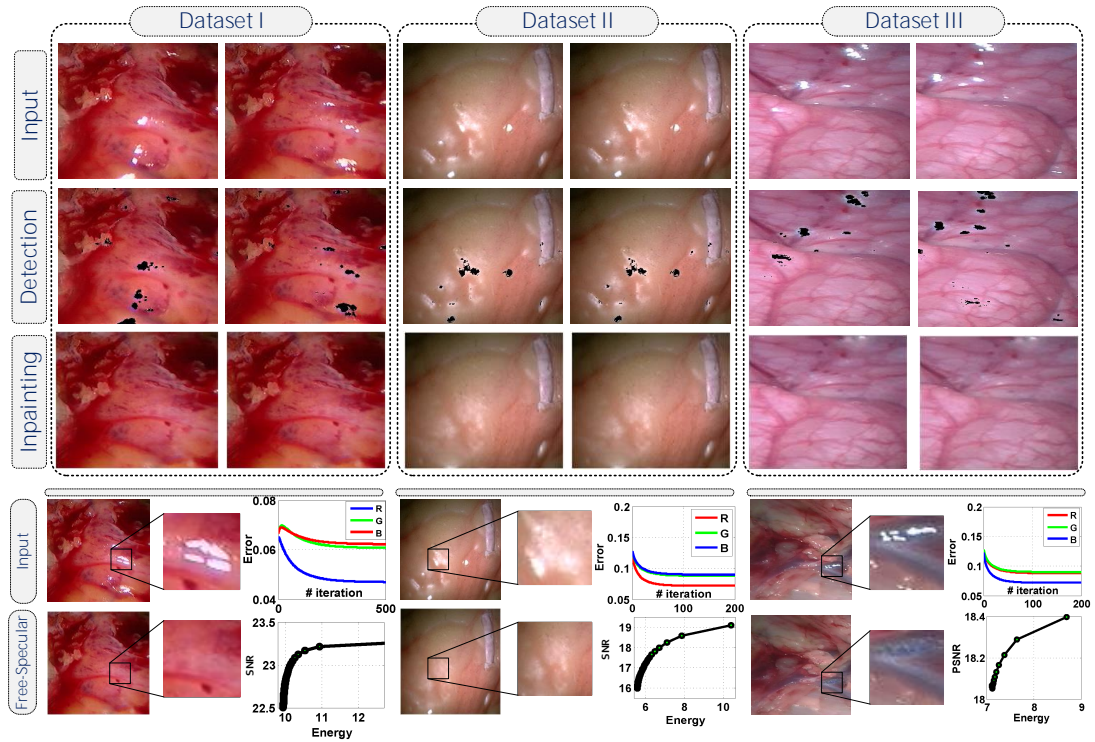


Fig. 5.10 Top part shows results from our specularity elimination approach on three different medical datasets while the bottom part shows zoom-in views of the inpainting results along with signal-to-noise ration SNR plots

- Numerical results of the specular highlights detection: bottom part of Fig. 5.10
- Illustration of the 3D heart surface reconstruction and inspection of the accumulated displacement field: Fig. 5.11
- Numerical visualization of our energy functional: Fig. 5.12
- Assessment of the heart motion recovery at a point of interest: Fig. 5.13
- Numerical visualization of heart motion using NARX, EKF and CRBM prediction: Figs. 5.14, 5.15, 5.16
- Error comparison of NARX, EKF and CRBM predictions schemes: Fig. 5.17
- Statistical comparison of NARX, EKF and CRBM predictions schemes using the Wilcoxon and Friedman tests.

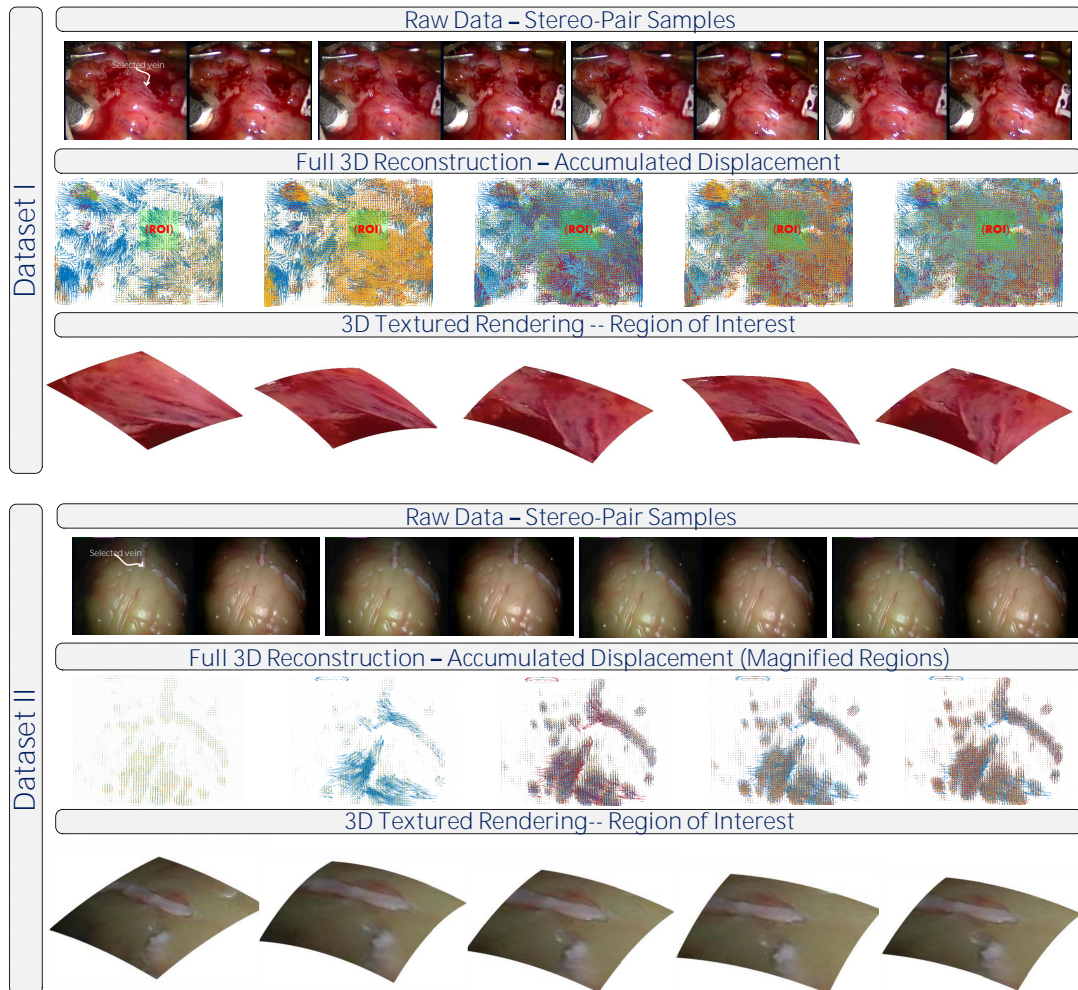


Fig. 5.11 For each dataset from top to bottom: example frames of the input raw data, accumulated displacement of the reconstructed 3D heart at different time instances, and visualization of the recovered region of interest.

5.3.3 Results and Discussion

This section describes in detail the experimentations that we conducted to validate the accuracy of the proposed solution.

Specular-Free Approach

We evaluated the performance of our specular highlight detection and elimination approach and show the results in Fig. 5.10. We tested our approach in three different datasets, the first two are cardiac sequences while the third are ureter and kidney sequences. We included the kidney dataset in our tests in order to evaluate the robustness of our approach and test its performance with organs other than the heart. To offer a quantitative evaluation of our detection approach,

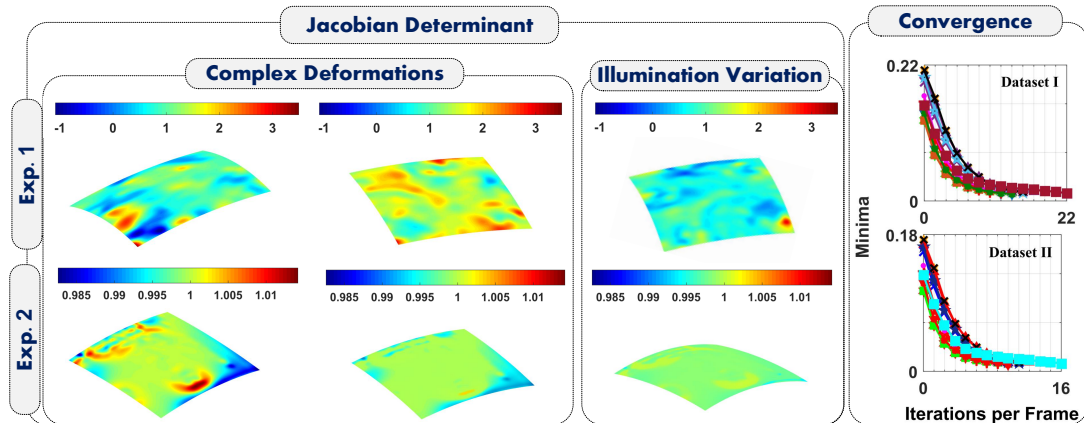


Fig. 5.12 Left side shows the Jacobian Determinant results of our vision based approach, with and without applying our topology preservation term, in two different cases: the retrieval of complex deformation and the under illumination variation. The right side shows the convergence results of our optimization process on the two datasets while using the topology preservation term

we used a ground truth from each of the sequences. The results showed that the specular highlight regions were detected with an $> 99\%$ accuracy in all datasets. Aside from this numerical evaluation, we also show detection and inpainting results on frames from each dataset in the top part of Fig. 5.10. For each frame, we show the original RGB image with specular artifacts along with the detection and inpainting results. From visual inspection, it is clear that our approach is able to adapt well to diverse color variations and specular reflections and accurately detect and eliminate their artifacts. The bottom part of Fig. 5.10 shows zoom-in visualizations of the inpainting results along with plots that represent Sobolev energy minimization and signal-to-noise (SNR) ratio improvement during the inpainting process.

Vision-based Cardiac Motion Cancellation

We start evaluating our vision-based approach (see Eq. 5.5) by recovering the heart’s motion. In Fig. 5.11, we show the resulted 3D reconstruction of the heart surface using Datasets I and II. The top rows of Fig. 5.11 of both datasets show stereo-pair image samples with the region to be repaired pointed out. The middle rows show the accumulated displacement field of the complete image domain. As it evidence by the images, unlike Dataset I which exhibits a strong homogeneity in the surface, Dataset II presents strong visual texture which provide more stable features during the tracking process of the region of interest. For visualization purposes, at the middle row of Dataset II in Fig. 5.11, we magnified those regions

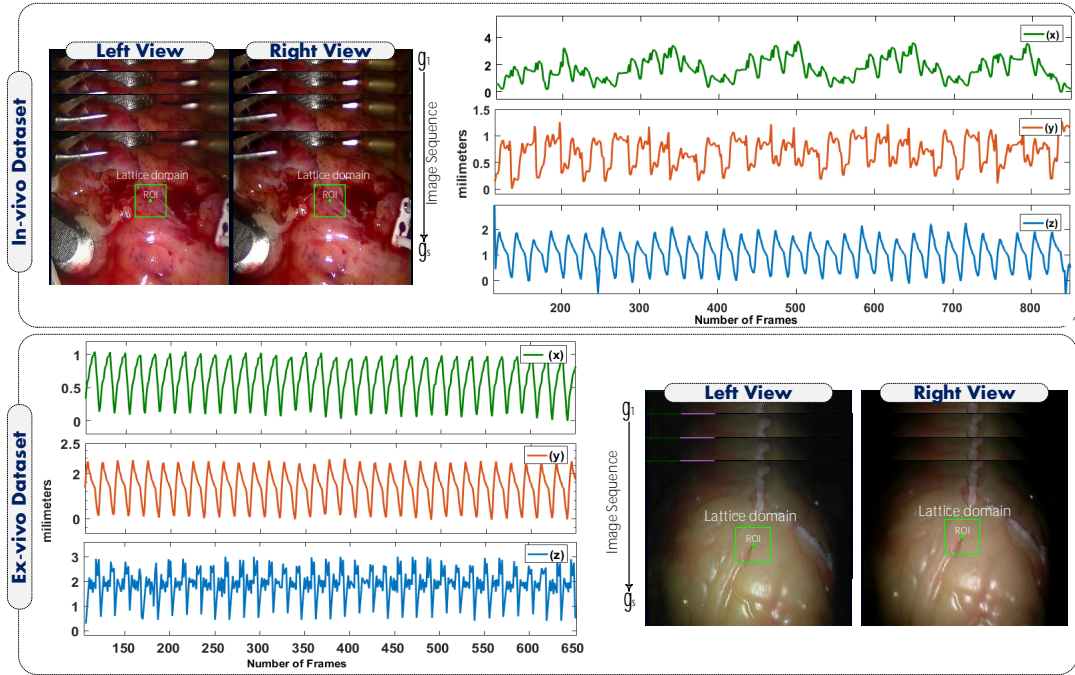


Fig. 5.13 The motion of a point of interest over time used in the prediction stage.

with potential visual features. The bottom rows from both datasets illustrate the 3D reconstruction of the region of interest (ROI), this information is the one used as an input to the next stage (prediction stage). We only use information from the ROI since the surgeon's attention is focused on the zone to be repaired and this allows decreasing the computational time to more than half the time. The plots at the bottom rows clearly show pleasant visual results of the 3D ROI with both phantom and in-vivo data.

