Computer Aided Medical Procedures
Prof. Dr. Nassir Navab

Dissertation

Multi-modal Registration and Robotic Imaging for Computer Assisted Surgery

Bernhard Fürst

Fakultät für Informatik
Technische Universität München

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin

Multi-modal Registration and Robotic Imaging for Computer Assisted Surgery

Bernhard Fürst

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r):	Prof. Dr.-Ing. Darius Burschka
Prüfer der Dissertation:	Prof. Nassir Navab, Ph.D.
	Assist. Prof. Greg Michael Osgood, M.D. The Johns Hopkins University

Die Dissertation wurde am 25.07.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 03.10.2016 angenommen.

Abstract

Imaging and visualization of anatomy and physiology are the very foundation of Computer Assisted Surgery, and crucially contribute to the success of interventions. The goal of the research presented in this thesis is the introduction of novel image registration techniques as means of transferring pre-operative imaging and planning to the surgical scenario, and intra-operative image acquisition to obtain images just in time and within the surgical site.

The contribution to image registration include the prediction of deformation, by generating a patient-specific anatomical biomechanical model for which spatial and temporal varying physiological properties are estimated based on observations from 4D Computed Tomography (CT) volumes.

Another contribution to image registration are similarity measures for multimodal image registration are proposed and evaluated. Linear Correlation for Linear Combinations (LC^2) are used to enable the alignment of images from modalities that measure different physical properties. This allows the combination of imaging data containing complementing information, but the alignment fails when the initial alignment is poor.

The second part of this thesis focuses on the interventional imaging, which is of special importance when large deformations render pre-operative data of little use, real-time imaging is required for image-guided interventions, or when imaging is used to simplify surgeries. One of the contributions is the introduction of minimally invasive robotic Single-Photon Emission Computed Tomography (SPECT) imaging for robot-assisted surgery, for which vision based tracking of a miniaturized gamma probe enables the SPECT reconstruction using Maximum Likelihood Expectation Maximization (MLEM).

Finally, this thesis concludes with the presentation of augmented reality application for to interventional imaging during orthopedic surgeries. The technical contributions aim at fusing Cone-Beam Computed Tomography (CBCT) volumes and live 3D color images. The usability of this novel concept is evaluated in a multi-user study using objective and subjective measures.

Zusammenfassung

Medizinische Bildgebung und Visualisierung von Anatomie und Physiologie sind die wesentlichen Fundamente der computerassistierten Chirurgie, und stellen einen unabdinglichen Beitrag zum Erfolg der Intervention dar. Das Forschungsziel dieser Doktoratarbeit ist die Verbesserung der Registrierung von medizinischen Bildern um präoperativen Bildinformationen während einer Intervention zur Verfügung zu stellen und die intraoperative Aufzeichung von Echtzeitbildinformationen zur Darstellung in Relation zum operativen Feld.

Die wissenschaftlichen Beiträge zur Bildregistrierung beinhalten unter anderem die Vorhersage der Gewebsdeformation mittels einem patientenspezifischen, anatomischen, biomechanischen Modell, welches die räumliche und zeitliche Veränderung der physiologischen Eigenschaften durch Beobachtungen von 4D-Computertomographie (CT) approximiert.

Ein weiterer Beitrag zur Bildregistrierung sind Ähnlichkeitsmaße für multimodale Bildregistrierung. In erster Line wird die Lineare Korrelation von Linearkombinationen verwendet, um den Abgleich von Bildern unterschiedlicher Modalitäten, welche verschiedene physikalische Eigenschaften messen, zu ermöglichen. Dies ermöglicht die Kombination von Bilddaten, die ergänzende Informationen beinhalten.

Der zweite Teil dieser Doktorarbeit beschäftigt sich mit interventioneller Bildgebung, welche von spezieller Bedeutung ist, wenn die Deformierung von Gewebe die Brauchbarkeit der prä-operativen Bildgebung einschränkt, Echtzeitbildgebung notwendig ist, oder wenn interventionelle Bildgebung den medizinischen Eingriff vereinfachen kann. Erstmalig wird minimal invasive, robotergestützte Einzelphotonen-Emissionscomputertomographie (en: SPECT) Bildgebung fuer robotergestützte Chirurgie präsentiert. Basierend auf optischen Bildern wird ein miniaturisierter Gammadetektor verfolgt, was eine SPECT Rekonstruktion ermöglicht.

Diese Doktorarbeit schließt ab mit der Präsentation von Erweiterter Realität für interventionelle Bildgebung für orthopädische Eingriffe. Die technischen Beiträge umfassen unter anderem die Fusion von Primärstrahlenkegelcomputertomografie und Echtzeit-3D Farbbildern. Die Umsetzbarkeit dieser neuartigen Technologie ist durch eine Nutzerstudie anhand objektiv quantifizierbaren und subjektiven Merkmalen evaluiert.

Acknowledgments

First and foremost, I would like to express my sincere appreciation and gratitude to Prof. Dr. Nassir Navab, who has provided continuous support and encouragement for my research in collaboration with the Computer Aided Medical Procedures group, not only at the Technical University in Munich,Germany, but also at Johns Hopkins University in Baltimore, Maryland, USA. To me, this is a great experience and undertaking, for which I am thankful of being part of.

Over the past years I have had the privilege of experiencing and enjoying Prof. Navab's affirmative academic supervision, but it must also be mentioned that his distinguished and yet unassuming manner has been exemplary for me and an inspiration for the further development of my own personality.

Moreover, I would like to thank my colleagues Javad Fotouhi, Christoph Hennersperger, Benjamin Frisch, Oliver Zettinig, Risto Kojcev, Wolfgang Wein, Richard Brosig, Kanishka Sharma, Ali Kamen, Amit Shah, Tommaso Mansi, Alison Morrow, and Martina Hilla, for their assistance and continuous support, and this gratitude also extends to all current and recent team members with whom I have had the privilege to work with. Please allow me to express my appreciation to Greg Osgood, Alex Johnson, Mehran Armand, Russ Taylor and Yair Amir for their great efforts and kindness.

And finally, I wish to say thank you to my parents Joyce and Wolfgang, my sister Julia, other family members and friends, for their continuous encouragement and support over the years. Last but not least, I am deeply grateful to my wife Nikola for her unbroken and untiring faith in me.

Contents

List of Abbreviations **5**

1 Introduction **7**
- 1.1 Medical Imaging Basics 8
  - 1.1.1 Anatomical/Structural Imaging 9
  - 1.1.2 Functional Imaging 14
- 1.2 Image-Guided Interventions 15

2 Medical Image Registration **17**
- 2.1 Definition of Registration and Image 17
- 2.2 Registration Process and Components 19
  - 2.2.1 Transformation 19
  - 2.2.2 Interpolator 23
  - 2.2.3 Similarity Measure 24
  - 2.2.4 Optimizer 26
- 2.3 Biomechanical Simulations for Deformation Estimation 27
  - 2.3.1 Related Work 27
  - 2.3.2 Contribution: Patient-Specific Biomechanical Model for the Prediction of Lung Motion From 4-D CT Images (IEEE TMI 2015) 27
- 2.4 Multi-Modal Image Similarity Measure 28
  - 2.4.1 Related Work 28
  - 2.4.2 Contribution: Automatic Ultrasound-MRI Registration for Neurosugery using the 2D and 3D LC^2 Metric (MedIA 2014) 30
- 2.5 Tissue Displacement Compensation for SPECT-guided Surgery . 31
  - 2.5.1 Related Work 31
  - 2.5.2 Contribution: Radiopositive Tissue Displacement Compenstation for SPECT-Guided Surgery (MICCAI 2015) . 31

3 Image-Guided Surgery and Intra-Operative Imaging **35**
- 3.1 Introduction and Definition 35
- 3.2 Robotic Single-Photon Emission Computed Tomography for Robot-Assisted Interventions 38
  - 3.2.1 Related Work 38

3.2.2 Contribution: First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping (IEEE TMI 2016) . . 38
3.3 Robotic Ultrasound Acquisition 41
3.4 RGBDX: Multi-Modal Imaging 44
3.4.1 Standard of Care and Related Work 47
3.4.2 Contribution: Vision-based Intra-Operative Cone-Beam CT Stitching for Non-overlapping Volumes (MICCAI 2015) 48
3.4.3 Contribution: Calibration of RGBD Camera and Cone-Beam CT for 3D Intraoperative Mixed Reality Visualization (IJCARS 2016) 49
3.4.4 Contribution: Pre-Clinical Usability Study of Multiple Augmented Reality Concepts for K-Wire Placement (IJCARS 2016) 50

A Patient-Specific Biomechanical Model for the Prediction of Lung Motion from 4D CT Images (IEEE TMI 2014) 69
A.1 Introduction 70
A.1.1 Clinical Motivation: Thoracic Imaging and Radiotherapy 70
A.1.2 Physiology of the Respiratory System 71
A.1.3 Image-Based Approaches for Motion Estimation 72
A.1.4 Biomechanical-Based Methods for Motion Estimation . . 73
A.1.5 Proposed Solution 73
A.2 Methods 74
A.2.1 Anatomical Model Generation 74
A.2.2 Computational Motion Model 75
A.2.3 Model Personalization 78
A.2.4 Implementation 79
A.3 Experiments and Results 79
A.3.1 Patient Data 79
A.3.2 Computation of Landmark and Surface Errors 79
A.3.3 Numerical Analysis 80
A.3.4 Validation of Motion Prediction on Patient Data 82
A.4 Discussion and Future Works 85

B Automatic US-MRI Registration for Neurosurgery using the 2D and 3D LC^2 Metric (MedIA 2014) 93
B.1 Introduction 94
B.2 Related Work 94
B.3 Method 95
B.3.1 Similarity Measure 95
B.3.2 Similarity Measure Computation 98
B.3.3 Optimization of Rigid Transformation 98
B.4 Experiments 99
B.4.1 Clinical Data and Experimental Setup 99
B.4.2 Numerical Analysis of LC^2 Configuration 99
B.4.3 Registration Results 100
B.4.4 Accuracy, Precision and Capture Range 100
B.4.5 Computation Time 102
B.5 Discussion and Conclusion 103
B.5.1 Choice of Method 103

B.5.2 Performance 103
B.5.3 Future Work 104

C Radiopositive Tissue Displacement Compensation for SPECT-Guided Surgery (MICCAI 2015) 111
C.1 Introduction 112
C.2 Materials & Methods 113
C.2.1 SPECT Imaging & Technical Background 113
C.2.2 2D γ-camera 113
C.2.3 Displacement Compensation 114
C.3 Experiments & Results 115
C.4 Discussion & Conclusion 116

D First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping (IEEE TMI 2015) 121
D.1 Introduction 122
D.1.1 Clinical Background and Motivation 122
D.1.2 Preoperative Imaging of Sentinel Lymph Nodes 123
D.1.3 Interventional Sentinel Lymph Node Mapping 124
D.1.4 Robot-Assisted Lymphadenectomy 125
D.1.5 Objective: Minimally Invasive SPECT Imaging 127
D.2 Materials and Methods 127
D.2.1 System Overview 127
D.2.2 Drop-In Gamma Probe 127
D.2.3 Robotic Surgical System 128
D.2.4 Probe Tracking 128
D.2.5 Robotic In-Patient SPECT Acquisition 129
D.2.6 SPECT Reconstruction 130
D.2.7 Augmented Reality Visualization for Surgical Navigation 130
D.3 Experiments and Results 131
D.3.1 Experimental Setup 131
D.3.2 Data Acquisition 131
D.3.3 Evaluation of Tracking Accuracy 131
D.3.4 SPECT Reconstruction Results 132
D.3.5 Augmented Reality Visualization 132
D.4 Discussion and Conclusion 133

E Vision-based Intraoperative Cone-Beam CT Stitching for Non-overlapping Volumes (MICCAI 2015) 145
E.1 Introduction 146
E.2 Materials and Methods 147
E.2.1 System setup and calibration 147
E.2.2 CBCT Volume and Video Acquisition 149
E.2.3 Two-Dimensional Feature Detection and Matching 149
E.2.4 Recovering Three-Dimensional Coordinates 150
E.2.5 Estimating 3D Transformation and CBCT Volume Stitching 150
E.3 Experiments and Results 150
E.4 Discussion and Conclusion 151

F Pre-Clinical Usability Study of Multiple Augmented Reality Concepts for K-Wire Placement (IJCARS 2016) 157
F.1 Introduction 158
F.2 Method 161
  F.2.1 Imaging Systems 161
  F.2.2 Evaluation Method 163
F.3 Experiments 164
F.4 Results 166
  F.4.1 Statistical Evaluation 168
F.5 Discussion and Conclusion 168

G Calibration of RGBD Camera and Cone-Beam CT for 3D Intraoperative Mixed Reality Visualization (IJCARS 2016) 175
G.1 Introduction 176
G.2 Method 179
  G.2.1 System setup 179
  G.2.2 Calibration phantom design, point cloud extraction and pre-processing 180
  G.2.3 Calibration of C-arm and the RGBD camera 182
  G.2.4 Mixed reality visualization of DRRs overlaid on the patient's surface 183
G.3 Experimental Validation and Results 183
  G.3.1 Mixed reality visualization 185
G.4 Discussion and Conclusion 185

Outline and Publications

In the following, a brief outline of the main chapters in this dissertation is presented.

Chapter 1: Introduction. The crucial importance of medical imaging for surgical treatment, and the techniques to provide image-guidance are presented. First, concepts and terminology are defined, followed by an overview of commonly used imaging techniques. Finally, this chapter concludes with the introduction to medical image registration and outlines the need of interventional imaging.

Chapter 2: Medical Image Registration. In this chapter, medical image registration and the components will be discussed. The contributions to deformation estimation will be discussed in the context of registration problems, as well as the contributions to similarity measures, such as Linear Correlation for Linear Combinations (LC^2).

Chapter 3: Image-Guided Surgery and Intra-Operative Imaging. In some surgical scenarios, the use of diagnostic imaging is not feasible, either due to limited imaging capabilities or the need of real-time imaging. In this chapter, the contributions towards robotic Single-Photon Emission Computed Tomography (SPECT) during robot-assisted surgery, and intra-operative imaging are presented and discussed.

Selected Publications in Chronological Order

[**Fotouhi *et al.*, 2016**] J. Fotouhi, **B. Fuerst**, S.C. Lee, M. Keicher, M. Fischer, S. Weidert, E. Euler, N. Navab and G. Osgood, "Interventional 3D Augmented Reality for Orthopedic and Trauma Surgery," 16th Annual Meetign of the International Society for Computer Assisted Orthopeadic Surgery (CAOS), 2016.

[**Fischer and Fuerst *et al.*, 2016**] M. Fischer, **B. Fuerst**, S. C. Lee, F. Fotouhi, S. Habert, S. Weidert, G. Osgood and N. Navab, "Pre-Clinical Usability Study of Multiple Augmented Reality Concepts for K-Wire Placement," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, 2016.

[**Lee and Fuerst *et al.*, 2016**] S. C. Lee, **B. Fuerst**, J. Fotouhi, M. Fischer, E. Euler, G. Osgood and N. Navab, "Calibration of RGBD Camera and Cone-Beam CT for 3D Intra-operative Mixed Reality Visualization," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, 2016.

[**Kojcev *et al.*, 2016**] R. Kojcev, **B. Fuerst**, O. Zettinig, J. Fotouhi, S. C. Lee, R. Taylor, E. Sinibaldi and Nassir Navab, "Dual-Robot Ultrasound-Guided Needle Placement: Closing the Planning-Imaging-Action Loop," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, 2016.

[**Graumann *et al.*, 2016**] C. Graumann, **B. Fuerst**, C. Hennersperger, F. Bork, and N. Navab, "Robotic Ultrasound Trajectory Planning for Volume of Interest Coverage," *IEEE International Conference on Robotics and Automation (CARS)*, 2016.

[**Zettinig *et al.*, 2016**] O. Zettinig, **B. Fuerst**, R. Kojcev, M. Esposito, M. Salehi, W. Wein, J. Rackerseder, B. Frisch, and N. Navab, "Toward Real-time 3D Ultrasound Registration-based Visual Servoing for Interventional Navigation," *IEEE International Conference on Robotics and Automation (CARS)*, 2016.

[**Fuerst *et al.*, 2016**] **B. Fuerst**, J. Sprung, F. Pinto, B. Frisch, T. Wendler, H. Simon, L. Mengus, N.S. van den Berg, H. G. van der Poel, F.W.B. van Leeuwen, and N. Navab, "First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping," *Medical Imaging, IEEE Transactions on*, 2016.

[**Fuerst *et al.*, 2015c**] **B. Fuerst**, J. Fotouhi, S. C. Lee, M. Fischer, R. Kojcev, N. Navab, M. Dejong, M. Holman, J. Richmon, and C. Fakhry, "Can Intraoperative Ultrasound Assist TransOral Robotic Surgery?," *Hopkins Imaging Conference*, Hopkins Imaging Initiative, 2015.

[**Bork *et al.*, 2015**] F. Bork, **B. Fuerst**, A. Schneider, F. Pinto, C. Graumann, and N. Navab, "Auditory and Visio-Temporal Distance Coding for 3-Dimensional Perception in Medical Augmented Reality," in *International Symposium on Mixed and Augmented Reality-ISMAR 2015.* 2015.

[**Fuerst *et al.*, 2015b**] **B. Fuerst**, J. Fotouhi, and N. Navab, ”Vision-based Intraoperative Cone-Beam CT Stitching for Non-overlapping Volumes,” in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015.* Springer, 2015.

[**Pinto and Fuerst *et al.*, 2015**] F. Pinto, **B. Fuerst**, B. Frisch, and N. Navab, ”Radiopositive Tissue Displacement Compensation for SPECT-Guided Surgery,” in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015.* Springer, 2015.

[**Fuerst *et al.*, 2015a**] **B. Fuerst**, T. Mansi, F. Carnis, M. Salzle, J. Zhang, J. Declerck, T. Boettger, J. Bayouth, N. Navab, and A. Kamen, ”Patient-Specific Biomechanical Model for the Prediction of Lung Motion From 4-D CT Images,” *Medical Imaging, IEEE Transactions on*, vol. 34, no. 2, pp. 599-607, 2015.

[**Fuerst *et al.*, 2014**] **B. Fuerst**, W. Wein, M. Müller, and N. Navab, ”Automatic Ultrasound-MRI Registration for Neurosurgery using the 2D and 3D LC^2 Metric,” *Medical Image Analysis*, vol. 18, no. 8, pp. 1312-1319, 2014.

[**Schoch *et al.*, 2013**] A. Schoch, **B. Fuerst**, F. Achilles, S. Demirci, and N. Navab, ”A Lightweight and Portable Communication Framework for Multimodal Image-Guided Therapy,” *MIDAS Journal-Systems and Architectures for Computer Assisted Interventions 2013*, 2013.

[**Wein *et al.*, 2013**] W. Wein, A. Ladikos, **B. Fuerst**, A. Shah, K. Sharma, and N. Navab, ”Global Registration of Ultrasound to MRI Using the LC^2 Metric for Enabling Neurosurgical Guidance,” in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013.* Springer, 2013, pp. 34-41.

[**Fuerst *et al.*, 2012b**] **B. Fuerst**, T. Mansi, J. Zhang, P. Khurd, J. Declerck, T. Boettger, N. Navab, J. Bayouth, D. Comaniciu, and A. Kamen, ”A Personalized Biomechanical Model for Respiratory Motion Prediction,” in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012.* Springer, 2012, pp. 566-573.

[**Fuerst *et al.*, 2012a**] **B. Fuerst**, T. Mansi, P. Khurd, J. Zhang, J. Declerck, T. Boettger, N. Navab, J. Bayouth, and A. Kamen, ”Patient-specific finite-element simulation of respiratory mechanics for radiotherapy guidance, a first evaluation study,” in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on.* IEEE, 2012, pp. 1212-1215.

Patents

T. Boettger, F. Carnis, D. Comaniciu, J. Declerck, **B. Fuerst**, A. Kamen, T. Mansi, J. Zhang et al., ”System and method for prediction of respiratory motion from 3d thoracic images,” Aug. 25 2014, US Patent App. 14/467,246.

N. Navab, N. Cowan, **B. Fuerst**, and E. S. Fortune, ”System and Method

---

for Bioelectric Localization and Navigation of Interventional Medical Devices," Mar. 13 2013, US Patent App. 13/798,637.

List of Abbreviations

4D CT	Four-Dimensional Computed Tomography	8
CBCT	Cone-Beam Computed Tomography	9
CR	Correlation Ratio	26
CT	Computed Tomography	7
CTa	Computed Tomography Angiography	10
DRR	Digitally Reconstructed Radiograph	25
FBP	Filtered Back-Projection	10
FDK	Feldkamp-Davis-Kress	11
GPU	Graphical Processing Unit	23
HU	Hounsfield units	11
IGS	Image-Guided Surgery	35
IGT	Image-Guided Therapy	35
LC^2	Linear Correlation for Linear Combinations	1
MI	Mutual Information	24
MLEM	Maximum Likelihood Expectation Maximization	15
MRI	Magnetic Resonance Imaging	7
NCC	Normalized Cross-Correlation	25
PET	Positron Emission Tomography	8
SAD	Sum of Absolute Differences	24
SLN	Sentinel Lymph Node	31
SLNB	Sentinel Lymph Node Biopsy	31
SPECT	Single-Photon Emission Computed Tomography	1
SSD	Sum of Squared Differences	24
US	Ultrasound	8

CHAPTER 1

Introduction

A Brief History Throughout history, medical knowledge and expertise have improved parallel to the development of knowledge about the human body. The earliest known medical literature dates back to the 17th century BC [28], and it is assumed that medical illustrations were implemented since the very beginning of medical treatments [9]. One of the most well known volumes on human anatomy are the seven books written by Andreas Vesalius in 1543 AD, entitled the *De humani corporis fabrica*, which included more than 600 printed images. However, knowledge was limited to macroscopically observable anatomy and physiology, and access to medical education remained difficult.

In the early 19th century, modern medicine started to be established based on nosological systems of diseases, reasoning based on sound pathology and physiology, and the acceptance of modern technology as a meaningful tool. Since this time, technology and medical imaging have truly revolutionized the capabilities of treating injuries and diseases, and have changed the understanding of the normal or deviant functions of the human body.

In the very late 19th century, X-rays were discovered and showed that the visualization of interior parts of a human body was possible by transmitting radiation. This was ground breaking and motivated several important developments during the 20th century, such as medical ultrasound imaging in 1949, X-ray based Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) in 1973. Based on these and other imaging systems which evolved in the 20th century, understanding of human anatomy and physiology was significantly improved.

In general, effective treatment depends on the correct diagnosis, which is the task of identifying the disease, illness or injury based on objectively quantifiable signs and symptoms. Nowadays, medical imaging is a vita tool for diagnosing. The use of imaging technology is paramount in confirming diagnoses, assisting and validating treatments, and monitoring and documenting the courses and possible recurrence of diseases. Due to the varying imaging properties, the choice of the imaging technology is highly dependent on the medical application, and also on the anatomy and physiology that needs to be observed. Section 1.1 outlines the most commonly used and most relevant imaging modalities.

Future Vision and Mission With confidence we can assume that future developments will lead to an increase of early disease detection and improved treatment efficiency in terms of duration of the surgery, surgical task load, patient outcome, and reduction of side effects. Imaging and interventional assistance systems will evolve to observe and predict the surgeon's actions to most efficiently provide the required information or tool at just the right moment during surgery. This will be closely followed by taking over simplistic tasks. To achieve the long term vision of a smart, multi-modal operating room companion, our goal is to take meaningful steps on this path to enable the use of new technology in tomorrow's interventions.

Contributions The main contributions presented in this thesis aim at *(i)* transferring information from pre-operative imaging to the interventional scenario (Chapter 2), and *(ii)* performing intra-operative imaging (Chapter 3):

(i) The prediction of tissue deformation, and the registration of pre- and intra-operative images are presented in Chapter 2. More specifically, the prediction of tissue deformation is based on biomechanical simulations of the anatomy and physiology, and the personalization based on Four-Dimensional Computed Tomography (4D CT) images. Next, contributions to the registration of MRI and ultrasound images are presented. As these modalities inherently display different tissue properties, the successful alignment of these images represents a significant achievement. Finally, the chapter on medical image registration concludes with an outline of the impact of monotonic similarity measures on the registration process.

(ii) Novel intra-operative imaging concepts are presented, such as the robotic SPECT imaging during robot-assisted surgery, robot-assisted ultrasound acquisition, and the fusion of medical and optical imaging to achieve intuitive visualization during orthopedic and trauma surgeries. These novel concepts and technologies are designed and developed to assist surgeons during interventions, smooth workflow, and increase patient safety. Chapter 3 presents the background and contributions in this field.

In the following sections, the basics for medical imaging modalities (Sec. 1.1) and image-guided intervention (Sec. 1.2) are outlined, which represents the scientific and technical background of the contributions presented later in this thesis.

1.1 Medical Imaging Basics

Starting in the past century the use of medical images has become increasingly frequent and has spread from health care to biomedical research. In 1895, the first application of X-ray imaging for diagnostic purposes was reported. Since then the dramatic development of technology has lead to impressive imaging quality, improved use and benefit of imaging modalities. Nowadays, X-ray CT, MRI, medical Ultrasound (US), multi-spectral or fluorescence imaging, SPECT, Positron Emission Tomography (PET), microscopic imaging, and optical coherence tomography are commonly used for diagnosis and guidance of treatment.

Early medical imaging systems, such as X-ray imaging, were able to visualize the anatomy of the patient, but not the underlying physiological processes. Following the categories of observable information, medical imaging modalities

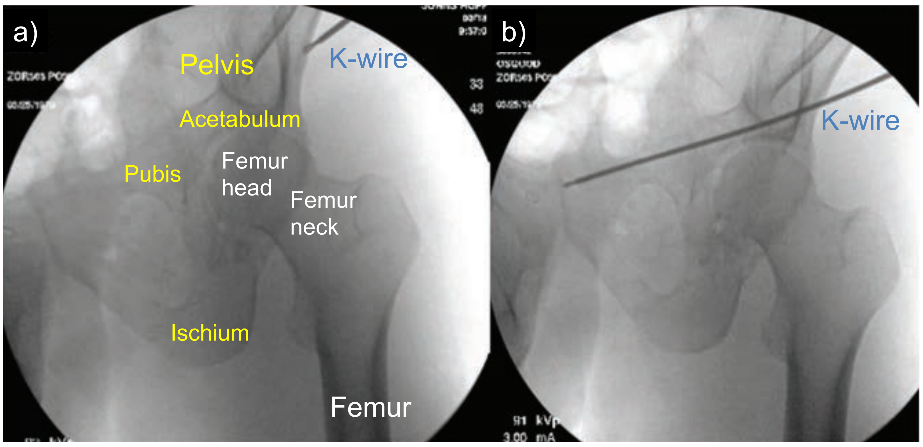


Fig. 1.1: Typical interventional X-ray images acquired using a C-arm. These images show the placement of a guide wire (k-wire) into the pubic ramus bone. First, the surgeon needs to determine the entry point through the muscle, see *a)*, followed by drilling the wire through the bone as seen in *b)*. Note that the surgeon requires a vast amount of X-ray images from various points of view to determine the precise three-dimensional trajectory relative to the bone.

are generally classified in either *anatomical/structural* or *functional* imaging. However, for some modern imaging modalities and their variants, such as MRI and functional MRI, this classification may not be absolute.

In this thesis the main focus is on the registration of MRI and US images, reconstruction of SPECT volumes, and use of Cone-Beam Computed Tomography (CBCT) as interventional imaging. Therefore, the following section presents the fundamental properties and concepts of each of these modalities, starting with anatomical/structural imaging, followed by functional imaging.

1.1.1 Anatomical/Structural Imaging

X-Ray Imaging Using a low dose of ionizing radiation, a projection image of a patient or object can be recorded. The radiation is generated using an X-ray tube, directed towards the subject and partially absorbed or scattered by tissue. Finally, the remaining radiation is recorded using a photostimulable phosphor plate, image intensifier or flat panel detector. As photoelectric absorption, Compton and Rayleigh scattering predominantly depend on the material properties of the tissue and X-ray photon energy, the projection image observed directly reflects the densities and material composition of the tissue. Modern imaging systems include stationary imaging systems specialized for static imaging, or mobile systems which may also provide live X-ray imaging (fluoroscopy). Typical interventional X-ray images acquired during a pelvic fracture reduction are shown in Fig. 1.1. Due to their shape factor, mobile X-ray machines are often referred to as *C-arm*, where one end of the C-shape holds the detector, and the X-ray source is mounted on the opposite end. Reconstruction of three-dimensional images is possible using a motorized or robotic system which can rotate around the patient and acquire 50 or more X-ray images (see Computed Tomography below).

X-Ray Computed Tomography (CT) By combining X-ray images taken from different angles, cross-sectional images and three-dimensional volumes of the subject can be reconstructed. In general, the reconstruction algorithms to compute the three-dimensional volume can be categorized in analytic or algebraic techniques. Most clinical CT scanners use a (proprietary) variation of the analytic method referred to as Filtered Back-Projection (FBP), which will be explained in the context of the more complex CBCT reconstruction (see below).

While the basic CT scanner comprises of an X-ray source and a detector, modern devices may incorporate multiple sources or detectors to enable simultaneous imaging using photons of different energy levels. The detectors are composed of up to 320 detector lines (Toshiba Aquilion One), but no published study can be identified whether diagnostics based on a 320 slice system yields a significant difference to commonly used 64 slice systems. By injecting contrast agents with higher X-ray attenuation, the vasculature and perfusion can be visualized. For instace, fast multi-slice detector CT scanners are used to image cardiac perfusion. Imaging of vasculartures is frequently referred to as Computed Tomography Angiography (CTa).

Cone-Beam Computed Tomography (CBCT) In contrast to conventional X-ray CT, the CBCT uses a source and detector similar to static X-ray imaging where the X-rays diverge and form a cone. CBCT enables intraoperative 3D imaging for various applications, for instance orthopedics [2], dentistry [52], or radiation therapy [4]. The reconstruction of the volume of interest is performed using the projection data form different angles. First, we will explore the problem statement and solution of the line detectors, and then expand the problem to the cone-beam shape. When an X-ray beam leaves the detector with an intensity I_0 it travels along a line j and is attenuated by an object following Beer's law for photon intensities:

$$I_j = I_0 * \exp\left(- \int_{\texttt{line}\ j} \mu(\vec{r})dl \right),$$

where $\mu(\vec{r})$ is the attenuation coefficient of the material along the line j at the position $\vec{r}$. The projection data p can now be expressed as:

$$p_j = -\ln \frac{I_j}{I_0} = \int_{\texttt{line}\ j} \mu(\vec{r})dl$$

The attenuation coefficient is the desired material property, which can be reconstructed by solving a set of integral equations for the attenuation coefficients μ based on the projection data p:

$$p = \mathbf{A}\mu, \tag{1.1}$$

where $\mathbf{A}$ is the system characterizing operator. This problem can be solved using analytical or algebraic methods. The algebraic method is mainly used when the acquisition geometry prohibits a uniform backprojection or the observed data can not be represented as projection data and will be presented in context of SPECT. The analytical solution is typically used to perform CT reconstructions and is based on the concept of a filtered backprojection:

$$\tilde{\mu}(x) = \int d\lambda \int du' k(u') p(\lambda, u - u'),$$

where $d\lambda$ is the spatial component of the backprojection, $du'k(u')$ the spatially varying filter and p remains the projection data. Taking the rotation of the source and detector around the patient along the trajectory Γ into account, the cone-beam transform is defined as:

$$D\mu(\vec{\alpha}(t), \vec{\beta}) = \int_{0}^{\infty} \mu\left(\vec{\alpha}(t) + s\vec{\beta}\right) ds, (\vec{\alpha}, \vec{\beta}) \in \Gamma \times S^2,$$

where $\vec{\alpha}(t)$ is the source position along trajectory Γ, and $\vec{\beta}$ is the unit vector pointing along a particular X-ray beam. For conventional CT reconstruction, the Radon transformation can be directly applied based on the 2D parallel-beam geometry. However, for CBCT, the projection data is linked to the three-dimensional Radon transformation, which describes the mapping of a function $\mu(\vec{r})$ onto its plane integrals:

$$\Re\mu(\rho, \vec{\theta}) = \int d^3r \delta(\vec{r} \cdot \vec{\theta} - \rho) \cdot \mu(\vec{r}),$$

where $\vec{\theta}$ is the normal vector on the plane for which the plane integral is computed, and ρ is the distance between origin and plane. The link back to the projection data is not trivial, and is performed through filtering and the 3D Fourier transformation, leading to the result of the derivations of the analytical solution:

$$\mu(\vec{r}) = -\frac{1}{8\pi^2} \int d\vec{\theta} \, \frac{\delta^2}{\delta\rho^2} \Re\mu(\rho, \vec{\theta})|_{\rho = \vec{r} \cdot \vec{\theta}},$$

where $d\vec{\theta}$ represents the backprojection, $\Re\mu(\rho, \vec{\theta})$ the plane integral of the object, and $\frac{\delta^2}{\delta\rho^2}$ is the inversion that - unlike the two-dimensional case - acts locally in the three-dimensional Radon space. However, this naive application of the three-dimensional Radon inversion formula is prohibited due to the long object problem (limiting the scanning region to the region of interest due to an increasing parallax effect) and the computational expense. Therefore, CBCT reconstruction algorithm resorts to simplifications to end up in an efficient and numerically stable shift-invariant one-dimensional filtered backprojection. The Feldkamp-Davis-Kress (FDK) algorithm, a generalization of the fan beam inversion formula to the cone-beam geometry, is a possible solution [13], as is the Z-smart approach [11].

Finally, the attenuation coefficients are normalized based on the attenuation coefficient of water, resulting in the CT number or Hounsfield units (HU). Comparing identical materials scanned using CBCT and CT, it can be observed that the HU are not identical. The short-scan FDK, especially limited angles, simplifications and filtering during reconstruction, may result in different HU compared to CT [39, and therein]. Therefore, the ability to asses tissue properties is limited.

However, the effectiveness of CBCT in orthopedic surgeries is bound by its limited field of view, resulting in small volumes. Intraoperative surgical planning and verification could benefit from an extended field of view. Long bone fracture surgeries could be facilitated by using three-dimensional absolute measurements and multi-axis alignment in the presence of large volumes, assisting the surgeon's mental alignment.

Magnetic Resonance Imaging (MRI) In contrast to X-ray based imaging, MRI does not use ionizing radiation and is therefore non-invasive. The information acquisition relies on the principle that spinning charged particles (protons) align their spins with a strong magnetic field. When the strong magnetic field changes, the spins follow and emit a weak field.

More specifically, medical MRI systems align spins of hydrogen atoms (only one electron and proton) using a strong, constant magnetic field (B0), which is called longitudinal magnetization. At this stage, the hydrogen nuclei in the patient's body align parallel or antiparallel to the field direction, but a slightly greater proportion aligns parallel. The precession of the hydrogen nuclei around the long axis of the magnetic field depends on the atom and the strength of the field. The precession rate is referred to as the Larmor frequency, where a set of protons can either precess together (in phase) or separately (out phase). Within the bore of the primary magnet there are three gradient coils, which are arranged in opposition to each other to create poles in order to enable the alteration of the primary magnetic field. By combining the three gradient magnetic fields, the strength and orientation of the magnetic field can be changed locally, which allows for spatial encoding and localization. Finally, RF coils are used near the patient's body to emit a second magnetic field (B1) with the same precession frequency, which makes some low energy parallel protons flip to be antiparallel, and protons become synchronized and precess in phase.

After the RF pulse is applied, the change of the net magnetic vector, which is the sum of longitudinal (T1) and transverse (T2) magnetization, is observed. The protons flip back to their original spin direction (relaxation) by giving off the energy to the environment at their local Larmor frequency. The temporal increase of magnetization due to the relaxation is contingent on the tissue composition, and is known as T1 relaxation. The transverse magnetization depends on the synchronization of the spins, and reduces over time. Based on the tissue composition, the time that the spins need to go from in phase to out phase spins (T2 relaxation) varies. The change of the net magnetic moment of the net magnetic vector results in free induction decay and induces an electrical signal in the RF coils. The reconstruction of the spatial image is performed using a Fourier transform.