For a more detailed quantitative analysis, we evaluated the global performance of our vision-based approach. The first question that we pose is – *How robust is our vision-based cardiac motion cancellation approach?* To respond to this question, we carried out two experiments under the following conditions:

- Experiment 1: We set $\delta_\varphi = 0$ in Eq. 5.5, i.e. we remove our topology preservation term.
- Experiment 2: We include our topology preservation term by setting $\varphi = 3 \cdot 10^{-3}$ in Eq. 5.5.

After running both experiments, we found that the average range $[\min, \max]$ of the Jacobian determinant for Exp. 1 was $[-2.5471, 3.0012]$ with an average residual error of the order of magnitude 10^{-2} , while for Exp. 2, the Jacobian

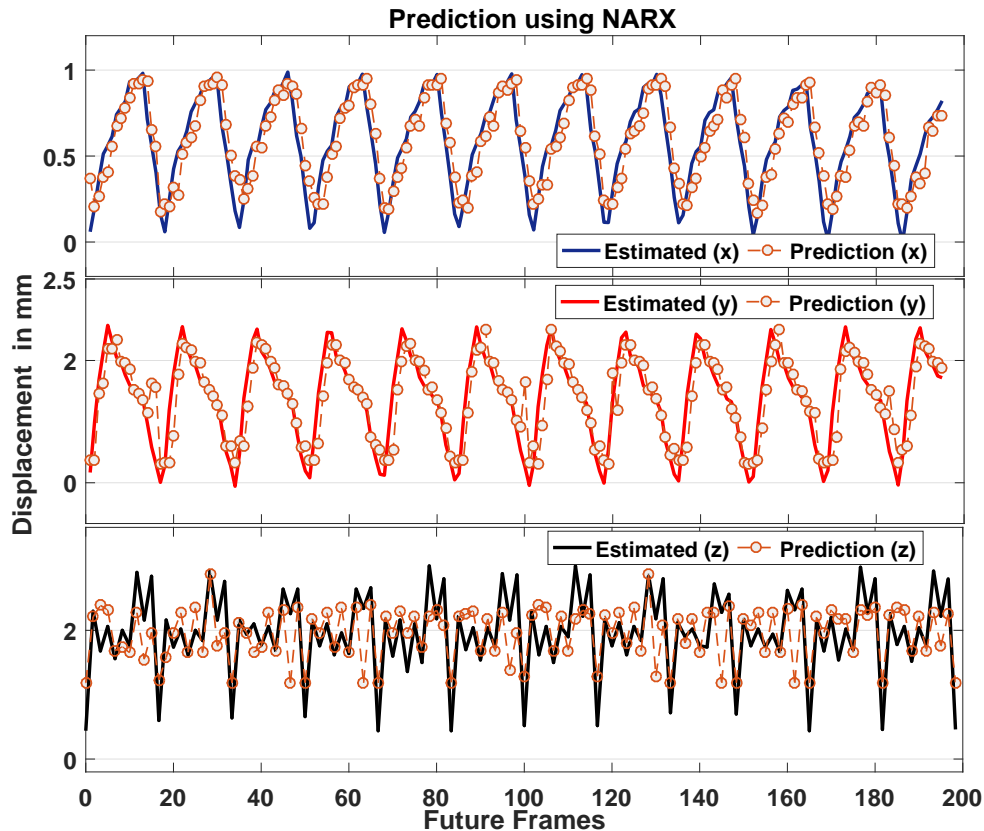


Fig. 5.14 Estimated vs predicted motions in the three direction using NARX predictor over 200 frames

exhibited stable values with an average range of $[0.9715, 1.015]$ yielding to an average minima in the order of magnitude of 10^{-7} . Some samples showing the Jacobian determinant over the region of interest, for both experiments, are displayed at the left part of Fig. 5.12. We used the nonparametric Wilcoxon test to show if there is a statistical significant difference between Exps. 1 and 2 in terms of the Jacobian determinant. According to the results, we found that the null hypothesis was rejected with $p < 0.05$ significance level, which lead us to conclude that there is a statistical significant difference if we use the topology preservation term to preserve the anatomy of the heart. This yields to a better result and stabilization of our approach in different cases such as complex deformation or illumination variation.

The results also showed that Exp. 2 needed an average of 22 (Dataset I) and 16 (Dataset II) iterations per frame to converge compared to the 31(Dataset I) and 27(Dataset II) needed in Exp. 1. Visualization of the convergence for Exp. 2 can be seen at the right side of Fig. 5.12.

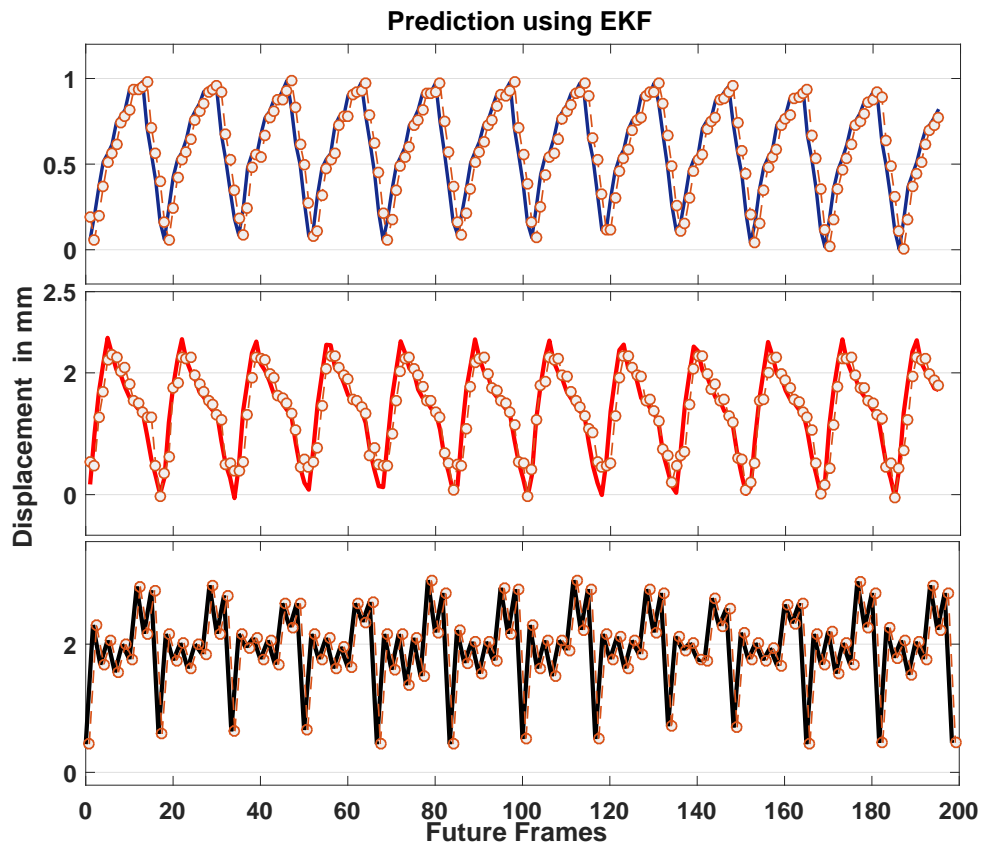


Fig. 5.15 Estimated vs predicted motions in the three direction using EKF predictor over 200 frames

Cardiac Motion Prediction

Since the ultimate goal of this work is to *accurately and robustly* retrieve the cardiac motion in order to synchronize it actively with the surgical instrument, we need to make sure that our solution is robust enough to different kind of disturbances. While at the beginning of this section, we coped with the problem of specular highlights, in this subsection we analyze the performance of our approach during partial occlusions. To do this, we first extracted the motion of a point of interest in (x,y,z) directions from both datasets as shown in Fig. 5.13. This data is the one used in the remaining of this section.

In order to offer a careful analysis of our prediction scheme, we took two well-known predictors from classic estimation theory: the NARX and EKF (for detailed description refer to Appendix B). We use these two predictors to check whether a statistical significant difference exists between those schemes and the one based on CRBM over 200 frames.

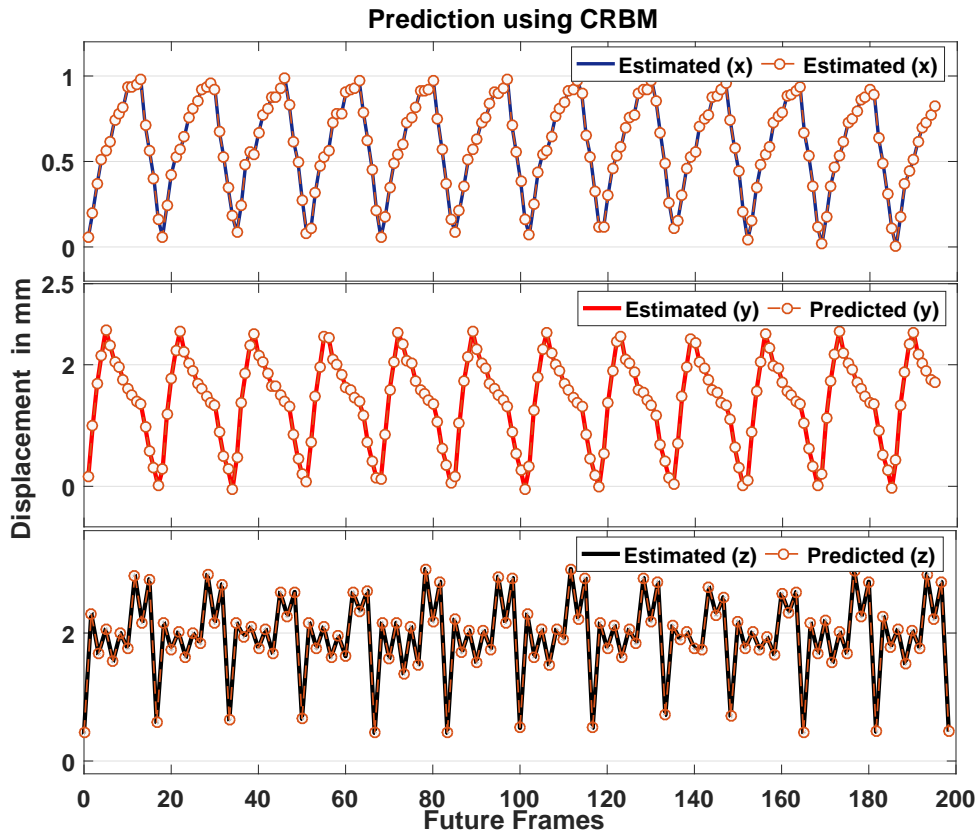


Fig. 5.16 Estimated vs predicted motions in the three direction using CRBM predictor over 200 frames

We begin by analyzing the NARX predictor and Fig. 5.14 shows the resulted prediction for x,y and z directions. From visual inspection, it is clear that for the x and y directions, the prediction was acceptable. However in the z direction, the predicted values were far from the target. This is further supported by the Root-Mean Square Error (RMSE) computed for all directions and plotted in the left side of Fig. 5.17. The RMSE shows that NARX was able to predict x and y direction within a maximum RMSE of $1.1mm$ while z was far to be retrieved accurately since it reached a maximum RMSE of $1.7mm$. The average RMSE for NARX is $0.69mm$.

We also evaluated the performance of the EKF, which is probably the most used well-known predictor. The results are reported in plot 5.15. A visual inspection shows that EKF overcame the NARX predictor in all directions. This is also evidenced by the RMSE reported in the middle of Fig. 5.17 which exhibits a concentration of error values lower than $0.2mm$. Particularly, the maximum errors for x,y and z are 0.38 , 0.43 and 0.27 mm respectively, and the average RMSE is $0.1153mm$.

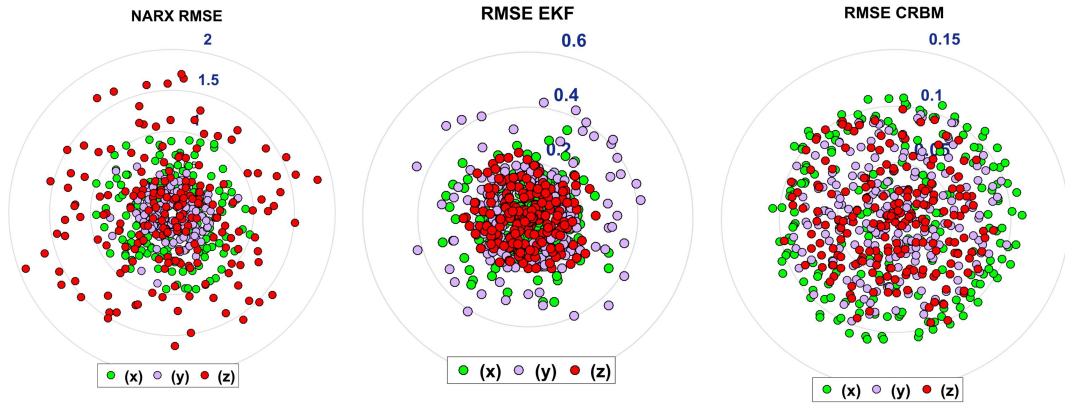


Fig. 5.17 Numerical comparison of the three prediction models in mm between the target and predicted values using RMSE

Finally, we evaluated the CRBM for predicting the cardiac motion. For the CRBM, we set the learning rate as 10^{-2} , a momentum value of 0.9, and 350 hidden units. The results from the prediction can be seen in 5.16. In a visual comparison of CRBM against NARX and EKF, one can see that the estimated values of the CRBM are closer to the target values. This is supported by the RMSE which offered a maximum value of 0.12 mm for all direction, and an average RMSE of 0.071mm.

But is there a significant difference in terms of prediction between NARX, EKF and CRBM? To answer this question, we computed the nonparametric Friedman test to detect differences across multiple tests and the results indicated statistically significant difference. This lead us to conclude that CRBM achieved a better prediction, from a statistical point of view, than NARX and EKF.

5.4 Conclusions

Cardiovascular diseases are the leading global cause of death and thanks to the recent technological advances; it has been possible to offer alternative solutions oriented to the patient's benefit. A clear example is the Robotic-assisted cardiac surgery that is performed through small incisions and while the heart is still beating. Although from a medical point of view, avoiding heart arrest offers clear benefits to the patient, from a technical point view it is very challenging to deal with a dynamic target, which compromise the surgery precision.

In this chapter, we offered a vision-based cardiac motion cancellation approach as an alternative solution to the mechanical stabilizers. To achieve a robust solution, we took into account different disturbances such as specular reflections

and occlusion events. Particularly, we proposed recovering the 3D cardiac motion by the means of a variation framework that guarantees diffeomorphic transformations with the aim of preserving the anatomical structure of the heart. Moreover, we incorporated a Sobolev based approach to achieve specular-free images for dealing with the singularities produced by the specular reflections.

Another key point of our solution is its robustness in terms of partial occlusions. To cope with this problem, we proposed restructuring the visual data to formulate a supervised learning problem with the aim of predicting the missing information. Specifically, we used a CRBM predictor.

Based on the results, we demonstrated that our visual approach reached an average minima in the order of magnitude of 10^{-7} while preserving the heart's anatomic structure and providing stable values for the Jacobian determinant ranging from 0.917 to 1.015. Moreover, we proved the accuracy of our specular reflection detector of 99% based on a ground truth. In terms of prediction, our approach reported the lower average RMSE of 0.071 in comparison with the NARX and EKF of 0.69 and 0.1153 respectively. We further supported this statement using a multiple comparison statistical test. The results pointed out significant difference in estimation between the three predictors, this together with the RMSE support the performance of the CRBM predictor.

Our approach avoids the risk of damaging the heart given by the mechanical stabilizers. Our solution can also be effective for acquiring the motion of organs other than the heart such as the lung or everyday dynamic objects.

“If a conclusion is not poetically balanced, it cannot be scientifically true.”

Isaac Asimov

6

Concluding Remarks

In this thesis, we dealt with a central topic in computer vision – *Estimating and Understanding Motion* with particular emphasis in clinical scenarios. Through three applications, we set the basis for achieving a realistic motion estimation with the aim of offering a better clinical understanding. We went beyond existing solutions from the state of the art and presented alternatives drawn from different areas such as mathematical modeling, machine learning, robust statistics, computer vision and psychology. Ours solutions were strongly evaluated through numerical experiments and a combination of statistical, graphical and perceptual analyses.

In this chapter we summarize the conclusions of this thesis as follows:

In **Chapter 2**, we took advantage of the high temporal resolution of a relatively new medical imaging modality *Ultrafast Ultrasound imaging* and used it to estimate the complex motion patterns of the heart. We used this modality since it offers a better temporal resolution compared with other modalities such as MRI, CT, SPECT or PET, which is important to capture different mechanical events of the heart. We estimated the cardiac motion using a variational framework in which we *highlighted* the synergy between our proposed topology preservation term and low-rank data representation. From a technical point of view, we reported a RMSE less than 1 mm while keeping the CPU time

low and a minima in the order of magnitude of 10^{-12} . From a clinical point of view, we offered to the physician objective clinical results of the strain profiles and displacements that can be provided either in a global or in a local manner by selecting a region of interest.

✓ *Our approach is promising for the analysis other organs experiencing complex motions such as the lungs in respiration or everyday deformable objects.*

In **Chapter 3**, we transitioned from medical diagnostic to medical robotics. In that Chapter, we tackled one of the major problems in medical robotics that is the lack of force feedback. Our key idea came from the conservation of continuum mechanics in which it is clear that the change in shape of an elastic object is directly proportional to the applied force. Based on this, we extended our variational framework from 2D to 3D to retrieve the deformation that tissues undergo and then we found the nonlinear relationship between deformation and force using a deep neural architecture. From a technical point of view, we reported a RMSE less than 0.2N for all our experiments. We also pointed out the advantages of our approach including robustness, accuracy, and stability over long periods of time. From a clinical point of view, we connected Chapters 3 and 4 by posing the question - *how to provide this information to the surgeon?*

In **Chapter 4**, we carried out a user study with twenty eight surgeons from three different hospitals to respond to the aforementioned question. We offered an extensive discussion of our findings and recommendations on the best options to display the force information in an efficient way based on the surgeon's preferences. We also proved the feasibility and potentials of using sensory substitution, vision modality in particular, in surgical systems. We reported that it is important to take into account the surgeon's mental model so that the information can be easily interpreted. Meaning that the design of the visual cues should fit the perceptual and cognitive principles of the end user.

✓ *Our approach can be useful in different situations in which knowing the applied forces makes a difference in the results, including: detection and prevention of diseases or abnormal behavior, needle-based procedures, microsurgery and knot tying. Also, in outdoor environment such as in robotic grasping/recollection or robot recognition and navigation.*

Finally, in **Chapter 5**, we coped with a still open problem in medical robotics which is cardiac motion cancellation for RAMIS. We used our variational framework and optimized it to compute the heart beat. We dealt with the real-time requirement by reducing the deformation structure to a region of

interest such as the area in which the target vein lies. Moreover, we ensured robustness of our approach by recovering the lost information caused by specular highlights and including a prediction stage to guarantee visual information over time in events such as occlusions or delays. After evaluating our solution using in-vivo and ex-vivo datasets, we concluded that our solution offered a RMSE less than 0.071mm in terms of prediction in comparison with classical algorithms from estimation theory such as the NARX and EKF of 0.69mm and 0.1153mm respectively.

✓ *Our solution can be used in different scenarios such as video stabilization, surface reconstruction or improving object recognition tasks. Also, our prediction solution can be extrapolated in other applications such as trajectory planning or walking prediction.*

6.1 Future Work

In this thesis, we proposed novel solutions for challenging problems in medical imaging and medical robotics. Particularly, we pointed out the necessity of estimating and understanding motion in our world that is composed of an inherent temporal dimension. Through these solutions, we set the basis of diverse tools that can be further investigated.

For example in Chapter 2, further investigation of the synergy between low-rank representation and topology preservation can answer important questions like – What is the effect of integrating low-rank approximation within the minimization procedure? Does it significantly affect the computation cost and minima? From a clinical point of view, we set a proof of concept which opens a new line of research for further clinical investigation. The next step should be to analyze statistically our approach in terms of variation across subjects. This requires a very large dataset of patients which is not an easy task particularly with UUS being a relatively new modality.

Another focus for further investigation could be an extended analysis of the solution presented in Chapter 3 using different tissues and/or other synthetic materials. – Is it possible to find optimal hyperparameters that accurately deal with diverse tissues using single training?, Is there a statistically significant difference between different tissues/materials in terms of force estimation?

Since we tackled the problem of estimating the interaction forces in Chapters 3 and 4 and achieved motion cancellation in Chapter 5, a natural extension is to put them together to get full feedback in RAMIS.