In summary, the image intensities depend on the composition of tissue that contains significant amounts of hydrogen atoms. Therefore, MRI is well suited to visualize soft tissue compositions, but yields poor imaging quality for tissue such as bones. Imaging of patients with magnetic implants is not possible. Performing interventions under live MRI-guidance is challenging due to the limited accessibility (bore size) and the required non-metallic medical instruments.

Ultrasound Imaging (US) Unlike MRI and CT systems, ultrasound imaging systems are mobile devices that allow interactive and real-time imaging.

Rather than performing tomographic reconstruction, ultrasound images represent the changes in acoustic properties of tissue. Today, the basic mode of operation provides a two-dimensional slice (*B-mode* image) of the patient where the intensities are correlated to the amount of ultrasonic reflections.

Imaging artifacts are common and may be inherent to the ultrasound beam characteristics, caused by multiple echo paths or changes in acoustic velocities within the tissue, or be based on attenuation errors. Random speckle (noise), comet tails, multiple echoes, mirror images and shadow regions are commonly found in the images.

To enable imaging, the transducer needs to be in direct contact with the patient's skin, using gel to achieve acoustic coupling and avoid air between transducer and tissue. In general, the sound pulses, at frequencies around 1 to 15 MHz, propagate through the tissue and are reflected at interfaces where the acoustic properties change, which usually represents a change in tissue, layers or organs. These reflections are recorded by the same transducer, and the image intensities are then correlated to the origins and intensities of the acoustic echoes. Naturally, the type, dimension and resolution depends on the ultrasound transducer itself.

Transducers generate the ultrasound beam through mechanical vibrations using piezoelectric elements of capacitive micro-machined ultrasonic transducers. The most common transducer models are categorized based on the arrangement of the emitting and receiving elements. For instance, linear array transducers have a flat interface (referred to as *aperture*) and the beam is shaped to be parallel. Consequently, the image is rectangular and represents the narrow stripe from aperture into the body. For certain clinical applications, this is sufficient and allows the imaging at a constant resolution, but the limited field of view motivates other beam shaping designs. Similar to the linear array transducer, the phased array transducer also has a flat aperture, but the beams are steered in an angulated fashion. This allows the imaging of a wider field of view, but due to limited space in the transducer, the number of elements can not be increased. To better fit more elements and the interface to the patient, the curvilinear transducers have a curved interface. These transducers are frequently used for abdominal examinations, where both the field of view, as well as the spatial resolution are of importance.

To acquire three-dimensional images, wobblers are used which usually encapsulate a curvilinear transducer within a casing and automatically swivel it inside. This allows the coverage of a three-dimensional volume without the user moving the transducer, but requires the mechanical motion inside the transducer. Modern phased array transducers can also shape the ultrasound beam to scan larger, three-dimensional areas. Both techniques allow for acquisition of a three-dimensional area over some time. The area is limited to the pose of the transducer held by the user. Other approaches to acquire a larger three-dimensional volume are based on the movement of the transducer along the surface of the patient. To later compound the individual two-dimensional slices to a large three-dimensional volume, tracking techniques are required. For instance, infrared optical, inertial or electromagnetic tracking have been used in commercial systems. Registration-based approaches try to assemble the volume by registering the images. This is partially successful when acquiring several three-dimensional volumes because the overlap is sufficient. However, systems require the user to move the transducer, which usually results in varying degrees of pressure applied to the surface throughout the scan. This causes registration based mosaicing to fail, and introduces step artifact in images compounded based on tracking.

A common issue concerning current ultrasound imaging is strong user dependence, due to the hand held nature of the device and difficulty of recognizing patterns, structures and certain tissue in ultrasound images. This motivates the use of advanced robotics to enable a precise scan of large areas, and the application of a continuous pressure profile over the entire scan. Thus, the ultrasound volumes acquired can be larger and without step artifacts. This enables the automatic registration with diagnostic imaging and therefore the automatic targeting of regions of interest based on pre-operative planning using diagnostic images.

1.1.2 Functional Imaging

In contrast to structural imaging, the goal of functional imaging is to reveal the physiological activities within tissue or organs. The modalities frequently require tracers, which are chemically similar to fluids in the body and are therefore dynamically distributed throughout the body. Nuclear imaging - an important group of functional imaging modalities - require the tracer to be radioactive, by either binding an isotope to a molecule, or dissolving it in a fluid. Common representatives of nuclear imaging are SPECT and PET. Other commonly utilized modalities are based on similar physical concepts like structural imaging, such as dynamic angiography/perfusion CT or functional MRI. As the scientific contributions are useful for gamma and SPECT-like imaging, the following will focus on these two modalities.

Gamma Detection and Imaging Gamma detectors are hand-held devices capable of detecting radiation and reporting the activity observed as *counts*. Tissue with no or little concentrations of tracer have low counts, while areas with higher uptakes of radioactive tracers yield higher counts. The primary use of gamma detectors is sentinel lymph node mapping, parathyroid surgery, and radioactive seed localization. Surgeons use gamma detectors by moving them over the tissue in a grid-like fashion until they localize areas causing higher counts. In 1998, a first patent on a mini gamma camera was filed [41], showing an array of detectors in a grid-like arrangement. This enables surgeons to obtain a simple, low-pixel image of the distribution of radioactivity, and subsequently decreases the scanning duration [71, and references therein].

Single-Photon Emission Computed Tomography SPECT imaging provides true three-dimensional volumes, where the image intensities are correlated with the activity of a radioactive tracer. This is helpful for, but not limited to tumor imaging, leukocyte imaging, thyroid monitoring, perfusion imaging, functional imaging, etc. The system provides an array of gamma detectors - a gamma camera - which rotates around the patient. The projection images are recorded and used to perform a tomographic reconstruction.

This nuclear imaging technique is frequently combined with anatomical imaging modalities, such as CT. The basic principle is based on the direct measurement of radioactivity by using an array of gamma detectors. However, rather than just recording a two-dimensional projection, one or more detector plates rotate around the patient in order to record multiple views. Simultaneously, CT volumes are acquired to correct attenuation by differently dense tissue. To reconstruct the three-dimensional SPECT volume, the problem is stated as a system of linear equations as in equ. 1.1, and an iterative solver is used to solve it. The elements a_{ij} of the system characterizing operator $\mathbf{A}$ are defined by the probability distribution that one voxel j contributed to the observed radioactivity on the detector pixel i. This is dependent on the camera response function (usually look-up tables are empirically recorded or simulated) to the known radioactive tracer, and the absorption/attenuation, scattering, and diffraction of surrounding tissue, where attenuation is corrected through the CT volume simultaneously acquired. The voxel values reflect the concentration of the radioactive tracer.

Using the *Algebraic Reconstruction Technique* (ART), the volume of interest μ is iteratively reconstructed:

$$\mu^{(n+1)} = \mu^{(n)} - \frac{\mathbf{a}_j \sum_i \left[\mu_i^{(n)} a_{ji} - p_j\right]}{\sum_j a_{ji}^2},$$

where $\mu^{(n)}$ is the volume at iteration n, a_{ji} are the elements of the system matrix $\mathbf{A}$, and p_j are the projections. The system matrix $\mathbf{A}$ also models the response function of the detector or camera, which is usually implemented as a look-up table. Similarly, Maximum Likelihood Expectation Maximization (MLEM) can be used to solve equ. 1.1 for SPECT reconstruction, where the update step is multiplicative rather than additive. This also allows for the incorporation of relaxation factors, which may represent attenuation, camera response properties, etc.

1.2 Image-Guided Interventions

The information on the anatomy and physiology of the patient that medical images provide, can enable guidance during the intervention itself. However, in many cases the surgeon needs to translate the information provided to the surgical site, plan and execute the actions. For example, during orthopedic interventions, the surgeon may look at X-ray images showing bones and fractures. He then needs to mentally align the anatomy seen in the images with the position of the patient, perform the incision and repair. In general, this mental alignment of medical images, patient, and medical tools proof to be difficult, especially when dealing with complex three-dimensional structures, tumor margins, close proximity to crucial structures (e.g. in spine or brain), etc.

Assistance systems support the surgeon by *a)* providing additional context to the images, *b)* providing the trajectory and deviation of it towards a pre-defined target, or *(c)* executing components of the surgery autonomously. Typical functions of surgical image-guided navigation systems are *(a)* and *(b)*. Similarly, scientific contributions (see Chapters 2 and 3) of this thesis also focus on providing context or visualization of the target. Additionally, robotic ultrasound contributes to the path of executing some of the tasks autonomously.

Image-guided navigation reduces the uncertainty of the association of structures and therefore contributes as a tool to improve accuracy. When establishing a precise spatial relationship between images and the surgical site, specific points in the image are matched with the corresponding real-world position. This is usually achieved by means of registration. In the course of this thesis, this registration of the pre-operative images and intra-operative information is performed by means of image registration, where intra-operative images are acquired in such a fashion that the relation between real-world positions and image coordinates is known. Therefore, the registration of pre-operative images and intra-operative images allows for the translation of pre-operative information to the real-world surgical site.

In general, the transformation mapping between images or coordinate frames requires at least three point correspondences, assuming the relation between two images or coordinate frames can be represented without deformations. Once this relationship is established, the surgeon can target a point in the patient using a tool with tracking targets, and the corresponding position will be visualized within the medical image. Additionally, the trajectory towards a pre-defined target and deviation from a ideal trajectory can also be visualized.

While current navigation systems add time to the overall occupancy of the operating room, they may improve patient outcome. However, a recent publication indicated that even this advantage may not be true for some applications [34]. Most navigation systems comprise of a tracking system, tracking targets on tools, devices and patient, and a cart containing the computer and monitors. These devices need to be introduced to the operating room, where they reduce the free space which is already very limited. Additionally, the devices require cleaning and storage, which adds overhead in terms of cost and time. Once the systems are set up, and the tools and devices calibrated, the surgery can begin. This additional setup time increases the overall duration of the surgery, and therefore limits the availability of the operating room.

During the actual procedure, optical tracking systems require line of sight between the tracking targets on the patient, tools and devices, and the tracking system. This limits the surgeon and staff in their moving space and access to the surgical site. Electromagnetic (EM) tracking systems are used infrequently, as metal in the tools, surgical table, and devices may reduce the tracking quality. Additionally, the distance between EM tracking system and targets is limited, rendering the tracking in a realistic OR setup impractical.

In the following chapters, image guided interventions are enabled by means of registration of pre- and intra-operative imaging (Chapter 2), and intra-operative imaging acquisition and fusion of camera views and medical images (Chapter 3). In contrast to the classic approach of image guided navigation, this concept of medical augmented reality provides an intuitive understanding of the relationship of patient anatomy, medical instruments and surgeon's hands, as all of these are visualized in the same view. Therefore, tracking of tools and tracking systems are not required as the establishment of the association of real-world positions and points in the image can be done with any tool, finger or object that is visible in the camera view. Furthermore, guidance along the trajectory and the deviation from the ideal path are implicitly shown as the current tool position is virtually overlaid on the current medical image.

CHAPTER 2

Medical Image Registration

In a clinical scenario, complementary medical images are acquired showing the relationship of anatomy and physiology as well as abnormalities. The clinician's task is to mentally align these images and combine the information from these sources to come to a meaningful conclusion. This is made more difficult when images are acquired using different imaging devices with dissimilar imaging capabilities. However, this complementary representation is of great relevance to a vast range of clinical applications, and the fusion of interrelated information provides a great benefit for improving diagnosis, treatment planning, image guidance during interventions, and post-operative validation.

For instance, MRI provides excellent information on soft tissue and tumor boundaries, but is extremely difficult to acquire intra-operatively. On the other hand, Medical US provides real-time images which can be acquired intra-operatively, but lacks the field of view or image quality of MRI. The fusion of both modalities enables the transfer of information present in MRI images to a common coordinate frame defined by the acquired ultrasound image. However, this registration presents a difficult and complex problem, as the representation of the same object is very different, which is largely based on the varying physical properties that are observed by the modalities. MRI intensities correlate to the relaxation times of the ^{1}H nuclei, while US images indicate the changes in acoustic impedance.

In addition to the different nature of images, dynamic or real-time imaging is often used during diagnosis or treatment, resulting in vast amounts of medical images. Therefore, an automatic image fusion is desired, which also aids the clinician in avoiding false alignments.

2.1 Definition of Registration and Image

In this context, image registration is defined as the automatic process of aligning, deforming or scaling two or more images in a common coordinate space, which results in the establishment of correspondences between the information contained in the different images. Medical image registration can be categorized based on the modalities (Sec. 1.1), registration components (Sec. 2.2), images, or observed subject(s).

Definition of Image For the process of image registration, the images must be available in a digital form, for which the information is discretized and digitalized. In most cases, the images are represented using an array of elements. The data acquired by commonly used imaging modalities can easily be represented using one-, two-, or three-dimensional arrays of tuples, usually singles or triples. In some cases temporal sequences of these images are obtained, increasing the dimensionality of the array. A tuple of the array is commonly referred to as picture element (*pixel*) or volume element (*voxel*), which have been used synonymously in various publications. In summary, we define an image to be a function f which operates from a sampled subset X of its domain to a subset Y of the codomain:

$$f[X] = \{y \in Y | y = f(x) \forall x \in X\}. \tag{2.1}$$

Note that this generalized definition does not limit the definition of an image to the aforementioned sampling of data as an array with tuples. For instance, X can be a set of three-dimensional points with Y being a unit set or representing color information.

Examples of Images CT and MRI images are usually represented as three-dimensional arrays with gray scale values. Additional information is provided by the imaging devices, such as a rigid transformation from a known point in the real world to the origin of the image, and the element spacing. US images are mostly two-dimensional arrays of gray scale values. Based on the transducer used (e.g. curved), the image may contain large areas with no information. Most US systems do not provide a relationship between the real world and the origin of the image. The pixel spacing often needs to be computed based on the user-defined depth setting. Three- or four-dimensional images may be acquired by moving a two-dimensional transducer. Alternatively, an oscillating or a phased array transducer can be deployed.

Registration In general, the image registration process is modeled in an algorithm that finds a mapping $\mathbf{H}$ from one image to another one. The majority of clinically relevant cases, two images are registered, which is also referred to as *pair-wise* registration. Registering more than two images can be achieved by performing multiple *pair-wise* registrations, or using one image and multiple references during a *group-wise* registration [40]. For most purposes we can reduce our registration process to a pair-wise registration, which could be expressed as maximization of the similarity between two images:

$$\mathbf{H}' = \underset{\mathbf{H}}{\operatorname{argmax}}\, S(I, \mathbf{H}(J)), \tag{2.2}$$

where $\mathbf{H}'$ is the optimal mapping from the first image I (*reference* or *fixed* image) to the transformed image $\mathbf{H}(J)$ (*template* or *moving* image). The similarity computed by the measure/function S expresses how well the images are aligned. In Section 2.2, the components of the registration process, such as S and $\mathbf{H}$, are presented in detail.

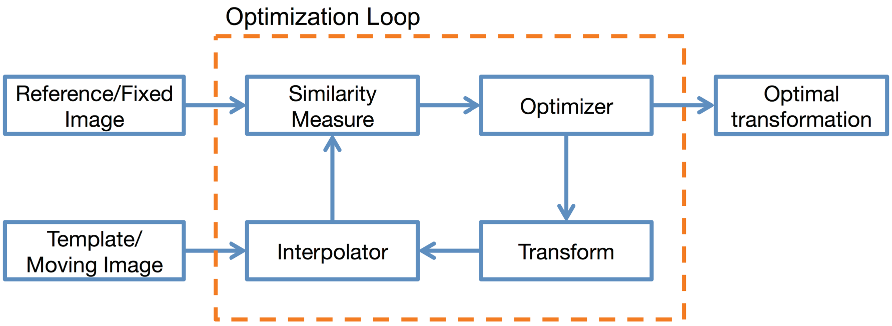


Fig. 2.1: The four major components of the image registration process are the similarity measure, optimizer, interpolator, and transformation.

2.2 Registration Process and Components

The registration process can be broken down into four major components, as outlined in Fig. 2.1. First, a *Transformation* is a mapping between two or more images. It describes how features in images relate to the common coordinate frame. Secondly, the *Interpolator* is used to apply the transformation to obtain a representation of the image in the common coordinate frame. Thirdly, the quality of the image alignment is to be quantified and expressed through an ordinal variable. We will refer to this component as *Similarity Measure*. Finally, to obtain the optimal alignment, the *Optimizer* changes the transformation based on the similarity measure until convergence has been reached.

2.2.1 Transformation

Transformations are a set of unknown variables, and are required or estimated when calibrating different sensors or medical imaging devices, navigating a tool or a medical device, monitoring or controlling a robot, augmenting a camera view, or registering different image data. The differences observed from one image to the next are based on three-dimensional changes of structure and physiology over time. Therefore, the transformation should be modeled as a four-dimensional function. However, based on the sampling and discretization of the images, we can approximate and simplify the transformation to be represented by a mapping of fewer dimensions.

In general, a transformation is a mapping, composed of a linear map and a translation, and is defined by its action on point coordinates. As a rule, transformations are classified based on the dimensionality of the underlying data, quantities or properties that are invariant, or on the consistency over the function domain. For instance, the registration of two undistorted two-dimensional images may only require a two-dimensional, rigid, linear transformation, while the registration of three-dimensional CT images of the thorax acquired at different phases during the respiratory cycle may require a three-dimensional, deformable, spatially varying transformation.

To achieve a matrix-vector product for all types of transformations *homogeneous coordinates* will be used for all points. This notation allows the representation of points and lines at infinity using finite coordinates by adding an additional dimension w to points. For projective transformations, this new dimension scales the coordinates. Otherwise it allows for the simple matrix-vector product to perform the transformation. The space with an additional dimension is now referred to as *projective space*. When $w = 1$, then it has no effect on the coordinates. For $0 < w < 1$ and $w > 1$, the object represented by points is scaled and appears smaller or larger respectively. However, if $w = 0$, then the coordinates represent a point at infinity.

In the following paragraphs the hierarchy of transformations in terms of the quantities or properties that are invariant will be laid out, as described in [21].

Rigid Transformation Transformations that preserve the Euclidean distance are referred to as rigid or Euclidean Transformations or Isometries. The Euclidean transformation can be written in block form as:

$$\mathbf{x}' = \mathbf{T}_{\mathsf{e}}\mathbf{x} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{x},$$

where $\mathbf{R}$ is an orthogonal rotation matrix, $\mathbf{t}$ is a translation vector, and $\mathbf{0}$ is a null vector. They are dependent on the dimensionality of the transformation. For instance for two-dimensional transformations (superscript $^{(2)}$) the transformation is defined as:

$$\mathbf{T}_{\mathsf{e}}^{(2)} = \begin{bmatrix} \mathbf{R}^{(2)} & \mathbf{t}^{(2)} \\ \mathbf{0}^{(2)\top} & 1 \end{bmatrix} = \begin{bmatrix} \epsilon\cos\theta & -\sin\theta & t_x \\ \epsilon\sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix},$$

where θ is the angle of rotation, $t_{x,y}$ is the translation along the x- and y-axis respectively, and $\epsilon \in \{-1, +1\}$ reverses the orientation. Note that the rotation matrix $\mathbf{R}^{(2)}$ only allows the rotation around one axis. To expand this concept for a three-dimensional transformation, the rotation matrix $\mathbf{R}^{(3)}$ needs to be composed by the clockwise (right hand rule) rotations around all three axes:

$$\begin{aligned} \mathbf{R}^{(3)} &= \mathbf{R}_z(\gamma)\,\mathbf{R}_y(\beta)\,\mathbf{R}_x(\alpha) = \\ &= \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}, \end{aligned}$$

where α, β and γ are the rotations (Euler angles) about the x, y and z axes respectively. The expansion of the translation vector is straightforward by adding the translation along the z-axis.

During image registration unknown parameters are estimated, and the computational effort of the registration process correlates directly with the degrees of freedom. Rigid transformations exhibit the least degrees of freedom compared to affine, projective or deformable transformations. Two-dimensional rigid transformations have one degree of freedom for rotation, and two for translation, resulting in a total of three. Three-dimensional rigid transformations have six degrees of freedom (three rotation, three translation).

Applying a two-dimensional rigid transformation, lengths, areas and angles remain unchanged, and are therefore the invariants. Volume and angles are invariants to the three-dimensional rigid transformation.

Similarity Transformations An isotropic scaling factor is introduced, resulting in a similarity transformation. The transformation matrix is composed by multiplying the rotation matrix by the scaling factor. The degrees of freedom of a two-dimensional and three-dimensional similarity transformation is four (one rotation, two translation and one isotropic scaling), and seven (three rotation, three translation and one isotropic scaling), respectively. In two dimensions, angles, rations of lengths along a line and parallel lines are not affected by this transformation. For three-dimensional transformations, angles are invariant.

Affine Transformations Non-isotropic scaling and sheering can be represented with an affine transformation. In practice, sheering is often not applied to medical images, while non-isotropic scaling can compensate for nearly uniformly applied pressure (e.g. brain shift or ultrasound transducer). In the block form, the non-singular affinity matrix $\mathbf{A}$ replaces the rotation matrix:

$$\mathbf{x}' = \mathbf{T}_{\mathtt{a}}\mathbf{x} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{x},$$

where $\mathbf{A}$ is a 3×3 matrix with nine degrees of freedom for a three-dimensional transformation $\mathbf{T}_{\mathtt{a}}$. $\mathbf{A}$ can be composed by a rotation matrix, a diagonal matrix with the non-isotropic scaling factors along the diagonal and a final rotation matrix[21]. To represent the geometrical operations defined through an affine transformation, we compose $\mathbf{A}$ using a separate scaling $\mathbf{T}_{\mathtt{scale}}$, shearing $\mathbf{T}_{\mathtt{shear}}$, rotation and translation (Euclidean) matrix $\mathbf{T}_{\mathtt{e}}$:

$$\mathbf{A}' = \mathbf{T}_{\mathtt{e}}\mathbf{T}_{\mathtt{shear}}\mathbf{T}_{\mathtt{scale}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} 1 & sh_{xy} & sh_{xz} & 0 \\ 0 & 1 & sh_{yz} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $sh_{xy,xz,yz}$ represent shearing on the $x - y$, $x - z$ and $y - z$ plane, and $s_{x,y,z}$ represent the non-isotropic scaling along x, y and z axes. Note that the order of $\mathbf{T}_{\mathtt{scale}}$, $\mathbf{T}_{\mathtt{shear}}$, and $\mathbf{T}_{\mathtt{e}}$ are chosen arbitrarily, but it will remain in this order throughout the thesis.

Parallel lines, rations of areas and the line at infinity are invariant to the two-dimensional affine transformation. Parallel planes, ratios of volumes and the plane at infinity are not affected by the three-dimensional affine transformation.

There are six degrees of freedom (two translation, four entries of $\mathbf{A}$) and twelve (three translation, nine entries of $\mathbf{A}$ or three for each translation, rotation, scaling and shearing) for two- and three-dimensional transformations. Note that it is common to reduce the degrees of freedom in the three-dimensional case to nine (three translation, three rotation, three scaling) for medical image registration.

Projective Transformations The projective transformation generalizes the affine transformation, and is represented as a non-singular linear transformation of homogeneous coordinates. Using the block form, the projective transformation can be written as:

$$\mathbf{x}' = \mathbf{T}_{\mathtt{p}}\mathbf{x} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\top & v \end{bmatrix} \mathbf{x},$$

where the vector **v** allows points at infinity to become finite points or the other way around. In the two-dimensional case we can not distinguish between orientation preserving and orientation inverting projections. The three-dimensional protective transformation has 15 degrees of freedom defined by 16 matrix entries up to scale.

Regarding projective transformations, the remaining invariants are intersections and tangency of surfaces in contact. Also, in the two-dimensional case, the order of contact and of ratio of lengths remain invariant.

Deformable Transformations If the motion of tissue is spatially varying within the image, a global Euclidean, affine or projective transformation may not be appropriate, reacquiring a *deformable* (non-rigid) transformation. Three of the most important applications of deformable image registration are *i)* multi-modal image registration, *ii)* longitudinal studies, and *iii)* inter-patient registration [61].

For the first group of applications, when image data is acquired using different imaging modalities, deformations often occur due to changes in patient positioning. Additionally, some image modalities, e.g. mammography and ultrasound imaging, require some external force to be applied to the patient's surface causing internal deformations. Multi-modal image registration may be the basis of diagnosis, treatment planning, or interventional surgical guidance. The latter may also require the deformation model to deal with deformations caused by the intervention itself, for instance resections, brain shift, etc.

For each part of the anatomy the deformation may be comprised of motion, change of size, or change of appearance and visibility. Furthermore, the interaction between different parts of anatomy is highly specialized, for instance bones are connected in joints with limited degrees of freedom, while lungs slide within the pleural cavity. The complexity of anatomical deformations is therefore very high, and this granularity may require the model to have a vast amount of parameters (degrees of freedom). However, finding the ideal alignment in terms of deformable registration requires finding the ideal set of these parameters under consideration of the similarity measure. Depending on the deformation model, the number of parameters can range from a mere 10 to millions, which shows the importance of the choice of deformation model.

The geometric transformations for non-rigid registration can be categorized into either *(i)* physically based models described by partial differential equations of mechanics, or *(ii)* a basis function expansion derived from interpolation [25]. While physical models may be able to estimate a realistic deformation, some simplifications (e.g. linear vs piece-wise linear, viso-elastic, or non-linear elasticity) cause the model to be only accurate for small deformations. On the other hand, increasing the complexity of the model yields a significantly higher computational effort. Future research may result in improved biomechanical models and a better understanding of continuum biomechanics. In contrast to physical models, the basis function expansions do not model the anatomy or physiology, but instead are a set of functions to interpolate or approximate the motion based on image data.

2.2.2 Interpolator

After the transformation is applied to the template image, an interpolation should be performed to estimate the pixel values of the template image at a coordinate determined by the reference image. This process is referred to as either interpolation or re-sampling. Depending on application, similarity measure, or transformation, it may be of benefit to interpolate the value of the reference image at given coordinates in the template image, especially when a deformable transformation causes sparsely populated areas in the template image. The most commonly used interpolation methods are nearest-neighbor, linear, and cubic convolution for two-dimensional images. In the following the interpolation techniques will be described for two-dimensional images. The expansion to three-dimensions is straight-forward using analogous concepts.

Nearest-Neighbor Assuming pixel coordinates are integer values, the value for a given floating-point coordinate can be found by rounding the x- and y-coordinates. This is a fast and computational inexpensive process. All values obtained through nearest-neighbor interpolation are part of the original image. In other words, this interpolation technique preserves the pixel values. However, for some transformations aliasing effects are caused. For instance, a horizontal line will exhibit clear step artifacts after being rotated by several degrees, as no blurring or smoothing will be performed.

Linear Interpolation Rather than choosing the value of the nearest neighbor, the linear interpolation approximates the value at a given floating-point coordinate from the weighted sum of intensities closest to the given coordinate. The four weights depend on the area of the rectangle defined by the closest image coordinates and the given floating-point coordinate. This technique eliminates most of the aliasing artifacts by introducing approximated pixel values, which have not been present in the original image. This technique presents a good balance between image quality and computational effort [30], and is also implemented on modern Graphical Processing Units (GPUs), which makes this interpolation very efficient.

Cubic Convolution This interpolation technique considers a grid (4×4) of neighboring pixels, and combines the row and column-wise one-dimensional interpolations. The interpolated image is visually similar to the image generated using bi-linear interpolation. After several interpolations, the image noise generated is significantly lower, at the cost of a significantly increased computational effort due to several intensity queries. Note that the theoretical computational complexity remains $O(n)$ for an image with n pixels - the same as for nearest-neighbor and bi-linear interpolation.

Other interpolation techniques include cubic basis splines and compactly supported radial functions. However, due to their computational effort and complexity they are infrequently used for image registration purposes. Especially for medical imaging applications, the balance between fast and clinically relevant imaging quality is of importance [30]. Furthermore, it may be vital to perform a simple interpolation in order to avoid the inadvertent creation of new pixel values or even structures. In the following, only nearest-neighbor and (bi/quadri)linear interpolations are considered as they are fully supported by modern GPUs, enabling a very fast and efficient computation of interpolated images.

2.2.3 Similarity Measure

In general, the similarity measure is a function that compares the reference and template image. After transformation and interpolation, pixels of the images are aligned at the same image coordinate, meaning that we can assume that the first pixel in image I and in image $\mathbf{H}(J)$ attempt to represent the same physical coordinate. The similarity measure is used to evaluate the quality of this alignment by comparing the two images, either based on the intensities or features. In some cases a preprocessing step is used to emphasize or remove certain properties of the images (e.g. gradient images or edge detection), or generate complementing data (e.g. confidence maps).

Desired properties of similarity measures are differentiability, continuity, robustness with respect to noise, and monotonicity towards one extrema which is reached at the ideal alignment. The similarity measure does not have to be an actual metric, nor does it actually need to fulfill the desired properties. In a real scenario, the images may show the same structure under different deformations, or based on different physical properties. Therefore, the similarity measure is usually specific to the application, and often exhibits multiple local extrema and noise.

Basics: Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), and Normalizied Cross-Correlation (NCC) A simple way of comparing images is to compute the absolute differences between each pixel in the reference and template image, and compute the sum of these difference, which is referred to as Sum of Absolute Differences (SAD). This works when images are very similar in contrast, brightness, and actual content. For larger images, this measure is increasingly robust towards outliers. Mathematically, it is not differentiable at zero (ideal alignment of equal images), which makes an analytical expression difficult.

Similarly to SAD this similarity measure computes the difference between the individual pixels, squares the values, and computes the sum, which is referred to as Sum of Squared Differences (SSD). For mono-modal images of similar brightness and contrast, the similarity value will be minimal at the correct alignment of the images. However, this measure assumes that the image intensities are identical, which cannot be fulfilled if a structure is images in one but not the other image. Therefore, the robustness towards outliers and noise is low.

Under the assumption that two images are random variables, Mutual Information (MI) can be used to determine the dependence between the images. The idea is that the correct alignment minimizes the quantity of information in a fused image. This measure was introduced for two-dimensional and three-dimensional image registration [67, 38], and assumes no prior functional relationship between images. The measure is based on the assumption of a statistical relationship that can be analyzed using the images' joint entropy:

$$MI(I,J) = \sum_i \sum_j p_{IJ}(i,j) \log \frac{p_{IJ}(i,j)}{p_I(i)p_J(j)},$$

where p_{IJ} is the joint probability distribution of pixels associated with images I and J, and p_I is the probability distribution of one image I. The entropy is highest when the alignment is poor. Consequently, this measure needs to be minimized in order to find the best alignment.

Normalized Cross-Correlation (NCC) allows to determine the similarity of two signals using a sliding dot product, where the result indicates the lag of one signal relative to another. If the signals are identical, then the cross-correlation is zero. For images of different brightness, this measure can be used by demeaning and dividing by the standard deviations:

$$NCC(I,J) = \frac{1}{n} \sum_{x,y} \frac{(I(x,y) - \overline{I})(J(x,y) - \overline{J})}{\sigma_I \sigma_J},$$

where n is the number of pixels, $I(x,y)$ and $J(x,y)$ are the pixel values, $\overline{I}$ and $\overline{J}$ are the averages, and σ_I and σ_J the standard deviations of I and J respectively. NCC is frequently used for mono-modal image registration of images with different brigthness, but the same underlying image features and properties. For instance, the registration of Digitally Reconstructed Radiographs (DRRs) and real X-ray images can be performed using NCC [26].

From Correlation Ratio (CR) to Linear Correlation of Linear Combination (LC^2) Multi-modal image registration presents a very different challenge, as the pixel values do not originate from the same physical properties, but the images may contain complementing information. In the following we will focus on registration of US with MRI, as this remains a difficult problem to solve.

Some approaches try to reduce the complexity by processing of both images to achieve a common representation of information [68], or by simulating one modality based on images from another modality. For instance, US/CT registration can be performed with ultrasound images simulated from CT images [69]. This is possible as there is a physical correspondence between acoustic properties and attenuation coefficients. However, there is little or no physical correlation between the concentration and compound of hydrogen and the acoustic properties. Therefore, simulations of ultrasound from MRI images would require the identification of tissue and the correct assignment of acoustic properties, which is currently not possible. Finally, the aforementioned two approaches to perform registration with ultrasound have a major drawback as ultrasound images are specific to the orientation of the ultrasound transducer, which requires the computationally expensive pre-processing step to be performed for each iteration during the registration process.

Although the structure, gray values, and boundaries of the anatomy are represented differently, the underlying anatomy remains the same. We can therefore assume that the MRI and ultrasound images exhibit some correlation, and under careful observation a relationship between structures in ultrasound images, MRI image gradients, and MRI image intensities can be identified for some regions of the image. This was first utilized by Roche et al. [58], by fitting a combination of MRI image intensities and image values to the ultrasound image intensities. This global fit allows for the computation of correlation of the MRI and ultrasound image, and defines the similarity measure Correlation Ratio (CR):

$$\eta^2(T) = 1 - \min_f \frac{\sum_{x,y} w^{(T)}_{x,y} [I(x,y) - f(J(T(x,y)))]^2}{n^{(T)} \sigma^2(I^{(T)})}, \tag{2.3}$$

where $w^{(T)}_{x,y}$ are linear interpolation weights for the neighborhood of the coordinates (x, y), $n^{(T)}$ is the number of pixels in this neighborhood, and $\sigma^2(I^{(T)})$ is the variance in this neighborhood, and f is the intensity mapping function. Therefore, this similarity measure correlates the image I with the transformed image $T(J)$ using an intensity mapping, local weighting, and a local constraint that the variance of I be large. The intensity mapping f is in the form of a linear combination of image intensities and image gradients with the reference image intensities. It is fitted for the entire image, and therefore does not consider locally varying changes in brightness or correlation. The neighborhood size is defined by the interpolation scheme used to determine the weights.

The similarity measure LC^2 follows a similar concept as CR, but generalizes the linear combination to allow for different overall brightness and areas where no fit is possible, and enables the locally varying fitting of this relationship function. This allows for dynamic compensation of changes in correlation of ultrasound and MRI images, for instance at tissue boundaries visible in both images, gray areas with different intensities, or shadow regions in ultrasound images. The overall cost function is then determined by computing a weighted average of the locally estimated correlations [17], see Appendix B.

2.2.4 Optimizer

The optimizer is the component performing the mathematical optimization, which is the determination of the best parameter values given a function. Our optimization parameters are the input values of the transformation, which then - given an interpolation technique - allow the computation of a cost value using the similairty measure. Based on pre-defined convergence criteria, the optimization process will yield and determine the best transformation parameters under consideration of the similarity value.

Most optimizers are either finitely terminating algorithms, converging iterative methods, or heuristics that approximate the ideal solution. For medical image registration, either optimizers or a combination of several techniques can be used. The most commonly deployed optimizers are gradient decent algorithm (slowly converging iterative method), conjugate gradient method (converging iterative method), Nelder-Mead method with calling gradients (also downhill simplex) method (finitely terminating or heuristic), Levenberg-Marquardt algorithm (converging iterative method) [46], BOBYQA (heuristic, iterative method) [54], NEWOA (heuristic, iterative method) [55], and many more. For multi-parameter optimization during image registration using a similarity measure resembling a quadratic function without the possibility of analytically computing the gradients, BOBYQA yields a good trade-off between the number of evaluations, speed, and accuracy [17].

2.3 Biomechanical Simulations for Deformation Estimation

Respiration remains a major source for imaging artifacts, as well as for complicated treatments, such as radiation therapy. As it is a crucial and involuntary motion, pausing respiration during imaging or treatment is often not a viable option. Compensating and predicting the motion, yields the potential to overcome these issues. The respiratory motion is especially complex, as it combines change of volume, pose of organs, nearly friction free lateral adjustments (sliding motion), and even inter-cycle variations.

2.3.1 Related Work

Image-based approaches attempt to cope with this difficult set of changes by introducing a set of regularization terms, which model normal and tangential motion [60]. Alternatively, piecewise-diffeomorphic registration techniques correct for the sliding interface [56]. However, it remains difficult to model the intra-cycle and inter-cycle changes of the motion by introducing regularizations for image-based predictions. Biomechanical models can be initialized based on patient-specific sequences of images, and constructed to mimic the physiology on a detailed level [43, 42]. Therefore, biomechanical modes can be generative, and enable the computation of non-observed motion and changes in breathing patterns. However, in previous work, the deformation was constrained by a fixed boundary condition and the physiology was not correctly modeled.