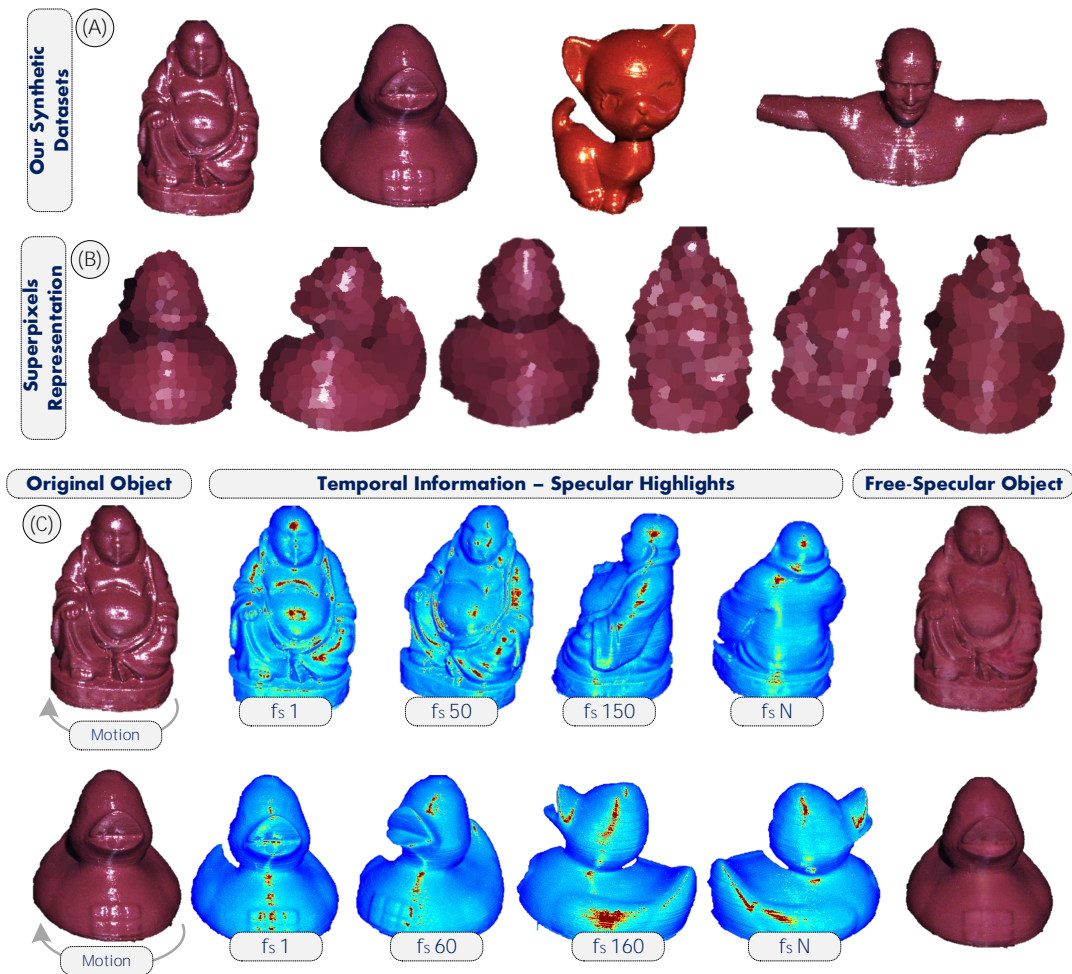


Fig. 6.1 (a) Our proposed approach for achieving a specular-free object was evaluated with synthetic data. (B) It first creates a superpixel representation of the image domain to reduce computational time. (C) By restricting the searching to only key areas in the temporal dimension, we can efficiently obtain a specular-free object.

6.2 Beyond Medical Applications

Estimating and Understanding Motion are fundamental components in a lot of applications. In this subsection, we want to highlight the adaptability of our solutions to different applications in domains other than the medical. For example, in our very recent work [6], we adapted our tools drawn from motion estimation and image inpainting to accurately capture and then remove the light reflections on dynamic objects, which remains a challenging problem in computer graphics. We proved that having specular-free objects allows improving tasks such as object recognition and visual tracking.

An illustration of this work can be seen in Fig. 6.1 in which we used synthetic data for evaluating our approach (see Fig. 6.1-(A)). Our main contribution in that work is the use of the temporal dimension to retrieve in a low-rank manner the specular features which takes into consideration N past frames. We reduced computational cost by reducing the search space using a superpixel representation (Fig. 6.1-(B)) of the image domain. After detecting the specular highlights, we proposed an inpainting process based on minimizing an energy function, which search the most consistent pixel for the missing information. The final result can be seen in (Fig. 6.1-(C)).

References

- [1] TP Abraham and RA Nishimura. Myocardial strain: can we finally measure contractility? *Journal of the American College of Cardiology*, 37(3):731–734, 2001.
- [2] Antonio Agudo, Francesc Moreno-Noguer, Begoña Calvo, and José María Martínez Montiel. Sequential non-rigid structure from motion using physical priors. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):979–994, 2016.
- [3] Chi Young Ahn and Jin Keun Seo. Myocardial motion tracking method integrating local-to-global deformation for echocardiography. *2012 IEEE International Ultrasonics Symposium (IUS)*, pages 1948–5719, 2012.
- [4] Samar M. Alsaleh, Angelica I. Aviles, Pilar Sobrevilla, Alicia Casals, and James K. Hahn. Automatic and robust single-camera specular highlight removal in cardiac images. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
- [5] Samar M. Alsaleh, Angelica I. Aviles, Pilar Sobrevilla, Alicia Casals, and James K. Hahn. Adaptive segmentation and mask-specific sobolev inpainting of specular highlights for endoscopic images. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016.
- [6] Samar M. Alsaleh, Angelica I. Aviles, Alicia Casals, and James K. Hahn. Let specular highlights perform!: Temporal image priors for free-specular recovery. *Submitted to ACM SIGGRAPH*, 2017.
- [7] Mehdi Ammi, Hamid Ladjal, and Antoine Ferreira. Evaluation of 3d pseudo-haptic rendering using vision for cell micromanipulation. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.
- [8] Michael A Arbib. *The handbook of brain theory and neural networks*. MIT press, 2003.
- [9] Alessandro Artusi, Francesco Banterle, and Dmitry Chetverikov. A survey of specular removal methods. In *Computer Graphics Forum*, volume 30, pages 2208–2230. Wiley Online Library, 2011.
- [10] John Ashburner and Karl J. Friston. Voxel-based morphometry—the methods. *NeuroImage*, pages 805 – 821, 2000.

- [11] John Ashburner, Jesper LR Andersson, and Karl J Friston. High-dimensional image registration using symmetric priors. *NeuroImage*, 9(6):619–628, 1999.
- [12] Angelica I. Aviles, A. Marban, Pilar. Sobrevilla, J. Fernandez, and Alicia Casals. A recurrent neural network approach for 3d vision-based force estimation. *IEEE International Conference on Image Processing Theory, Tools and Applications*, 2014.
- [13] Angelica I. Aviles, Sobrevilla Pilar, and Casals Alicia. An approach for physiological motion compensation in robotic-assisted cardiac surgery. *Journal in Experimental and Clinical Cardiology*, 2014.
- [14] Angelica I. Aviles, Pilar Sobrevilla, and Alicia Casals. Unconstrained ℓ_1 -regularized minimization with interpolated transformations for heart motion compensation. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5109–5112, 2014.
- [15] Angelica I. Aviles, Samar M. Alsaleh, Eduard Montseny, and Alicia Casals. V-ANFIS for dealing with visual uncertainty for force estimation in robotic surgery. *The 16th World Congress of the International Fuzzy Systems Association and the 9th Conference of the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT)*, 2015.
- [16] Angelica I. Aviles, Samar M. Alsaleh, Pilar Sobrevilla, and Alicia Casals. Sensorless force estimation using a neuro-vision-based approach for robotic-assisted surgery. *IEEE EMBS International Conference on Neural Engineering*, pages 86–89, 2015.
- [17] Angelica I. Aviles, Samar M. Alsaleh, Pilar Sobrevilla, and Alicia Casals. Force-feedback sensory substitution using supervised recurrent learning for robotic-assisted surgery. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
- [18] Angelica I. Aviles, Samar M. Alsaleh, James K. Hahn, and Alicia Casals. Towards retrieving force feedback in robotic-assisted surgery: A supervised neuro-recurrent-vision approach. *IEEE Transactions on Haptics*, 2016.
- [19] Angelica I. Aviles, Samar M. Alsaleh, Eduard Montseny, Pilar Sobrevilla, and Alicia Casals. A deep-neuro-fuzzy approach for estimating the interaction forces in robotic surgery. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1113–1119, 2016.
- [20] Angelica I. Aviles, Tomas Widlak, Alicia Casals, and Habib Ammari. Towards estimating cardiac motion using low-rank representation and topology preservation for ultrafast ultrasound data. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016.
- [21] Angelica I. Aviles, Samar M. Alsaleh, and Alicia Casals. 3d diffeomorphic deformation with mixture components as visual stimuli for perceiving interaction forces in robotic-assisted surgery. *Submitted to IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2017.

- [22] Angelica I. Aviles, Samar M. Alsaleh, James Hahn, and Alicia Casals. Sliding to predict: Improving vision-based cardiac motion cancellation by modeling temporal interactions. *Submitted to The International Journal for Computer Assisted Radiology and Surgery*, 2017.
- [23] Angelica I. Aviles, Tomas Widlak, Maartje M. Nillesen, Alicia Casals, and Habib Ammari. Robust cardiac motion estimation using ultrafast ultrasound data: A low-rank-topology-preserving approach. *Physics in Medicine and Biology*, 2017.
- [24] Paul Bach-y Rita and Stephen W. Kerce. Brain mechanisms in sensory substitution. *Academic Press*, 1972.
- [25] Paul Bach-y Rita and Stephen W. Kerce. Sensory substitution and the human – machine interface. *TRENDS in Cognitive Sciences*, pages 541–546, 2003.
- [26] Paul Bach-y Rita, Carter C. Collins, Frank Saunders, Benjamin White, and Lawrence Scadden. Vision substitution by tactile the image projection. *Nature*, pages 963—964, 1969.
- [27] Wael Bachtá, Pierre Renaud, Edouard Laroche, Antonello Forgione, and Jacques Gangloff. Design and control of a new active cardiac stabilizer. *International Conference on Intelligent Robots and Systems*, pages 404–409, 2007.
- [28] Konstantin V Baev. *Biological neural networks: hierarchical concept of brain function*. Springer Science & Business Media, 2012.
- [29] B. Bayle, M. Joinie-Maurin, L. Barbe, J. Gangloff, and M. de Mathelin. Robot interaction control in medicine and surgery: Original results and open problems. *Book Chapter in Computational Surgery and Dual Training*, pages 196–191, 2014.
- [30] Ozkan Bebek and M Cenk Cavusoglu. Whisker-like position sensor for measuring physiological motion. *IEEE/ASME transactions on mechatronics*, 13(5):538–547, 2008.
- [31] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.
- [32] Sliman Bensmaïa, Yuk-Yuen Leung, Steven S Hsiao, and Kenneth O Johnson. Vibratory adaptation of cutaneous mechanoreceptive afferents. *Journal of neurophysiology*, 94(5):3023—3036, 2005.
- [33] Jeremy Bercoff. Ultrafast ultrasound imaging. *INTECH Open Access Publisher*, 2011.
- [34] Brian T Bethea, Allison M Okamura, Masaya Kitagawa, Torin P Fitton, Stephen M Cattaneo, Vincent L Gott, William A Baumgartner, and David D Yuh. Application of haptic feedback to robotic surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 14(3):191–195, 2004.

- [35] Evgeniya Bogatyrenko, Pascal Pompey, and Uwe D Hanebeck. Efficient physics-based tracking of heart surface motion for beating heart surgery robotic systems. *International journal of computer assisted radiology and surgery*, 6(3):387–399, 2011.
- [36] William M Boothby. An introduction to differentiable manifolds and riemannian geometry. *Academic Press, Second Edition*, 2003.
- [37] Peter Bovendeerd, Wilco Kroon, and Tammo Delhaas. Determinants of left ventricular shear strain. *American Journal of Physiology-Heart and Circulatory Physiology*, 297(3):1058–1068, 2009.
- [38] Donald Eric Broadbent. *Perception and communication*. Elsevier, 2013.
- [39] Wolfgang Broll, Irma Lindt, Jan Ohlenburg, Iris Herbst, Michael Wittkamper, and Thomas Novotny. An infrastructure for realizing custom-tailored augmented reality user interfaces. *IEEE transactions on visualization and computer graphics*, 11(6):722–733, 2005.
- [40] C. Bruneel, R. Torguet, K. M. Rouvaen, E. Bridoux, and B. Nongaillard. Ultrafast echotomographic system using optical processing of ultrasonic signals. *Applied Physics Letter*, vol. 30, no. 8, pages 371–373.
- [41] Jennifer L. Burke, Matthew S. Prewett, Ashley A. Gray, Liuquin Yang, Frederick R. B. Stilson, Michael D. Coovert, Linda R. Elliot, and Elizabeth Redden. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. *Proceedings of the 8th International Conference on Multimodal Interfaces*, pages 108–117, 2006.
- [42] Eduardo R. Caianiello. Outline of a theory of thought-processes and thinking machines. *Journal of Theoretical Biology*, 1(2):204 – 235, 1961.
- [43] E.J. Candes and Y. Plan. Matrix completion with noise. *IEEE Proceeding*, 98(6):925–936, 2010.
- [44] Emmanuel J Candes, Carlos A Sing-Long, and Joshua D Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on Signal Processing*, pages 4643–4657, 2013.
- [45] Gary E Christensen, Richard D Rabbitt, and Michael I Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 1996.
- [46] Brendan Chwyl, Audrey G Chung, Alexander Wong, and David A Clausi. Specular reflectance suppression in endoscopic imagery via stochastic bayesian estimation. In *International Conference Image Analysis and Recognition*, pages 385–393. Springer, 2015.
- [47] Maja Cikes, Ling Tong, George R Sutherland, and Jan D’hooge. Ultrafast cardiac ultrasound imaging: Technical principles, applications, and clinical benefits. *JACC: Cardiovascular Imaging*, 7(8):812–823, 2014.
- [48] Robert W Cootney. Ultrasound imaging: Principles and applications in rodent research. *ILAR Journal*, pages 233–247, 2001.

- [49] L. Cuthbert, B. Duboulay, D. Teather, B. Teather, M. Sharples, and G. Duboulay. Expert/novice differences in diagnostic medical cognition- a review of the literature, 1999.
- [50] Bernard Dacorogna and Jürgen Moser. On a partial differential equation involving the Jacobian determinant. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 7:1–26, 1990.
- [51] V. De Luca, G. Székely, and G. Tanner. Estimation of large-scale organ motion in b-mode ultrasound image sequences: A survey. *Journal in Ultrasound in Medicine and Biology*, pages 3044 – 3062, 2015.
- [52] B. Delannoy, R. Torguet, C. Bruneel, E. Bridoux, J. M. Rouaven, and H. Lasota. Acoustical image reconstruction in parallel-processing analog electronic systems. *Applied Physics*, vol. 50, no. 5, pages 3153–3159, 1979.
- [53] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.
- [54] M. Diana and J. Marescaux. Robotic surgery. *British Journal of Surgery*, pages 15–28, 2015.
- [55] Iñaki Díaz, Josune Hernantes, Ignacio Mansa, Alberto Lozano, Diego Borro, Jorge Juan Gil, and Emilio Sánchez. Influence of multisensory feedback on haptic accessibility tasks. *Virtual Reality*, 10(1):31–40, 2006.
- [56] Thomas G Dietterich. Machine learning for sequential data: A review. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 15–30. Springer, 2002.
- [57] S.P. DiMaio and S.E. Salcudean. Needle insertion modeling and simulation. *IEEE Transactions on Robotics and Automation*, 19(5):864–875, 2003.
- [58] Jon Driver. A selective review of selective attention research from the past century. *British Journal of Psychology*, 92(1):53–78, 2001.
- [59] Qi Duan, SL Herz, CM Ingrassia, KD Costa, JW Holmes, A Laine, E Angelini, O Gerard, and S Homma. Dynamic cardiac information from optical flow using four dimensional ultrasound. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the IEEE*, pages 4465–4468, 2005.
- [60] Nicolas Duchateau, Bart Bijnens, Jan D’hooge, and Marta Sitges. Cardiac motion and deformation. *Chapter Book in 3D Echocardiography, Second Edition*, 2013.
- [61] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. Applying hci principles to ar systems design. 2007.

- [62] Igor Dydenko, Denis Friboulet, Jean-Marie Gorce, Jan D’hooge, Bart Bijmens, and Isabelle E Magnin. Towards ultrasound cardiac image segmentation based on the radiofrequency signal. *Medical Image Analysis*, 3:353 – 367, 2003.
- [63] Roger Dzwonczyk, L Carlos, Chittoor Sai-Sudhakar, John H Sirak, Robert E Michler, Benjamin Sun, Nicole Kelbick, and Michael B Howie. Vacuum-assisted apical suction devices induce passive electrical changes consistent with myocardial ischemia during off-pump coronary artery bypass graft surgery. *European journal of cardio-thoracic surgery*, 30(6):873–876, 2006.
- [64] Nima Enayati, Elena De Momi, and Giancarlo Ferrigno. Haptics in robot-assisted surgery: Challenges and benefits. *IEEE Reviews in Biomedical Engineering*, 9, 2016.
- [65] Volkmar Falk. Manual control and tracking—a human factor analysis relevant for beating heart surgery. *The Annals of thoracic surgery*, 74(2): 624–628, 2002.
- [66] Angela Faragasso, Joao Bimbo, Yohan Noh, Allen Jiang, Sina Sareh, Hongbin Liu, Thrishantha Nanayakkara, Helge A Wurdemann, and Kaspar Althoefer. Novel uniaxial force sensor based on visual information for minimally invasive surgery. *IEEE International Conference on Robotics and Automation*, pages 2934–2939, 2014.
- [67] Wendy Faulkner, James Fleck, and Robin Williams. *Exploring Expertise: Issues and Perspectives*, pages 1–27. Palgrave Macmillan UK, 1998.
- [68] David Feygin, Madeleine Keehner, and Frank Tendick. Haptic guidance: Experimental evaluation of a haptic training method for a perceptual motor skill. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, HAPTICS ’02, pages 40–47, 2002.
- [69] Bernd Fischer and Jan Modersitzki. A unified approach to fast image registration and a new curvature based registration technique. *Linear Algebra and its applications*, 380:107–124, 2004.
- [70] Katerina Fragkiadaki, Sergey Levine, Panna Felsen, and Jitendra Malik. Recurrent network models for human dynamics. In *IEEE International Conference on Computer Vision*, pages 4346–4354, 2015.
- [71] Yuan-Cheng Fung. Mathematical representation of the mechanical properties of the heart muscle. *Journal of Biomechanics*, 3(4):381 – 404, 1970.
- [72] Julien Gagne, Wael Bachta, Pierre Renaud, Olivier Piccin, Édouard Laroche, and Jacques Gangloff. Beating heart surgery: Comparison of two active compensation solutions for minimally invasive coronary artery bypass grafting. In *Computational Surgery and Dual Training*, pages 203–210. Springer, 2014.

- [73] H. Gao, N. Bijnens, D. Coisne, M. Lugiez, Rutten M., and D’hooge J. 2-d left ventricular flow estimation by combining speckle tracking with navier–stokes-based regularization: An in silico, in vitro and in vivo study. *Journal in Ultrasound in Medicine and Biology*, pages 99 – 113, 2015.
- [74] Ravi Garg, Anastasios Roussos, and Lourdes Agapito. A variational approach to video registration with subspace constraints. *International journal of computer vision*, 104(3):286–314, 2013.
- [75] Michael A Greminger and Bradley J Nelson. Modeling elastic objects with neural networks for vision-based force measurement. *IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, pages 1278–1283, 2003.
- [76] Michael A Greminger and Bradley J Nelson. Vision-based force measurement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pages 290–298, 2004.
- [77] Audrunas Gruslys, Remi Munos, Ivo Danihelka, Marc Lanctot, and Alex Graves. Memory-efficient backpropagation through time. In *Advances in Neural Information Processing Systems*, pages 4125–4133, 2016.
- [78] Kaiwen Guo, Feng Xu, Yangang Wang, Yebin Liu, and Qionghai Dai. Robust non-rigid motion tracking and surface reconstruction using l0 regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3083–3091, 2015.
- [79] E. Haber and J. Modersitzki. Image registration with guaranteed displacement regularity. *Int. J. Comput. Vision*, 71(3):361–372, 2007.
- [80] J. Hadamard. Lectures on the cauchy problems in linear partial differential equations. *Yale University Press, New Haven*, 1923.
- [81] B. Haeffele, E. Young, and R. Vidal. Structured low-rank matrix factorization: Optimality, algorithm, and applications to image processing. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 2007–2015, 2014.
- [82] Mitsuhiro Hayashibe, Naoki Suzuki, and Yoshihiko Nakamura. Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management. *Medical Image Analysis*, 10(4):509–519, 2006.
- [83] Kristen L Helton, Buddy D Ratner, and Natalie A Wisniewski. Biomechanics of the sensor–tissue interface—effects of motion, pressure, and design on sensor performance and the foreign body response—part I: Theoretical framework. *Journal of Diabetes Science and Technology Vol. 5*, pages 632–646, 2011.
- [84] Kristen L Helton, Buddy D Ratner, and Natalie A Wisniewski. Biomechanics of the sensor-tissue interface—effects of motion, pressure, and design on sensor performance and the foreign body response—part i: theoretical framework, 2011.

- [85] Michiel Hermans and Benjamin Schrauwen. Training and analysing deep recurrent neural networks. *Advances in Neural Information Processing Systems 26*, 2013.
- [86] B. Heyde, M. Alessandrini, J. Hermans, D. Barbosa, P. Claus, and J. D’hooge. Anatomical image registration using volume conservation to assess cardiac deformation from 3d ultrasound recordings. *IEEE Transactions on Medical Imaging*, 2016.
- [87] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- [88] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [89] Lars Hoff, Ole Jakob Elle, MJ Grimnes, Steinar Halvorsen, Hans Jørgen Alker, and Erik Fosse. Measurements of heart motion using accelerometers. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 2049–2051, 2004.
- [90] Robert R. Hoffman. *How Can Expertise be Defined? Implications of Research from Cognitive Psychology*. Palgrave Macmillan UK, 1998.
- [91] B.D. Hoit. Strain and strain rate echocardiography and coronary artery disease. *Circ Cardiovasc Imaging*, (4):179–190, 2011.
- [92] John M. Hollerbach. Some current issues in haptics research. In *IEEE International Conference on Robotics and Automation ICRA*, volume 1, pages 757–762, 2000.
- [93] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [94] Rui Hua, Jose M. Pozo, Zeike A. Taylor, and Alejandro F. Frangi. Multiresolution extended free-form deformations (xffd) for non-rigid registration with discontinuous transforms. *Medical Image Analysis*, 36: 113 – 122, 2017.
- [95] X. Huang, D.P. Dione, C.B. Compas, Papademetris X., Lin B.A., A. Bregasi, A.J. Sinusas, L.H. Staib, and J.S. Duncan. Contour tracking in echocardiographic sequences via sparse representation and dictionary learning. *Medical Image Analysis*, pages 253 – 271, 2014.
- [96] John F. Hughes, Andries Van Dam, Morgan McGuire, David F. Sklar, James D. Foley, Steven K. Feiner, and Kurt Akeley. *Computer graphics: principles and practice (3rd ed.)*. Addison-Wesley Professional, 2013.
- [97] P.J. Hunter, A.D. McCulloch, and H.E.D.J. ter Keurs. Modelling the mechanical properties of cardiac muscle. *Progress in Biophysics and Molecular Biology*, 69(2-3):289 – 331, 1998.