2.3.2 Contribution: Patient-Specific Biomechanical Model for the Prediction of Lung Motion From 4-D CT Images (IEEE TMI 2015)

By introducing a personalized and pressure driven biomechanical model, we have enabled the prediction of lung motion (see Appendix A). The overall goal of this work is to allow for predictions of any internal lung deformations based on simple surrogate signals, such as lung volume or readings from a thorax pressure belt. The model is no longer being bound to multiple image data and does not rely on a limiting geometry.

A patient-specific biomechanical model is presented, which enables the prediction of respiratory motion. For each patient, the lungs, mediastinum, and diaphragm are automatically segmented for each CT volume during a respiratory cycle, resulting in a set of binary masks representing the lungs, thorax, mediastinum, and the area below the diaphragm. The anatomical model is created, based on the segmentations at the end of expiration. This model is a discretized representation of the volume, and is comprised of nearly unilateral tetrahedrons. The personalization of the spatially varying force field is performed using segmentations from images at the end of inspiration. The force field is defined by automatically clustering regions on the lungs and diaphragm in patches. The force amplitudes are estimated through a multivariate optimization, which implicitly accounts for patient-specific and spatially-varying material properties. Additionally, the surface patches on the lung and diaphragm are used to later enable the nearly friction free sliding motion. The personalized force field and behavior properties are referred to as physiological model. Finally, the Finite Element Method simulation is run, and estimates the prediction based on a surrogate signal, such as the respiratory lung volume. Appendix A outlines the entire framework to predict the lung deformation during respiration. Results indicate that the accuracy is comparable to an intensity-based image registration approach, while at the same time remaining generative and allowing to simulate deformations which were not observed during the CT scan.

2.4 Multi-Modal Image Similarity Measure

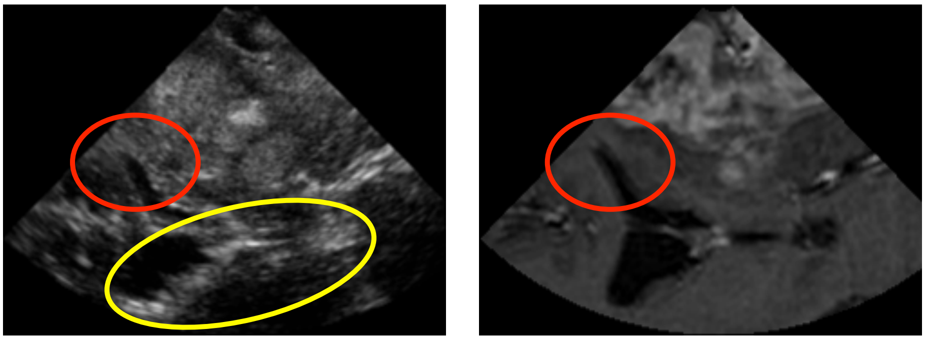
Image registration is especially complex when either different information is contained in each image, or the same information is represented differently. However, the fusion of different imaging modalities may yield a significant advantage for diagnosis or treatment. For instance, soft tissue or perfusion can be imaged using MRI, while good contrast for bony structures is achieved using CT. Another example is the use of pre- and intra-operative imaging. MRI and CT imaging provide good pre-operative imaging quality and rich contrast for certain anatomical or pathological structures. Due to logistical and technical reasons, these modalities are infrequently deployed during interventions. However, large deformations caused by patient positioning, changes in anatomy, or the intervention itself may render the pre-interventional images partially useless. Interventional X-ray or US imaging may lack the detail and quality needed, which motivates the fusion of pre- and intra-operative images.

Fusion of CT and US images has been widely investigated recently (for instance in [69, 27]). Most approaches map the Houndsfield units to backscatter, attenuation and characteristic acoustic impedance. This allows for the simulation of directional dependent ultrasound images from CT volumes. Therefore, the complex registration of these modalities can be reduced to registration of simulated and real ultrasound images.

On the other hand, the registration of MRI and US imaging remains a challenge, due to the fact that the observed material properties vary strongly. To generate the MRI image, the response and relaxation times of hydrogen atoms after a magnetic excitation are measured. This is fairly uncorrelated with the relative changes of acoustic impedance measured for US imaging. For instance, the free hydrogen concentration in bony tissue is low, resulting in nearly unimaged (black) areas in MRI volumes, while the surface of bones are usually represented by high image intensities in ultrasound images. In general, some soft tissue may be represented similar in both modalities, while in other regions the ultrasound image yields a closer resemblance to the gradient magnitudes of the MRI volume. Additionally, some regions do not seem to correlate at all. Fig. 2.2 illustrates the different tissue representations in MRI and US images. In contrast to CT/US registration, a simulation of US from MRI may therefore not be possible.

2.4.1 Related Work

The nature and origin of the intensity values in MRI and US images are very different, and some structures may only be visible in one or the other image. Additionally, details in MRI may lay in shadow regions in US images, and



(a) Trans-dural US of brain

(b) Corresponding slice of MRI

Fig. 2.2: Both images show the same 3D slice of the brain, where a prominent tumor is visible in the top center and the mid brain in the bottom center region. Some representations of tissue regions are represented similarly in the US image and MRI slice (red circles), while in other regions the MRI gradient magnitudes seem to correlate with the US image intensities (yellow circle). Other regional representations, such as the tumor, seem to relate, but not correlate linearly.

certain materials, such as in calcifications, are not visible in MRI imaging, but are clearly visible in US imaging. This causes registrations using similarity measures based on sum of squared distances, NCC, MI and normalized Mutual Information (nMI) to tend to fail [27]. Therefore, the focus is on similarity measures specific to MRI and US registration, but at the same time not organ-specific. Additionally, no significant effort should be introduced during pre-processing. For instance, US/MRIregistration has been proposed using higher-dimensional Mutual Information [57, 24]. Although it works theoretically, it is not practical in terms of implementation effort or computation time. The use of high confidence gradient and their alignment was presented in [10]. It is an interesting approach as anatomical structures are mostly characterized by their boundaries both in MRI and US images. However, the gradients in the US image are strongly dependent on the orientation of the transducer. Therefore, the lack of the use of intensity values suggests that this method requires a close initialization.

Another set of similarity measures attempts to utilize the neighborhood relationships. For instance, the modality independent neighborhood descriptor (MIND) [22] and its extension self-similarity context (SSC) [23, 6] utilize predefined neighborhood descriptors. Self-similarity as a similarity measure was first presented in [1], and does not rely on the assumption of a global intensity relation. Modality specific artifacts are not considered and the neighborhood descriptors tend to abstract the image, which may yield a lower overall registration accuracy. Furthermore, the computational effort for pre-processing is high due to the generation of voxel-wise neighborhood descriptors.

2.4.2 Contribution: Automatic Ultrasound-MRI Registration for Neurosugery using the 2D and 3D LC^2 Metric (MedIA 2014)

Considering the modality specific image properties, the most promising general strategy for robust US-MRI registration is to compare US to both the MRI intensity and its gradient, which does not require application-specific pre-processing or segmentation. The global fitting of a polynome to model the relationship between image intensities and image gradients has been initially presented in [58], referred to as CR. In [69], the similarity measure LC^2 is introduced for US/CT registration, which allows the local fitting of the polynomial and exhibits local invariance. This motivates a new way of comparing US and MRI images: Correlating a combination of MRI image intensities and MRI image gradients to the US image.

The main contribution is the presentation of a novel similarity measure to compare US and MRI data and enable intensity-based registration. The construction of the similarity measure LC^2, to align either freehand US slices or US volumes with MRI images is thoroughly explained, and applied to 2D and 3D data sets. The main scientific contribution is the novel pre-processing of the images, and the locally normalized LC^2 measure for images I and J, the transformation T, the neighborhood position $\mathbf{x}$ and neighborhood size s:

$$LC_l^2(I, J, T, \mathbf{x}, s) = 1 - \frac{\sum_{\mathbf{y} \in \Omega(\mathbf{x},s)} \left(I(\mathbf{y}) - f(J(T(\mathbf{y})))\right)^2}{|\Omega(\mathbf{x}, s))|\, Var(I(\mathbf{y} \,|\, \mathbf{y} \in \Omega(\mathbf{x}, s)))},$$

where the relationship function f provides an optimal fit of the linear combination of MRI image intensities and image gradients with the ultrasound image, and the neighborhood $\Omega(\mathbf{x}, s)$ is either a 2D or 3D neighborhood function. To compute the overall similarity measure, a weighted averaged of the locally normalized similairty measure is computed using the neighborhood local standard deviation within image I.

Both the registration of 2D sets of US images with MRI volumes, and the registration of 3D US volumes with MRI volumes, allow an automatic and robust registration, while the three dimensional method yields a significantly improved percentage of optimally aligned registrations. The extensive evaluation comprises a convergence study and a validation of accuracy. First, the influence of the ultrasound slice spacing is analyzed, resulting in a convergence in terms of correct alignments for spacing in the same range of the MRI volume spacing, and is further reduced for deformable registration to avoid missing anatomical details. Next, the influence of neighborhood size in 2D and 3D registration is analyzed in terms of accuracy and capture range (Lipschitz continuity), providing a good estimate of an appropriate neighborhood size. Finally, the influence of the pre-processing of the MRI image is analyzed and gradient magnitudes and directed gradients are compared. Interestingly the use of gradient magnitudes results in a larger capture range, which can be explained by the increased robustness of the method compared to the directed gradients.

Based on the convergence studies, we have performed a detailed registration study by simulating 2800 randomly initialized registrations. The use of the 3D similarity measure statistically significantly outperforms the use of the 2D similarity measure. Using real patient data, this methodology can achieve an accuracy of 2.51 mm, precision of 0.85 mm and capture range (Lipschitz continuity) of 15 mm (>95% convergence).

The registration of 2D ultrasound images with the MRI volume using the 2D LC^2 measure does not require the generation of a volume, which allows the registration to start during the scanning motion. However, the generation of a 3D ultrasound volume allows for an increase capture range when registering with the MRI volume. In general, the computation of these similarity measure requires the GPU to support double floating-point precision. The optimized use of the GPU allows for 100-500 similarity evaluations per second.

Appendix B presents the details on the implementation and results.

2.5 Tissue Displacement Compensation for SPECT-guided Surgery

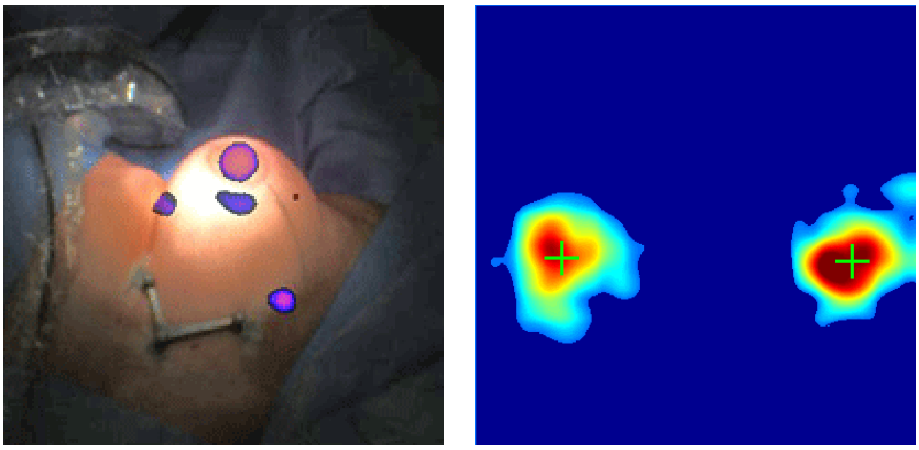
For melanoma [72], breast cancer [37, 33] and vulvar cancer [64, 66] Sentinel Lymph Node Biopsy (SLNB) is the standard of care. The involvement of lymph nodes is one of the most important prognostic factors for survival of the patient. Fortunately, many patients do not exhibit lymph node involvement, which can only be determined after the biopsy. To locate the Sentinel Lymph Node (SLN), a radioactive tracer (usually a ^{99m}Tc nanocolloid) is injected near the tumor site, which then drains through the lymphatic system towards the SLN. Under guidance of a gamma detector, gamma imaging, or freehand SPECT imaging, the surgeon then performs the resection or biopsy. Freehand SPECT imaging is an interventional imaging technique, which can be used to generate three-dimensional SPECT images by acquiring several thousand gamma activity recordings over a region of interest with a spatially tracked detector [70, 48].

2.5.1 Related Work

To compensate for motion, the current clinical approach is to either revert to a hand-held one-dimensional gamma probe, perform mental mapping and lose the benefits of image-guided surgery; or to re-acquire the entire SPECT volume, which leads to a significant delay. An alternative approach was presented in [65], where a one-dimensional probe was used to perform a 1D/3D registration. This, however, requires a precise model of the detector and several hundred tracked gamma activity readings. It was shown that the position of a 1 cm radiopositive node could be updated with an accuracy of 8 mm, which may prove insufficient.

2.5.2 Contribution: Radiopositive Tissue Displacement Compenstation for SPECT-Guided Surgery (MICCAI 2015)

The use of augmented reality freehand SPECT imaging is shown in Fig. 2.3a), while the representation of radioactive sources using a gamma camera is illustrated in Fig. 2.3b). The SPECT imaging provides clear advantages, such as an understanding of depth and three-dimensional structure, the images may be



(a) Augmented Reality (AR) SPECT view (b) Two sources seen by gamma camera

Fig. 2.3: (a) SPECT system showing several SLNs via Augmented Reality. Courtesy SurgicEye GmbH. (b) Interpolated gamma camera output for two sources, showing the auto-marked centers.

outdated once the surgery begins and the lymph nodes are displaced. To allow for compensation of lymph node motion, a registration-based approach is presented in [53], see Appendix C. During resection of SLNs, the radiopositive tissue is displaced through the intervention itself. For the first time, this paper presents a methodology to update pre- or intra-operative SPECT volumes using a 2D gamma camera. The algorithm utilizes tracking of the gamma camera and the known response function to dynamically estimate tissue displacement.

After an initial SPECT volume acquisition using a gamma camera, the volume reconstruction is performed. Next, the individual hotspots are automatically segmented using thresholding. The changes in the 2D gamma camera image allows for the estimation of the 3D motion of the lymph nodes, which is comprised of motion parallel to the detector plane (intensities do not vary), and motions towards of away from the detector. The changes in intensities can be mapped to distances using the gamma camera response function or look-up table. After the estimation of the transformation of the individual nodes, the SPECT volume can be updated by adjusting the positions of the individual lymph nodes.

The methodology has been evaluated using Cobalt point sources that emit similar radiation as lymph nodes, but are save to handle. In-plane and out-plane motion can be accurately estimated. The achieved accuracy of < 1 mm yields sufficient information to perform image-guided surgery. A possible clinical system is demonstrated in Fig. 2.4. Contributions are outlined in Appendix C.

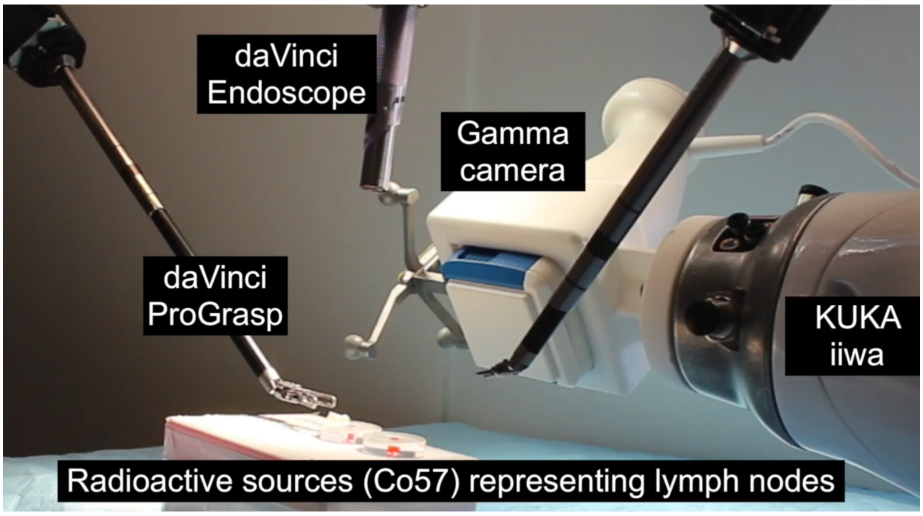

daVinci Endoscope
Gamma camera
daVinci ProGrasp
KUKA iiwa
Radioactive sources (Co57) representing lymph nodes

Fig. 2.4: Gamma camera mounted on a KUKA iiwa industrial lightweight robot, allows for automatic tracking of the displacement of radiopositive lymph nodes. In this setup, the KUKA iiwa will move the gamma camera so that it follows the movement performed by the da Vinci® surgical system while maintaining a perpendicular view of the radioactive sources.

CHAPTER 3

Image-Guided Surgery and Intra-Operative Imaging

3.1 Introduction and Definition

The use of imaging to support localization and targeting of regions of interest, and monitoring and controlling the treatment is frequently referred to as Image-Guided Therapy (IGT), and primarily applies to Image-Guided Surgery (IGS) and radiation therapy. The original concept utilized diagnostic imaging to inform the surgeon about the patient's anatomy. First computer aided systems combined diagnostic imaging, planning and intra-operative tracking techniques to guide the surgeon during the interventions. The goal is to use imaging to complement or replace the direct view of the surgical site to increase understanding of anatomy, increase accuracy, be less invasive, and lead to a better outcome. From the original static scenario, IGT has evolved over time and gained complexity. It now integrates a large variety of technology and image sources, and is centered around the pre-operative planning, monitoring of the procedure and the dynamic adaptation of the plan during the intervention itself. Intra-operative imaging does not only allow for this dynamic update of the plan, but also yields several other obvious advantages, including the visualization of deformations caused by patient positioning or the surgical intervention.

Image-Guided Surgery (IGS) In this thesis, the focus is on image-guided surgery, which is the subcategory of image-guided therapy dealing with guidance during surgery. One could furthermore distinguish between image-guided surgery and image-guided interventions, where the latter should emphasize that the interventions are performed without the traditional surgical access [7]. However, this differentiation is difficult in the context of surgical procedures, and in the following image-guided surgery and interventions are used interchangeably. Image-guided interventions can be categorized by the assistance provided by technology:

(i) Pre-operative or diagnostic imaging allows the observation of the state before the intervention. No context to the current surgical site is provided.

(ii) Intra-operative imaging enables images to be acquired during the in-

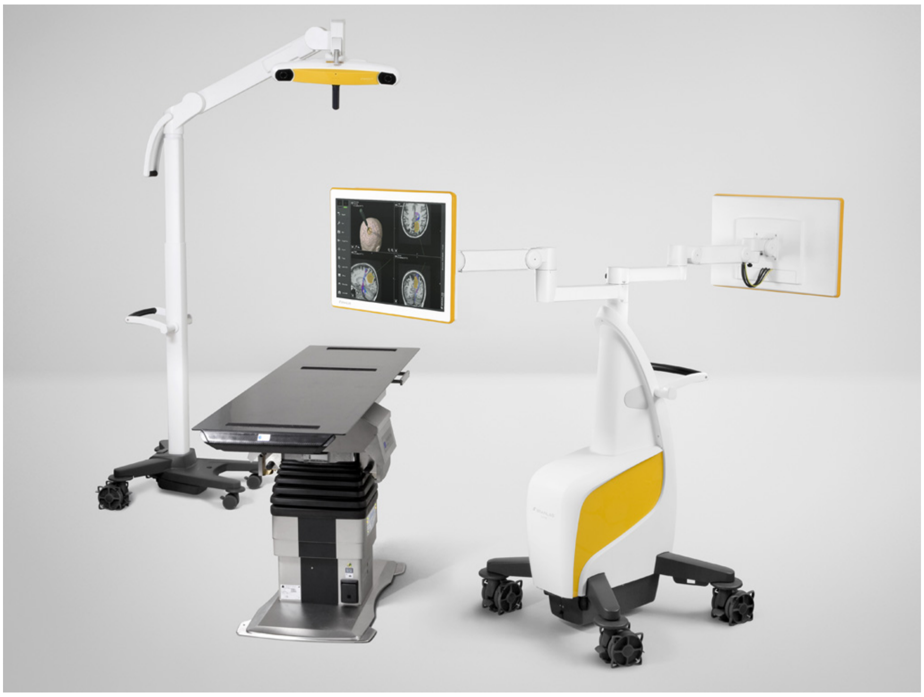


Fig. 3.1: A typical surgical navigation system uses pre-operative images, a tracking system (top left: stereo camera), a method to register the images with the surgical site, and tracked tools to guide the surgeon. While the accuracy improves for specific applications, common drawbacks include the additional hardware introduced to the operating room, limitations of registration and tracking, and increased duration. Image showing Brainlab Curve, courtesy of Brainlab AG.

terventions and presented to the surgeon. The placement of references or tools within the image provides some context and assists with the mental alignment of images and patient.

(iii) Surgical navigation systems use medical images, a tracking system, and methods to register the pre-operative images with the surgical site. These systems frequently rely on pre-operative images which are outdated once deformations are caused by the intervention. The relationship between tools and medical images provided by these systems may yield an increased surgical accuracy. Fig. 3.1 shows a typical surgical navigation system.

(iv) Augmented reality for surgery allows the surgeon to intuitively establish a relationship between medical images, patient's surface, medial images, surgical tools, and the surgeon's hands.

Surgical Navigation Systems Surgical navigation systems are commercially available and have been deployed in surgeries for approximately 20 years.

Today the use of navigation is standard of care for specific surgeries, such as neurosurgical interventions. In these specific applications, it can be shown to increase accuracy, or enable interventions which were previously not possible. However, in other applications, such as pedicle screw placement, the advantages may not be present [34].

Commonly reported drawbacks of surgical navigation systems are the increased setup time, which adds to the overall occupancy of the operating room. Furthermore, additional hardware in the operating room reduces the free moving space of the personnel, especially considering that most systems rely on optical tracking system that require a non-occluded line of sight between the cameras and the surgical site. Tracking targets on the patient and tools furthermore limit the access, and may require specially calibrated tools.

A typical system relies on pre-operative images, tracked surgical instruments, and a procedure to register the pre-operative images to the surgical site (e.g. tracking target, tracked stylus, image registration, etc.), see Fig. 3.1 [7]. Finally, the surgical navigation system can visualize the medical image data and render the tools, targets or references in the same view. This allows the surgeon to access a wide variety of views from different or complementing points of view.

Surgical Robotics The current generation of IGT integrates the use of robots to assist and support the surgeon. The focus of robot-assisted surgical systems are applications in which the introduction of complex mechanical systems poses a significant benefit to the surgeon and patient. In general, the applications of robots range from mimicking the surgeon's movements to performing tasks which would not otherwise be possible. The most prominent examples are the da Vinci® surgical system (Intuitive Surgical) and the Artis zeego (Siemens Healthcare). The first system is a teleoperator which helps the surgeon to perform an intervention by increasing the dexterity in areas that would not be accessible using traditional tools. For instance, transoral robotic surgeries allow the resection of areas of the tongue base or lingual tonsils through a minimally invasive procedure in contrast to the traditional approach which requires splitting the jaw. In contrast to these delicate interventions, the Artis zeego enables the dynamic movement of an entire C-arm around the surgical table. This permits imaging and reconstruction of arbitrary regions of the patient, adaptation of the imaging trajectory to reduce artifacts caused by metallic tools or implants, and the accurate repositioning of the C-arm to monitor the progression of an intervention.

The following sections will focus on novel approaches to intra-operative imaging during surgeries. First, the robotic SPECT will be addressed to guide robot assisted sentinel lymph node resections (Sec. 3.2. Next, in Sec. 3.3 contributions to robotic ultrasound for generic medical applications will be presented. Finally, this chapter will conclude with the fusion of color, depth, and X-ray imaging in combination with digitally rendered radiographs from CBCT volumes (Sec. 3.4, which is a representative of medical augmented reality systems.

3.2 Robotic Single-Photon Emission Computed Tomography for Robot-Assisted Interventions

The number of cases of cervical cancer has been increasing over the past decade. Modern technology, especially robotic assistants, enables minimally invasive resection of the tumor, which then leads to faster and better recovery. However, the involvement of lymph nodes in patients with early stage cervical cancer is one of the most important factors of survival. Fortunately, in up to 70% of the women surgically treated for this cancer, no involvements of the lymph nodes are present, which can only be determined after the resection of the lymph nodes. The currently performed radical removal of all lymph nodes is associated with a substantial treatment-related morbidity. Rather than removing all nodes, the resection can safely be limited to nodes in the direct drainage path, which referred to as SLNs. Finding these SLNs is difficult, but can be efficiently assisted using a nuclear imaging technique known as SPECT.

3.2.1 Related Work

3.2.2 Contribution: First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping (IEEE TMI 2016)

Sentinel lymph node biopsy of deeply seated or difficult to access is increasingly performed using tele-operated robot-assisted surgical systems. A typical operating room is shown in Fig. 3.2. In contrast to semi or fully autonomous robotic systems, tele-operated devices are fully controlled by the surgeon who sits at a console. The surgeon observes the surgical site via a laparoscope or endoscope (depending on the application), and his motions are replicated by the patient-side manipulators. The most commonly available robot-assisted system is the da Vinci® (Intuitive Surgical).

For the very first time, robotic SPECT for minimally invasive SLN mapping is presented, which now allows intra-operative SPECT reconstruction in the surgical site, taking advantage of the robot-assisted surgical systems for flexible in-patient data acquisition. The system integrates the novel miniaturized drop-in gamma detector, which can be picked up using a surgical robot da Vinci® . A novel combination of mechanical tracking using the robot's kinematics and vision-based tracking allows for a robust pose estimated. When combining the tracking information, a simple model of response function of the gamma probe, and the activity counts the volume can be reconstructed based on a MLEM. As the relationship between camera and reconstructed volume is already known, the image data can be thresholded and overlaid onto the optical camera view of the surgeon to guide the surgeon towards the important sentinel lymph nodes, as illustrated in Fig. 3.3. In conclusion, the surgeon can now use the system, scan the inside of the patient, and see the lymph nodes that may contain tumor cells. This is a cutting-edge and ground breaking novel approach which has been a *featured article* in the journal IEEE Transactions on Medical Imaging [18], see Appendix D for details.

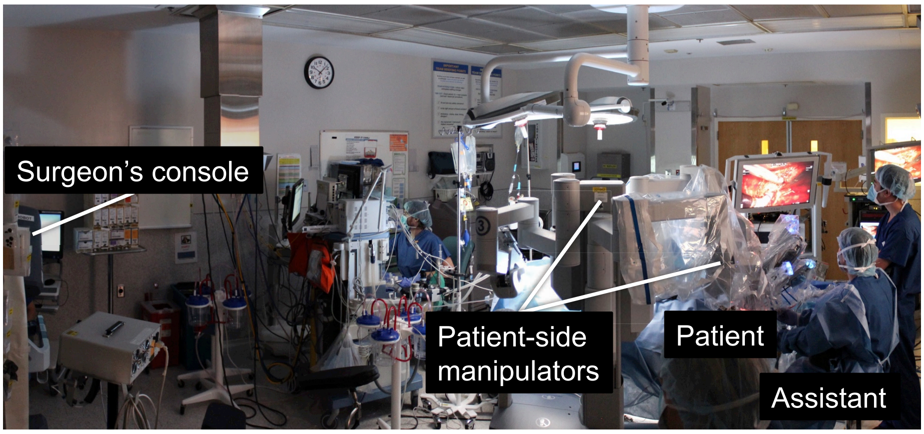

Surgeon's console
Patient-side manipulators
Patient
Assistant

Fig. 3.2: In an operating room with a robot-assisted surgical system, the surgeon's console is usually set up in a corner, the patient-side manipulators are mobile and moved towards the patient once the trocars are placed. In many interventions, a surgical assistant sits at the surgical table to support the primary surgeon using rigid laparoscopic tools or exchanging the robotic tools. The surgery can be observed by every surgical team member via large screens showing the same views as through the surgeon's console.

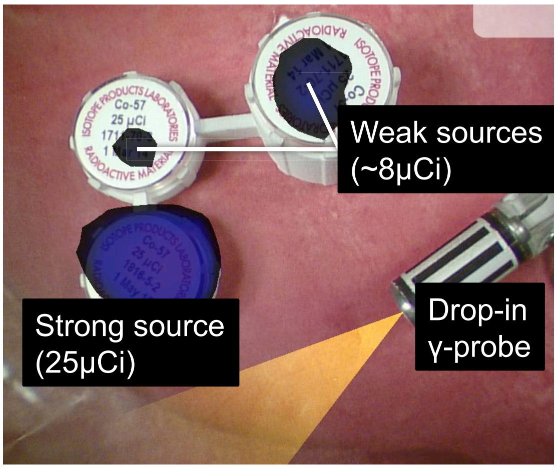

Weak sources (~8μCi)
Strong source (25μCi)
Drop-in γ-probe
Co-57
25 μCi
ISOTOPE PRODUCTS LABORATORIES
RADIOACTIVE MATERIAL

Fig. 3.3: The robotic SPECT reconstruction results can be directly overlaid on the surgeon's view of the surgical site, creating an immerse augmented reality visualization, and an intuitive surgical image-guidance.

This distinct combination of in-patient SPECT image acquisition and robotic surgery is a major contribution towards simplifying surgery, and improving patient care and safety. In conclusion, this approach improves patient care, patient safety, and the diagnosis of lymph node involvement, while simultaneously dramatically reducing the side effects of the intervention itself.

3.3 Robotic Ultrasound Acquisition

Ultrasound has become one of the standard medical imaging techniques and is widely used within diagnostic and interventional applications, especially because of its low cost, lack of ionizing radiation and ease of use in interventional settings. As described in Sec. 1.1, ultrasound systems usually comprise a handheld transducer, and the processing and visualization machine. The acquired 2D images are referred to as *B-mode images*, and represent a single plane. The pixel intensities correlate with the acoustic impedance, changes in impedance, and are overlaid with significant speckle patterns and other artifacts. Ultrasound imaging provides real-time information is is therefore especially suited to guide interventions.

Using native 3D transducers (*phased array transducer*) or tracking and compounding 2D images, 3D US images can be obtained [19]. Although they contain valuable information, the visualization of 3D ultrasound volumes is problematic resulting in mostly visualizing single slices of the volume. In clinical practice, 3D probes are not yet widely available. Using modern image registration techniques, the ultrasound images can be overlaid onto pre-operative diagnostic images, allowing the planning to be transferred to the surgical site.

US imaging also contains major drawbacks, for example, imaging quality is dependent on pressure, acoustic coupling, tissue density, angle and many other properties controlled by the user. Therefore, it requires expertise and adds a significant task load during surgery. This leads to a recently established field of research focusing on the simplification and automation of the ultrasound image acquisition.

Observing US imaging performed by physicians, the following requirements for an automated robotic ultrasound acquisition can be identified:

1. **Determining the region of interest:** Finding the patient's surface and region of interest, and planning the acquisition are complex tasks. This requires the identification and tracking of the patient's surface, the knowledge of the relation between the region of interest and the patient's surface, and the methodology to plan the transducer trajectory to cover the volume of interest. Tele-operated robotic ultrasound imaging requires the user to perform these tasks. Regarding the contributions, solutions proposed include acquisition of the patient's surface using depth sensors, image registration to align diagnostic imaging and current observations, and trajectory planning to attempt to acquire information of the entire volume of interest and its neighborhoods.
2. **Establishing acoustic coupling:** The key components consist of ultrasound gel, probe angle, and sufficient pressure. Some medical applications may limit the angle or not require the application of pressure, for instance transrectal ultrasound [44]. For other applications, the robotic system needs to be equipped with force/torque sensors to automatically apply the pressure required. Semi-automatic systems may allow the user to position the transducer, and simply use the robot as a mount. Most robotic systems still require a manual application of ultrasound gel.
3. **Acquiring the ultrasound volume:** After planning the trajectory and establishing the acoustic coupling, the image acquisition is performed. The first critical task is to compensation patient motion, which can be

performed using advanced robotic systems and a force controller. The second important task includes the continuous monitoring of the imaging quality, which is significantly more difficult to perform. First, ultrasound image quality is not well defined, nor is the relationship between good image quality and acquisition formally described. Second, using measures, such as confidence maps [29], the acquisition process must be updated in real-time.

4. **Repositioning to monitor progress:** Finally, the acquired images will be used to perform the intervention, which requires the repositioning of the robot. This task depends strongly on the application, and may include the positioning of a needle guide, following the surgical tool, or moving safely away from the surgical site. This task is frequently not addressed in published articles.

Current literature provides several attempts towards the robotic ultrasound acquisition. First approaches were published in the late 1990s, and focused on tele-operating ultrasound acquisition to provide a more comprehensive availability of expert controlled ultrasound imaging [59]. The goal was to examine of carotid arteries, for which a six degree of freedom robot was used. Using stepper motors, the transducer can be moved in a controlled fashion [14], but this approach does not solve the problem of acoustic coupling or controlled pressure application. The imaging volume is limited by the kinematics of the stepper motor. Utilizing an advanced robotic system with torque sensors, a tele-operated system with haptic feedback was presented in [8]. This system utilized the KUKA lightweight robot which has an additional degree of freedom of the kinematics and can estimate the applied forces using sensors in every joint. Beyond these approaches of generalizing the robotic ultrasound acquisition, several application dependent solutions are available, such as for transrectal ultrasound [44].

While freehand ultrasound scanning facilitates a fast and dynamic scanning and screening of several anatomies, modern compact and lightweight robotic arms can support the physician, i.e. by maneuvering a second probe or tool to the same position as the doctor does in real-time. By doing so, they can provide additional advantages, such as pressure compensation, hand tremor filtering, or automatic servoing and data acquisition based on a registration to tomographic imaging data.

To enable real-time guidance electromagnetic or optical tracking of the hand-held US transducer has been proposed. This allows the establishment of the spatial relationship of individual images, reconstruction of the US volume, and the registration with pre-interventional images. This has been presented for example in [45, 5, 62, 73]. The established registration enables the image-based guidance using US images by transferring features from pre-interventional imaging. Through this approach the physician is required to manually ensure good image quality during the intervention. The main limitations are lower imaging quality, small field of view, and difficult image interpretation [74]. Furthermore, the manual navigation and coordination with the medical device (e.g. biopsy needle) is cumbersome and requires designated training [51].

In contrast to hand-held US, robotic US image acquisition can incorporate vision-based servoing and potentially ensure good image quality, coverage of the entire volume of interest, and automatic re-positioning for guidance during the intervention. Such autonomous acquisitions could also facilitate clinical acceptance by simplifying the procedure as well as by reducing scanning time and the necessary amount of manual work. A prerequisite for automatic servoing is, however, a prior planning of a target point or area of interest. In a clinical setup, such a planning should always be performed based on anatomical data, as provided by MRI and CT, which results in the necessity of incorporating that imaging information directly into the planning procedure. While the planning can then be done in a minimum amount of time, time-consuming acquisitions of (several) 3D-ultrasound datasets can be done autonomously without requiring the presence of medical staff. Although there have been some general attempts to incorporate tomographic image information into robotic systems to improve the visualization of ultrasound image information to physicians, to our knowledge an integration of the data to enable an *automatic* planning and acquisitions of 3D-ultrasound has not considered so far.

Autonomous Robotic Ultrasound Acquisition: Volume Coverage A crucial component of the autonomous ultrasound acquisition is the estimation of the trajectory the robot needs to follow in order to cover the volume of interest. A possible solution is the modeling of the ultrasound image and the definition of a trajectory function with very few parameter [20]. This allows a fast and efficient optimization of the volume of coverage. To achieve this, the target volume is selected in pre-operative images, the patient's surface is extracted from the medical images and matched to the view of a three-dimensional camera. The algorithm further fits a parameterizable function to the surface and performs an optimization to ensure coverage of the volume.

Autonomous Robotic Ultrasound, Servoing and Manual Needle Insertion In [75], the focus is on manual needle injections into the lumbar spine [3].The proposed system comprises a motorized three-dimensional US transducer, and a force-controlled robotic arm. A novel visual servoing technique that combines three-dimensional image registration and real-time robotic servoing is introduced. To enable the transfer of the pre-operative plan, the system allows for image registration using the similarity measure LC^2 [17]. The needle trajectory plan can then be updated by performing fast incremental registration updates. Under realistic conditions a sub-millimeter targeting accuracy can be achieved.

Dual Robot Intervention: Ultrasound and Needle Insertion The introduction of a second robot to perform the image-guided needle insertion was presented in [31]. The contribution closes the gap between imaging, servoing, planning and action. While one robot performs US imaging in such a manner that the needle and target are always visible, the second robot introduces the needle and updates the trajectory to guarantee accurate targeting. One of the main contributions in [31] is the novel needle tracking technique. Due to temporal filtering of possible needle candidates in the US image, a robust and reliable needle tracking can be performed even in realistic materials.

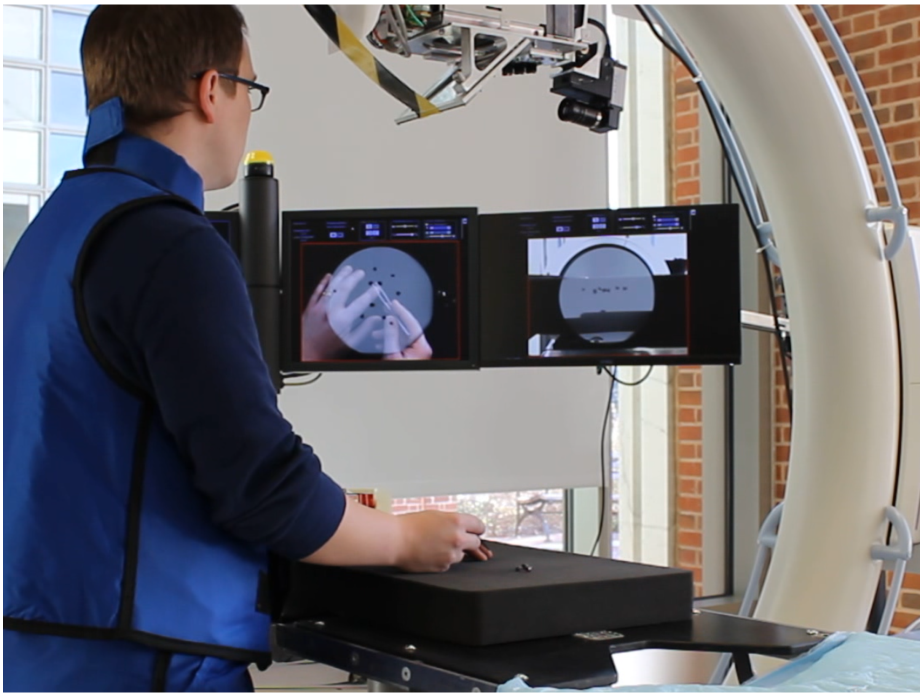


Fig. 3.4: The Camera-Augmented Mobile C-arm (CAMC) allows the overlay of X-ray images on live video images [50]. To achieve this the radiation source is unconventionally placed on top of the patient, enabling the optical centers of camera and X-ray to be virtually aligned through a mirror. The system allows the user to intuitively establish a relationship between hands and X-ray image.

3.4 RGBDX: Multi-Modal Imaging

One of the most difficult procedures during trauma surgeries is the placement of screws to reduce complex fractures. Using a vast amount of X-ray images (in some cases up to 246 images) the surgeon needs to drill a guide wire (k-wire) through the bone fragments, which later guides the screw to the correct placement. Rather than performing this placement in an open surgery, the wire is placed percutaneously through the muscle and other tissue into the bone (e.g. for pelvis). This placement of medical instruments deep inside soft tissue is a common, crucial and difficult task during orthopedic surgeries. Often the K-wire needs to be inserted through skin and muscle, and reach a bone at a pre-defined entry point and in a specific orientation. Once the K-wire is placed, screws are inserted following the guide wires and the operation can be completed.

In a typical scenario, the screw placement is first planned using CT images. The placement of the K-wire deep inside soft tissue is then performed using X-ray images which guide the surgeon to avoid damaging important structures such as spinal nerves, trunci and fasciculi, vegetative nerve system plexus, ureter, canalis inguinalis, and vena/arteria illiaca interna and externa which are located

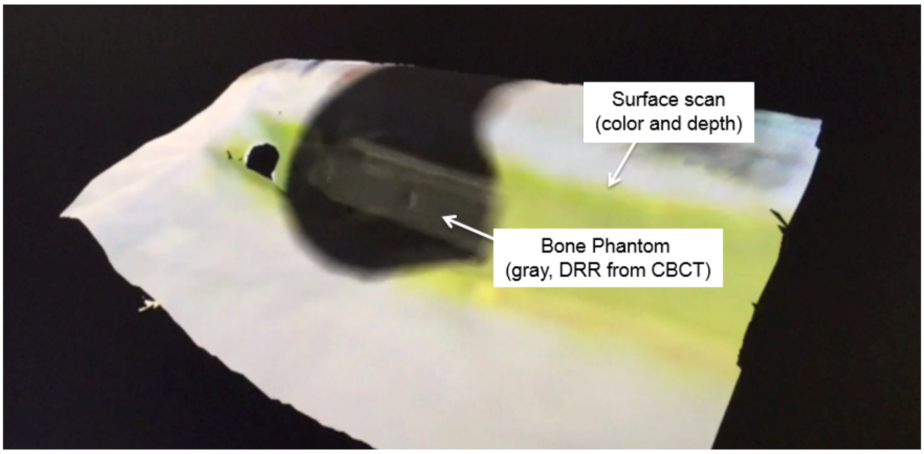

Surface scan
(color and depth)
Bone Phantom
(gray, DRR from CBCT)

Fig. 3.5: The three-dimensional view of patient's surface can be combined with the three-dimensional CBCT volume. This allows for a dynamic viewing of the patient's surface and anatomy.

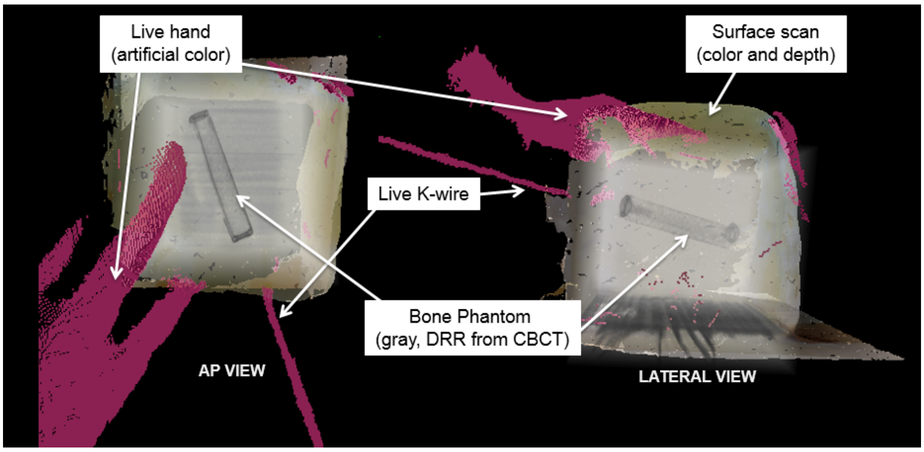

Live hand
(artificial color)
Surface scan
(color and depth)
Live K-wire
Bone Phantom
(gray, DRR from CBCT)
AP VIEW
LATERAL VIEW

Fig. 3.6: The platform for augmented reality for orthopedic and trauma interventions allows the surgeon to simultaneously see the patient's surface and anatomy (from CBCT) together with a live representation of tools and hands. Additionally, the surgeon can choose as many different views as desired. This image shows the simultaneous anterior-posterior (AP) and lateral view of the same scene.

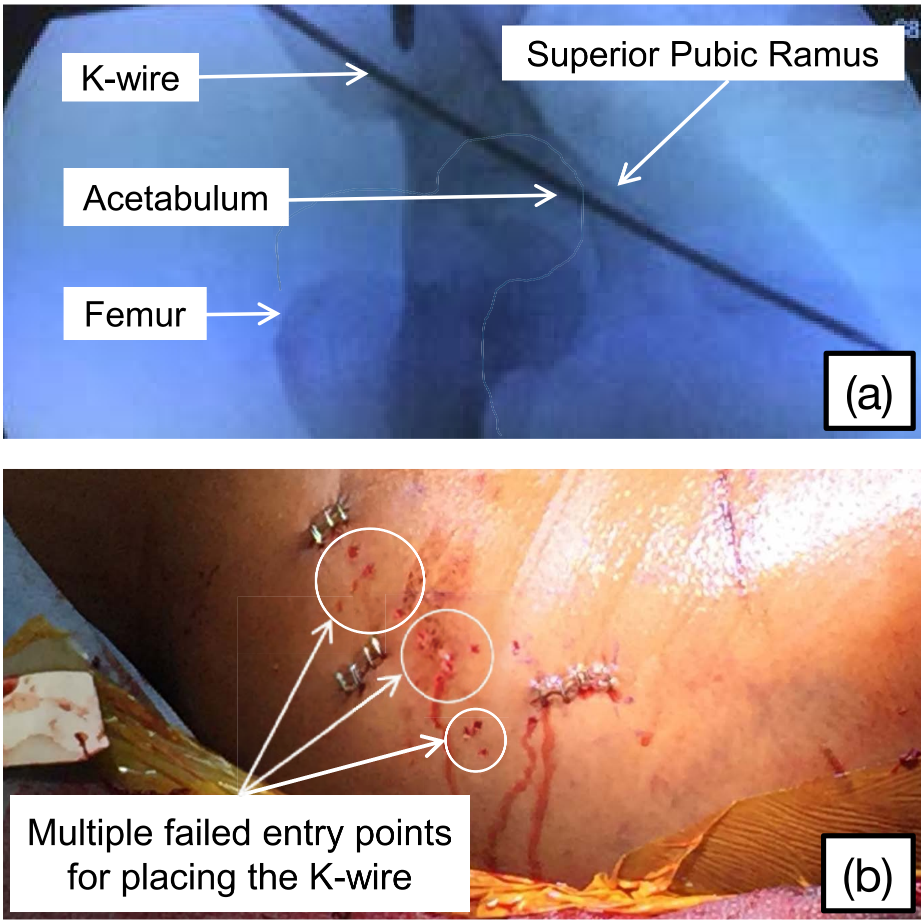

K-wire
Superior Pubic Ramus
Acetabulum
Femur
(a)
Multiple failed entry points for placing the K-wire
(b)

Fig. 3.7: The correct placement of guide wires in the pelvis is one of the most difficult steps during minimally invasive interventional fracture reduction. The placement is currently performed under X-ray guidance, see one image in *a)*. To achieve the correct trajectory, several attempts may be required to determine the correct entry point, as one can see in *b)* where multiple failed attempts are visible.

in the pelvic cavity. Because of the challenging three-dimensional placement of surgical tools using two-dimensional imaging such as fluoroscopy, the C-arm needs to be moved several times to allow the acquisition of images from different points of view. Therefore, patients in these surgical cases are often exposed to high radiation doses as surgeons may require many shots to correctly align and insert K-wires and screws. In conclusion, the issues of cumulative radiation dose, accuracy of inserted surgical materials, the operating time, and linked risk of infection are a risk to the patient's safety. Repeated exposure to radiation, long operating hours, and frustrating surgical work flow constitute a burden to surgical staff. Fig. 3.7 illustrates the challenges associated with placing K-wires during orthopedic surgeries.

3.4.1 Standard of Care and Related Work

The role of image guidance for K-wire placement in orthopedic surgeries has been vastly appreciated as crucial component of the standard of care. Nevertheless, the challenge of mental alignment together with hand-eye coordination is the primary cause of the need for continuous image validation. This does not only result in high X-ray exposure, but also longer surgical durations, more effort and frustration for the surgeon and the technical staff, longer operation room occupancy, and time consuming training requirement for surgical residents.

During computer-assisted interventions, surgical navigation systems are utilized in order to support the surgeon's mental alignment and guide him during surgery. Most current systems use pre-operative X-ray and CT volumes, infrared optical tracking systems, and tracking targets on tools and the patient, to provide a visualization of the tool relative to the patient's anatomy. Some systems include the use of a C-arm and enable tracking of tools relative to intraoperatively acquired X-ray images [32]. After the guided procedure, the validation of the placement is performed using conventional X-ray imaging. Beyond additional cost, time consumption, spatial requirements, and work flow interruptions, traditional navigation systems do not provide an intuitive relation to reality as only pre-interventional images and tools are visualized. Imagine driving your car using only a navigation system with precomputed maps, without looking outside, but with the constraint to follow the map down to the inch. When using surgical navigation systems, the surgeons are expected to accept systems without further enhancement through X-ray imaging onto the dynamic patient's anatomy. Navigation systems provide excellent accuracy, but they are not systematically developed to complement pre- and intra-operative data. Therefore, a large portion of orthopedic surgeries are still performed without computer-assisted navigation systems, and the surgeons rely on acquiring many intraoperative images to observe every minute change and to ensure accuracy.

Alternatively, an intuitive guidance can be achieved using a mirror and video camera mounted onto the C-arm [50, 47]. This establishes an augmented reality view of the surgical site, and is illustrated in Fig. 3.4. After a single calibration procedure, the video image can be warped to fit the X-ray geometry. Subsequently, the video view can be augmented with the X-ray images, which provides an intuitive optical feedback to the surgeon. Cadaver studies show a reduction of radiation dose and duration, while the accuracy improves [36, 63]. Similar improvements can be observed during orthopedic and trauma procedures [49, 12]. A major drawback it the system's limitation to two-dimensional video views

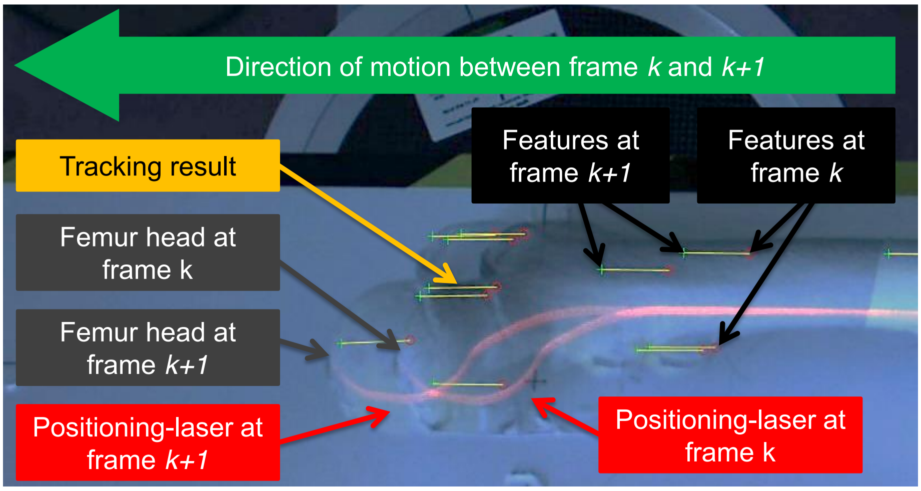

Direction of motion between frame *k* and *k+1*
Tracking result
Features at frame *k+1*
Features at frame *k*
Femur head at frame k
Femur head at frame *k+1*
Positioning-laser at frame *k+1*
Positioning-laser at frame k

Fig. 3.8: The displacement of the patient can be estimated based on color images. After co-registration of camera and X-ray origin, this information can be used to align multiple CBCT volumes. This figure blends two frames and shows feature correspondences to estimate the movement of the patient. From both frames, the positioning-laser (red) and natural surface features are extracted. The tracking results are illustrated as yellow lines.

and the significantly reduced free moving space of the surgeon due to the mirror construction.

3.4.2 Contribution: Vision-based Intra-Operative Cone-Beam CT Stitching for Non-overlapping Volumes (MICCAI 2015)

The combination of a mobile X-ray C-arm and color camera also allows for intra-operative alignment of CBCT volumes. In contrast to state of art, the novel method is applicable to non-overlapping volumes and does not require additional markers or references. Interventional X-ray imaging is widely in use, and stitching techniques for 2D images have been presented in the past decade. In this paper, we present a system and methodology to align two or more Cone-Beam CT (CBCT) volumes, by only utilizing an optical camera which is rigidly mounted to the C-arm. First, a one-time calibration needs to be performed to recover the transformation between camera origin and positioning laser mounted in the C-arm base. Placing a checkerboard pattern in such a manner that the laser plane intersects the checkerboard origin allows to obtain points on the laser plane in camera coordinates. Using Random Sample Consensus (RANSAC), a plane can be fitted to the set of points observed. This allows for an accurate calibration of C-arm base to camera.

While moving the C-arm from one position to the next, the algorithm is designed to track features on the patient surface, and recover the scale by identifying the positioning laser which are built in the X-ray machine. First, the two-dimensional features are detected. As the features on the skin are natural shapes and resemble blob-like structures, Speeded Up Robust Features (SURF) are used. Outliers are rejected by estimating the fundamental matrix **F**. Feature correspondences and the laser line are shown in Fig. 3.8. Next, the three-dimensional coordinates are recovered by estimating the scale based the laser line observed on the surface. This can be done after the three-dimensional coordinates of the dilated laser line are recovered by projecting them back using the Moore-Penrose pseudo-inverse of the camera projecting matrix and intersecting the points up to scale with the virtual laser plane. Finally, the three-dimensional transformation is estimated based on the three-dimensional feature points and a plausibility test is performed by matching the laser lines observed in both images.

Experimental validation using a bone and fiducial phantom confirms the validity of the methodology and shows sub-millimeter accuracy in the experiments performed. This is novel concept allows the alignment of CBCT volumes without the introduction of patient-side reference markers, while the traditional surgical workflow remains intact. Further details can be found in Appendix E and E.

3.4.3 Contribution: Calibration of RGBD Camera and Cone-Beam CT for 3D Intraoperative Mixed Reality Visualization (IJCARS 2016)

Interventional three-dimensional augmented reality is the future of intraoperative visualization and the path to simplifying surgery. It provides intuitive visualization of human tissue, bones, and muscles extracted from CBCT data to be overlaid on real views of surgical site (see Fig. 3.5 and 3.6). We have developed a prototype for augmented reality tool placement inside bone. To achieve the calibration, the surface data is recorded using a three-dimensional color camera while the C-arm automatically rotates around the patient to obtain the projection images. An example of a calibration phantom can be seen in Fig. 3.9a).

Our technology consists of a software platform that registers Fast Point Feature Histogram (FPFH) descriptors from point clouds and patient data characteristics extracted from CBCT volumes utilizing the SAmple Consensus Initial Alignment (SAC-IA). However, this process requires extensive pre-processing of both data sources in order to obtain artifact-free surfaces. For the three-dimensional data recorded using the camera, the a model of the calibration phantom is fit to the data and outliers are removed. Alternatively, the outliers can be removed by filtering the point clouds. Processing of the CBCT volume includes simple thresholding and detection of the most outer surface of the phantom. The calibration is then obtained after an Iterative Closest Point (ICP) refinement [35].

The methodology is evaluated using different three-dimensional color cameras as well as different distances between the camera and the calibration phantom. For each configuration the target registration error (TRE) is computed. It can be shown that the quality of depth information has an significant impact on the overall calibration accuracy, which ranges between 2.58 mm to 7.42 mm depending on which camera is used. Furthermore, different calibration phantoms and their impact on calibration is evaluated. The results indicate that smooth,

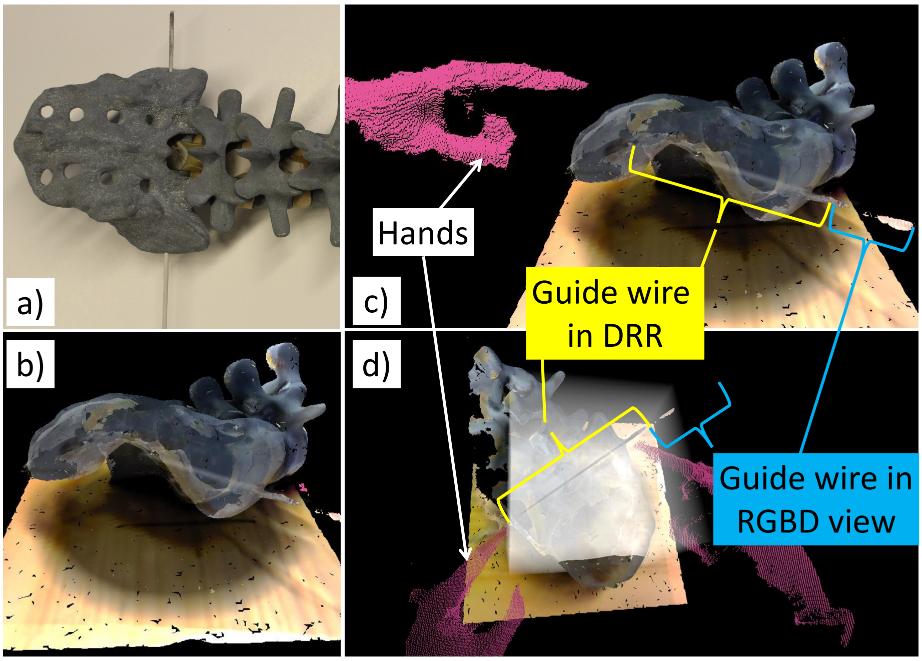


Fig. 3.9: The calibration phantom (a) can be an arbitrary object, such as this spine model. The calibration allows the overlay of simulated X-ray images and the real phantom surfaces, as shown in (b). In (c) and (d) the three-dimensional surface reconstruction, simulated X-ray images and the live three-dimensional information showing the user's hand are shown.

round objects with a matte finish allow for better depth images, which directly influences the calibration accuracy.

Finally, the mixed reality perception is achieved by dynamically adjusting the opacities and blending the DRRs, patient's surface, surgeon's hand and tools; as illustrated in Fig. 3.9c) and d). All algorithms are implemented on GPU, allowing real-time performance. The contribution towards novel calibration of RGBD and CBCT imaging is presented in [35], and in Appendix G.

3.4.4 Contribution: Pre-Clinical Usability Study of Multiple Augmented Reality Concepts for K-Wire Placement (IJCARS 2016)

The technology proposed reduces the time required for complex surgical procedures and the highly risky trial and error approach for tool placement. Most crucially, it improves patient safety by reducing the possibility of complications (such as nerve and muscle damage, injury of the intestine, vascular damage, clotting and bleeding) and lessening the number of attempts to place a guide wire. In contrast to the current systems, with our technology, the surgeon sees the intraoperatively acquired X-ray or CBCT images, the patient's surface, and the surgeon's hands and tools in real-time (Fig. 3.6). In short, we provide the surgeon with a unique and intuitive view of the patient's anatomy in relation to the surgeon's hands and necessary tools, improving targeting and adjustment of trajectories. This immediate direct feedback results in reduced time under anesthesia, exposure to radiation, and time in the operating room for both surgeon and patient [15, 16]. The simulated procedure for K-wire placement was performed by 7 trained surgeons utilizing superior pubic ramus phantoms, showing significantly reduced duration, radiation dose, and number of X-ray images, which is illustrated in Fig. 3.10 [15, 16].

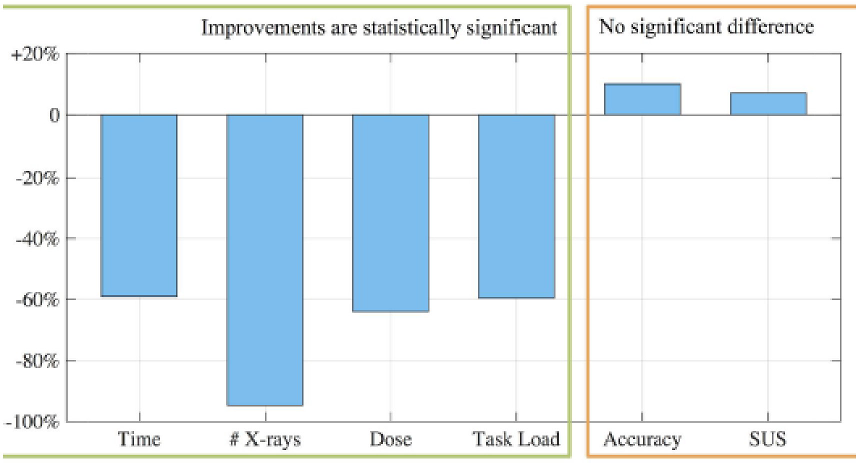


Fig. 3.10: The three-dimensional Augmented Reality visualization enables surgeons to perform K-wire placements faster (Time), with less X-ray images (# X-rays), and an overall lower dose (Dose). Additionally, the surgical task load index (Task Load) was reduced, while the accuracy remained nearly constant.

Our system comprises a traditional X-ray machine (C-arm), a three-dimensional camera mounted on this X-ray machine, and generally available CT images to guide the surgeon. Rather than seeing simple 2D Xray images, our system shows the surgeon a three-dimensional view of the bones, the drill, the patient surface and even the surgeon's hands in realtime. This *Superman* view, referred to interventional three-dimensional augmented reality, has shown to reduce duration, radiation dose, number of X-ray images, and complications in our pre-clinical experiments. In summary, our system increases patient safety and represents the future of interventional X-ray imaging. The various contributions are presented in [15, 16], and in Appendix F.

Bibliography

[1] Antoni Buades, Bartomeu Coll, and J-M Morel. “A non-local algorithm for image denoising”. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE. 2005, pp. 60–65.

[2] B. Carelsen, R. Haverlag, D.Th. Ubbink, J.S.K. Luitse, and J.C. Goslings. “Does intraoperative fluoroscopic 3D imaging provide extra information for fracture surgery?” English. In: *Archives of Orthopaedic and Trauma Surgery* 128.12 (2008), pp. 1419–1424. ISSN: 0936-8051.

[3] Pain Center and Laxmaiah Manchikanti. “A systematic review and best evidence synthesis of effectiveness of therapeutic facet joint interventions in managing chronic spinal pain”. In: *Pain physician* 18 (2015), E535–E582.

[4] Jenghwa Chang, Lili Zhou, Song Wang, and K. S. Clifford Chao. “Panoramic cone beam computed tomography”. In: *Medical Physics* 39.5 (2012), pp. 2930–2946.

[5] Elvis CS Chen, Parvin Mousavi, Sean Gill, Gabor Fichtinger, and Purang Abolmaesumi. “Ultrasound guided spine needle insertion”. In: (2010), pp. 762538–762538.

[6] Amalia Cifor, Laurent Risser, Mattias P Heinrich, Daniel Chung, and Julia A Schnabel. “Rigid Registration of Untracked Freehand 2D Ultrasound Sweeps to 3D CT of Liver Tumours”. In: *Abdominal Imaging. Computation and Clinical Applications*. Springer, 2013, pp. 155–164.

[7] Kevin Cleary and Terry M Peters. “Image-guided interventions: technology review and clinical applications”. In: *Annual review of biomedical engineering* 12 (2010), pp. 119–142.

[8] François Conti, Jaeheung Park, and Oussama Khatib. “Interface design and control strategies for a robot assisted ultrasonic examination system”. In: *Experimental Robotics*. Springer. 2014, pp. 97–113.

[9] Frank M Corl, Melissa R Garland, and Elliot K Fishman. “Role of computer technology in medical illustration”. In: *American Journal of Roentgenology* 175.6 (2000), pp. 1519–1524.

[10] D. De Nigris, D. Collins, and T. Arbel. “Fast and Robust Registration Based on Gradient Orientations: Case Study Matching Intra-Operative Ultrasound to Pre-Operative MRI in Neurosurgery”. In: *IPCAI* (2012), pp. 125–134.

[11] Frank Dennerlein, Frédéric Noo, Wolfgang Härer, Joachim Hornegger, and Günter Lauritsch. “Constriction of cone-beam artifacts by the Z-smart reconstruction method”. In: 6 (2007), pp. 4090–4096.

[12] Benoit Diotte, Pascal Fallavollita, Lejing Wang, Simon Weidert, Peter-Helmut Thaller, Ekkehard Euler, and Nassir Navab. “Radiation-free drill guidance in interlocking of intramedullary nails”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 18–25.

[13] LA Feldkamp, LC Davis, and JW Kress. “Practical cone-beam algorithm”. In: *Journal of the Optical Society of America A* 1.6 (1984), pp. 612–619.

[14] Aaron Fenster, Donal B Downey, and H Neale Cardinal. “Three-dimensional ultrasound imaging”. In: *Physics in medicine and biology* 46.5 (2001), R67.

[15] Marius Fischer, Bernhard Fuerst, Sing Chun Lee, Javad Fotouhi, Severine Habert, Simon Weidert, Ekkehard Euler, Greg Osgood, and Nassir Navab. “Preclinical usability study of multiple augmented reality concepts for K-wire placement”. In: *International Journal of Computer Assisted Radiology and Surgery* (2016), pp. 1–8.

[16] Javad Fotouhi, Bernhard Fuerst, Sing Chun Lee, Matthias Keicher, Marius Fischer, Simon Weidert, Ekkehard Euler, Nassir Navab, and Greg Osgood. “Interventional 3D Augmented Reality for Orthopedic and Trauma Surgery”. In: (2016).

[17] Bernhard Fuerst, Wolfgang Wein, Markus Müller, and Nassir Navab. “Automatic ultrasound–MRI registration for neurosurgery using the 2D and 3D LC2 Metric”. In: *Medical image analysis* (2014).

[18] Bernhard Fuerst, Julian Sprung, Francisco Pinto, Benjamin Frisch, Thomas Wendler, Herve Simon, Laurent Mengus, Nynke van den Berg, Henk van der Poel, Fijs van Leeuwen, et al. “First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping”. In: *IEEE Transactions on Medical Imaging* 35.3 (Mar. 2016), pp. 830–838. ISSN: 0278-0062. DOI: `10.1109/TMI.2015.2498125`.

[19] Andrew Gee, Richard Prager, Graham Treece, and Laurence Berman. “Engineering a freehand 3D ultrasound system”. In: *Pattern Recognition Letters* 24.4 (2003), pp. 757–777.

[20] Christoph Graumann, Bernhard Fuerst, Christoph Hennersperger, Felix Bork, and Nassir Navab. “Robotic ultrasound trajectory planning for volume of interest coverage”. In: (2016), pp. 736–741.

[21] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[22] Mattias P Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V Gleeson, Sir Michael Brady, and Julia A Schnabel. "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration". In: *Medical Image Analysis* 16.7 (2012), pp. 1423–1435.

[23] Mattias Paul Heinrich, Mark Jenkinson, Bartlomiej W Papiez, Michael Brady, and Julia A Schnabel. "Towards realtime multimodal fusion for image-guided interventions using self-similarities". In: *Medical Image Computing and Computer-Assisted Intervention.* Springer, 2013, pp. 187–194.

[24] Mattias Paul Heinrich, Mark Jenkinson, Bartlomiej W. Papież, Sir Michael Brady, and Julia A. Schnabel. "Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part I". In: (2013). Ed. by Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, pp. 187–194. DOI: `10.1007/978-3-642-40811-3_24`. URL: `http://dx.doi.org/10.1007/978-3-642-40811-3_24`.

[25] M. Holden. "A Review of Geometric Transformations for Nonrigid Body Registration". In: *IEEE Transactions on Medical Imaging* 27.1 (Jan. 2008), pp. 111–128. ISSN: 0278-0062. DOI: `10.1109/TMI.2007.904691`.

[26] D. H. Hristov and B. G. Fallone. "A grey level image alignment algorithm for registration of portal images and digitally reconstructed radiographs". In: *Medical Physics* 23.1 (1996), pp. 75–84.

[27] X. Huang, N. Hill, J. Ren, G. Guiraudon, D. Boughner, and T. Peters. "Dynamic 3D Ultrasound and MR Image Registration of the Beating Heart". In: *Medical Image Computing and Computer-Assisted Intervention 2005* (2005), pp. 171–178.

[28] J Trevor Hughes. "The Edwin Smith Surgical Papyrus: an analysis of the first case reports of spinal cord injuries". In: *Spinal Cord* 26.2 (1988), pp. 71–82.

[29] A. Karamalis, W. Wein, and N. Navab. "Fast Ultrasound Image Simulation Using the Westervelt Equation". In: *Medical Image Computing and Computer-Assisted Intervention 2010* (2010), pp. 243–250.

[30] Athanasios Karamalis, Wolfgang Wein, Oliver Kutter, and Nassir Navab. "Fast hybrid freehand ultrasound volume reconstruction". In: *Proc. of SPIE Vol.* Vol. 7261. 2009, pp. 726114–1.

[31] Risto Kojcev, Bernhard Fuerst, Oliver Zettinig, Javad Fotouhi, Sing Chun Lee, Benjamin Frisch, Russell Taylor, Edoardo Sinibaldi, and Nassir Navab. "Dual-robot ultrasound-guided needle placement: closing the planning-imaging-action loop". In: *International journal of computer assisted radiology and surgery* (2016), pp. 1–9.

[32] Everine B van de Kraats, Theo van Walsum, Lance Kendrick, Niels J Noordhoek, and Wiro J Niessen. "Accuracy evaluation of direct navigation with an isocentric 3D rotational X-ray system". In: *Medical image analysis* 10.2 (2006), pp. 113–124.

[33] David N Krag, Stewart J Anderson, Thomas B Julian, Ann M Brown, Seth P Harlow, Joseph P Costantino, Takamaru Ashikaga, Donald L Weaver, Eleftherios P Mamounas, Lynne M Jalovec, Thomas G Frazier, R Dirk Noyes, Andre Robidoux, Hugh MC Scarth, and Norman Wolmark. "Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial". In: *The Lancet Oncology* 11.10 (Oct. 2010).

[34] Michael Kraus, Julia Weiskopf, Jens Dreyhaupt, Gert Krischak, and Florian Gebhard. "Computer-Aided Surgery Does Not Increase the Accuracy of Dorsal Pedicle Screw Placement in the Thoracic and Lumbar Spine: A Retrospective Analysis of 2,003 Pedicle Screws in a Level I Trauma Center". In: *Global spine journal* 5.2 (2015), p. 93.

[35] Sing Chun Lee, Bernhard Fuerst, Javad Fotouhi, Marius Fischer, Greg Osgood, and Nassir Navab. "Calibration of RGBD camera and cone-beam CT for 3D intra-operative mixed reality visualization". In: *International journal of computer assisted radiology and surgery* 11.6 (2016), pp. 967–975.

[36] Roberto Londei, Marco Esposito, Benoit Diotte, Simon Weidert, Ekkehard Euler, Peter Thaller, Nassir Navab, and Pascal Fallavollita. "Intraoperative augmented reality in distal locking". In: *International journal of computer assisted radiology and surgery* (2015), pp. 1–9.

[37] G. H. Lyman, A. E. Giuliano, M. R. Somerfield, A. B. Benson, D. C. Bodurka, H. J. Burstein, A. J. Cochran, H. S. Cody, S. B. Edge, S. Galper, J. A. Hayman, T. Y. Kim, C. L. Perkins, D. A. Podoloff, V. H. Sivasubramaniam, R. R. Turner, R. Wahl, D. L. Weaver, A. C. Wolff, and E. P. Winer. "American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer". In: *Journal of Clinical Oncology* 23.30 (Oct. 2005), pp. 7703–7720.

[38] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. "Multimodality image registration by maximization of mutual information". In: *Medical Imaging, IEEE Transactions on* 16.2 (1997), pp. 187–198.

[39] P Mah, TE Reeves, and WD McDavid. "Deriving Hounsfield units using grey levels in cone beam computed tomography". In: *Dentomaxillofacial Radiology* (2014).

[40] JB Maintz and M.A. Viergever. "A survey of medical image registration". In: *Medical Image Analysis* 2.1 (1998), pp. 1–36.

[41] Stanislaw Majewski, Andrew G Weisenberger, and Randolph F Wojcik. *Mini gamma camera, camera system and method of use.* US Patent 6,271,525. Aug. 2001.

[42] A. Al-Mayah, J. Moseley, and K Brock. "Contact surface and material nonlinearity modeling of human lungs". In: *Physics in Medicine and Biology* 53 (2008), pp. 305–317.

[43] A. Al-Mayah, J. Moseley, M. Velec, and KK Brock. "Sliding characteristic and material compressibility of human lung: Parametric study and verification". In: *Medical Physics* 36 (2009), pp. 4625–4633.

[44] Omid Mohareri, Guy Nir, Julio Lobo, Richard Savdie, Peter Black, and Septimiu Salcudean. “A System for MR-Ultrasound Guidance during Robot-Assisted Laparoscopic Radical Prostatectomy”. In: (2015), pp. 497–504.