- [98] Stephan MD Jacobs, David Holzhey, Gero MD Strauss, Oliver Burgert, and Volkmar Falk. The impact of haptic learning in telemanipulator-assisted surgery. *Surg. Laparosc. Endosc.*, 17(5):402–406, 2007.
- [99] J-SR Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3):665–685, 1993.
- [100] S.F. Johnsen, S. Thompson, M.J. Clarkson, M. Modat, Y. Song, J. Totz, K. Gurusamy, B. Davidson, Z.A. Taylor, D.J. Hawkes, and S. Ourselin. Database-based estimation of liver deformation under pneumoperitoneum for surgical image-guidance and simulation. *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2016.
- [101] Timothy N. Judkins, D. Oleynikov, and Nicholas Stergiou. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surgical Endoscopy and Other Interventional Techniques*, no. 3, 29: 590–597, 2009.
- [102] B. Karacali and C. Davatzikos. Estimating topology preserving and smooth displacement fields. *Medical Imaging, IEEE Transactions on*, 23(7):868–880, 2004.
- [103] Fatemeh Karimirad, Sunita Chauhan, and Bijan Shirinzadeh. Vision-based force measurement using neural networks for biological cell microinjection. *Journal of Biomechanics, Volume 47*, pages 1157–1163, 2014.
- [104] Amy E Kerdok, Stephane M Cotin, Mark P Ottensmeyer, Anna M Galea, Robert D Howe, and Steven L Dawson. Truth cube: Establishing physical standards for soft tissue simulation. *Medical Image Analysis*, 2003.
- [105] Jungsik Kim, Farrokh Janabi-Sharifi, and Jung Kim. A haptic interaction method using visual information and physically based modeling. *IEEE Transactions on Mechatronics, Vol. 15, No. 4*, pages 636–645, 2010.
- [106] Wooyoung Kim, Sungmin Seung, Hongseok Choi, Sukho Park, Seong Young Ko, and Jong-Oh Park. Image-based force estimation of deformable tissue using depth map for single-port surgical robot. *International Conference on Control, Automation and Systems*, pages 1716–1719, 2012.
- [107] Masaya Kitagawa, Daniell Dokko, Allison M. Okamura, Brian T. Bethea, and David D. Yuh. Effect of sensory substitution on suture manipulation forces for surgical teleoperation. *Studies in Health Technology and Informatics*, 98:157–163, 2004.
- [108] Masaya Kitagawa, Daniell Dokko, Allison M. Okamura, and David D. Yuh. Effect of sensory substitution on suture-manipulation forces for robotic surgical systems. *The Journal of Thoracic and Cardiovascular Surgery*, 129(1):151 – 158, 2005.
- [109] Jacqueline K. Koehn and Katherine J. Kuchenbecker. Surgeons and non-surgeons prefer haptic feedback of instrument vibrations during robotic surgery. *Surgical Endoscopy*, 29(10):2970–2983, 2015.

- [110] Kenneth J Kokjer. The information capacity of the human fingertip. *IEEE Transactions on Systems, Man, & Cybernetics*, 1987.
- [111] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *Proceedings. Eighth IEEE International Conference on Computer Vision*, volume 2, pages 508–515, 2001.
- [112] Árni Kristjánsson, Alin Moldoveanu, Ómar Jóhannesson, Oana Balan, Simone Spagnol, Vigdís Vala Valgeirsdóttir, and Rúnar Unnpórsson. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience*, (Preprint):1–19, 2016.
- [113] Matthew Kroh and Sricharan Chalikonda. Essentials of robotic surgery. *Springer*, 2015.
- [114] Suha Kwak, Woonhyun Nam, Bohyung Han, and Joon Hee Han. Learning occlusion with likelihoods for visual tracking. In *Proceedings. Eighth IEEE International Conference on Computer Vision*, pages 1551–1558, 2011.
- [115] J. Kybic, M. Unser, and J. Marques. Multidimensional elastic registration of images using splines. *International Conference on Image Processing (ICIP)*, pages 455–458, 2000.
- [116] William W Lau, Nicholas A Ramey, Jason J Corso, Nitish V Thakor, and Gregory D Hager. Stereo-based endoscopic tracking of cardiac surface deformation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 494–501. Springer, 2004.
- [117] C. Le Guyader, D. Apprato, and C. Gout. On the construction of topology-preserving deformation fields. *Image Processing, IEEE Transactions on*, 21(4):1587–1599, 2012.
- [118] M.J. Ledesma-Carbayo, J. Kybic, M. Desco, A. Santos, M. Sühling, P. Hunziker, and M Unser. Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *Medical Imaging, IEEE Transactions on*, 24(9):1113–1126, 2005.
- [119] Massimo Lemma, Andrea Mangini, Alberto Redaelli, and Fabio Acocella. Do cardiac stabilizers really stabilize? experimental quantitative analysis of mechanical stabilization. *Interactive CardioVascular and Thoracic Surgery*, pages 222–226, 2005.
- [120] Charles Lenay, Olivier Gapenne, Sylvain Hanneton, Catherine Marque, and Christelle Geouelle. Sensory substitution: limits and perspectives. *Touching for Knowing, Cognitive psychology of haptic manual perception*, pages 275–292, 2013.
- [121] Thomas Sean Lendvay, Blake Hannaford, and Richard M Satava. Future of robotic surgery. *The Cancer Journal Vol. 19, No. 2*, pages 109–119, 2013.
- [122] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 1944.

-
- [123] Annan Li, Min Lin, Yi Wu, Ming-Hsuan Yang, and Shuicheng Yan. Nus-pro: A new visual tracking challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):335–349, 2016.
- [124] Min Li, Hongbin Liu, Jichun Li, L.D. Seneviratne, and K. Althoefer. Tissue stiffness simulation and abnormality localization using pseudo-haptic feedback. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5359–5364, 2012.
- [125] Yunpeng Ling, Liming Bao, Wei Yang, Yu Chen, and Qing Gao. Minimally invasive direct coronary artery bypass grafting with an improved rib spreader and a new-shaped cardiac stabilizer: results of 200 consecutive cases in a single institution. *BMC cardiovascular disorders*, 16(1):42, 2016.
- [126] Benny Lo, Adrian J Chung, Danail Stoyanov, George Mylonas, and Guang-Zhong Yang. Real-time intra-operative 3d tissue deformation recovery. In *IEEE International Symposium on Biomedical Imaging*, pages 1387–1390, 2008.
- [127] R.G. Lopata, M.M. Nillesen, J.M. Thijssen, L. Kapusta, and C.L. de Korte. Three-dimensional cardiac strain imaging in healthy children using rf-data. *Ultrasound in Medicine & Biology*, 9:1399 – 1408, 2011.
- [128] A. Lopez-Perez, R. Sebastian, and J.M. Ferrero. Three-dimensional cardiac computational modelling: methods, features and applications. *BioMedical Engineering OnLine*, pages 1–31, 2015.
- [129] LokMing Lui and TszChing Ng. A splitting method for diffeomorphism optimization problem using beltrami coefficients. *Journal of Scientific Computing*, 63(2):573–611, 2015.
- [130] Mohsen Mahvash and Allison Okamura. Friction compensation for enhancing transparency of a teleoperator with compliant transmission. *IEEE Transactions on Robotics*, 23(6):1240–1246, 2007.
- [131] A. Mang and G. Biros. Constrained h1 regularization schemes for diffeomorphic image registration. *arXiv:1503.00757 [math.OA]*, pages 1–29, 2015.
- [132] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, pages 431–441, 1963.
- [133] Michael J Massimino and Thomas B Sheridan. Sensory substitution for force feedback in teleoperation. *Analysis, Design and Evaluation of Man-Machine Systems*, 1992.
- [134] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

- [135] Stephen McKinley, Animesh Garg, Siddarth Sen, Rishi Kapadia, Adithyavairavan Murali, Kirk Nichols, Susan Lim, Sachin Patil, Pieter Abbeel, Allison M Okamura, and K. Goldberg. A single-use haptic palpation probe for locating subcutaneous blood vessels in robot-assisted minimally invasive surgery. *IEEE International Conference on Automation Science and Engineering CASE*, 2015.
- [136] Giuseppe Meccariello, Federico Faedi, Saleh AlGhamdi, Filippo Montevocchi, Elisabetta Firinu, Claudia Zanotti, Davide Cavaliere, Roberta Gunelli, Marco Turchini, Andrea Amadori, and Claudio Vicini. An experimental study about haptic feedback in robotic surgery: may visual feedback substitute tactile feedback? *Journal of Robotic Surgery*, 10(1): 57–61, 2016.
- [137] Medtronic. Beating heart technologies. <http://www.medtronic.com>. 2017.
- [138] C.T. Metz, S. Klein, M. Schaap, T. Van Walsum, and W.J. Niessen. Nonrigid registration of dynamic medical imaging data using nd + t b-splines and a groupwise optimization approach. *Medical Image Analysis*, (15):238 – 249, 2011.
- [139] F. Milletari, Ahmadi S.A., D. Kroll, C. Hennersperger, F. F Tombari, A. Shah, A. Plate, K. Boetzel, and N. Navab. Robust segmentation of various anatomies in 3d ultrasound using hough forests and learned data representations. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (15):111–118, 2015.
- [140] D.J. Mirota, M. Ishii, and G.D. Hager. Vision-based navigation in image-guide interventions. *Ann. Rev. Biomedical Engineering*, pages 297–319, 2013.
- [141] Kwang Moo Yi, Kimin Yun, Soo Wan Kim, Hyung Jin Chang, and Jin Young Choi. Detection of moving objects with non-stationary cameras in 5.8 ms: Bringing motion detection to your mobile device. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 27–34, 2013.
- [142] R. Morin, A. Basarab, S. Bidon, and D. Kouamé. Motion estimation-based image enhancement in ultrasound imaging. *Elsevier Journal in Ultrasonics*, (60):19–26, 2015.
- [143] Peter Mountney, Danail Stoyanov, and Guang-Zhong Yang. Three-dimensional tissue deformation recovery and tracking: Introducing techniques based on laparoscopic or endoscopic images. *IEEE Signal Processing Magazine*, pages 14–24, 2010.
- [144] Peter Mountney, Danail Stoyanov, and Guang-Zhong Yang. Three-dimensional tissue deformation recovery and tracking. *IEEE Signal Processing Magazine*, 27(4):14–24, 2010.
- [145] Ahmad Mozaffari, Saeed Behzadipour, and Mehdi Kohani. Identifying the tool-tissue force in robotic laparoscopic surgery using neuro-evolutionary

- fuzzy systems and a synchronous self-learning hyper level supervisor. *Elsevier Journal in Applied Soft Computing*, pages 1278–1283, 2014.
- [146] Matthew C. Murphy, Amy C. Nau, Christopher Fisher, Seong-Gi Kim, Joel S. Schuman, and Kevin C. Chan. Top-down influence on the visual cortex of the blind during sensory substitution. *NeuroImage*, 125:932–940, 2016.
- [147] O. Musse, F. Heitz, and J.P. Armpach. Topology preserving deformable image matching using constrained hierarchical parametric models. *Image Processing, IEEE Transactions on*, pages 1081–1093, 2001.
- [148] Saskia K. Nagel, Christine Carl, Tobias Kringe, Robert Märtin, and Peter König. Beyond sensory substitution—learning the sixth sense. *Journal of Neural Engineering*, 2(4), 2005.
- [149] Yoshihiko Nakamura, Kousuke Kishi, and Hiro Kawakami. Heartbeat synchronization for robotic cardiac surgery. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 2014–2019. IEEE, 2001.
- [150] C. Nappi, W. Acampa, T. Pellegrino, M. Petretta, and A. Cuocolo. Beyond ultrasound: advances in multimodality cardiac imaging. *Journal in Internal and Emergency Medicine*, 10:9–20, 2015.
- [151] S. Nekolla, C. Rischpler, and K. Kunze. Pet/mri for cardiac imaging: Technical considerations and potential applications. *Book Chapter in Molecular and Multimodality Imaging in Cardiovascular Disease*, pages 29–48, 2015.
- [152] Aaron Netz and Margarita Osadchy. Recognition using specular highlights. *IEEE transactions on pattern analysis and machine intelligence*, 35(3): 639–652, 2013.
- [153] A. Newson, M. Tepper, and G. Sapiro. Matrix completion with noise. In *British Machine Vision Conference, BMVC*, pages 1–12, 2015.
- [154] Andrew Y Ng. Feature selection, l_1 vs. l_2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78, 2004.
- [155] S.I. Nikolov and J.A. Jensen. Virtual ultrasound sources in highresolution ultrasound imaging. in *Proc. SPIE vol. 3*, pages 395—405, 2002.
- [156] M. Nillesen, A. Saris, H. Hansen, S. Fekkes, F.J. Van Slochteren, P. Bovendeerd, and C. De-Korte. Cardiac motion estimation using ultrafast ultrasound imaging tested in a finite element model of cardiac mechanics. *Functional Imaging and Modeling of the Heart*, pages 207–214, 2015.
- [157] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armpach. 3-d deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *Image Processing, IEEE Transactions on*, pages 553–566, 2005.