[45] John Moore, Colin Clarke, Daniel Bainbridge, Chris Wedlake, Andrew Wiles, Danielle Pace, and Terry Peters. “Image guidance for spinal facet injections using tracked ultrasound”. In: (2009), pp. 516–523.

[46] Jorge J Moré. “The Levenberg-Marquardt algorithm: implementation and theory”. In: (1978), pp. 105–116.

[47] N. Navab, S.-M. Heining, and J. Traub. “Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications”. In: *Medical Imaging, IEEE Transactions on* 29.7 (July 2010), pp. 1412–1423. ISSN: 0278-0062.

[48] Nassir Navab, Tobias Blum, Lejing Wang, Asli Okur, and Thomas Wendler. “First Deployments of Augmented Reality in Operating Rooms”. In: *Computer* 99 (2012), pp. 48–55. ISSN: 0018-9162. DOI: `http://doi.ieeecomputersociety.org/10.1109/MC.2012.75`.

[49] Nassir Navab, Tobias Blum, Lejing Wang, Aslı Okur, and Thomas Wendler. “First deployments of augmented reality in operating rooms”. In: *Computer* 7 (2012), pp. 48–55.

[50] Nassir Navab, Matthias Mitschke, and Oliver Schütz. “Camera-augmented mobile C-arm (CAMC) application: 3D reconstruction using a low-cost mobile C-arm”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI’99*. Springer. 1999, pp. 688–697.

[51] Getúlio Rodrigues de Oliveira Filho. “The construction of learning curves for basic skills in anesthetic procedures: an application for the cumulative sum method”. In: *Anesthesia & Analgesia* 95.2 (2002), pp. 411–416.

[52] Ruben Pauwels, Kazuyuki Araki, JH Siewerdsen, and Saowapak S Thongvigitmanee. “Technical aspects of dental CBCT: state of the art”. In: *Dentomaxillofacial Radiology* 44.1 (2015).

[53] Francisco Pinto, Bernhard Fuerst, Benjamin Frisch, and Nassir Navab. “Radiopositive Tissue Displacement Compensation for SPECT-guided Surgery”. In: (2015). Ed. by Nassir Navab, Joachim Hornegger, M. William Wells, and F. Alejandro Frangi, pp. 536–543. DOI: `10.1007/978-3-319-24571-3_64`. URL: `http://dx.doi.org/10.1007/978-3-319-24571-3_64`.

[54] Michael JD Powell. “The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives”. In: *Cambridge Report NA2009/06, University of Cambridge* (2009).

[55] Michael JD Powell. “The NEWUOA software for unconstrained optimization without derivatives”. In: (2006), pp. 255–297.

[56] L Risser, F Vialard, H Baluwala, and J Schnabel. “Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions”. In: *Medical Image Analysis* 17.2 (2013), pp. 182–193.

[57] Hassan Rivaz and D Collins. “Self-similarity Weighted Mutual Information: A New Nonrigid Image Registration Metric”. In: *Medical Image Computing and Computer-Assisted Intervention* (2012), pp. 91–98.

[58] A. Roche, X. Pennec, G. Malandain, and N. Ayache. “Rigid Registration of 3-D Ultrasound with MR Images: A New Approach Combining Intensity and Gradient Information”. In: *Medical Imaging, IEEE Transactions on* 20.10 (2001), pp. 1038–1049.

[59] SE Salcudean, G Bell, Simon Bachmann, Wen-Hong Zhu, Purang Abolmaesumi, and Peter D Lawrence. “Robot-assisted diagnostic ultrasound–design and feasibility experiments”. In: (1999), pp. 1062–1071.

[60] A. Schmidt-Richberg, R. Werner, H. Handels, and J. Ehrhardt. “Estimation of Slipping Organ Motion by Registration with Direction-Dependent Regularization”. In: *Medical Image Analysis* 16 (2011), pp. 150–159.

[61] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. “Deformable medical image registration: A survey”. In: *Medical Imaging, IEEE Transactions on* 32.7 (2013), pp. 1153–1190.

[62] Denis Tran, Allaudin A Kamani, Elias Al-Attas, Victoria A Lessoway, Simon Massey, and Robert N Rohling. “Single-operator real-time ultrasound-guidance to aim and insert a lumbar epidural needle”. In: *Canadian Journal of Anesthesia/Journal canadien d’anesthésie* 57.4 (2010), pp. 313–321.

[63] Joerg Traub, Tim Hauke Heibel, Philipp Dressel, Sandro Michael Heining, Rainer Graumann, and Nassir Navab. “A multi-view opto-Xray imaging system”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*. Springer, 2007, pp. 18–25.

[64] A. G. Van der Zee, M. H. Oonk, J. A. De Hullu, A. C. Ansink, I. Vergote, R. H. Verheijen, A. Maggioni, K. N. Gaarenstroom, P. J. Baldwin, E. B. Van Dorst, J. Van der Velden, R. H. Hermans, H. van der Putten, P. Drouin, A. Schneider, and W. J. Sluiter. “Sentinel node dissection is safe in the treatment of early-stage vulvar cancer”. In: *Journal of Clinical Oncology* 26.6 (Feb. 2008), pp. 884–889.

[65] Christoph Vetter, Tobias Lasser, Asli Okur, and Nassir Navab. “1D-3D Registration for Intra-Operative Nuclear Imaging in Radio-Guided Surgery”. In: *IEEE Transactions on Medical Imaging* 34.2 (Feb. 2015), pp. 608–617. DOI: `10.1109/tmi.2014.2363551`. URL: `http://dx.doi.org/10.1109/TMI.2014.2363551`.

[66] Sergi Vidal-Sicart, Beatriz Doménech, Blanca Luján, Jaume Pahisa, Aureli Torné, Sergio Martínez-Román, Juan Antonio Lejárcegui, Pere Fusté, Jaume Ordi, Pilar Paredes, and Francesca Pons. “Sentinel node in gynaecological cancers. Our experience”. In: *Revista Española de Medicina Nuclear (English Edition)* 28.5 (Sept. 2009), pp. 221–228. DOI: `10.1016/s1578-200x(09)70022-1`. URL: `http://dx.doi.org/10.1016/S1578-200X(09)70022-1`.

[67] Paul Viola and William M. Wells III. “Alignment by Maximization of Mutual Information”. In: *International Journal of Computer Vision* 24.2 (1997), pp. 137–154. ISSN: 1573-1405. DOI: `10.1023/A:1007958904918`. URL: `http://dx.doi.org/10.1023/A:1007958904918`.

[68] Christian Wachinger and Nassir Navab. “Entropy and Laplacian images: Structural representations for multi-modal registration”. In: *Medical Image Analysis* 16.1 (2012), pp. 1–17.

[69] W. Wein, S. Brunke, A. Khamene, M.R. Callstrom, and N. Navab. “Automatic CT-Ultrasound Registration for Diagnostic Imaging and Image-Guided Intervention”. In: *Medical Image Analysis* 12.5 (2008), p. 577.

[70] Thomas Wendler, Alexander Hartl, Tobias Lasser, Joerg Traub, Farhad Daghighian, Sybille Ziegler, and Nassir Navab. “Towards intra-operative 3D nuclear imaging: reconstruction of 3D radioactive distributions using tracked gamma probes”. In: *Medical Image Computing and Computer-Assisted Intervention.* LNCS 4792.2 (Oct. 2007). Ed. by N. Ayache, S. Ourselin, and A. Maeder, pp. 909–917.

[71] Thomas Wendler, Ken Herrmann, Andreas Schnelzer, Tobias Lasser, Joerg Traub, Olivier Kutter, Alexandra Ehlerding, Klemens Scheidhauer, Tibor Schuster, Marion Kiechle, et al. “First demonstration of 3-D lymphatic mapping in breast cancer using freehand SPECT”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 37.8 (2010), pp. 1452–1461.

[72] Sandra L Wong, Charles M Balch, Patricia Hurley, Sanjiv S Agarwala, Timothy J Akhurst, Alistair Cochran, Janice N Cormier, Mark Gorman, Theodore Y Kim, Kelly M McMasters, et al. “Sentinel lymph node biopsy for melanoma: American Society of Clinical Oncology and Society of Surgical Oncology joint clinical practice guideline”. In: *Journal of Clinical Oncology* 30.23 (2012), pp. 2912–2918.

[73] Charles XB Yan, Benoît Goulet, Julie Pelletier, Sean Jy-Shyang Chen, Donatella Tampieri, and D Louis Collins. “Towards accurate, robust and practical ultrasound-CT registration of vertebrae for image-guided spine surgery”. In: *International journal of computer assisted radiology and surgery* 6.4 (2011), pp. 523–537.

[74] Steve H Yoon, Sarah Lee OBrien, and Mike Tran. “Ultrasound guided spine injections: advancement over fluoroscopic guidance?” In: *Current Physical Medicine and Rehabilitation Reports* 1.2 (2013), pp. 104–113.

[75] Oliver Zettinig, Bernhard Fuerst, Risto Kojcev, Marco Esposito, Mehrdad Salehi, Wolfgang Wein, Julia Rackerseder, Edoardo Sinibaldi, Benjamin Frisch, and Nassir Navab. “Toward real-time 3D ultrasound registration-based visual servoing for interventional navigation”. In: (2016), pp. 945–950.

List of Figures

1.1 Typical interventional X-ray images acquired using a C-arm. These images show the placement of a guide wire (k-wire) into the pubic ramus bone. First, the surgeon needs to determine the entry point through the muscle, see *a)*, followed by drilling the wire through the bone as seen in *b)*. Note that the surgeon requires a vast amount of X-ray images from various points of view to determine the precise three-dimensional trajectory relative to the bone. 9

2.1 The four major components of the image registration process are the similarity measure, optimizer, interpolator, and transformation. 19

2.2 Both images show the same 3D slice of the brain, where a prominent tumor is visible in the top center and the mid brain in the bottom center region. Some representations of tissue regions are represented similarly in the US image and MRI slice (red circles), while in other regions the MRI gradient magnitudes seem to correlate with the US image intensities (yellow circle). Other regional representations, such as the tumor, seem to relate, but not correlate linearly. 29

2.3 (a) SPECT system showing several SLNs via Augmented Reality. Courtesy SurgicEye GmbH. (b) Interpolated gamma camera output for two sources, showing the auto-marked centers. 32

2.4 Gamma camera mounted on a KUKA iiwa industrial lightweight robot, allows for automatic tracking of the displacement of radiopositive lymph nodes. In this setup, the KUKA iiwa will move the gamma camera so that it follows the movement performed by the da Vinci® surgical system while maintaining a perpendicular view of the radioactive sources. 33

3.1 A typical surgical navigation system uses pre-operative images, a tracking system (top left: stereo camera), a method to register the images with the surgical site, and tracked tools to guide the surgeon. While the accuracy improves for specific applications, common drawbacks include the additional hardware introduced to the operating room, limitations of registration and tracking, and increased duration. Image showing Brainlab Curve, courtesy of Brainlab AG. 36
3.2 In an operating room with a robot-assisted surgical system, the surgeon's console is usually set up in a corner, the patient-side manipulators are mobile and moved towards the patient once the trocars are placed. In many interventions, a surgical assistant sits at the surgical table to support the primary surgeon using rigid laparoscopic tools or exchanging the robotic tools. The surgery can be observed by every surgical team member via large screens showing the same views as through the surgeon's console. 39
3.3 The robotic SPECT reconstruction results can be directly overlaid on the surgeon's view of the surgical site, creating an immerse augmented reality visualization, and an intuitive surgical image-guidance. 39
3.4 The Camera-Augmented Mobile C-arm (CAMC) allows the overlay of X-ray images on live video images [50]. To achieve this the radiation source is unconventionally placed on top of the patient, enabling the optical centers of camera and X-ray to be virtually aligned through a mirror. The system allows the user to intuitively establish a relationship between hands and X-ray image. . 44
3.5 The three-dimensional view of patient's surface can be combined with the three-dimensional CBCT volume. This allows for a dynamic viewing of the patient's surface and anatomy. 45
3.6 The platform for augmented reality for orthopedic and trauma interventions allows the surgeon to simultaneously see the patient's surface and anatomy (from CBCT) together with a live representation of tools and hands. Additionally, the surgeon can choose as many different views as desired. This image shows the simultaneous anterior-posterior (AP) and lateral view of the same scene. 45
3.7 The correct placement of guide wires in the pelvis is one of the most difficult steps during minimally invasive interventional fracture reduction. The placement is currently performed under X-ray guidance, see one image in *a)*. To achieve the correct trajectory, several attempts may be required to determine the correct entry point, as one can see in *b)* where multiple failed attempts are visible. 46
3.8 The displacement of the patient can be estimated based on color images. After co-registration of camera and X-ray origin, this information can be used to align multiple CBCT volumes. This figure blends two frames and shows feature correspondences to estimate the movement of the patient. From both frames, the positioning-laser (red) and natural surface features are extracted. The tracking results are illustrated as yellow lines. 48

3.9 The calibration phantom (a) can be an arbitrary object, such as this spine model. The calibration allows the overlay of simulated X-ray images and the real phantom surfaces, as shown in (b). In (c) and (d) the three-dimensional surface reconstruction, simulated X-ray images and the live three-dimensional information showing the user's hand are shown. 50

3.10 The three-dimensional Augmented Reality visualization enables surgeons to perform K-wire placements faster (Time), with less X-ray images (# X-rays), and an overall lower dose (Dose). Additionally, the surgical task load index (Task Load) was reduced, while the accuracy remained nearly constant. 51

A.1 CT image showing the respiratory system of a patient. The diaphragm and intercostal muscles expand the thoracic cavity to inflate the lung. The pleura (not visible in CT slice) keeps the lung in close proximity to the thoracic cavity surface. Both lung and diaphragm can slide along the rib cage during respiration. . 71

A.2 Pipeline of the proposed motion model. The initial input are CT images at end-expiration (EE) and end-inspiration (EI), the segmentations and the definition of the anatomical models are perfomed automatically. Furthermore, the thoracic and diaphragmatic pressure values are estimated automatically, allowing the prediction of the respiratory motion. 72

A.3 The steps before simulation comprise of (a) reading the CT and corresponding segmentations for EE and EI, (b) automatically defining the anatomy model based on the segmentations, detection of the sub-diaphragm area and generating the tetrahedral mesh, and (c) defining the muscular contact zones on the surface of the thoracic cavity. The vectors for the computation of the collision and sliding interations are shown in (d). 76

A.4 Effect on the anatomical model of the modified closing operation performed on the height map, which is used to generate the sub-diaphragm area. 77

A.5 (a) The landmark and surface errors converge around 2.6k. The goal of a convergence study is to investigate at which point a parameter will no longer influnce the entire method. Therefore, not the resolution yielding the smallest error is of interest, but for which resolution the system converges. Note that the computational effort grows exponetially. (b) The convergence in terms of temporal resolutions was reached at $dt = 1$ ms, suggesting that a simulation using temporal resolution of less than 1 ms is indistinguishable from a simulation using 1 ms. 81

A.6 (a) Evaluating the personalization with respect to different numbers of pressure zones showed that the coarse-to-fine strategy results in lower errors, while converging after the same number of iterations compared to a direct personalization. (b) Furthermore, the coarse-to-fine strategy results in smoother, more-realistic pressure distrubutions. 83

A.7 (a) The lung motion is computed during exhale and shown as a solid lung. The ground truth from CT for each phase is overlayed as wireframe. (b) Average landmark error during exhale prediction for three experiments: personalization with 20 pressure zones using c_1, c_2 and c_3 as cost function. 84
A.8 Average landmark error during exhale prediction for three experiments: personalization with 20 pressure zones using c_1, c_2 and c_3 as cost function. 85

B.1 The intensity values of the US image need to be expressed by a locally varying relationship function. This is indicated by the red circle, which depicts a high correspondence between US intensity values as MRI intensity values (a and b), and the yellow ellipse, which marks regions of high correspondance between the US image and the MRI gradient magnitudes (a and c). The locally estimated coefficients (for neighboorhood $\Omega(\mathbf{x}, s)$) of the linear combination in the relationship function are visualized in (d). . . 97
B.2 Relationship between convergence (blue), accuracy (green) and the neighborhood size s for the three different LC^2 methods. Differences between the two 2D implementations (a and b) are caused by different numerical behaviour in the implementation (e.g. on the CPU everything is done with double precision, on the GPU only the final matrix inversion), not by differences in the methodology. The convergence analysis indicates that a neighborhood size of $s = 9$ ($m = 361$ pixels) yields a fair trade-off between convergence, accuracy and computation time. The typical behavior of the 3D approach is shown in (c), indicating $s = 3$ ($m = 343$ voxels) being a good choice of the neighborhood size. Note that patient 2 and 1 show the typical behavior of the 2D and 3D approaches respectively and are therefore plotted here. . 105
B.3 Superimposed US on axial MRI slice to show the registration result of patient 6. The FRE improves by 0.11 mm from rigid to deformable registration. 106
B.4 Registration experiments with 100 randomly initialized transformations for all patient data sets. The converged results (2D: magenta, 3D: blue) are clearly separated from the failed registration attempts (outliers: 2D: green, 3D: red). Convergence is defined based on typical patient-specific FRE values as reported in Tab. B.1. 107
B.5 The capture range describes the relation between outliers and initial FRE. For selected patients this relation is plotted in (a) and (b) for the 2D and 3D LC^2 approach resprectively. The total capture range of the algorithm for all patients is plotted in (c). It can be seen that the the capture range (less than 5% outliers) for 2D LC^2 is 9 mm, while it is increased to 15 mm for the 3D approach. 108

C.1 γ-detectors and respective outputs. 112
C.2 The workflow of the proposed solution 114

C.3 Workflow for updating the SPECT AR overlay based on 2D γ-camera acquisitions. We start with original state **A**. When the tracked γ-camera's (off-screen) output changes as in **B**, the overlay is updated as in **C**. 118

D.1 Lymph node locations in the pelvic area. Determination of the lymph node(s) to which possible tumor spread has occurred (e.g. identification of the sentinel lymph node(s)), can improve staging of e.g. cervical cancer. 123

D.2 Laparoscopic freehand SPECT imaging has been performed during robot-assisted surgery, yielding clinically valuable images. This figure shows the acquisition procedure in which a freehand SPECT volume is obtained using the declipse®SPECT imaging system (SurgicEye GmbH, Munich, Germany) in combination with a laparoscopic γ-probe. To enable freehand SPECT acquisition and navigation, one robotic arm had to be removed, significantly modifying the work-flow. 126

D.3 The drop-in γ-probe can be picked up using the da Vinci® ProGrasp™ Forceps (a), and used to scan a region of interest containing radioactive isotopes, such as $^{99\mathrm{m}}$Tc in a SLN or ^{57}Co in our SPECT phantom (b). 135

D.4 The probe tip is tracked in using a federated Kalman filter that combines the vision-based (yellow: $^{\mathrm{probe}}\mathbf{T}_{\mathrm{base}}$) and mechanical tracking ($^{\mathrm{forceps}}\mathbf{T}_{\mathrm{base}}$) streams. The movements of the patient and da Vinci® base are compensated for using an infrared tracking system. 136

D.5 The radioactive sources have been scanned using a freehand γ-camera and the drop-in γ-probe held by the da Vinci® . The reconstructions are visualized in (a) and (b) respectively. Due to the limited angles and limited access from only two sides, the acquisition using the drop-in γ-probe and da Vinci® exhibit some artifacts along the borders of the volume (not illustrated). However, in both cases the reconstruction results clearly indicate two sources and the centroid distances are comparable. The ground truth (GT) distance is 30 mm in both cases. 137

D.6 This augmented reality view is shown to the surgeon in parallel to the original endoscopic view. It allows the surgeon to see the reconstructed radioactivity distribution (manually windowed SPECT image is visualized in blue) semi-transparently overlaid on the endoscopic view. The stripe pattern on the probe is used for vision-based tracking, which allows for the augmentation of the probe field of view (yellow). During the acquisition, the γ source disks are mounted on a holder (white, 3D printed) to ensure that the ground truth is known. The background is designed to represent the interior of a patient, which contributes to a realistic, ex situ setup. 138

D.7 To evaluate the accuracy of the reconstruction, three sources with varying intensities have been scanned. The distances between the sources were evaluated on a three-dimensional planar slice through all three centroids (a). The augmentation of the endoscopic view is shown in panel (b). 139

E.1 The 3D misalignment of bones (red lines) may be difficult to quantify using 2D images. CBCT contributes as a valuable tool for interventions in which the 3D alignment is of importance, for instance in acute fracture treatment or joint replacement. Background image courtesy of BodyParts3D, Center for Life Science, Japan. 147

E.2 A mobile C-arm, the positioning-laser (red), and an optical camera (blue) are illustrated. The mirror (purple) aligns the optical camera and X-ray source centers. The patient motion relative to the C-arm is estimated by observing both the positioning-laser and natural features (green) on the patient's surface. The 3D positions of the features are estimated using the depth of the nearest positioning-laser on the patient (black dotted line intersecting green line), of which the depth is based on calibration. . . 148

E.3 The figure shows the overlay of two frames to illustrate the feature correspondences to estimate the movement of the patient. From both frames, the positioning-laser (red) and natural surface features are extracted. The tracking results of the matched features in frame k (+) and frame $k + 1$ (∘) are illustrated as yellow lines. 149

E.4 Absolute distance of the aligned sub-volumes in (a) is measured (415.37 mm), and compared to the real world measurements (415 mm) of the femur phantom in (b). Similarly, a fiducial phantom was scanned and the vision-based stitching (c) compared to the real world object (d). For visualization purposes and volumetric appearance in (a) and (c), multiple parallel slices are averaged. . 153

F.1 a) Lateral view of the hip in pelvic surgery: several skin punctures demonstrate the number of failed attempts to place the K-wire. b) Anteroposterior X-ray image of hip region in pelvic surgery. The narrow superior pubic ramus requires a precise placement of the K-wire, especially considering that a misplacement could cause severe damage to external iliac artery or vein. 159

F.2 The same stage in the K-wire placement has been recreated using the different image-guidance systems. In a), the K-wire is placed under conventional C-arm guidance, which requires frequent imaging and may result in a higher radiation dose for the surgeon. b) The X-ray image is augmented onto a live video stream, and the surgeon can update the X-ray image at his discretion. c) The use of an RGBD camera and DRR computed from a CBCT allows for the simultaneous visualization of the patient from different views. The surgeon can choose which views should be displayed, which will be updated using live RGBD information. 162

F.3 During the experiments the surgeons drilled a K-wire into a phantom. This figure shows the experimental setup during a procedure guided by the 3D RGBD and DRR visualization (S3). The RGBD camera is mounted on the C-arm X-ray detector, and the surgeon is watching the live 3D RGBD and DRR views on the monitor while drilling into the phantom containing the bone model. 165
F.4 This plot illustrates duration of the intervention, number of X-ray images taken, radiation dose, K-wire placement error and surgical task load, where each bar shows the accumulated values using one of the systems (conventional X-ray, RGB/X-ray fusion, or RBGD/DRR). Each measure is normalized relative to the maximum value observed. The $\star$ symbols indicate significant differences. 167

G.1 The first step during a minimally invasive orthopedic surgery is the entry point localization. The surgical site, containing patient and tools, is shown in (a). Panel (b) shows an X-ray image during entry point localization. 177
G.2 The offline calibration of RGBD camera to CBCT origin is performed by introducing an arbitrary object into the common views of both devices. Before an intervention begins, CBCT and surface scans of the patient are acquired simultaneously. During the intervention, the fused visualization of patient's surface, surgeon's hands and tools, together with simulated X-ray images (DRR) are displayed to assist the surgeon. 179
G.3 System setup: A depth camera is rigidly mounted on the detector, so that the field of view, and depth of view cover the CBCT volume. The transformation ${}^{c}\mathbf{T}_{CBCT}$ is recovered by registering the point clouds $\mathbf{P}^{c}$ and $\mathbf{P}^{CBCT}$ of the calibration phantom. . . . 180
G.4 (a) The calibration phantom comprised 3 pipes in different heights, lengths, and orientations which represent an unambiguous 3D object. (b) Raw surface point cloud reconstruction from RGBD data, and (c) filtered surface points after the cylinder fitting. 181
G.5 The steps to acquire the smooth surface point cloud from CBCT for calibration start with the CBCT volume acquisition (a). The CBCT surface segmentation is performed by thresholding (b). The ambient quality of CBCT points is assigned to each point (c), where white is a higher score (outer surface), and red is a lower scores (inner surface). Based on the ambient score, the outer CBCT surface is defined (d). 182
G.6 The surface reconstruction (a) and the DRR rendering of the calibration phantom (b) are overlaid using the opacities 0.4 and 0.7 in (c) and (d), respectively. This fused view may assist surgeons during orthopedic interventions. 183
G.7 The calibration accuracy is evaluated using a phantom (a) with radiopaque markers which are visible in the RGBD view. (b) Test phantom in CBCT (DRR rendering), (c) point clouds of the phantom obtained using the Intel RealSense F200 camera, and (d) point clouds of the phantom from the Kinect camera. 185

G.8 (a) The spine phantom for visualization experiment, (b) the overlay of DRR and phantom surface, (c) the overlay with a point cloud of a hand, and (d) the overlay from a different view. Depending on the structure of interest, the user can change the windowing (level and width of intensity range) and transparency of the DRR dynamically. This is demonstrated by the different renderings in (b) and (d). 186
G.9 A shrapnel phantom (a) is designed to perform foreign body removal under mixed reality visualization, shown in (b) and (c). . . 186

APPENDIX A

Patient-Specific Biomechanical Model for the Prediction of Lung Motion from 4D CT Images (IEEE TMI 2014)

Bernhard Fuerst[1,2], Tommaso Mansi[1], Francois Carnis[1], Martin Saelzle[1], Jingdan Zhang[1], Jerome Declerck[3], Thomas Boettger[4], John Bayouth[5], Nassir Navab[2] and Ali Kamen[1]

[1] Siemens Corporation, Corporate Technology, Imaging and Computer Vision, Princeton, NJ, United States.

[2] Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstrae 3, 85748 Garching b. München, Germany

[3] Siemens Molecular Imaging, United Kingdom

[4] Siemens Healthcare, Germany

[5] Department of Human Oncology at University of Wisconsin - Madison, School of Medicine & Public Health

Copyright © 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

Contributions A generative biomechanical model of the respiratory system is proposed. In contrast to previous approaches, the model is driven by optimized patient-specific thoracic and diaphragmatic spatially varying pressure, and the motion is not constrained by any fixed boundary condition. The model generation and personalization, and motion prediction are shown using 4D CT data sets.

PATIENT-SPECIFIC BIOMECHANICAL MODEL FOR THE PREDICTION OF LUNG MOTION FROM 4D CT IMAGES (IEEE TMI 2014)

Abstract This article presents an approach to predict the deformation of the lungs and surrounding organs during respiration. The framework incorporates a computational model of the respiratory system, which comprises an anatomical model extracted from Computed Tomography (CT) images at end-expiration (EE), and a biomechanical model of the respiratory physiology, including the material behavior and interactions between organs. A personalization step is performed to automatically estimate patient-specific thoracic pressure, which drives the biomechanical model. The zone-wise pressure values are obtained by using a trust-region optimizer, where the estimated motion is compared to CT images at end-inspiration (EI). A detailed convergence analysis in terms of mesh resolution, time stepping and number of pressure zones on the surface of the thoracic cavity is carried out. The method is then tested on five public datasets. Results show that the model is able to predict the respiratory motion with an average landmark error of 3.40 ± 1.0 mm over the entire respiratory cycle. The estimated 3D lung motion may constitute as an advanced 3D surrogate for more accurate medical image reconstruction and patient respiratory analysis.

A.1 Introduction

A.1.1 Clinical Motivation: Thoracic Imaging and Radiotherapy

Thoracic imaging and radiotherapy suffer from complications caused by the complex respiratory motion, which is a source of artifacts in images and makes it difficult to determine critical information for radiotherapy, such as a lung tumor's shape, size, position and surrounding tissue [6]. Therefore, there is a need for methods to predict the 3D lung deformation during regular and irregular breathing cycles. The accurate estimation of the 3D lung deformation is difficult and currently approximated by one- or multi-dimensional signals from devices such as spirometers, abdominal pressure belts, external markers, or image modalities [7, 31, 22]. These signals are referred to as surrogates and only part-wise reflect the complexity of lung deformation during a respiratory cycle. For instance, during thoracic image acquisition, a 4D computed tomography (CT) data set is compounded by image segments sorted and combined (*binning*) either based on the amplitude or the phase-angle of a respiratory surrogate [16]. Because the signal is assumed to be periodic, difficulties arise when the breathing pattern changes [37], resulting in imaging artifacts due to the combination of different breathing states. In a second approach, images are acquired at a specific instance of the respiratory cycle by triggering the imaging modality according to the surrogate signal [33, 39]. This is referred to as *gating* and is commonly used for nuclear imaging, such as PET [24]. In radiotherapy gating is used to apply the ionizing radiation during pre-defined respiratory states only [6]. Both approaches have drawbacks, such as the increase of radiation dose to achieve oversampling, or increase of treatment or imaging time. Furthermore, for image reconstruction, interpolation due to the lack of information between phases can cause step artifacts. For this purpose, motion models have been proposed to estimate or predict the deformation of lungs, liver or other organs during breathing.

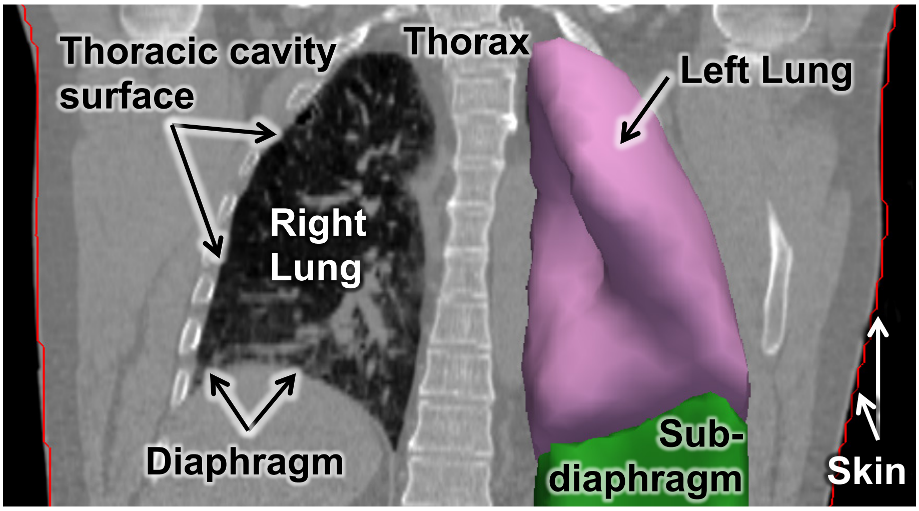


Fig. A.1: CT image showing the respiratory system of a patient. The diaphragm and intercostal muscles expand the thoracic cavity to inflate the lung. The pleura (not visible in CT slice) keeps the lung in close proximity to the thoracic cavity surface. Both lung and diaphragm can slide along the rib cage during respiration.

A.1.2 Physiology of the Respiratory System

The respiratory motion is complex, as the lungs do not just compress and deform, but also slide along the thoracic cavity thanks to the pleura (see anatomy and explanations in Fig. A.1), which is filled with a serous fluid and does not change its volume during respiration. The anatomical properties of the pleura allow nearly friction free sliding of the lungs and diaphragm along the thoracic cavity. The motion is caused by two major groups of muscles: the diaphragm and the intercostal muscles [11, 15]. Their contraction enlarges the thoracic cavity creating a negative pressure causing air to flow into the lungs. Respiratory motion may vary from cycle to cycle, as the contributions of the muscle groups differ. The patient's orientation and the breathing pattern (shallow, deep, abdominal or thoracic) can have major influence on the lung deformation. Three types of lung motion variations can be identified [23]: *i) Intra-cycle* variations describe changes within one cycle, and are mostly caused by different paths during inhale and exhale. In [16, 13, 12, 14, 30] models of intra-cycle variations have been proposed. However, these approaches cannot deal with *ii) inter-cycle* variations, which are often caused by changes in breathing patterns. Models, such as those presented in [22] and [18], compensate for such variations, but the authors are interested in reducing artifacts during imaging, rather than in modeling the physiological correct motion. A generative model was not required and their model was based on observations and free-form deformable registration. To deal with *iii) inter-patients* variations, atlas-based methods have been used to identify motion patterns varying between patients [8]. However, these methods can only describe motion which has been observed and are therefore not generative. By creating a patient-specific biomechanical motion model, our technique attempts to compensate all three motion variations.

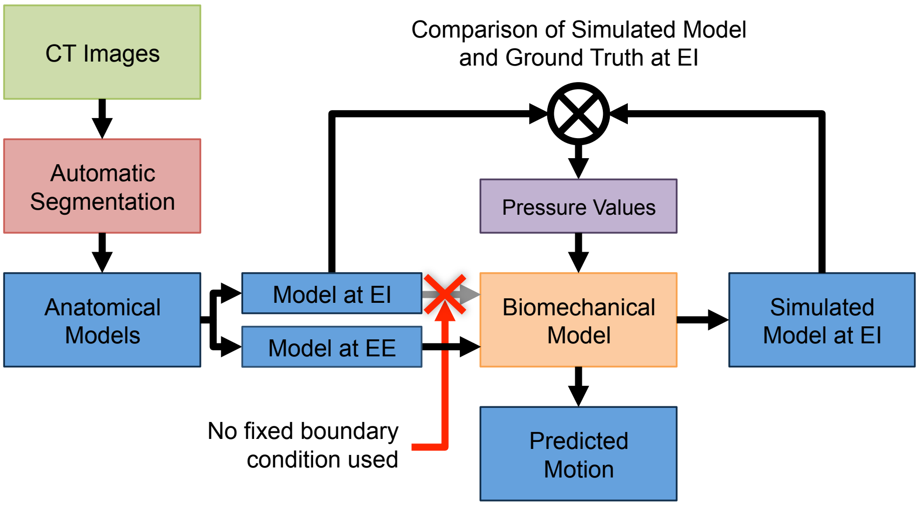


Fig. A.2: Pipeline of the proposed motion model. The initial input are CT images at end-expiration (EE) and end-inspiration (EI), the segmentations and the definition of the anatomical models are perfomed automatically. Furthermore, the thoracic and diaphragmatic pressure values are estimated automatically, allowing the prediction of the respiratory motion.

A.1.3 Image-Based Approaches for Motion Estimation

Image-based models describe the respiratory motion using dense deformation fields. This is achieved by registering the images at different respiratory phases to a phase of reference. Optical flow techniques have been proposed [7, 5], in which the cost function incorporates the temporal difference between the images and the spatial image gradient. However, the regularization process of optical flow techniques is not suited for the sliding motion of the lung, causing wrong deformations close to the thorax/lung interface. To cope with this limitation, new regularization terms have been introduced that distinguish between normal and tangential motion [30].

More sophisticated diffeomorphic image registration approaches have been proposed [29, 28] in particular piecewise-diffeomorphic registration techniques [28], to correctly capture the sliding interface. This approach smooths the deformation field in regions without sliding motion, and allows non-smooth deformation between lung and thorax. Though yielding good results, it is not versatile enough to fully model the sliding of the diaphragm along the rib cage, as the piecewise regularization is only applied on the lung surface, not the boundary between diaphragm and thorax. Therefore, gaps between diaphragm and thorax remain during exhale.

To model *inter-patient* variations, mean motion models have been presented in [8]. Intuitively, an anatomical atlas is first estimated by averaging the thoracic images acquired at a specified time of the respiratory cycle. Then, for each patient of the database, the respiratory motion is computed using image registration and encoded in terms of deformation fields. These deformation fields are finally transferred to the atlas and averaged to obtain the mean motion model.

When a new case needs to be processed, the mean motion model is transferred to its coordinate system and applied for lung motion prediction. However, this approach is not totally generative as it can capture only what is observed in the database. In particular, sudden respiratory changes or disease lung motion are more complex to capture.

A.1.4 Biomechanical-Based Methods for Motion Estimation

Image-based approaches cannot fully take into account the variability of the respiratory motion, as they rely on observations and are therefore not generative. To overcome these limitations due to variations in respiratory motion and allow the patient-specific prediction of respiratory motion, biomechanical models have been proposed [21, 34, 12, 13]. Biomechanical models for the lung have been initially presented by West *et al.* [35], which showed the simulation of a half thorax and the deformation of the lung under its own weight. Current computational models create an anatomical model from patient images and simulate the physiological deformations during respiration cycles [21, 34].

A standard strategy relies on image data, and directly deforms the model by projecting the triangulated surface nodes of the inhale lungs onto the surface of the exhale phase. The deformation is constrained by a *fixed boundary condition* and is not driven by pressure forces generated by the thorax [20, 21].

Another approach is to define a negative pressure on the lung surfaces and constrain the inflation by lung surfaces extracted from another respiratory phase [32]. Hence, these and similar approaches are still not fully predictive since they rely on a boundary condition defined by a secondary geometry and do not model the physiology to compute the motion.

A first step towards more accurate biomechanical modeling has been presented in [27], where the authors propose a method to automatically estimate the inhomogeneous material properties in patients. While the hyper-elastic material properties of lung have been identified early [35, 15, 38], it is nowadays not clear whether using non-linear elastic material properties would give significant improvements in terms of motion prediction [21, 20] compared to linear elasticity [34] due to their higher number of parameters to estimate and model complexity.

A.1.5 Proposed Solution

To cope with the above-mentioned limitations, we propose in this manuscript a generative biomechanical model of the respiratory system, which, contrary to previous approaches, is driven by patient-specific thoracic and diaphragmatic pressure force fields (Sec. A.2.2). The motion is not constrained by any fixed boundary condition. The pressure force reflects the muscular forces generated by the thoracic cavity and is transferred to the lung surface through a novel thorax/diaphragm/lung interaction model.

Fig. A.2 illustrates the various steps of our method. First, a comprehensive anatomical model of the respiratory system is computed from an image at end-exhale (EE) (Sec. A.2.1). Using the anatomical model, a biomechanical model is employed to simulate the lung deformation during respiration (Sec. A.2.2) based on the thoracic and diaphragmatic pressure. These pressure values can not be measured. Therefore, we estimate the values automatically using a trust region optimizer. During this *personalization* step the pressure values are iteratively improved with respect to the differences between simulated lung and the end-inspiration (EI) image (Sec. A.2.3). In contrast to our previous work on the estimation of the respiratory motion using a direct parameter estimation [13], this manuscript introduces a novel patch-wise coarse-to-fine optimization strategy during personalization.

In contrast to other approaches, the lung at EI is never used as a boundary condition. The anatomical model at EE, and the personalized thoracic and diaphragmatic pressure are the only components used to predict the respiratory cycle. Varying the amplitude of the personalized pressure can potentially enable the simulation of respiratory motion which was not observed during imaging, which is not possible when using a fixed boundary condition such as in [21, 20, 32].

Furthermore, a convergence analysis in terms of spatial and temporal resolution is presented. The model predictions are evaluated by predicting exhale deformations in five DIR-Lab datasets (Sec. A.3). We show that modeling the diaphragmatic and thoracic movement decoupled enables the estimation to achieve an average error of 3.40 ± 1.0 mm in predicted landmark positions during a respiratory cycle. Sec. A.4 concludes the manuscript.

A.2 Methods

A.2.1 Anatomical Model Generation

Our detailed anatomical model of the respiratory system comprises the lungs, thorax, and a sub-diaphragm region grouping abdominal organs including the diaphragm. This allows the individual sliding of lungs and diaphragm along the thorax. The anatomical model is generated from the end of exhale phase of a thoracic 4D CT through three steps herein described: segmentation, mesh generation, and mesh post-processing.

Segmentation

3D CT images are segmented using an automatic multi-organ technique based on a machine learning approach with a level-set optimization [19] (Fig. A.3a). While the lung meshes are directly generated from the segmentation, the thorax is based on the skin and lung segmentations. In some patients, the strong diaphragm curvature and deep belly breathing causes a sliding movement of the diaphragm along the rib cage, which can be observed in a 4D CT image set. To allow this movement, the diaphragm must be segmented independently from the rib cage. The sub-diaphragm area is synthetically generated by casting the lung downwards (Fig. A.3b). Due to the non-convex nature of the lung, a simple downward projection of the lung will cause outliers and sub-diaphragm wedges between lung and thorax, as depicted in Fig. A.4. To address this difficulty, we first compute the height map of the lower third of the inferior surface (z-axis) of the lung. Then, a modified closing operation is applied: erosion and dilation are only performed on voxels in the height map if more than 60% (set experimentally) of the neighboring voxels have a lower or equal height value.

Therefore, anatomically incorrect ridges will be removed in case the height map has a pot-like shape, while preserving the borders and height if diaphragm is represented correctly.

Mesh Generation

The volumetric meshes are generated through a 3D-triangulation[1] based on a feature preserving Delaunay refinement. The algorithm explicitly samples corners and edges from the input image, which is a 3D binary volume representing the segmentation and constraining the refinement to preserve these features [3]. After Delaunay refinement, a mesh optimization phase is performed to remove slivers and achieve a good mesh quality, resulting in three tetrahedral meshes for thorax, lung and sub-diaphragm (Fig. A.3b showing lung and sub-diaphragm). The Computational Geometry Algorithms Library (CGAL) provides an implementation.

Muscular Contact Zones

To capture the heterogeneous muscle forces, thoracic and diaphragmatic pressures are estimated regionally according to pressure zones. For the thorax, the pressure zones are defined automatically by sub-dividing the inner surface in evenly spaced horizontal rings, which are themselves further subdivided into patches based on the relative angular position of each surface triangle (Fig. A.3c). The angle is defined to be 0^o along the negative x-axis (dexter), and 90^o along the negative y-axis (anterior). For the sub-diaphragm, its superior surface (namely the diaphragm interface) is split into zones, based on the relative position of each triangle on the antero-posterior (y) and dextro-sinister (x) axis.

A.2.2 Computational Motion Model

The computational model deforms the anatomical model (Sec. A.2.1) by computing the deformation under a set of pressure forces applied on the pressure zones (Sec. A.2.1).

Biomechanical Model

To solve the *dynamics equation* of motion of a structure [36], the finite element method is applied, which incorporates the mechanical parameters, the respiratory forces, and a specialized collision model. We define the internal stress and strain (deformation) of the lungs, diaphragm, and thorax to be equal to the sum of interaction forces and applied pressure force. Therefore, the deformations are computed by simultaneously solving the *dynamics equations* (Eq. (A.1)) for the lungs (l), thorax (t), and sub-diaphragm region (d). Superscripts (^{l}, ^{t}, or ^{d}) are omitted if not necessary.

$$\begin{cases} \mathbf{M}^l\ddot{\mathbf{U}}^l + \mathbf{C}^l\dot{\mathbf{U}}^l + \mathbf{K}^l\mathbf{U}^l = \mathbf{F}_c^{t\to l} + \mathbf{F}_c^{d\to l} \\ \mathbf{M}^t\ddot{\mathbf{U}}^t + \mathbf{C}^t\dot{\mathbf{U}}^t + \mathbf{K}^t\mathbf{U}^t = \mathbf{F}_c^{l\to t} + \mathbf{F}_c^{d\to t} + \mathbf{F}_p^t \\ \mathbf{M}^d\ddot{\mathbf{U}}^d + \mathbf{C}^d\dot{\mathbf{U}}^d + \mathbf{K}^d\mathbf{U}^d = \mathbf{F}_c^{l\to d} + \mathbf{F}_c^{t\to d} + \mathbf{F}_p^d \end{cases}, \tag{A.1}$$

---

[1] 3D-triangulation refers to the partition of a volume into tetrahedra

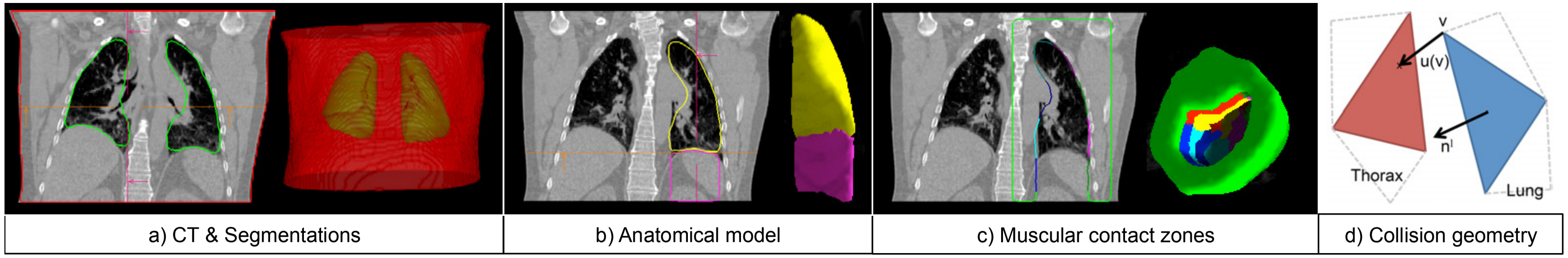

a) CT & Segmentations
b) Anatomical model
c) Muscular contact zones
d) Collision geometry

Fig. A.3: The steps before simulation comprise of (a) reading the CT and corresponding segmentations for EE and EI, (b) automatically defining the anatomy model based on the segmentations, detection of the sub-diaphragm area and generating the tetrahedral mesh, and (c) defining the muscular contact zones on the surface of the thoracic cavity. The vectors for the computation of the collision and sliding interations are shown in (d).

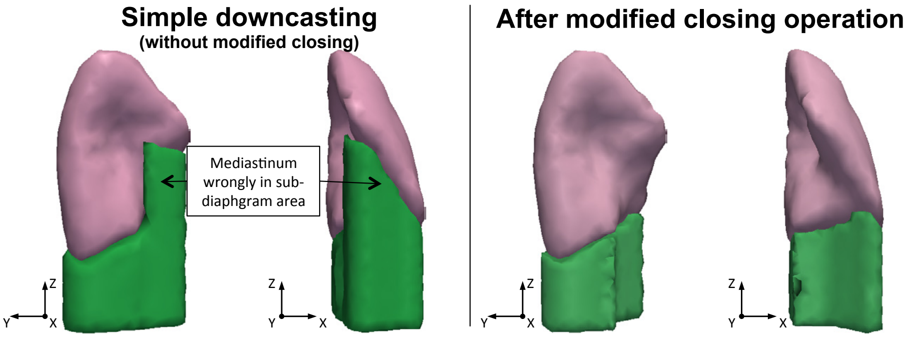

Simple downcasting
(without modified closing)
Mediastinum wrongly in sub-diaphgram area
After modified closing operation
Z
Y
X
Z
Y
X
Z
Y
X
Z
Y
X

Fig. A.4: Effect on the anatomical model of the modified closing operation performed on the height map, which is used to generate the sub-diaphragm area.

where the acceleration, velocities and positions of the free nodes of each part are gathered in the vectors $\ddot{\mathbf{U}}$, $\dot{\mathbf{U}}$, and $\mathbf{U}$. The lumped mass matrix $\mathbf{M}$ is computed according to the mass density $\rho^l = 1.05$ g/mL and $\rho^{\{t,d\}} = 1.50$ g/mL. The stiffness matrix $\mathbf{K}$ describes the internal elastic forces. The damping matrix $\mathbf{C}$ represents the Rayleigh damping with coefficients of 0.1 for mass and stiffness. The right-hand side terms of (A.1) represent the forces of which the lungs, thorax and sub-diaphragm are subject to. The pressure forces F_p represent the physiological forces driving the respiratory motion (Sec. A.2.2), while the interaction forces F_c model the sliding interaction between the organs (Sec. A.2.2). An implicit Euler scheme is used to integrate Eq. (A.1) in time since it allows larger time steps.

Tissue Model

In this work, the non linear, heterogeneous material properties [32, 27] of lung, thorax and muscles are simplified and represented by a linear elastic model like in [21]. A co-rotational formulation is used to cope with large deformations and rotations [25]. The Young's modulus Y and the Poisson's ratio ν define tissue stiffness and compressibility respectively. The sub-diaphragm and thorax tissue is assumed to be equal and fairly stiff with $Y^{\{t,d\}} = 7800$ Pa, while the lung is softer with $Y^l = 900$ Pa [34]. The thorax and sub-diaphragm are more incompressible ($\nu^{\{t,d\}} = 0.43$) than the lungs ($\nu^l = 0.4$) [21].

Respiratory Forces

The lungs are deformed passively by the surrounding thoracic muscles (see A.1.2). Our model represents this behavior by applying pressures on the automatically pre-defined thoracic pressure zones. For every zone z^i, the pressure p^i is applied as force $\mathbf{f}_p^i = p^i\,\mathbf{n}\,dS$, where $\mathbf{n}$ is the unit normal of the surface element dS.

Collision and Sliding Interaction

A collision model of pleural behavior is proposed to transfer the thoracic pressure force field to the lungs. The collision model attempts to keep the distance d between thorax and lung greater than the contact distance $d_c = 1$ mm, as the typical pleural thickness is reported to be $1 - 2$ mm [17]. When the distance is smaller than the alarm distance $d_a = 5$ mm, a collision is detected and the contact force $\mathbf{F}_c^{m \to m}$ is applied from mesh m_1 to mesh m_2. To avoid decoupling or interpenetration, the force is defined to keep the meshes at an optimal distance $d_o = \frac{1}{2}(d_a + d_c)$:

$$\begin{cases} \mathbf{F}_c^{m \to m}(\mathbf{v}) = -\mathbf{n}^m k_s \left(\mathbf{u}(\mathbf{v}) \cdot \mathbf{n}^m\right), \text{ if } d_o \leq d \leq d_a \\ \mathbf{F}_c^{m \to m}(\mathbf{v}) = +\mathbf{n}^m k_s \left(\mathbf{u}(\mathbf{v}) \cdot \mathbf{n}^m\right), \text{ if } d_c \leq d \leq d_o \\ \mathbf{F}_c^{m \to m}(\mathbf{v}) = \mathbf{0}, \text{ otherwise} \end{cases}$$

where $\mathbf{u}$ is the vector between the vertex $\mathbf{v}$, which belongs to the triangle $\mathcal{T}^m$ on mesh m_2, and the corresponding collision point on mesh m_1 (see Fig. A.3d), and is used to compute the current distance $d = \|\mathbf{u}(\mathbf{v})\|$. Furthermore, $\mathbf{n}^m$ is the normal of the triangle $\mathcal{T}^m$ and k_s is the penalty force stiffness coefficient, set to 0.1 N/m in this study. The interactions F_c between all three objects are defined in a similar way.

Three Stopping Criteria

When forces are applied, the biomechanical model converges towards an equilibrium where the dynamics equations (Eq. (A.1)) are balanced (steady state). For the optimization process (see Sec. A.2.3), the steady state needs to be detected automatically in order to compute the cost function. This is achieved by testing three criteria at every iteration of the computational model. The first criterion is met when the user-defined maximum simulation time (e.g. $T_{max} = 1$ s) is reached, T_{max} in normalized time. The second and third criterion are based on the velocities and computed at every time step i. The second criterion is met when the simulation becomes unstable: if the velocity of any node in the lung becomes physiological impossible, the simulation is aborted. The test is performed by comparing each magnitude of the velocity vector to infinity ($> 10^{10}$ m/s). However, we could not obverse speeds beyond 1 m/s in stable cases. The third criterion is met when the sliding average (window size of $n = 50$ time steps) of the mean velocity of all nodes in the lung falls below a given threshold $\epsilon = 2.5 \cdot 10^{-1}$ mm/s (set experimentally) for robustness with respect to potentially slight numerical instabilities. The values in the sliding window are initialized with zeros and the test is enabled after $i > n$.

A.2.3 Model Personalization

The pressure necessary to load the lung from EE to EI is estimated by minimizing a multi-variate cost function using Powell's NEWUOA algorithm [26], a trust-region method that does not explicitly calculate cost function gradients. Three different cost functions are investigated, defined as

$$c_1 = d_S, \; c_2 = d_{LM}, \text{ and } c_3 = d_S + d_{LM}, \tag{A.2}$$

where d_S is the mean Hausdorff surface-to-surface distance between the deformed EE lung surface at system equilibrium and the segmented lung surface at EI, and d_{LM} is the average Euclidian distance between internal landmarks at EI and their corresponding EE landmarks moved according to the internal deformation provided by the biomechanical models. Note that during personalization only landmarks at EI are compared to the simulated landmarks, while for the evaluation of the respiratory motion, the landmarks between EI and EE are used. To minimize risks of local minima, the following coarse-to-fine strategy is employed. First, the personalization is performed with each one zone on the sub-diaphragm and the thorax with an initial value of 0. Then, the thorax zones are split horizontally and vertically into four equal zones, the pressure values are set to the previously obtained values, and the personalization is restarted. This is repeated until the desired number of pressure zones is reached.

A.2.4 Implementation

The anatomical modeling pipeline is implemented in C++ and includes the automatic segmentation, detection of the anatomical model, and meshing based on CGAL. The biomechanical model and the collision algorithm are implemented within the Simulation Open Framework Architecture (SOFA) framework [1] using CUDA. Finally, the optimization framework, incorporating NEWUOA, calls SOFA as cost function to estimate the pressure distribution.

A.3 Experiments and Results

A.3.1 Patient Data

Our framework was evaluated using 4D CT data sets from DIR-Lab [4], where the entire thorax was visible (cases 6 to 10), image resolution of $0.97 \times 0.97 \times 2.50$ mm, average image dimension of $512 \times 512 \times 128$, an average of 414 landmarks was available, but no respiratory trace was provided. For evaluation the lung volume at each phase needs to be computed in order to allow the synchronization of the predicted lung with observations. The segmentation of the lungs and skin was performed automatically for the CT phases 0 to 5 (end-inspiration to end-expiration), and our pre-processing pipeline automatically defined the lung, thorax and sub-diaphragm regions. When applying the modified closing operation, parts of the mediastinum were merged to the thorax instead of connected to the sub-diaphragm, as illustrated on case 6 in Fig. A.4.

A.3.2 Computation of Landmark and Surface Errors

The *landmark error* is the euclidean distance between the landmark's simulated position and the position at the target phase. To compute the simulated position, first the surrounding tetrahedra and the corresponding barycentric coordinates were computed during initialization. The simulated position is then computed with respect to the deformed tetrahedra using barycentric mapping. The *surface error* is the average distance between two triangle meshes. We use the symmetric Hausdorff distance defined as the average bi-directional point to triangle distance.

Table A.1: Mean errors between simulation and ground truth with different spatial resolutions for the lung, suggesting a system convergence at mesh resolutions of 2.6k elements.

Number of Tetrahedra Elements	0.8k	1.4k	2.6k	5.7k
EI surface error (mm)	2.81	2.19	2.18	2.00
EI landmark error (mm)	3.50	3.40	3.52	3.52
Standard deviation (mm)	1.92	1.85	1.97	1.95
90%-tile landmark error (mm)	6.93	5.06	5.57	5.74
Average computation time (s)	122	132	184	281

A.3.3 Numerical Analysis

All convergence analyses were performed using the same patient data (DIR-Lab data set case 6, see A.3.1) and the cost function c_3 (Eq. (A.2)), with, if not otherwise specified, four pressure zones on the diaphragm and 16 pressure zones along the thoracic cavity. Four different mesh configurations were generated with 0.8k, 1.4k, 2.6k and 5.7k tetrahedra elements for the lung (Fig. A.5a). The number of elements in the thorax and sub-diaphragm were kept constant with 6.8k and 0.9k respectively.

Convergence analysis in terms of spatial resolution

In virtue of the spatial discretization, the accuracy of predicted landmark position and quality of the surface representation may be directly related to the number of tetrahedra elements. However, the number of elements also affects collision accuracy and computation time.

Therefore, we perform the convergence analysis to assess the number of elements necessary to reach the convergence of the system. As metric, we use the landmark distance with respect to the ground truth. As depicted in Fig. A.5a) and Tab. A.1, the observed landmark and surface errors do not change significantly once the mesh yields 2.6k or more elements. At this point, the errors are no longer influenced by the resolution, but other components of the method (i.e. temporal resolution or coarse-to-fine personalization strategy). It should be noted that we are interested in finding the number of elements for which the system converges, which may not be the optimal one in terms of landmark error.

Convergence analysis in terms of temporal resolution

Similarly, we analyzed the effect of the time step on the prediction accuracy, as illustrated in Fig. A.5b). Results suggested that time steps lower than 1 ms did not improve the results, while a higher time step significantly worsened the results, mainly because of poor collisions (the time resolution is not sufficient to detect collisions). Therefore, the time step was set to 1 ms for all subsequent experiments as a compromise between accuracy and computation time.

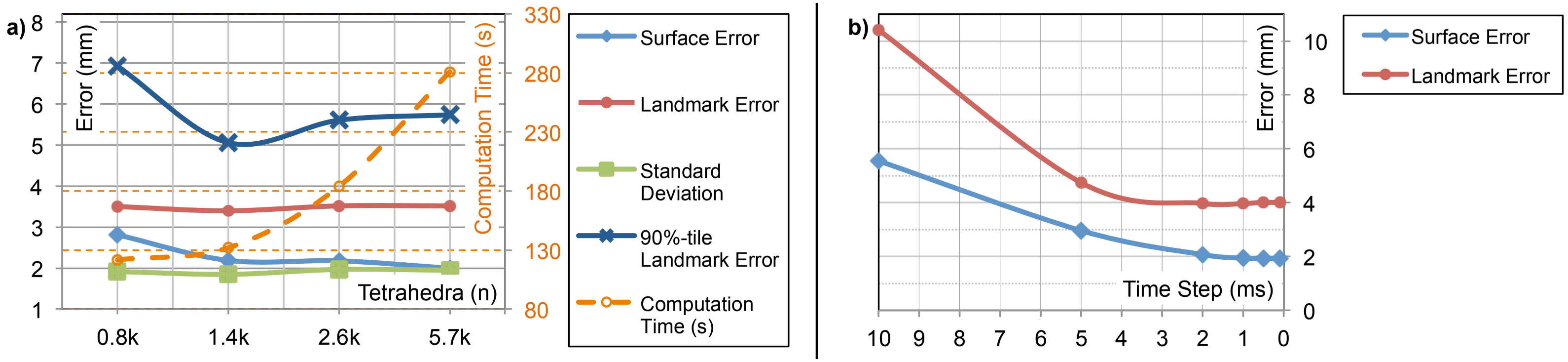


Fig. A.5: (a) The landmark and surface errors converge around 2.6k. The goal of a convergence study is to investigate at which point a parameter will no longer influnce the entire method. Therefore, not the resolution yielding the smallest error is of interest, but for which resolution the system converges. Note that the computational effort grows exponetially. (b) The convergence in terms of temporal resolutions was reached at $dt = 1$ ms, suggesting that a simulation using temporal resolution of less than 1 ms is indistinguishable from a simulation using 1 ms.

Evaluation of the coarse-to-fine personalization strategy

The proposed hierarchical personalization strategy aims at automatically finding the optimal trade-off between number of surface zones, accuracy and computational efficiency. Fig. A.6a) shows a comparison between a direct strategy and the coarse-to-fine strategy as presented in Sec. A.2.3. Three configurations of pressure zones were used:

- *Low:* 1 zone on the sub-diaphragm. 1 zone on the thorax.
- *Medium:* 2 zones on the sub-diaphragm. 2 rings and 2 zones per ring on the thorax. The coarse-to-fine strategy used the result of the *low* optimization as starting point.
- *High:* 4 zones on the sub-diaphragm. 4 rings and 4 zones per ring on the thorax. The coarse-to-fine strategy used the result of the *medium* optimization as starting point.

As one can see, stability and more accurate solutions were gained when there were 4 zones on the diaphragm and 16 on the thoracic surface. Furthermore, the NEWUOA software converged as fast using the coarse-to-fine strategy (295 iterations in total including *low*, *medium* and *high*), as compared to the direct personalization (*high* configuration, 292 iterations), while getting more robust and accurate solutions using the coarse-to-fine strategy. Furthermore, estimated pressure fields were more spatially consistent when using the coarse-to-fine strategy, as illustrated in Fig. A.6b). For the following experiments, the optimal profile (*high*), as well as the coarse-to-fine strategy were employed. To conclude, we used the following settings for the patient experiments:

- 2.6k tetrahedra elements for the lung.
- a time step of 1 ms.
- a coarse-to-fine strategy.
- 4 and 16 pressure zones on the sub-diaphragm and thoracic cavity respectively.

A.3.4 Validation of Motion Prediction on Patient Data

Generation of Anatomical Model

Using the patient data described in Sec. A.3.1, the meshing resulted in an average over all patients of 3.4k, 18.6k and 2.3k tetrahedra for the lung, thorax and sub-diaphragm respectively. In the following sections we present the results for the left lung only. It should be stressed though that our approach is generic and can handle both lungs irrespectively.

Biomechanical Model Personalization

For each proposed cost function (Eq. (A.2)), the personalization automatically estimated pressure values per muscular contact zone, which enabled the deformation of the lung from end-expiration (EE) to end-inspiration (EI) (Tab. A.2,

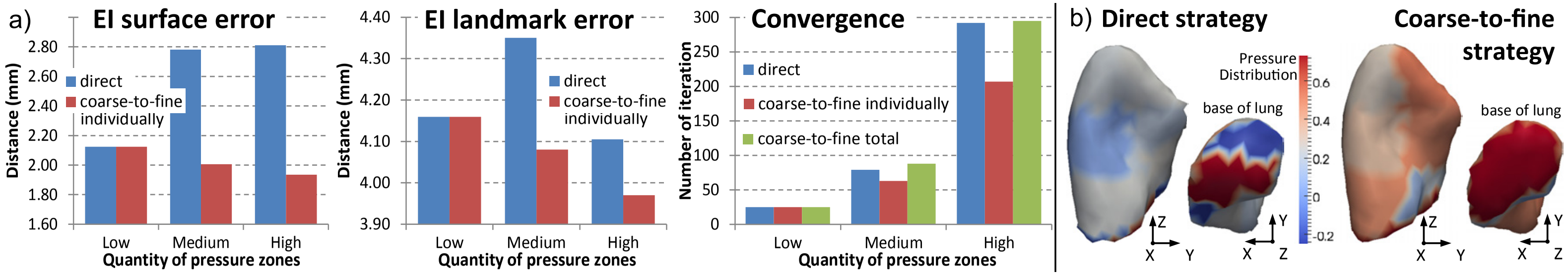


Fig. A.6: (a) Evaluating the personalization with respect to different numbers of pressure zones showed that the coarse-to-fine strategy results in lower errors, while converging after the same number of iterations compared to a direct personalization. (b) Furthermore, the coarse-to-fine strategy results in smoother, more-realistic pressure distrubutions.

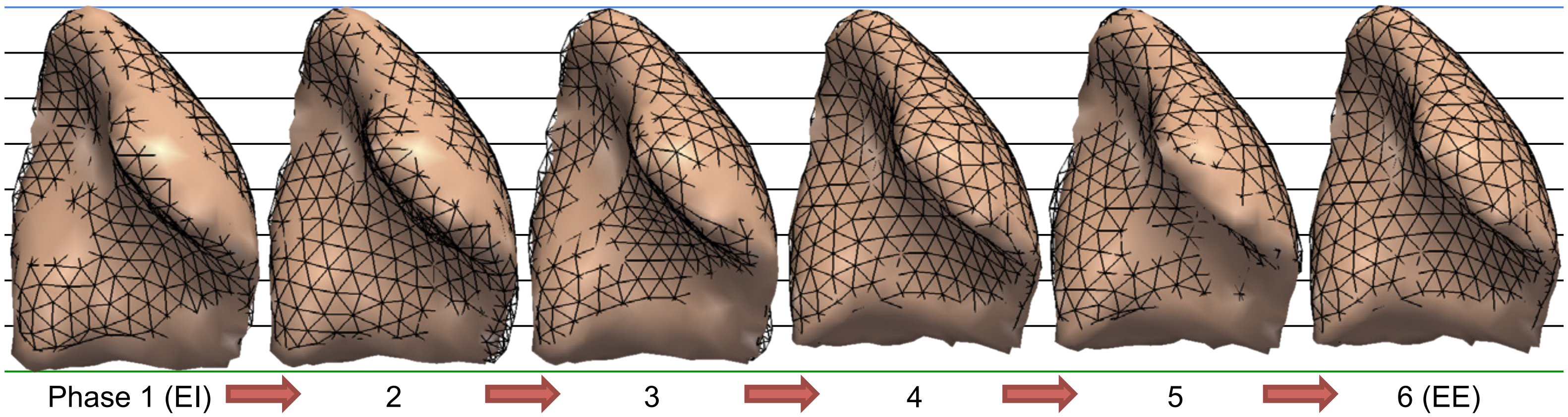


Fig. A.7: (a) The lung motion is computed during exhale and shown as a solid lung. The ground truth from CT for each phase is overlayed as wireframe. (b) Average landmark error during exhale prediction for three experiments: personalization with 20 pressure zones using c_1, c_2 and c_3 as cost function.

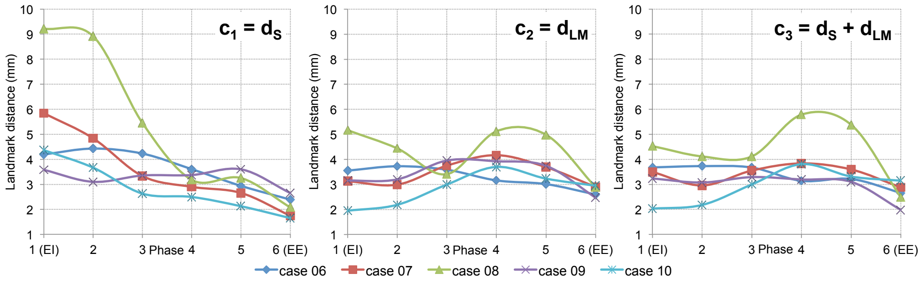


Fig. A.8: Average landmark error during exhale prediction for three experiments: personalization with 20 pressure zones using c_1, c_2 and c_3 as cost function.

blue rows indicate the surface and landmark distances after personalization). Depending on the number of pressure zones, the optimization algorithm converged after an average over all patients of 24, 55, 198 and 277 iterations for low, medium, high and total number of zones respectively while using the coarse-to-fine strategy. Each optimized iteration ran a full simulation of a lung motion based on a set of pressure values, where the average computation time[2] is 2 min. The quality of each personalization was then evaluated by predicting the exhale (EI to EE), without any image information.

Evaluation of Motion Model Prediction

A full cycle was simulated to evaluate the quality of the prediction, based on the three cost functions (Eq. (A.2)). The simulation of inhale was driven by the personalized pressure values, while exhale was simulated by setting the pressure values to zero. The surface and landmark errors were computed for all corresponding phases during exhale, where the synchronization was performed by means of the lung volume and therefore represented a real-world scenario (binning and gating is already performed on basis of lung volume). The quality of the exhale prediction was evaluated by comparing the landmarks for every intermediate phase, which are shown in Fig. A.7. As Tab. A.2 and Fig. A.8 show, the model based on cost function c_3 performed best, and despite simplifications, a mean landmark error over all landmarks and phases of 3.40 ± 1.0 mm was obtained.

A.4 Discussion and Future Works

This manuscript presented a framework to predict respiratory motion in patients based on a personalized biomechanical model. To the best of our knowledge, our pipeline is the first to segment the lungs, generate the anatomical models, personalize the model to reflect the patient-specific physiology, and predict the deformation without being explicitly driven by image forces. The personalized model allowed the generative prediction of the motion, where the obtained results were of the same order of magnitude as state-of-the-art respiratory motion

---

[2] Configuration: single-threaded simulation with CUDA based collision detection, Intel Xeon 3.07GHz, 3.5GB RAM and NVIDIA Quadro 2000

Table A.2: Results indicate that the personalization based on the surface and landmark errors allows a better prediction of the internal lung deformation. surface and average landmark errors after personalization (blue); average over all landmarks in all available 4DCT phases (white)

Case	6	7	8	9	10	Mean
Experiment 1: $c_1 = d_S$						
EI surface error (mm)	1.75	2.03	2.09	2.16	1.36	1.88 ± 0.52
EI landmark error (mm)	4.20	5.84	9.24	3.58	4.36	5.44 ± 3.80
Mean landmark error during prediction (mm)	3.63	3.55	5.35	3.27	2.82	3.72 ± 1.63
Experiment 2: $c_2 = d_{LM}$						
EI surface error (mm)	2.35	3.46	2.93	2.44	2.32	2.70 ± 0.76
EI landmark error (mm)	3.55	3.13	5.17	3.14	1.95	3.38 ± 1.78
Mean landmark error during prediction (mm)	3.27	3.43	4.33	3.40	2.83	3.45 ± 0.88
Experiment 3: $c_3 = d_S + d_{LM}$						
EI surface error (mm)	2.09	2.78	3.28	2.21	2.05	2.48 ± 0.80
EI landmark error (mm)	3.67	3.49	4.53	3.23	2.03	3.39 ± 1.36
Mean landmark error during prediction (mm)	3.34	3.38	4.40	2.97	2.91	3.40 ± 1.00

models [21, 34, 10], but in contrast to our method, previous techniques rely on a fixed boundary condition, are not generative, and do not model the physically correct pressure and interactions between lungs and thorax. Furthermore, the method outperformed our previously presented biomechanical models for full cycle prediction [13, 12]. We have only presented results on the left lung, but the extension to both lungs is straightforward, as for the anatomical model, personalization, and prediction of motion, and the lungs can be treated individually for each lung. However, as the real motion of the lungs is coupled, the modeling of constraints defining the coupling should be investigated.

As tissue model linear-elastic material properties were chosen. Previous publications (e.g. [21]) indicate that hyper-elastic material properties may yield better simulation results. However, a closer look at the experiments in [21] shows that the change of errors between models using linear and hyper-elastic properties is far less significant compared to models that allow nearly friction-free sliding motion. This coincides with our observations that the linear-elastic properties in combination with nearly friction-free sliding are sufficient to simulate the respiratory motion, and the proposed model is still predictive in terms of internal landmark position, as demonstrated in five cases. However, our efforts to improve the model and make it more realistic will include further investigations into the possible benefits and potential difficulties of using hyper-elastic material properties. Note that the tissue properties are especially complex, as the real properties of lung and thorax depend strongly on the patient's health, age, gender, and physical condition. Therefore, several approaches would need to be combined, such as the hyper-elastic material properties and the automatic estimation of spatially varying tissue properties such as presented in [27]. Currently, the simplifications in terms of linear material properties were partly compensated for by the personalized and spatially varying thoracic pressure values. These pressures, indeed, may not correspond to the actual forces exercised in the patient, but rather a "lump" force to cope with model simplifications. We have observed that the applied pressures directly correspond to the linear-elastic material properties (e.g. higher stiffness leads to increased pressure estimation). Therefore, personalization of pressures partially compensates for the rough approximation of the Young's modulus of lung, thorax and sub-diaphragm.

In summary, the personalization is based on three different cost functions. We have shown that in most patients the direct surface distance between simulated and observed EE phase is not sufficient to predict the pressure distribution. On the contrary, in this manuscript we have shown that a model which has been personalized based on landmark and surface distance (c_3) yields a better prediction of the internal deformation of the lung during the respiratory cycle. The improved behavior of the model is caused by the personalization considering the sum of surface and landmark errors, instead of relying on one or the other separately. This results in a better approximation of the physiological correct pressure values, which in turn provides an improved model. These findings make it necessary to provide landmarks together with the segmentations. Therefore, automatic detection of landmarks in the lung, such as in [2, 9], would need to be added to the framework. We have not investigated the influence of tumors and have considered these to behave as the surrounding lung. This simplification will be removed in future. While our previous papers [13, 12] relied on a direct estimation of each pressure value, the novel coarse-to-fine personalization strategy during personalization results in a smoother and physically plausible pressure distribution. Due to the lack of a respiratory trace for the patient data, the model and the observations were synchronized based on the lung volume, which was computed from the images through automatic segmentation of all 4D phases. This actually represents a real-world scenario, as the trace during imaging is often not available, or due to inter-cycle variations, hard to acquire. However, once available, a possible extension of the framework would be to introduce time-varying pressure values and improve the personalization process by considering the intermediate respiratory phases as well. While this is an interesting approach, the amount of unknown parameters to estimate would significantly increase. We therefore propose to further investigate the correlation between the 1D trace and the pressure values.