- [158] Ehsan Noohi, Sina Parastegari, and Miloš Žefran. Using monocular images to estimate interaction forces during minimally invasive surgery. *IEEE International Conference on Intelligent Robots and Systems*, pages 4297–4302, 2014.
- [159] Irina Nurutdinova and Andrew Fitzgibbon. Towards pointless structure from motion: 3d reconstruction and camera parameters from general 3d curves. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2363–2371, 2015.
- [160] Shogo Okamoto, Masashi Konyo, and Satoshi Tadokoro. Vibrotactile stimuli applied to finger pads as biases for perceived inertial and viscous loads. *IEEE Transactions on Haptics*, 4(4):307–315, 2011.
- [161] Allison M. Okamura, Lawton N. Verner, Tomonori Yamamoto, James C. Gwilliam, and Paul G. Griffiths. *Force Feedback and Sensory Substitution for Robot-Assisted Surgery*, pages 419–448. Springer US, 2011.
- [162] A.M. Okamura, L.N. Verner, C.E. Reiley, and M. Mahvash. Haptic feedback in robot-assisted minimally invasive surgery. *Book Chapter in Robotics Research, Springer Tracts in Advanced Robotics*, pages 361–372, 2011.
- [163] O. Oktay, A. Schuh, M. Rajchl, K. Keraudren, A. Gomez, M.P. Heinrich, G. Penney, and D. Rueckert. Structured decision forests for multi-modal ultrasound image registration. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (15):363–371, 2015.
- [164] Tobias Ortmaier, Martin Groger, Dieter H Boehm, Volkmar Falk, and Gerd Hirzinger. Motion estimation in beating heart surgery. *IEEE Transactions on Biomedical Engineering*, 52(10):1729–1740, 2005.
- [165] B.F. Osmanski, C. Martin, G. Montaldo, P. Lanière, F. Pain, M. Tanter, and H. Gurden. Functional ultrasound imaging reveals different odor-evoked patterns of vascular activity in the main olfactory bulb and the anterior piriform cortex. *NeuroImage 95*, pages 176–184, 2014.
- [166] K. Owen, M.I. Fuller, and J.A. Hossack. Application of x-y separable 2-d array beamforming for increased frame rate and energy efficiency in handheld devices. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 59, no. 7, pages 1332–1343, 2012.
- [167] C. Pacchierotti, L. Meli, F. Chinello, M. Malvezzi, and D. Prattichizzo. Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems. *The International Journal of Robotics Research*, 34(14):1773–1787, 2015.
- [168] Claudio Pacchierotti. *Cutaneous Haptic Feedback in Robotic Teleoperation*. 2192-2977. Springer International Publishing, 2015.
- [169] Claudio Pacchierotti, Asad Tirmizi, and Domenico Prattichizzo. Improving transparency in teleoperation by means of cutaneous tactile force feedback. *ACM Transactions on Applied Perception*, 2014.

- [170] C. Papadacci, M. Pernot, M. Couade, M. Fink, and M. Tanter. High-contrast ultrafast imaging of the heart. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, pages 288—301, 2014.
- [171] X. Papademetris, A.J. Sinusas, D.P. Dione, and J.S. Duncan. 3d cardiac deformation from ultrasound images. *Medical Image Computing and Computer-Assisted Intervention*, 1999.
- [172] Elizabeth W Paxton, Robert S Namba, Gregory B Maletis, Monti Khatod, Eric J Yue, Mark Davies, Richard B Low, Ronald WB Wyatt, Maria CS Inacio, and T Ted Funahashi. A prospective study of 80,000 total joint and 5000 anterior cruciate ligament reconstruction procedures in a community-based registry in the united states. *J Bone Joint Surg Am*, 92(Supplement 2):117–132, 2010.
- [173] Gary Perlman. Electronic surveys. *Behavior Research Methods, Instruments, & Computers*, 17(2):203–205, 1985.
- [174] Matteo Pettinari, Emiliano Navarra, Philippe Noirhomme, and Herbert Gutermann. The state of robotic cardiac surgery in europe. *Annals of Cardiothoracic Surgery*, 6(1):1–8, 2017.
- [175] M. Pfingsthorn and A. Birk. Generalized graph slam: Solving local and global ambiguities through multimodal and hyperedge constraints. *The International Journal of Robotics Research*, 2016.
- [176] Srinivas K. Prasad, Masaya Kitagawa, Gregory S. Fischer, Jason Zand, Mark A. Talamini, Russell H. Taylor, and Allison M. Okamura. A modular 2-dof force-sensing instrument for laparoscopic surgery. *Lecture Notes in Computer Science*, 2878:279–286, 2003.
- [177] J Provost, V.T.-H. Nguyen, D Legrand, S. Okrasinski, A. Costet, A. Gambhir, H. Garan, and E.E. Konofagou. Electromechanical wave imaging for arrhythmias. *Physics in Medicine and Biology* 56, pages 1–11, 2011.
- [178] J. Provost, C. Papadacci, J.E. Arango, M. Imbault, M. Fink, J.L. Gennisson, M. Tanter, and M. Pernot. 3d ultrafast ultrasound imaging in vivo. *Physics in Medicine and Biology*, 59(19), 2014.
- [179] M. Prummer, J. Hornegger, G. Lauritsch, L. Wigstrom, E. Girard-Hughes, and R. Fahrig. Cardiac c-arm ct: A unified framework for motion estimation and dynamic ct. *IEEE Transactions on Medical Imaging*, pages 1836–1849, 2009.
- [180] Josef P. Rauschecker. Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends in Neurosciences*, 18(1):36–43, 1995.
- [181] Carol E. Reiley, Takintope Akinbiyi, Darius Burschka, David C. Chang, Allison M. Okamura, and David D. Yuh. Effects of visual force feedback on robot-assisted surgical task performance. *The Journal of Thoracic and Cardiovascular Surgery*, 135(1):196 – 202, 2008.

- [182] Laurent Renier and Anne G. De Volder. Cognitive and brain mechanisms in sensory substitution of vision: A contribution to the study of human perception. *Journal of Integrative Neuroscience*, 04(04):489–503, 2005.
- [183] Rogério Richa, Philippe Poignet, and Chao Liu. Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model. *The International Journal of Robotics Research*, 29(2-3):218–230, 2010.
- [184] Ellen T Roche, Robert Wohlfarth, Johannes TB Overvelde, Nikolay V Vasilyev, Frank A Pigula, David J Mooney, Katia Bertoldi, and Conor J Walsh. A bioinspired soft actuated material. *Advanced Materials*, pages 1–7, 2013.
- [185] Giulio Rognini and Olaf Blanke. Cognetics: Robotic interfaces for the conscious mind. *Trends in Cognitive Sciences*, 20(3):162 – 164, 2016. ISSN 1364-6613.
- [186] T. Rohlfing, C. Maurer, D. Bluemke, and M. Jacobs. olume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, pages 730–741, 2003.
- [187] Robert W. Root and Steve Draper. Questionnaires as a software evaluation tool. In *ACM Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 83–87, 1983.
- [188] J. Royuela-del Val, L. Cordero-Grande, F. Simmross-Wattenberg, M. Martín-Fernández, and C. Alberola-López. Nonrigid groupwise registration for motion estimation and compensation in compressed sensing reconstruction of breath-hold cardiac cine mri. *Magnetic Resonance in Medicine*, 2015.
- [189] L. Rubbert, P. Renaud, W. Bacht, and J. Gangloff. Compliant mechanisms for an active cardiac stabilizer: lessons and new requirements in the design of a novel surgical tool. *Mechanical Sciences*, 2(1):119–127, 2011.
- [190] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4): 259–268, 1992.
- [191] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.
- [192] Ozere S. and C. Le Guyader. Topology preservation for image-registration-related deformation fields. *COMMUN. MATH. SCI.*, 13(5):1135–1161, 2015.
- [193] S. Salles, A.J.Y. Chee, D. Garcia, A.C.H. Yu, D. Vray, and H. Liebgott. 2-d arterial wall motion imaging using ultrafast ultrasound and transverse oscillations. *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, 62(6):1047–1058, 2015.

- [194] C. Santiago, J.C. Nascimento, and J.S. Marques. Automatic 3-d segmentation of endocardial border of the left ventricle from ultrasound images. *IEEE Journal of Biomedical and Health Informatics*, (42):399–348, 2015.
- [195] Mickaël Sauvée, Aurélien Noce, Philippe Poignet, Jean Triboulet, and Etienne Dombre. 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, 2007.
- [196] Ryan E. Schoonmaker and Caroline G.L. Cao. Vibrotactile feedback enhances force perception in minimally invasive surgery. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 1029–1033, 1992.
- [197] T. Schubert, K. Eggenesperger, A. Gkogkidis, F. Hutter, T. Ball, and W. Burgard. Automatic bone parameter estimation for skeleton tracking in optical motion capture. *IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2016.
- [198] Mohsen Shahinpoor and Siavash Gheshmi. Introduction to surgical robots' general configurations. *Book Chapter, Robotic Surgery - Smart Materials, Robotic Structures, and Artificial Muscles*, Taylor & Francis Group, 2015.
- [199] James Shanteau. *The Psychology of Experts An Alternative View*, pages 11–23. Springer US, 1992.
- [200] D. Shattuck, M. Weinshenker, S. Smith, and O. von Ramm. Explososcan: A parallel processing technique for high speed ultrasound imaging with linear phased arrays. *Journal of the Acoustical Society of America vol. 75, no. 4*, pages 1273—1282, 1984.
- [201] Ben Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India, 2010.
- [202] K. Shung. *Ultrasound: Imaging and blood flow measurements*, second edition. Taylor and Francis Group, 2015.
- [203] W. Slaughter. *The linearized theory of elasticity*. Springer Science+Business Media, LLC, 2002.
- [204] Pilar Sobrevilla and Eduard Montseny. Fuzzy sets in computer vision: An overview. *Mathware and Soft Computing*, 10:71–83, 2003.
- [205] S. Sokhanvar, J. Dargahi, S. Najarian, and S. Arbatani. Clinical and regulatory challenges for medical devices tactile sensing and displays. *Haptic Feedback for Minimally Invasive Surgery and Robotics*, Wiley Publications, 2012.
- [206] Nathaniel J Soper, Lee L Swanström, and Steve Eubanks. *Mastery of endoscopic and laparoscopic surgery*. Lippincott Williams & Wilkins, 2008.
- [207] K. Spencer. Focused cardiac ultrasound. *Book Chaper in ASE's Comprehensive Echocardiography*, 2015.

- [208] Giuseppe Spinoglio, Alessandra Marano, and Giampaolo Formisano. Robotic surgery: Current applications and new trends. *Springer*, 2015.
- [209] Charles V. Stewart. Robust parameter estimation in computer vision. *SIAM Rev.*, 41(3), 1999.
- [210] Danail Stoyanov, George P Mylonas, Fani Deligianni, Ara Darzi, and Guang Zhong Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 139–146. Springer, 2005.
- [211] Danail Stoyanov, Marco Visentini Scarzanella, Philip Pratt, and Guang-Zhong Yang. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 275–282. Springer, 2010.
- [212] Minghui Sun, Xiangshi Ren, and Xiang Cao. Effects of multimodal error feedback on human performance in steering tasks. *Journal of Information Processing*, 18:284–292, 2011.
- [213] M. Tanter. The advent of ultrafast imaging in biomedical ultrasound. *Med. Phys.*, (42), 2015.
- [214] M. Tanter and M. Fink. Ultrafast imaging in biomedical ultrasound. *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, 61(1):102–119, 2014.
- [215] M. Tanter, J. Bercoff, A. Athanasiou, T. Deffieux, J-L Gennisson, G. Montaldo, M. Muller, A. Tardivon, and M. Fink. Quantitative assessment of breast lesion viscoelasticity: initial clinical results using supersonic shear imaging. *Ultrasound in Medicine and Biology* 34, pages 1373—1386, 2014.
- [216] Graham W Taylor and Geoffrey E Hinton. Factored conditional restricted boltzmann machines for modeling motion style. In *Proceedings of the 26th annual international conference on machine learning*, pages 1025–1032. ACM, 2009.
- [217] Graham W Taylor, Geoffrey E Hinton, and Sam T Roweis. Two distributed-state models for generating high-dimensional time series. *Journal of Machine Learning Research*, 12:1025–1068, 2011.
- [218] D. Tenbrinck, S. Schmid, Jiang X., Schäfers K., and J. Stypmann. Histogram-based optical flow for motion estimation in ultrasound imaging. *Journal Math Imaging Vis*, pages 138–150, 2013.
- [219] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

- [220] Lorenzo Torresani, Aaron Hertzmann, and Chris Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE transactions on pattern analysis and machine intelligence*, 30(5):878–892, 2008.
- [221] Pinyo uangmali, Hongbin Liu, Lakmal D Seneviratne, Prokar Dasgupta, and Kaspar Althoefer. Miniature 3-axis distal force sensor for minimally invasive surgical palpation. *IEEE Transactions on Mechatronics*, pages 646–656, 2012.
- [222] Michael Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, vol.16, no.6, pages 22–38, 1999.
- [223] O.A.J. Van der Meijden and M.P. Schijven. The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review. *Surgical Endoscopy*, pages 1180–1190, 2009.
- [224] S. Vyas, J.S. Gammie, and P. Burlina. Computing cardiac strain from variational optical flow in four-dimensional echocardiography. *2014 IEEE 27th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 149–152, 2014.
- [225] C.R. Wagner, N. Stylopoulos, and R.D. Howe. The role of force feedback in surgery: analysis of blunt dissection. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, pages 68–74, 2002.
- [226] Jack M Wang, David J Fleet, and Aaron Hertzmann. Gaussian process dynamical models for human motion. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):283–298, 2008.
- [227] L. Wang, A. Basarab, P.R. Girard, P. Croisille, P. Clarysse, and Delachartre P. Analytic signal phase-based myocardial motion estimation in tagged {MRI} sequences by a bilinear model and motion compensation. *Medical Image Analysis*, pages 149–162, 2015.
- [228] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., 2004.
- [229] P. J. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, pages 1550–1560.
- [230] P. J. Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Harvard University*, pages 1550–1560, 1974.
- [231] Christopher D Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.
- [232] R. Williams and D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, pages 270–280, 1989.

- [233] Erik B Wilson, Hossein Bagshahi, and Vicky D Woodruff. Overview of general advantages, limitations, and strategies. *Book Chapter, Robotics in General Surgery, Springer*, 2014.
- [234] Wai-Keung Wong, Bo Yang, Chao Liu, and Philippe Poignet. A quasi-spherical triangle-based approach for efficient 3-d soft-tissue motion tracking. *IEEE/ASME Transactions on Mechatronics*, 18(5):1472–1484, 2013.
- [235] Bo Yang, Wai-Keung Wong, Chao Liu, and Philippe Poignet. 3d soft-tissue tracking using spatial-color joint probability distribution and thin-plate spline model. *Pattern recognition*, 47(9):2962–2973, 2014.
- [236] Bo Yang, Chao Liu, Wenfeng Zheng, and Shan Liu. Motion prediction via online instantaneous frequency estimation for vision-based beating heart tracking. *Information Fusion*, 35:58–67, 2017.
- [237] I. Yanovsky, P.M. Thompson, S. Osher, and A.D. Leow. Topology preserving log-unbiased nonlinear image registration: Theory and implementation. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.
- [238] Michael C Yip, Shelten G Yuen, and Robert D Howe. A robust uniaxial force sensor for minimally invasive surgery. *IEEE Transactions on Biomedical Engineering, Vol. 57, No. 5*, pages 1008–1011, 2010.
- [239] Shelten G Yuen, Daniel T Kettler, Paul M Novotny, Richard D Plowes, and Robert D Howe. Robotic motion compensation for beating heart intracardiac surgery. *The International journal of robotics research*, 28(10): 1355–1372, 2009.
- [240] Shelten G Yuen, Daniel T Kettler, Paul M Novotny, Richard D Plowes, and Robert D Howe. Robotic motion compensation for beating heart intracardiac surgery. *The International journal of robotics research*, 28(10): 1355–1372, 2009.
- [241] Matthew D Zeiler, Graham W Taylor, Nikolaus F Troje, and Geoffrey E Hinton. Modeling pigeon behavior using a conditional restricted boltzmann machine. In *ESANN*, 2009.
- [242] Z. Zhang, D. Sahn, and X. Song. Diffeomorphic cardiac motion estimation with anisotropic regularization along myofiber orientation. *WBIR 2012. LNCS*, 2012.
- [243] Z. Zhang, M. Ashraf, D.J. Sahn, and X. Song. Temporally diffeomorphic cardiac motion estimation from three-dimensional echocardiography by minimization of intensity consistency error. *Medical Physics*, pages 1–16, 2014.



Mathematical Proofs

Proof 1 (*Singular Value Decomposition, SVD*) Consider the $n \times n$ matrix $\mathbf{A}^T \mathbf{A}$. It is symmetric and positive semidefinite, and therefore its eigenvalues are all nonnegative. By the spectral theorem $\mathbf{A}^T \mathbf{A}$ admits an eigenvalue decomposition $\mathbf{A}^T \mathbf{A} = V \Lambda V^T$ where V is a $n \times n$ matrix with orthonormal basis at the columns and a diagonal matrix $\Lambda = (\lambda_1, \dots, \lambda_r, 0, \dots, 0)$. Let U be a orthogonal matrix of $m \times m$ where

$$u_i = \frac{1}{\sigma_i A v_i} \text{ for } i = 1, \dots, r \quad (\text{A.1})$$

By orthogonalization procedure, the set of vectors u_{r+1}, \dots, u_m form an orthogonal matrix $U = (u_1, \dots, u_m) \in \mathbf{R}^m$. Then by showing that $U^T A V^T = \tilde{S} := \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$ by

$$(U^T A V)_{ij} = u_i^T \mathbf{A} v_j = \begin{cases} \sigma_j u_i^T u_j & \text{if } j \leq r \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.2})$$

then can be claimed the proof of Theorem 1.

Proof 2 (*Eckart-Young*) Having $U^T A_k V = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$ it follows that \mathbf{A}_k is rank k . Consider $\mathbf{B} = k$ for some $\mathbf{B} \in \mathbf{R}^{m \times n}$. It can find orthornomal

vectors x_1, \dots, x_{n-k} so $\text{null}(\mathbf{B}) = \text{span}x_1, \dots, x_{n-k}$. A dimension argument shows that

$$\text{span}x_1, \dots, x_{n-k} \cap \text{span}v_1, \dots, v_{k+1} \neq 0 \quad (\text{A.3})$$

Let z be a unit 2-norm vector in this intersection. Since $\mathbf{B}z = 0$ and

$$\mathbf{A}z = \sum_{i=1}^{k+1} \sigma_i(v_i^T z)u_i \quad (\text{A.4})$$

it results

$$\|A - B\|_2^2 \geq \|(A - B)_z\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2(v_i^T z)^2 \geq \sigma_{k+1}^2 \quad (\text{A.5})$$

B

Estimation Theory

Extended Kalman Filter (EKF)

Consider the following non-linear dynamic system:

$$\begin{aligned} X_t &= f_t(X_{t-1}, W_t), \\ Y_t &= h_t(X_t, V_t), \end{aligned} \tag{B.1}$$

where X_t is the vector of *system state*, Y_t is the vector of *observation*, and W_t and V_t are the process and observation noises, with covariance Q_t and R_t respectively. The functions f_t and h_t are nonlinear and differentiable. The Extended Kalman Filter (EKF) calculates an approximation of the conditional expectation $\hat{x}_{t|t} = \mathbb{E}[X_t|Y_{1:t}]$ by an appropriate linearization of the state transition and observation models. The following is a summary of the EKF algorithm:

Initialization:

$$\hat{x}_{0|0} = \mathbb{E}[X_0], P_{0|0} = \text{cov}(X_0). \tag{B.2}$$

Prediction:

$$\hat{x}_{t|t-1} = f(\hat{x}_{t-1|t-1}, 0), \tag{B.3}$$

$$\tilde{Y}_t = Y_t - h(\hat{x}_{t|t-1}, 0), \tag{B.4}$$

$$P_{t|t-1} = F_X P_{t-1|t-1} F_X^T + F_W Q_t F_W^T. \quad (\text{B.5})$$

Update:

$$S_t = H_X P_{t|t-1} H_X^T + H_V R_t H_V^T, \quad (\text{B.6})$$

$$K_t = P_{t|t-1} H_X^T S_t^{-1}, \quad (\text{B.7})$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \tilde{Y}_t, \quad (\text{B.8})$$

$$P_{t|t} = (I - K_t H_X) P_{t-1|t-1}. \quad (\text{B.9})$$

where $F_X = \partial_X f(\hat{x}_{t-1|t-1}, 0)$ and $F_W = \partial_W f(\hat{x}_{t-1|t-1}, 0)$ are the partial derivatives of f (with respect to the system state and the process noise) evaluated at $(\hat{x}_{t-1|t-1}, 0)$ and $H_X = \partial_X h(\hat{x}_{t|t-1}, 0)$ and $H_V = \partial_V h(\hat{x}_{t|t-1}, 0)$ are the partial derivatives of h (with respect to the system state and the observation noise) evaluated at $(\hat{x}_{t|t-1}, 0)$.

Nonlinear ARX Model with multiple steps prediction

The Nonlinear Auto-Regressive eXogenous (NLARX) model for the measured system input and output $y(t)$ and $u(t)$ at discrete time instants $t = 1, 2, 3, \dots$ is described as:

$$\begin{aligned} y(t) = f[y(t-1), \dots, y(t-n_a), \\ u(t-n_k), \dots, u(t-n_k-n_b+1)] + e(t) \end{aligned} \quad (\text{B.10})$$

where n_a and n_b are the number of the model's past outputs and inputs, n_k is the pure input delay, f is a nonlinear functions, and $e(t)$ represents the modeling error.

The estimation of the NLARX model is typically achieved by minimizing some criterion based on the error sequence $e(t)$ with respect to some parameterization vector of the nonlinear function f . While the one step prediction of the system output at time instant t from past measured outputs is obtained by omitting the error $e(t)$ as follows:

$$\begin{aligned} \hat{y}(t) = f[y(t-1), \dots, y(t-n_a), \\ u(t-n_k), \dots, u(t-n_k-n_b+1)] \end{aligned} \quad (\text{B.11})$$

For multiple steps prediction, the same formula is recursively applied as follows. Let $\hat{y}^{(k)}(t)$ denote the k step prediction of the output $y(t)$. For $k = 1$,

the one step prediction $\hat{y}^{(1)}(t) \triangleq \hat{y}(t)$ is applied as defined in B.11. Then for $k = 2, 3, 4, \dots$, the k step prediction is recursively computed by:

$$\begin{aligned} \hat{y}^{(k)}(t) = f[\hat{y}^{(k-1)}(t-1), \dots, \hat{y}^{(k-1)}(t-n_a), \\ u(t-n_k), \dots, u(t-n_k-n_b+1)] \end{aligned} \quad (\text{B.12})$$

where the sequence $\hat{y}^{(1)}(t)$ (for $t = 1, 2, 3, \dots$) must be computed first, then $\hat{y}^{(2)}(t)$, then $\hat{y}^{(3)}(t)$, and so on.