Our anatomical model represents the lung, thorax, and sub-diaphragm area. The separation into these three parts allows the sliding of the diaphragm and lung along the thoracic cavity, which was a major problem in the method presented in previous work [12] and for image based methods, even for most recent methods such as [28]. The mechanical decoupling of thorax and diaphragm is therefore not only a major contribution of this manuscript, but indicates possible extensions to other image based and non-image based methods. Indeed, the real anatomy is more complex than currently represented. Enhancing the model to include further details such as the tethering of the lungs by the airways, or the definition of a spatially varying tissue model to represent inhomogeneity for regions such as bones, muscles, tumors and organs may constitute potential areas for future research. In conclusion, our method may provide novel, 4D surrogate information for thoracic image reconstruction and analysis.

Acknowledgment

This work was performed with partial support from NIH grant U01-CA140206.

References

[1] Jérémie Allard, Stéphane Cotin, François Faure, Pierre-Jean Bensoussan, François Poyer, Christian Duriez, Hervé Delingette, and Laurent Grisoni. "SOFA an Open Source Framework for Medical Simulation". In: *Medicine Meets Virtual Reality*. Long Beach, USA, Feb. 2007.

[2] Margrit Betke, Harrison Hong, Deborah Thomas, Chekema Prince, and Jane P Ko. "Landmark detection in the chest and registration of lung surfaces with an application to nodule registration". In: *Medical Image Analysis* 7.3 (2003), pp. 265–281.

[3] Dobrina Boltcheva, Mariette Yvinec, and Jean-Daniel Boissonnat. "Mesh generation from 3D multi-material images". In: *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2009, pp. 283–290.

[4] Edward Castillo, Richard Castillo, Josue Martinez, Maithili Shenoy, and Thomas Guerrero. "Four-dimensional deformable image registration using trajectory modeling". In: *Physics in Medicine and Biology* 55.1 (2010), pp. 305–313.

[5] M. Dawood, F. Buther, X. Jiang, and K.P. Schafers. "Respiratory motion correction in 3-D PET data with advanced optical flow algorithms". In: *IEEE Transactions on Medical Imaging* 27.8 (2008), pp. 1164–1175.

[6] L.A. Dawson and D.A. Jaffray. "Advances in image-guided radiation therapy". In: *Journal of Clinical Oncology* 25.8 (2007), pp. 938–946.

[7] J. Ehrhardt, R. Werner, D. Säring, T. Frenzel, W. Lu, D. Low, and H. Handels. "An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing". In: *Medical Physics* 34 (2007), pp. 711–721.

[8] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, B. Schulz, and H. Handels. "Generation of a mean motion model of the lung using 4D-CT image data". In: *Eurographics Workshop on Visual Computing for Biomedicine*. 2008, pp. 69–76.

[9] Jan Ehrhardt, Rene Werner, Alexander Schmidt-Richberg, and Heinz Handels. "Automatic landmark detection and non-linear landmark-and surface-based registration of lung CT images". In: *Medical Image Analysis for the Clinic-A Grand Challenge* (2010), pp. 165–174.

[10] J. Eom, C. Shi, X. Xu, and S. De. "Modeling Respiratory Motion for Cancer Radiation Therapy Based on Patient-Specific 4DCT Data". In: *Medical Image Computing and Computer-Assisted Intervention* (2009), pp. 348–355.

[11] A. Fauci, D. Kasper, Longo D., E. Braunwald, and J. Jameson. *Harrison's Principles of Internal Medicine. Vol. 2.* Mc Graw Hill Medical Companies, 2008.

[12] B. Fuerst, T. Mansi, P. Khurd, J. Zhang, J. Declerck, T. Boettger, N. Navab, J. Bayouth, and A. Kamen. "Patient-specific finite-element simulation of respiratory mechanics for radiotherapy guidance, a first evaluation study". In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on.* IEEE. 2012, pp. 1212–1215.

[13] B. Fuerst, T. Mansi, J. Zhang, P. Khurd, J. Declerck, T. Boettger, N. Navab, J. Bayouth, D. Comaniciu, and A. Kamen. "A Personalized Biomechanical Model for Respiratory Motion Prediction". In: *Medical Image Computing and Computer-Assisted Intervention* (2012), pp. 566–573.

[14] M. Georg, R. Souvenir, A. Hope, and R. Pless. "Simultaneous data volume reconstruction and pose estimation from slice samples". In: *IEEE Conference on Computer Vision and Pattern Recognition 2008*. IEEE. 2008, pp. 1–6.

[15] H. Gray, T. Pick, and R. Howden. *Gray's Anatomy: Anatomy, descriptive and surgical.* Philadelphia: Running Press, 1974.

[16] J. Hinkle, P. Fletcher, B. Wang, B. Salter, and S. Joshi. "4D map image reconstruction incorporating organ motion". In: *Information Processing in Medical Imaging*. Springer. 2009, pp. 676–687.

[17] Jung-Gi Im, WR Webb, A Rosen, and G Gamsu. "Costal pleura: appearances at high-resolution CT." In: *Radiology* 171.1 (1989), pp. 125–131.

[18] AP King, C. Buerger, C. Tsoumpas, P. Marsden, and T. Schaeffter. "Thoracic respiratory motion estimation from MRI using a statistical model and a 2-D image navigator". In: *Medical Image Analysis* (2011).

[19] T. Kohlberger, M. Sofka, J. Zhang, N. Birkbeck, J. Wetzl, J. Kaftan, J. Declerck, and S. Zhou. "Automatic multi-organ segmentation using learning-based segmentation and level set optimization". In: *Medical Image Computing and Computer-Assisted Intervention 2011* (2011), pp. 338–345.

[20] A. Al-Mayah, J. Moseley, and K Brock. "Contact surface and material nonlinearity modeling of human lungs". In: *Physics in Medicine and Biology* 53 (2008), pp. 305–317.

[21] A. Al-Mayah, J. Moseley, M. Velec, and KK Brock. "Sliding characteristic and material compressibility of human lung: Parametric study and verification". In: *Medical Physics* 36 (2009), pp. 4625–4633.

[22] J.R. McClelland, J.M. Blackall, S. Tarte, A.C. Chandler, S. Hughes, S. Ahmad, D.B. Landau, and D.J. Hawkes. "A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy". In: *Medical Physics* 33 (2006), pp. 3348–3358.

[23] JR McClelland, David J Hawkes, Tobias Schaeffter, and Andrew P King. "Respiratory motion models: A review". In: *Medical image analysis* 17.1 (2013), pp. 19–42.

[24] SA Nehmeh, YE Erdi, CC Ling, KE Rosenzweig, OD Squire, LE Braban, E. Ford, K. Sidhu, GS Mageras, SM Larson, and J. L. Humm. "Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer". In: *Medical Physics* 29 (2002), pp. 366–371.

[25] M. Nesme, Y. Payan, F. Faure, et al. "Efficient, physically plausible finite elements". In: *Eurographics 2005 Short Papers* (2005).

[26] M.J.D. Powell. "Developments of NEWUOA for minimization without derivatives". In: *IMA journal of numerical analysis* 28.4 (2008), pp. 649–664.

[27] P. Risholm, J. Ross, G. Washko, and W. Wells. “Probabilistic Elastography: Estimating Lung Elasticity”. In: *Information Processing in Medical Imaging*. Springer. 2011, pp. 699–710.

[28] L Risser, F Vialard, H Baluwala, and J Schnabel. “Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions”. In: *Medical Image Analysis* 17.2 (2013), pp. 182–193.

[29] A. Schmidt-Richberg, J. Ehrhardt, R. Werner, and H. Handels. “Integrierte Segmentierung und Trajektorienberechnung mittels diffeomorpher Registrierung in räumlich-zeitlichen CT-Bildfolgen”. In: *Bildverarbeitung für die Medizin 2009* 16 (2009), pp. 182–186.

[30] A. Schmidt-Richberg, R. Werner, H. Handels, and J. Ehrhardt. “Estimation of Slipping Organ Motion by Registration with Direction-Dependent Regularization”. In: *Medical Image Analysis* 16 (2011), pp. 150–159.

[31] SS Vedam, PJ Keall, VR Kini, H. Mostafavi, HP Shukla, and R. Mohan. “Acquiring a four-dimensional computed tomography dataset using an external respiratory signal”. In: *Physics in Medicine and Biology* 48 (2003), pp. 45–62.

[32] P.F. Villard, M. Beuve, B. Shariat, V. Baudet, and F. Jaillet. “Simulation of lung behaviour with finite elements: Influence of bio-mechanical parameters”. In: *Medical Information Visualisation-Biomedical Visualisation 2005*. IEEE. 2005, pp. 9–14.

[33] C. Wachinger, M. Yigitsoy, E.J. Rijkhorst, and N. Navab. “Manifold learning for image-based breathing gating in ultrasound and MRI”. In: *Medical Image Analysis* (2011), pp. 806–818.

[34] R. Werner, J. Ehrhardt, R. Schmidt, and H. Handels. “Patient-specific finite element modeling of respiratory lung motion using 4D CT image data”. In: *Medical Physics* 36 (2009), pp. 1500–1511.

[35] J.B. West and F.L. Matthews. “Stresses, strains, and surface pressures in the lung caused by its weight”. In: *Journal of Applied Physiology* 32.3 (1972), pp. 332–345.

[36] E Wilson and K BATHE. *Numerical methods in finite element analysis*. Prentice-Hall, 1976.

[37] N.M. Wink, C. Panknin, and T.D. Solberg. “Phase versus amplitude sorting of 4D-CT data”. In: *Journal of Applied Clinical Medical Physics* 7.1 (2006), pp. 77–85.

[38] Y Zeng, D Yager, and Y Fung. “Measurement of the mechanical properties of the human lung tissue”. In: *Journal of Biomechanical Engineering* 109.2 (1987), pp. 169–174.

[39] T. Zhang, R. Jeraj, H. Keller, W. Lu, G Olivera, T Mackie, and B. Paliwal. “Treatment plan optimization incorporating respiratory motion”. In: *Medical Physics* 31 (2004), pp. 1576–1586.

APPENDIX B

Automatic US-MRI Registration for Neurosurgery using the 2D and 3D LC^2 Metric (MedIA 2014)

Bernhard Fuerst[1,2], Wolfgang Wein[3], Markus Müller[1,3], Nassir Navab[1,2]

[1] Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstrae 3, 85748 Garching b. München, Germany

[2] Computer Aided Medical Procedures (CAMP), Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

[3] ImFusion GmbH, Agnes-Pockels-Bogen 1, 80992 München, Germany

Copyright © 2014 Elsevier B.V. All rights reserved.

Contribution The main contribution is the presentation of a novel similarity measure to compare ultrasound and MRI data and therefore enable intensity-based registration. The construction of the similarity measure is thoroughly explained, and applied to 2D and 3D data sets. The extensive evaluation comprises a convergence study and a validation of accuracy.

Abstract To enable image guided neurosurgery, the alignment of pre-interventional magnetic resonance imaging (MRI) and intra-operative ultrasound (US) is commonly required. We present two automatic image registration algorithms using the similarity measure Linear Correlation of Linear Combination (LC^2) to align either freehand US slices or US volumes with MRI images. Both approaches allow an automatic and robust registration, while the three dimensional method yields a significantly improved percentage of optimally aligned registrations for randomly chosen clinically relevant initializations. This study presents a detailed description of the methodology and an extensive evaluation showing an accuracy of 2.51 mm, precision of 0.85 mm and capture range of 15 mm (>95% convergence) using 14 clinical neurosurgical cases.

B.1 Introduction

Medical image registration is the process of spatially aligning images in a common coordinate space and aligning related features which exist in all images. It has been a widely investigated area in the past few decades, however remains challenging in particular for multi-modal registration. Often, different modalities complement each other well, which is relevant to a vast range of clinical applications for improving diagnosis, treatment planning, interventions, procedure follow-up, and screening. In a neurosurgical scenario which mainly motivates this work, MRI provides a good visualization of the anatomy and tumors, while US is inexpensive and allows for intra-operative use to detect and correct for brain shift after opening the skull. However, registering US and MRI images is a complex and difficult process, largely because represented information originates from very different physical properties. MRI intensities correlate with the relaxation times of the ^{1}H nuclei, while the US intensity values represent the changes in acoustic impedance, overlaid by a significant speckle noise and various direction-dependent artifacts.

In this article, we present a new powerful set of methods based on the previously proposed LC^2 similarity measure [20], which allows for globally convergent, automatic registration of MRI and US data with clinically acceptable computation times.

B.2 Related Work

During MRI and US registration a transformation is searched for which the alignment of the images is optimal. This requires a measure to evaluate the current alignment of the images, which is referred to as *cost functions* or *similarity measure*. Ideally, this function exhibits one distinctive extremum when the images are aligned optimally, and a nearly monotonous shape to provide optimal support in finding this extremum. In this section we will discuss several similarity measures which have been utilized for MR and US registration.

The MRI and US registration approaches using similarity measures based on sum of squared distances, Normalized Cross-Correlation (NCC), Mutual Information (MI) and normalized Mutual Information (nMI) tend to fail [8]. This is caused by the very different nature of the intensity values and by structures that are not visible in one or the other imaging modality. For instance, details in MRI may lay in US shadow regions or certain materials can not be visualized by MRI (e.g. calcifications, air). Therefore, we focus on similarity measures specific to the application during MRI and US registration, which are not organ-specific and do not introduce a significant effort due to pre-processing, such as for liver vasculature presented by [14].

Higher-dimensional Mutual Information (α-MI) is theoretically suited to assess US-MRI alignment based on both intensity and gradient information (in fact, an arbitrary number of features may be used). However, current approaches are neither practical in terms of implementation effort nor computation time [16, 7].

[5] presented an interesting approach utilizing the alignment of high confidence gradient orientations. Anatomical boundaries characterized by the gradient orientations from the MRI and US images are used, while small regions

with a high confidence for identifying anatomical boundaries were only selected from one image. However, the lack of the use of intensity values suggests that this method requires either nearly optimal data or a close initialization. Also the appearance of dominant gradients in one but not the other image, such as the skull in MRI but not in US, may lead to a poor alignment.

A different powerful method is the modality independent neighborhood descriptor (MIND) [6] and its extension self-similarity context (SSC) [7, 2], which utilize the differences of pre-defined neighborhood descriptors. They are based on a self similarity measure initially presented by [1], and do not rely on the assumption of a global intensity relation. However, modality specific artifacts can not be considered and the computational effort for pre-processing is high due to the generation of voxel-wise neighborhood descriptors. Also, such self-similarity approaches tend to strongly abstract the image data, which might impact its accuracy as opposed to methods using the original image information.

Instead of comparing images from different modalities, pseudo-US images may be generated using segmented structures from MRI [3, 4, 10, 11]. In light of the modality-specific considerations, the most promising general strategy for robust US-MRI registration, without relying on application-specific pre-processing or segmentation, is to compare US to both the MRI intensity and its gradient, as pioneered by [17], where a global polynomial intensity relationship is fitted during registration. The alternating optimization of the rigid pose and the polynomial coefficients, as well as the fact that it is a global mapping, limit the convergence range though (the requirement for a local intensity mapping is explained in detail in [20] section 2.2.4). Powerful tools for image registration are similarity measures which are invariant to local changes, such as local normalized cross-correlation (invariant wrt. local brightness and contrast). [20] introduced the similarity measure Linear Correlation of Linear Combination (LC^2), which exhibits local invariance to how much two channels of information contribute to an ultrasound image. The entire method has been specially designed for US-CT registration, where a strong correlation between X-ray attenuation coefficients and acoustic impedance is known, which allows a simulation of ultrasound effects from CT. These incorporate estimates of the acoustic attenuation, multiple reflections, and shadowing, which can not directly be estimated from MRI.

In this work, we adapt the LC^2 formulation to the case of MRI-US registration in neurosurgery, and evaluate it on a publicly available database of 14 patients. We had presented initial results of this method in [22]. Here, we provide a more thorough description and evaluation, and also add two novel alternative implementations of the LC^2 similarity, namely a 2D GPU version and a novel natively three-dimensional approach.

B.3 Method

B.3.1 Similarity Measure

The similarity measure Linear Correlation of Linear Combination (LC^2) is used to search for a transformation T which aligns two images I and J. Due to the different nature of the images, a relationship function f is required to allow a mapping of the intensity values. As a first step towards LC^2 a general cost function is defined, which applies the relationship function f to one of the images:

$$\min_{T} \sum_{\mathbf{x} \in \Phi} \Big(I(\mathbf{x}) - f(J(T(\mathbf{x}))) \Big)^2, \tag{B.1}$$

where $\mathbf{x}$ is a pixel or voxel position in the image domain Φ. If f is the identity function it can be seen that (B.1) represents sum of squared differences (SSD). In case of LC^2, the relationship function f is defined to be a linear combination such as $f(J(\mathbf{y})) = \sum_{i=1\dots n} c_i g_i(J(\mathbf{y}))$, where c_i are coefficients, and $\mathbf{y}$ an arbitrary position in image J. The functions g_i describe an arbitrary pre-processing of the transformed image J, for instance the computation of gradients or simulation of pseudo-ultrasound images. The choice of relationship function and pre-processing strongly depends on the nature of the underlying images. For mono-modal image registration, one could assume that a simple linear function, such as $f(J(\mathbf{y})) = \alpha J(\mathbf{y}) + \beta$, would be sufficient, which basically represents a windowing function.

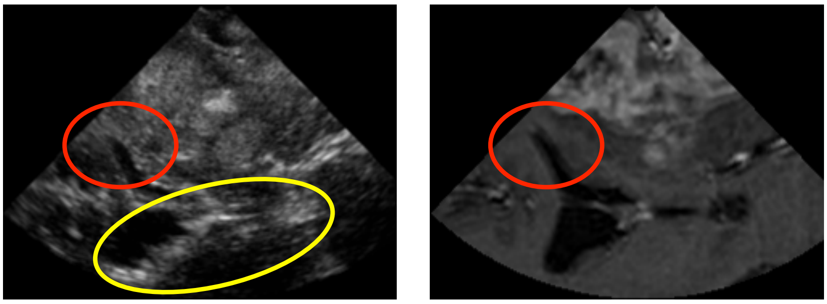
In case of a multi-modal image registration the relationship is more complex. As shown by [20], the relationship between X-ray computed tomography (CT) and ultrasound images can be motivated from a physics standpoint, which results in f being a linear combination of ultrasound reflection and ultrasound echogeneity simulated from CT. However, in case of MRI and US registration such a simulation can not be obtained as the physical properties that are responsible for the intensity values have little in common. Therefore, the relationship function f is based on plausible observations. We assume that the US intensity value u_i for pixel/voxel i is either correlated with the MRI intensity value $p_i = J(T(\mathbf{x_i}))$ or with the MRI image gradient magnitude $g_i = |\nabla p_i|$. The resulting relationship function is therefore $f(J(\mathbf{x})) = \alpha p_i + \beta g_i + \gamma$. This caters to the fact that ultrasound intensity values on one hand may depict different soft tissue properties (due to the varying tissue inhomogeneities and echogeneity), and on the other hand represent tissue interfaces or gradients, as illustrated in figure B.1. Of course LC^2 is not limited to linear combinations of only two components, but for MRI/US registration we currently do not see the need for additional components.

The coefficients $c = \{\alpha, \beta, \gamma\}$ of the linear combination in the relationship function are computed during every similarity measure evaluation, as shown by [21, 20]. Under the assumption that the relationship function $f(J(T\mathbf{x}))$ shall be an optimal fit to the image $I(\mathbf{x})$, its coefficients can be implicitly estimated based on an ordinary least squares formulation:

$$\hat{c} = (M^T M)^{-1} M^T U,$$
$$\text{where } M = \begin{pmatrix} p_1 & g_1 & 1 \\ \vdots & \vdots & \vdots \\ p_m & g_m & 1 \end{pmatrix}, \; U = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \tag{B.2}$$

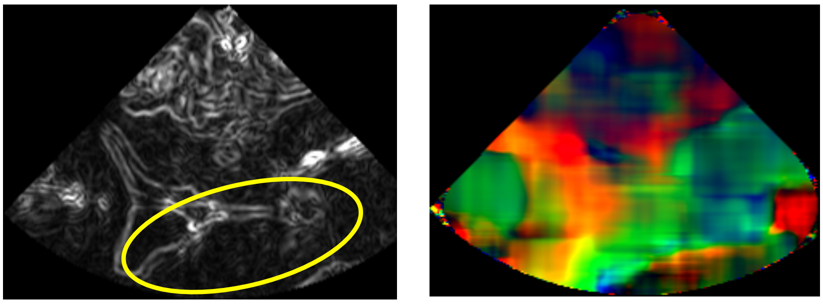
where m is the number of pixels/voxels in the domain, for instance $m = |\Phi|$.

It is now possible to compute a residual between the US and MRI images, but this proves to be problematic in regions in which the images do not overlap or in which the ultrasound image does not contain any structural information (e.g. shadow regions). Therefore, the cost function (B.1) is modified to penalize



(a) US image

(b) Corresponding MRI slice (p)



(c) Gradient magnitudes of MRI slice (g)

(d) Locally estimated coefficients $\hat{y}$

Fig. B.1: The intensity values of the US image need to be expressed by a locally varying relationship function. This is indicated by the red circle, which depicts a high correspondence between US intensity values as MRI intensity values (a and b), and the yellow ellipse, which marks regions of high correspondance between the US image and the MRI gradient magnitudes (a and c). The locally estimated coefficients (for neighboorhood $\Omega(\mathbf{x}, s)$) of the linear combination in the relationship function are visualized in (d).

such regions by introducing the variance of the ultrasound image, as done by [18, 19, 17], and is formulated as a similarity measure:

$$LC^2(I, J, T) = 1 - \frac{\sum_{\mathbf{x} \in \Phi} (I(\mathbf{x}) - f(J(T(\mathbf{x}))))^2}{|\Phi| \, Var(I)}. \tag{B.3}$$

When plugging simple relationship functions in (B.3), it can be shown that other cost functions, such as correlation ratio (η) or normalized cross-correlation (NCC), can be derived.

Locally normalized LC^2 The similarity measure (B.3) implies an image-wide constant relationship between US and MRI intensity values and MRI image gradient magnitudes. This, however, is not true in most cases, as visualized in Fig. B.1. Therefore, we compute (B.3) for every pixel position $\mathbf{x} \in \Phi$ using a neighborhood $\Omega(\mathbf{x}, s) \subset \Phi$ of size s, and consequently replace the domain on which the coefficients $\hat{y}$ are estimated (Eq. B.2) with $\Omega(\mathbf{x}, s)$. The size s of the neighborhood around an image position $\mathbf{x}$ defines a patch of $(2s+1)^2$ pixels or a volume $(2s+1)^3$ voxels in 2D or 3D respectively. This results in local similarity measures

$$LC(I, J, T, \mathbf{x}, s) = 1 - \frac{\sum (I(\mathbf{y}) - f(J(T(\mathbf{y}))))}{|\Omega(\mathbf{x}, s))| \, Var(I(\mathbf{y} \mid \mathbf{y} \in \Omega(\mathbf{x}, s)))}, \tag{B.4}$$

where the weighted average is computed in order to obtain the image similarity measure. The weights are the local standard deviation of the image I in the neighborhood ($\sigma = \sqrt{Var(I(y|y \in \Omega(\mathbf{x}, s)))}$).

B.3.2 Similarity Measure Computation

Two-dimensional LC^2 This approach is computed on the original tracked freehand ultrasound slices, for which a 3D transformation is given. The extraction of the corresponding MRI intensity values and MRI gradient magnitudes from the given 3D volume is efficiently performed on the GPU using its hardware tri-linear interpolation capability. The accumulation of all the neighborhood intensities required for equation B.4 is performed using a sliding-window approach on the CPU, parallelized over rows and columns of the images on all threads of the multi-core processor. The computation time is therefore independent of the chosen LC^2 neighborhood size. While this method has been initially used in [22], we have investigated an alternative full GPU implementation, where each shader accumulates the neighborhood information independently. The latter approach results in a computation dependency with respect to the neighborhood size of $O(s^2)$, and therefore offers superior performance for small neighborhood sizes.

Three dimensional LC^2 In order to investigate advantages of slice versus volume-based LC^2 computation, we have also implemented the similarity measure with three-dimensional neighborhood blocks on the GPU. Here, the intensity accumulation is implemented in a separable fashion for every dimension, such that the computation time scales linearly $O(s)$ with the neighborhood size s. Before the registration, the 3D freehand ultrasound data is reconstructed into a volume grid in a similar fashion as in [9], using a quadrilateral interpolation for a good trade-off of performance and image quality.

B.3.3 Optimization of Rigid Transformation

An analytic derivation of LC^2 is difficult due to the least-squares fitting in (B.2) which is computed for every position in the US image. Therefore we use Bound Optimization BY Quadratic Approximation (BOBYQA) [15], which internally creates own derivative approximations. This results in fewer evaluations than most other search methods, and is therefore used throughout this paper. However, clinical requirements on capture range may ask for other techniques. In particular, global optimization techniques may be useful to perform a more excessive search within the specified bounds.

Deformable registation After rigid registration, a free-from deformation (FFD) model using cubic splines can be fitted, where the deformation is applied on the MRI data J within the same GPU kernel which extracts MRI intensity and gradient magnitude. For that purpose, we place a configuration of 2×2×4 control points within the bounding box of the registered ultrasound sweep. Then the same BOBYQA algorithm is used to optimize the displacement vectors for all control points.

B.4 Experiments

B.4.1 Clinical Data and Experimental Setup

To evaluate our method and compare the results to other publications, we used a publicly available database containing Brain Images with Tumors for Evaluation from Montreal Neurological Institute [12], with pre-operative T1-weighted MRI and pre-resection 3D freehand US from 14 patients. The pre-resection ultrasound has been acquired before opening the dura, and therefore only little deformation has occurred. Initial transformations and corresponding landmarks for each US-MRI pair are included (Tab. B.1, lines 1 and 2). Therefore, we can provide ground truth evaluations, and denote the average Euclidean distance of the landmarks as Fiducial Registration Error (FRE).

Two dimensional LC^2 The MRI volumes were used as provided, while the higher-resolution US images were down-sampled such that their pixel sizes is smaller than twice the size of an MRI voxel. This guarantees that information provided by MRI voxels is never discarded when the tri-linear interpolation is used. Furthermore, US slices were skipped to avoid overlapping planes, resulting in an average distance between the slices of $<$ 1.5 mm or less due to slower scanning in the areas of interest.

Three dimensional LC^2 For the 3D experiments the freehand ultrasound data was first reconstructed into a cartesian volume grid with an isometric resolution of 0.3 mm, and afterwards further re-sampled as needed. We found that down-sampling the US volumes by the factor of three provides a good trade-off between fast computation times, a total load of <2GB memory on GPU, a high capture range and good accuracy. In addition, speckle noise is removed due to the smoothing effect. The used US volumes yield an isometric resolution of 0.9 mm. The MRI volumes were used as provided.

System specifications All registration attempts were performed utilizing the parallel processing capabilities of the Graphics Processing Unit (GPU) of a workstation with an Intel i7-3770 CPU with 8 threads and a Nvidia GeForce GTX Titan GPU with 2688 cores and 6GB memory.

B.4.2 Numerical Analysis of LC^2 Configuration

All convergence analyses were performed by carrying out 100 randomly initialized transformations (±10 mm/° in all 6 parameters) for each choice of a parameter.

Convergence analysis in terms of US slice spacing Consistently good results are obtained when performing a registration using the 2D approach with an *slice spacing* of $<$ 5 mm, where *slice spacing* refers to the average euclidean distance between the centers of tracked ultrasound slices. For deformable registration, we chose < 1.5 mm to make sure we are not missing even smallest structures.

Convergence analysis in terms of neighborhood size The sensitivity of the 2D method has been investigated using both CPU and GPU implementation of the similarity measure. For neighborhood sizes 2 to 24 the accuracy is similar, while the percentage of successful registration attempts peak around 8 and 9, as depicted in Fig B.2 (a) and (b). Furthermore, it can be seen that the two implementations yield nearly the same results. Therefore, we used neighborhood $s = 9$ (hence $m = (2{\cdot}9{+}1)^2 = 361$ pixels) as a fair trade-off between convergence and accuracy for all further experiments. Reported results are computed using the CPU implementation.

The 3D method requires an independent convergence analysis, as the neighborhood size describes a volume rather than a 2D patch. The accuracy is nearly constant for neighborhood sizes 2 to 7, which also exhibits the highest percentage of successful registrations (Fig. B.2c). To balance computation time, which scales linearly with the neighborhood size, and the performance, a neighborhood size $s = 3$ (hence $m = (2 \cdot 3 + 1)^3 = 343$ voxels) is used for all further experiments. Overly large patches result in a global mapping of MRI intensity and gradient, removing the main advantage of LC^2 over other methods (robustness wrt. local changes of intensity-gradient relationship).

Gradient magnitudes vs. directed gradients Finally, we have investigated the effect of using the dot product of the MRI gradient g with the US beam direction, instead of g directly. This reduces the influence of vertical gradients, similar to the US simulation presented by [20]. Interestingly, this results in $10 - 25\%$ more outliers (the cost function becomes more non-linear due to the added directional dependance).

B.4.3 Registration Results

The resulting errors for all 14 patient data sets are nearly the same for registrations using gradient orientation alignment [5], 2D LC^2, or 3D LC^2, as depicted in Tab. B.1, lines 4, 6, 12 respectively. [16] report slightly higher errors when applying a costly deformable registration which requires several hours (line 5). The increased FRE coincides with our findings when using the faster 2D LC^2 during deformable registrations (line 10). This indicates that performing deformable registration does not provide any significant benefit when applied to mostly rigid data sets. However, we are convinced that the change of landmark errors induced by deformable registration or the difference between the 2D and 3D approach of LC^2 lay within the range of the fiducial localization error (FLE) of the data. Examples of initially aligned and registered images are shown in Fig. B.3.

B.4.4 Accuracy, Precision and Capture Range

Some initial alignments yield significant errors (e.g. patients 2, 3, 9, 13), which are reduced by all algorithms listed in Tab. B.1. Therefore, an analysis of

Dataset Overview and Related Methods																
	Patient #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	mean
1	# of Tags	37	35	40	32	31	37	19	23	21	25	25	21	23	23	-
2	Initial FRE (mm)	4.93	6.30	9.38	3.93	2.62	2.30	3.04	3.75	5.09	2.99	1.52	3.70	5.15	3.77	4.18±5.20
3	US Spacing (mm)	0.24	0.42	0.23	0.20	0.25	0.17	0.24	0.18	0.18	0.22	0.16	0.18	0.21	0.19	0.22±0.20
4	FRE in [5] (mm)	4.89	1.79	2.73	1.68	2.12	1.81	2.51	2.63	2.7	1.95	1.56	2.64	3.47	2.94	2.53±0.87
5	FRE in [16] (mm)	-	2.05	2.76	1.92	2.71	1.89	2.05	2.89	2.93	2.75	1.28	2.67	2.82	2.34	2.57±0.82
Registration Results using 2D LC																
6	FRE (mm)	4.82	1.73	2.76	1.96	2.14	1.94	2.33	2.87	2.81	2.06	2.18	2.67	3.58	2.48	2.52 ±0.87
7	SD (mm)	0.01	0.01	0.01	0.01	0.02	0.01	0.05	0.30	0.02	0.00	0.03	0.15	0.05	0.04	0.05 ±0.08
8	Duration (sec)	5.9	8.3	11.1	5.7	7.1	8.2	18.2	8.6	6.0	23.4	17.3	25.8	8.1	7.0	11.5 ±6.8
9	Convergence (%)	54	66	44	60	75	38	51	19	54	76	82	94	53	44	57.9 ±19.6
10	FRE Def. (mm)	4.95	1.64	2.43	1.91	2.26	2.2	2.52	3.64	2.65	2.09	1.76	2.45	3.71	2.76	2.64 ±0.9
11	Time Def. (sec)	158	141	279	92	133	166	563	312	76	675	597	93	106	282	262 ±204
Registration Results using 3D LC																
12	FRE (mm)	4.86	1.70	2.55	1.73	1.96	1.83	2.32	2.68	2.74	2.09	1.81	2.71	3.44	2.45	2.49 ±0.84
13	SD (mm)	0.01	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.01	0.01	0.02	0.03	0.02	0.01	0.01 ±0.01
14	Duration (sec)	2.61	2.91	2.89	1.30	2.15	2.08	2.87	2.68	1.49	2.58	2.23	2.64	1.24	2.79	2.32 ±0.59
15	Convergence (%)	96	85	88	89	90	80	81	76	80	91	100	94	86	92	87.7 ±6.8

Table B.1: Overview of clinical data [12], previous published results [5, 16], and results using our method for rigid and deformable registration including computation times. The presented registration results using 2D and 3D LC^2 are based on 100 randomly initialized registration attempts for each patient and method.

the capability of our algorithm to reach the optimum under all conditions is necessary.

Trials with each 100 randomly initialized transformations ($\pm 10 mm/^\circ$ in all 6 parameters) were performed for all 14 patients using the 2D and 3D LC^2 approach, resulting in a total of 2800 registration attempts. Comparing the LC^2 similarity measure with the final FRE shows in all cases that the best transformation corresponds to the highest similarity and that the misalignments are cleary separated yielding a significantly lower similarity. This demonstrates that both LC^2 algorithms allow for global registration in a realistic clinical setup. Fig. B.4 shows the results, including the percentages of the converged optimizations. The average errors (*accuracy*) and standard deviation (SD; *precision*) are listed in Tab. B.1.

The *capture range* describes the range of initial FRE values for which $> 95\%$ of the registration approaches are successful. When using the 2D approach, it can be observed that 95% of the experiments converge within an initial FRE of 9 mm. This capture range is significantly increased to 15 mm when using the 3D approach. Both sets of experiments are based on the aforementioned randomly initialized studies using BOBYQA. Figure B.5 (a) and (b) depict the capture range for selected patients, while (c) shows the total percentage of outliers vs. initial FRE for both algorithms.

Since the gradient orientation alignment (GOA) method [5] yields similar FRE values, we implemented it to the best of our knowledge and re-ran the aforementioned randomized trials with it. We obtain $> 90\%$ outliers and further investigation into the cost function properties revealed that only a minor local optimum is present. A possible explanation is, that without further heuristics the GOA method would line up strong gradients from e.g. dura mater or skull; besides, using only gradients larger than a threshold limits the image content considered. While we believe these to be general issues, it has to be acknowledged that better results would probably be obtained by the original authors, e.g. by changing implementation details such as resolution, smoothing and interpolation. Precision and capture range were not reported in their work though, as unfortunately too often the case.

The randomized experiments with the 2D and 3D LC^2 versions have been compared using the Mann-Whitney U-test. Generally, a p-value of ≤ 0.01 is considered *very significant* [13]. In fact, the 3D approach shows such a *very significant* increased convergence in all but one patients (patient 12 unchanged at 94%). With the exception for patient 12, all computed p-values are below $6 \cdot 10^{-3}$, indicating that the chance of the improvement being coincidentally observed is less than 0.6%.

B.4.5 Computation Time

To compare the performance of the 2D CPU, 2D GPU and 3D GPU approaches, we have measured the average computation times. All randomly initialized registrations have been run on the same hardware. The reported times do not include the file loading, ultrasound volume reconstruction and down-sampling.

The first, original 2D implementation uses the GPU merely to perform the MRI slice extraction, while the similarity is computed using a sliding-window approach on the CPU. The average run time is 11.6 sec. In the second 2D implementation we have performed the similarity evaluation using a GPU based filtering strategy, which reduces the average run time by only 5%, with quadratic dependency on the LC^2 neighborhood size. Here, the similarity measure computation does not utilize the full capabilities of the GPU yet, since only a parallelization within the (small) ultrasound frames is used.

Finally, we measured the average registration time for the GPU based 3D approach. Due to the implementation as separable filter, the performance scales linearly with the neighborhood size. Using a neighborhood size $s = 3$ an average run time of 2.32 ± 0.59 sec can be observed.

B.5 Discussion and Conclusion

B.5.1 Choice of Method

The LC^2 computation on the original freehand ultrasound 2D images yields a number of advantages. First, an offline volume reconstruction step is avoided, which might reduce the quality of the original ultrasound image information. Considering the limited voxel resolution of the MRI data this, however, does not pose a problem since we need to further down-sample the ultrasound data anyway. More importantly, a slice-based approach may immediately start looking for the correct alignment in real-time once the first frames are obtained. Besides, an optimization of the calibration transformation or compensation of tracking errors may be computed on-the-fly, which would otherwise require recomputation of the 3D ultrasound volume.

The 3D volume-based approach yields a superior capture range, which can be explained by the fact that the LC^2 neighborhood stretches into the third dimension, therefore allowing even more consistent local matching of the relationship function f. This comes with the trade-off of higher computing resource requirements. While our implementation as fully parallelized separable filters in every dimension on the GPU ensures efficient computation, it requires a significant memory footprint (< 2 GB for US voxel size 0.9 mm, $2 - 3$ GB for US voxel size 0.6 mm).

Finally, a comparison of the same 2D approach on CPU and GPU depends very much on the actual implementation. In our case, the CPU method is completely independent of the neighborhood size s, whereas the GPU version has a squared dependence. The latter has room for improvement in terms of both a separable computation and parallelization over the US frames. It is also important to note that numerically equally stable behavior on the GPU is only obtained when the actual computation of equation (B.2) is performed with double floating-point precision. This in turn mandates a certain choice (and cost) of GPU hardware.

B.5.2 Performance

As opposed to other proposed methods such as [5, 7], the LC^2 method uses the full amount of available image information from both modalities, by locally matching ultrasound intensities to both MRI intensity and gradient magnitude. A direct comparison is unfortunately not possible, since other studies lack an evaluation of precision and capture range (which is crucial for developing an automatic registration in a given clinical context). The accuracy in terms of the point-based registration error is consistent with other studies, and is always < 1 mm higher than the best possible rigid fit of the point correspondences themselves. Deformable registration does neither significantly increase or decrease the errors, however the visual alignment generally improves (see Fig. B.3).

The computation time of the 2D and 3D GPU implementations allow for $100 - 500$ cost function evaluations per second. Hence global registration within a clinically realistic bounding box of the pose parameters is possible. To our knowledge, this has not been shown in related work to date.

The LC^2 similarity measure has a single parameter, namely the neighborhood size s. We have shown that both the 2D and 3D variants work well within a fairly large range of s. However, as common in image registration scenarios, further tunable parameters arise in the pre-processing (e.g. down-sampling, volume reconstruction) of the image data, as well as the choice and configuration of a non-linear optimization algorithm.

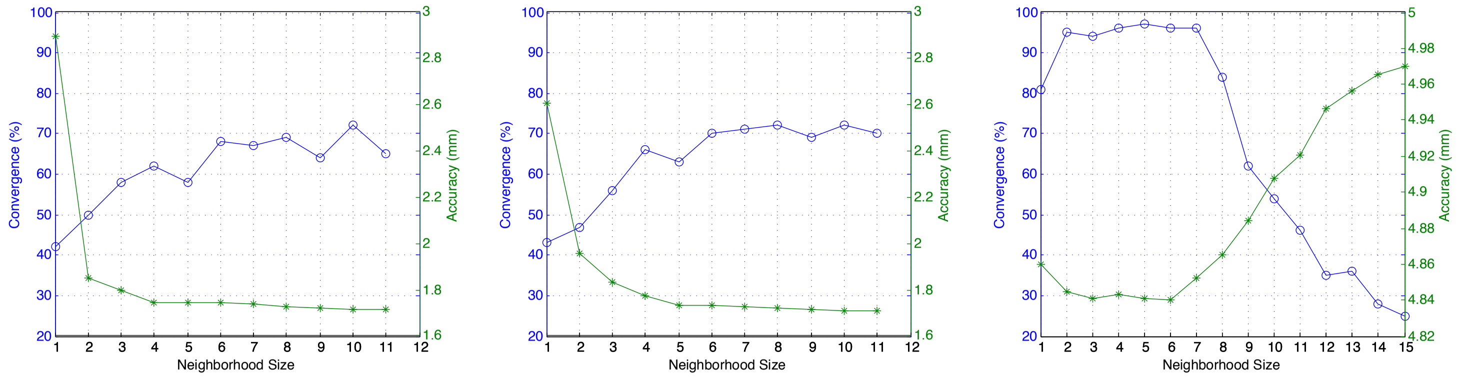
B.5.3 Future Work

While our approach adequately solves the problem of image-based MRI-US registration for correcting brain shift, some further work is required to address resection follow-up. Missing anatomical correspondence in the area of a tumour resection site might result in unstable registration results of our straight-forward free-form deformation model, when looking at the registration of pre- and post-resection ultrasound data. In order to provide a clinically meaningful solution in this context, however, a detailed discussion with physicians will be required in the first place, to establish how such data before and after the procedure shall be transformed and compared.

In other clinical application areas, such as oncological diagnosis in the abdomen, it would be desirable to look into extensions of our deformation model to incorporate physical constraints of the complex deformations due to respiratory motion and patient positioning. The same holds true for potential applications in the case of prostate MRI-US fusion [23]. The incorporation of bio-mechanical deformation models might be particularly suited to address such difficult non-linear registration problems.

Acknowledgements

The authors affiliated with Technische Universität München are partially supported by the EU 7th Framework Program (FP7/2007-2013 and FP7/ICT-2009-6) under Grant Agreements No. 256984 (EndoTOFPET-US) and No. 270460 (ACTIVE).

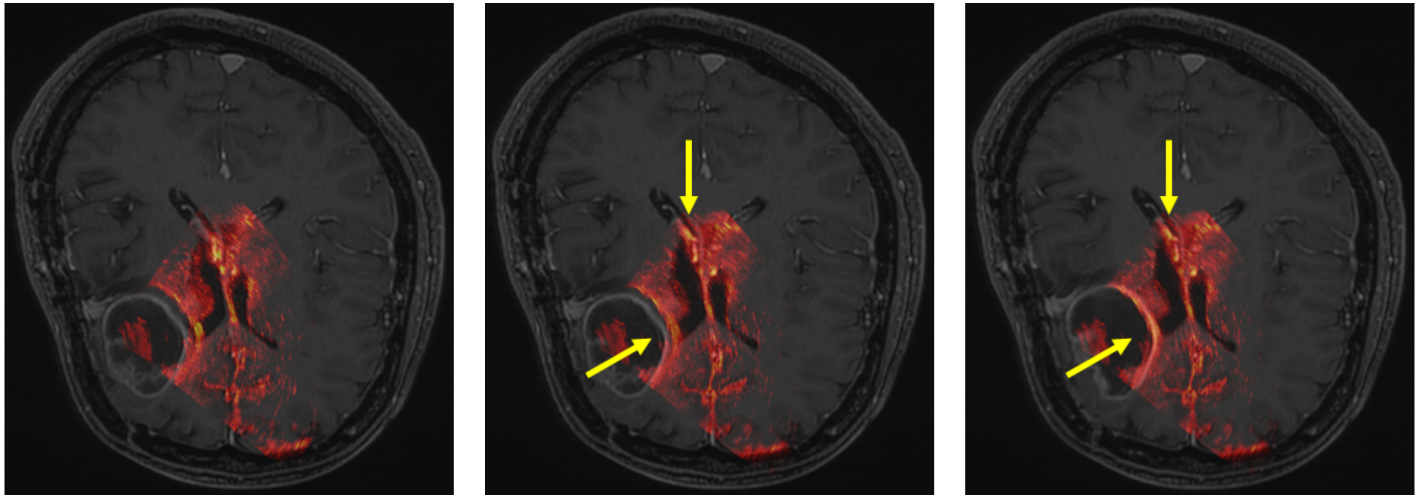


(a) 2D LC on CPU (Patient 2)

(b) 2D LC on GPU (Patient 2)

(c) 3D LC (Patient 1)

Fig. B.2: Relationship between convergence (blue), accuracy (green) and the neighborhood size s for the three different LC^2 methods. Differences between the two 2D implementations (a and b) are caused by different numerical behaviour in the implementation (e.g. on the CPU everything is done with double precision, on the GPU only the final matrix inversion), not by differences in the methodology. The convergence analysis indicates that a neighborhood size of $s = 9$ ($m = 361$ pixels) yields a fair trade-off between convergence, accuracy and computation time. The typical behavior of the 3D approach is shown in (c), indicating $s = 3$ ($m = 343$ voxels) being a good choice of the neighborhood size. Note that patient 2 and 1 show the typical behavior of the 2D and 3D approaches respectively and are therefore plotted here.



(a) Initial Alignment

(b) Rigid Registration

(c) Deformable Registration

Fig. B.3: Superimposed US on axial MRI slice to show the registration result of patient 6. The FRE improves by 0.11 mm from rigid to deformable registration.

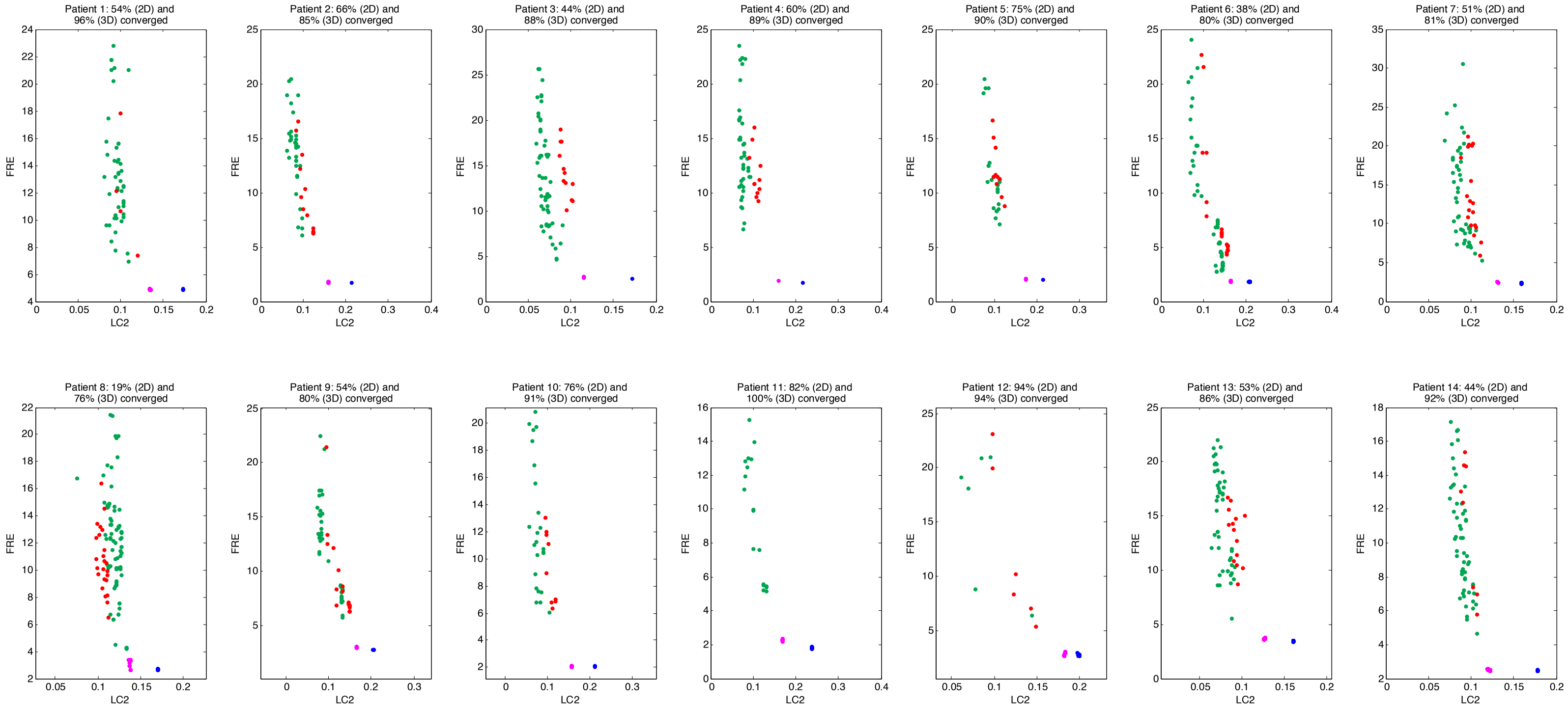
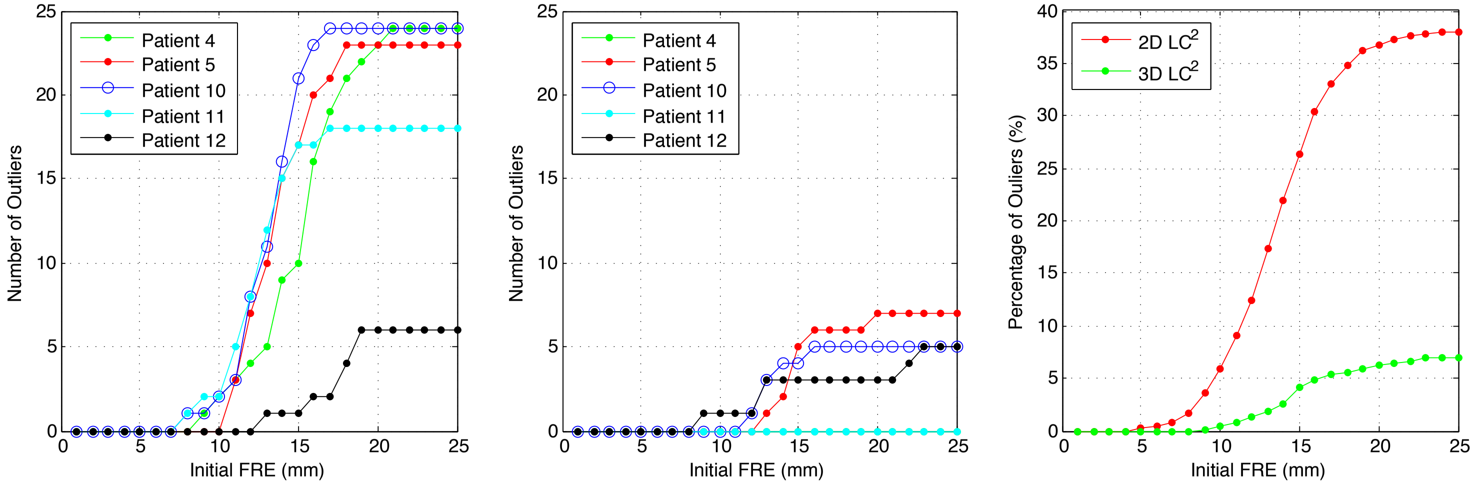


Fig. B.4: Registration experiments with 100 randomly initialized transformations for all patient data sets. The converged results (2D: magenta, 3D: blue) are clearly separated from the failed registration attempts (outliers: 2D: green, 3D: red). Convergence is defined based on typical patient-specific FRE values as reported in Tab. B.1.



(a) 2D LC capture range

(b) 3D LC capture range

(c) Total capture range

Fig. B.5: The capture range describes the relation between outliers and initial FRE. For selected patients this relation is plotted in (a) and (b) for the 2D and 3D LC^2 approach resprectively. The total capture range of the algorithm for all patients is plotted in (c). It can be seen that the the capture range (less than 5% outliers) for 2D LC^2 is 9 mm, while it is increased to 15 mm for the 3D approach.

References

[1] Antoni Buades, Bartomeu Coll, and J-M Morel. “A non-local algorithm for image denoising”. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE. 2005, pp. 60–65.

[2] Amalia Cifor, Laurent Risser, Mattias P Heinrich, Daniel Chung, and Julia A Schnabel. “Rigid Registration of Untracked Freehand 2D Ultrasound Sweeps to 3D CT of Liver Tumours”. In: *Abdominal Imaging. Computation and Clinical Applications*. Springer, 2013, pp. 155–164.

[3] Roch M Comeau, Abbas F Sadikot, Aaron Fenster, and Terry M Peters. “Intraoperative Ultrasound for Guidance and Tissue Shift Correction in Image-Guided Neurosurgery”. In: *Medical Physics* 27 (2000), p. 787.

[4] P. Coupé, P. Hellier, X. Morandi, and C. Barillot. “3D Rigid Registration of Intraoperative Ultrasound and Preoperative MR Brain Images Based on Hyperechogenic Structures”. In: *Journal of Biomedical Imaging* 2012 (2012), p. 1.

[5] D. De Nigris, D. Collins, and T. Arbel. “Fast and Robust Registration Based on Gradient Orientations: Case Study Matching Intra-Operative Ultrasound to Pre-Operative MRI in Neurosurgery”. In: *IPCAI* (2012), pp. 125–134.

[6] Mattias P Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V Gleeson, Sir Michael Brady, and Julia A Schnabel. “MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration”. In: *Medical Image Analysis* 16.7 (2012), pp. 1423–1435.

[7] Mattias Paul Heinrich, Mark Jenkinson, Bartlomiej W Papiez, Michael Brady, and Julia A Schnabel. “Towards realtime multimodal fusion for image-guided interventions using self-similarities”. In: *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 187–194.

[8] X. Huang, N. Hill, J. Ren, G. Guiraudon, D. Boughner, and T. Peters. “Dynamic 3D Ultrasound and MR Image Registration of the Beating Heart”. In: *Medical Image Computing and Computer-Assisted Intervention 2005* (2005), pp. 171–178.

[9] Athanasios Karamalis, Wolfgang Wein, Oliver Kutter, and Nassir Navab. “Fast hybrid freehand ultrasound volume reconstruction”. In: *Proc. of SPIE Vol.* Vol. 7261. 2009, pp. 726114–1.

[10] A.P. King, K.S. Rhode, Y. Ma, C. Yao, C. Jansen, R. Razavi, and G.P. Penney. “Registering Preprocedure Volumetric Images With Intraprocedure 3-D Ultrasound Using an Ultrasound Imaging Model”. In: *IEEE Trans. Med. Imag.* 29.3 (2010), pp. 924–937.

[11] M. Kuklisova-Murgasova, A. Cifor, R. Napolitano, A. Papageorghiou, G. Quaghebeur, J. Noble, and J. Schnabel. “Registration of 3D Fetal Brain US and MRI”. In: *Medical Image Computing and Computer-Assisted Intervention* (2012), pp. 667–674.

[12] L. Mercier, R.F. Del Maestro, K. Petrecca, D. Araujo, C. Haegelen, and D.L. Collins. “Online Database of Clinical MR and Ultrasound Images of Brain Tumors”. In: *Medical Physics* 39 (2012), p. 3253.

[13] Regina Nuzzo. “Scientific method: statistical errors”. In: *Nature* 506.7487 (2014), pp. 150–152.

[14] GP Penney, JM Blackall, MS Hamady, T. Sabharwal, A. Adam, DJ Hawkes, et al. “Registration of Freehand 3D Ultrasound and Magnetic Resonance Liver Images”. In: *Medical Image Analysis* 8.1 (2004), pp. 81–91.

[15] Michael JD Powell. “The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives”. In: *Cambridge Report NA2009/06, University of Cambridge* (2009).

[16] Hassan Rivaz and D Collins. “Self-similarity Weighted Mutual Information: A New Nonrigid Image Registration Metric”. In: *Medical Image Computing and Computer-Assisted Intervention* (2012), pp. 91–98.

[17] A. Roche, X. Pennec, G. Malandain, and N. Ayache. “Rigid Registration of 3-D Ultrasound with MR Images: A New Approach Combining Intensity and Gradient Information”. In: *Medical Imaging, IEEE Transactions on* 20.10 (2001), pp. 1038–1049.

[18] Alexis Roche, Gregoire Malandain, Xavier Pennec, and Nicholas Ayache. “The correlation ratio as a new similarity measure for multimodal image registration”. In: *Medical Image Computing and Computer-Assisted Interventation*. Springer, 1998, pp. 1115–1124.

[19] Alexis Roche, Xavier Pennec, Michael Rudolph, DP Auer, Grégoire Malandain, Sébastien Ourselin, Ludwig M Auer, and Nicholas Ayache. “Generalized correlation ratio for rigid registration of 3D ultrasound with MR images”. In: *Medical Image Computing and Computer-Assisted Intervention*. Springer. 2000, pp. 567–577.

[20] W. Wein, S. Brunke, A. Khamene, M.R. Callstrom, and N. Navab. “Automatic CT-Ultrasound Registration for Diagnostic Imaging and Image-Guided Intervention”. In: *Medical Image Analysis* 12.5 (2008), p. 577.

[21] W. Wein, A. Khamene, D.A. Clevert, O. Kutter, and N. Navab. “Simulation and Fully Automatic Multimodal Registration of Medical Ultrasound”. In: *Medical Image Computing and Computer-Assisted Intervention 2007* (2007), pp. 136–143.

[22] Wolfgang Wein, Alexander Ladikos, Bernhard Fuerst, Amit Shah, Kanishka Sharma, and Nassir Navab. “Global Registration of Ultrasound to MRI Using the LC2 Metric for Enabling Neurosurgical Guidance”. In: *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 34–41.

[23] C. Weiss, M. Seitz, K.A. Herrmann, A. Graser, B. Kiefer, M. Requardt, J. Fehre, R. Nanke, M. Diallo, A. Kamen, and W. Wein. “MR-US Fusion for Targeted Prostate Biopsy”. In: *International Society for Magnetic Resonance in Medicine (ISMRM) Proceedings*. May 2010.

APPENDIX C

Radiopositive Tissue Displacement Compensation for SPECT-Guided Surgery (MICCAI 2015)

Francisco Pinto[1,2*] and Bernhard Fuerst[1,2*], Benjamin Frisch[2], and Nassir Navab[1,2]

[*] F. Pinto and B. Fuerst are joint first authors.

[1] Computer Aided Medical Procedures (CAMP), Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

[2] Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstrae 3, 85748 Garching b. München, Germany

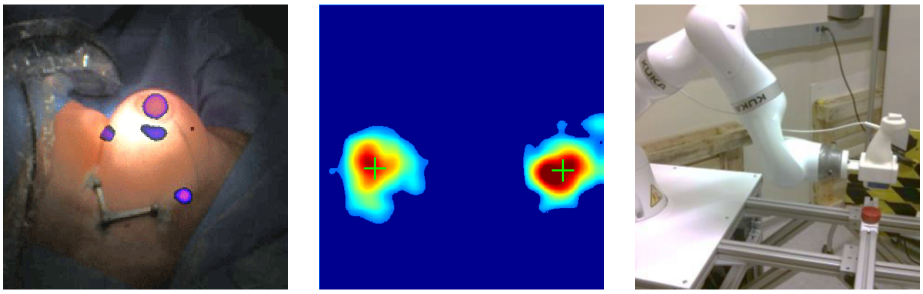
Copyright © Springer International Publishing Switzerland 2015.

Contribution During resection of Sentinel Lymph Nodes (SLN), the radiopositive tissue is displaced through the intervention itself. For the first time, this paper presents a methodology to update pre- or intra-operative SPECT volumes using a 2D gamma camera. The algorithm utilizes tracking of the gamma camera and the known response function to dynamically estimate tissue displacement.

Abstract We present a new technique to overcome a major disadvantage of SPECT-guided surgery, where a 3D image of the distribution of a radiotracer augments the live view of the surgical situs in order to identify radiopositive tissue for resection and subsequent histological analysis. In current systems, the reconstructed SPECT volume is outdated as soon as the situs is modified by further surgical actions, due to tissue displacement. Our technique intraoperatively estimates the displacement of radiopositive tissue, which enables the update of the SPECT image augmentation. After the initial SPECT reconstruction is complete, we deploy a 2D γ-camera along with a technique to optimize its placement. We automatically establish a correspondence between regions of interest in the reconstructed volume and the near real-time 2D γ images. The 3D displacement of the radiopositive nodules is then continuously estimated based

on the processing of the aforementioned γ-camera's output. Initial results show that we can estimate displacements with ±1 mm accuracy.

C.1 Introduction



(a) SPECT system showing several SLNs via Augmented Reality. Courtesy SurgicEye GmbH.

(b) Interpolated γ-camera output for two sources, showing the auto-marked centers.

(c) γ-camera mounted on a KUKA iiwa industrial lightweight robot. An off-screen PC collects data.

Fig. C.1: γ-detectors and respective outputs.

The Sentinel Lymph Node Biopsy (SLNB) is part of the standard of care for the treatment of melanoma [15], breast cancers [6, 5] and vulvar cancers [10, 13]. It further has demonstrated clinical value in the staging of head and neck [3], gastric [8], prostate [11], and cervical [4] cancers. In an SLNB, a radioactive tracer (usually a ^{99m}Tc nanocolloid) or a colored dye is injected close to a tumor, under the assumption that it primarily drains to the sentinel lymph nodes. Lymph nodes identified that way are resected and sent for subsequent histological analysis. In this paper, we focus on radioguided SLNB, which relies on either pre-interventional whole-body Single Photon Emission Computerized Tomography (SPECT) imaging or the use of a γ-detector for the live intraoperative identification of radiopositive tissue. Types of γ-detectors are 1D γ-probes and 2D γ-cameras that provide the surgeon with a planar view of the radioactivity distribution in the area (see Fig. C.1b). Brouwer et al. combine SPECT/CT with the intraoperative use of a laparoscopic γ-probe and γ-camera in their dual detector approach, improving SLN detection by 20% as compared to the sole use of pre-interventional information [2].

The intraoperative acquisition of several thousand γ-activity recordings over a region of interest with a spatially tracked detector allows for the 3D reconstruction of the nuclear information and its display via Augmented Reality (AR) overlays, a technique called freehand SPECT [14, 9]. The latter is applied in open and laparoscopic surgeries, such as SLNB for head and neck [7] or breast [1] cancer.

The AR visualization of SPECT information becomes outdated as soon as the tissue is manipulated, such as during any incision, when the radioactively marked tissue gets displaced. The acquisition of a new intraoperative SPECT volume to replace the previously augmented one is problematic as it requires prolonged handling of a γ-detector, delaying the procedure. Another solution is the use of a 1D or a 2D detector to acquire additional information about the γ distribution, losing the benefit of AR and disrupting the workflow. An alternate solution proposed in [12] is the registration of a pre-acquired 3D SPECT volume to an intraoperative 1D γ-probe signal, which requires a γ-probe, a model of its behavior (sensitivity, collimator aperture, etc.) and several hundred tracked γ-activity recordings. In a clinical scenario where lymph nodes can have a diameter of less than 1 cm, their demonstrated accuracy of 8 mm may prove insufficient.

Proposed Solution: Displacement Compensation. We propose a new method for continuous SLN displacement compensation. We utilize a 2D γ-camera rather than a 1D detector, and focus on minimally invasive SLN biopsies. The solution provides an update of the intraoperative SPECT image, by placing the γ-camera relative to the SLNs and estimating their displacement. Results are presented based on displacements in ex-vivo experiments and show an average accuracy of under 1 mm.

C.2 Materials & Methods

The proposed displacement compensation technique requires a SPECT volume, acquired by preoperative SPECT/CT [2] or intraoperative freehand SPECT [14], and a tracked γ-camera.

C.2.1 SPECT Imaging & Technical Background

The system we propose is based on a two step intraoperative workflow (see Fig. C.2). The first step is the acquisition of a SPECT volume. In the case of preoperative SPECT/CT, the volume has to be registered to the patient e.g. by using fiducials. In the case of freehand SPECT, registration is unnecessary as the reconstruction is done relative to an optically tracked reference target fixed to the patient's body. The SPECT volume is then visualized intraoperatively using an AR overlay, as in Fig. C.1a. To display the SPECT AR overlay and its continuous updates, we employ the commercially available declipse®SPECT (SurgicEye GmbH, Munich, Bavaria, Germany). The second step is to use the tracked γ-camera to calculate the displacement of each of the segmented lymph nodes.

C.2.2 2D γ-camera

The CrystalCam (Crystal Photonics GmbH, Berlin, Germany) is a handheld, miniaturized γ-camera capable of producing images with a resolution of 16×16 pixels. The selected collimator, the Tungsten-based LEHS (Crystal Photonics GmbH, Berlin, Germany), has dimensions of $44 \times 44 \times 11.5$ mm, with each square pixel having 2.16 mm long sides. This is a parallel collimator; The produced images can be compared to a parallel projection of the radioactivity in the observed area. Based on a single image, a 0.925 MBq (25 Ci) ^{57}Co source is identifiable within distances up to 15 cm.

We use an infrared tracking system (henceforth designated as *IR*) to track the γ-camera and patient. The system we chose, due to the convenience of

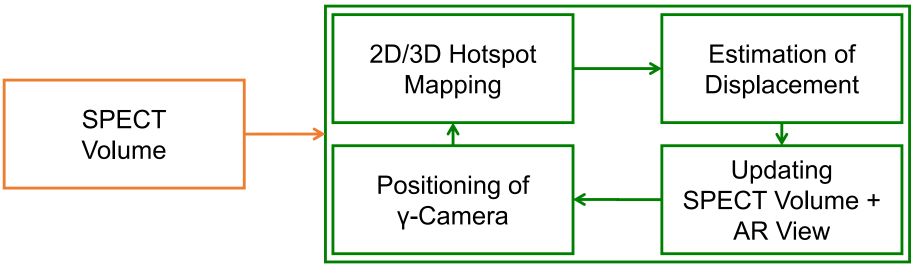


Fig. C.2: The workflow of the proposed solution

it being perfectly integrated into the declipse®SPECT system, is the Polaris Vicra (Northern Digital Inc., Waterloo, Ontario, Canada).

C.2.3 Displacement Compensation

Positioning of γ-camera: With the ${}^{\gamma\text{-camera}}\mathbf{T}_{\text{patient}}$ transformation computed as in (C.1), and given that the SPECT volume is reconstructed relative to the IR patient target (in the case of freehand SPECT) or registered to the patient (in the case of preoperative SPECT/CT), we can keep the γ-camera focused on the global centroid of the lymph nodes. A possible robotic-assisted solution to the γ-camera positioning problem is described in Sec. C.4.

$${}^{\gamma\text{-camera}}\mathbf{T}_{\text{patient}} = {}^{\gamma\text{-camera}}\mathbf{T}_{\text{IR}} \cdot {}^{\text{IR}}\mathbf{T}_{\text{patient}} \tag{C.1}$$

2D/3D Hotspot Mapping: We segment the 3D image, via thresholding, into several 'hotspots' each corresponding to a different lymph node. The end objective is to translate these hotspots according to their displacement and update the AR overlay. The tissue displacement has translational and rotation components. However, since each nodule is well approximated as an anatomical structure with spherical geometry, we are interested in its 3D position. Our 2D sensor can estimate well its movement parallel to the sensor plate. We also estimate the motion orthogonal to the sensor using γ-camera's Lookup Tables. As described within the discussion section, we also propose to guide the motion of the γ-camera based on the tracked motion of the surgical tool as the tissue displacement mostly occurs due to the tool tissue interaction. In this way, we propose a system which intelligently positions the sensor in order to dynamically optimize the estimation of radiopositive tissue displacement taking the surgical action into account.

We post process the 2D image produced by the γ-camera (see Fig. C.1b for an example) by first segmenting regions of interest where the reported radioactivity is both a local and one of few global maxima. The distribution of radioactivity in the hotspot can be approximated using a Gaussian distribution. Therefore, we employ Gaussian fitting, which provides us with an estimated center for the resulting 2D blob.

Knowing both the γ-camera's pose and the location of each lymph node computed before, we can compare the 3D SPECT hotspot centroids and 2D blob centers. This can be done by backprojecting the 3D hotspots onto the γ-camera's collimator view plane. The same Gaussian fitting method can be applied to the backprojection and we can then establish a 3D hotspot to 2D blob mapping by comparing the distance between centers.

Estimation of Displacement: With continuous γ-camera acquisitions, we update the SPECT volume accordingly, albeit with translations solely parallel to the γ-camera's collimator view plane. Z-direction translations are recovered with a lesser degree of accuracy by referring to the γ-camera's Lookup Tables (LUT), which allows the user to estimate relative, but not absolute, depth changes.

Updating SPECT Volume: After quantification of displacement is achieved as explained above, we feed back the corrections into the SPECT system, where the AR overlay (see Fig. C.1a) is duly adjusted to reflect the SLNs' translations. These are applied independently of the underlying SPECT volume's voxel grid, thereby allowing for translations of virtually arbitrary magnitude.

C.3 Experiments & Results

Experimental Setup: To validate our workflow and algorithm, we perform freehand SPECT reconstruction of a pair of sealed point-like 0.925 MBq ^{57}Co sources, which were chosen due to their similarity to ^{99m}Tc, the radioactive element usually present in the injectable nanocolloid for SPECT. ^{57}Co's longer half-life and safer handling makes it more suitable for laboratory use. The sources are placed on a specially designed mount so that we know their relative position with high certainty. After the reconstruction is complete, we translate the sources by several defined amounts and calculate the displacement as described in Sec. C.2.3. We estimate the translations applied to each lymph node and compare them to the ground truth.

Evaluation: For our first ex-vivo experiment (Table C.1), the γ-camera was kept static and placed so that the collimator's plane was parallel to the direction of source movement. The source was moved to 5 different points ($p\{0-4\}$), which are colinear and 5 mm apart from each neighbor.

Table C.1: Results for displacement estimation when $p\{0-4\}$ interdistance is 5 mm.

↱	$p1 = 5$ mm	$p2 = 10$ mm	$p3 = 15$ mm	$p4 = 20$ mm
$p0$	5.44 mm	10.92 mm	16.31 mm	21.54 mm
$p1$		5.49 mm	10.87 mm	16.11 mm
$p2$			5.39 mm	10.62 mm
$p3$				5.24 mm

$\bar{\epsilon} \approxeq 0.79$ mm, $\sigma \approxeq 0.43$ mm

Our second ex-vivo experiment (Table C.2) was done under the same circumstances, with the sole difference being that the γ-camera was angled at 45° relative to the direction of source motion.

Table C.2: Results for displacement estimation when $p\{0-4\}$ interdistance is $5\text{ mm} * \sin(45^{o}) \approxeq 3.54$ mm.

↱	$p1 = 3.54$ mm	$p2 = 7.07$ mm	$p3 = 10.61$ mm	$p4 = 14.14$ mm
$p0$	3.40 mm	7.06 mm	11.25 mm	14.46 mm
$p1$		3.66 mm	7.85 mm	11.07 mm
$p2$			4.19 mm	7.41 mm
$p3$				3.21 mm

$\bar{\epsilon} \approxeq 0.38$ mm, $\sigma \approxeq 0.25$ mm

Multiple sources in view are accurately handled, as in Fig. C.1b.

Performance: The algorithm operates at a frequency of 2 Hz. γ-camera output is computed using compound exposures on a sliding-window basis. The window's width is configurable and varies, but is usually between 1 to 5 seconds. Assuming 2 seconds of exposure, the end result is that displacements are presented to the user with a 2.5 second delay in the worst case, which is acceptable for an intraoperative scenario.

C.4 Discussion & Conclusion

Stable 2D-3D Matching: In our 2D-3D matching algorithm described in Sec. C.2.3, we focus on deriving a 2D blob to 3D hotspot correspondence. This correspondence is not guaranteed to be stable in practice, i.e., a 2D blob may not always correspond to the *same* 3D hotspot. One example of this behavior occurring is if the lymph nodes move into close proximity with each other, and then apart again. However, this does not represent a major clinical problem as all radiopositive lymph nodes are treated as equivalent and the end result is their resection.

Robotic Platform for Automated Tool Tracking: Our ideal operating room setup is composed of one of the scenarios described in Sec. C.2 and an additional lightweight industrial robot such as the KUKA LBR iiwa (KUKA Roboter GmbH, Augsburg, Bavaria, Germany) that would be programmed to automatically track either the tip of a Minimally Invasive Surgery (MIS) robot's tool (as displacement is most likely to occur in its vicinity) or the centroid of the SPECT-identified lymph nodes, depending on the scenario. An example of the proposed setup is depicted in Fig. C.1c.

Wide Field-of-View γ-cameras: One of the upper bounds for the quality of displacement estimation presented in this paper is the area covered by our current γ-camera. Wide field-of-view collimators would allow for improved results requiring smaller displacement of the γ-camera. Alternatively, the robotically controlled γ-camera could follow displacement of SLNs, resulting in radioguided visual servoing.

Conclusion: In this paper, we introduced a novel technique for radiopositive tissue displacement compensation. Our initial ex-vivo experiments show that estimation accuracy is in the sub-millimeter range, which motivates full integration of our approach into existing intra-operative freehand SPECT acquisition systems system and further feasibility studies.

Acknowledgements The authors would like to thank Anton Deguet for robotics hardware support; and Intuitive Surgical Inc. for loaning the da Vinci® surgical system used for testing.

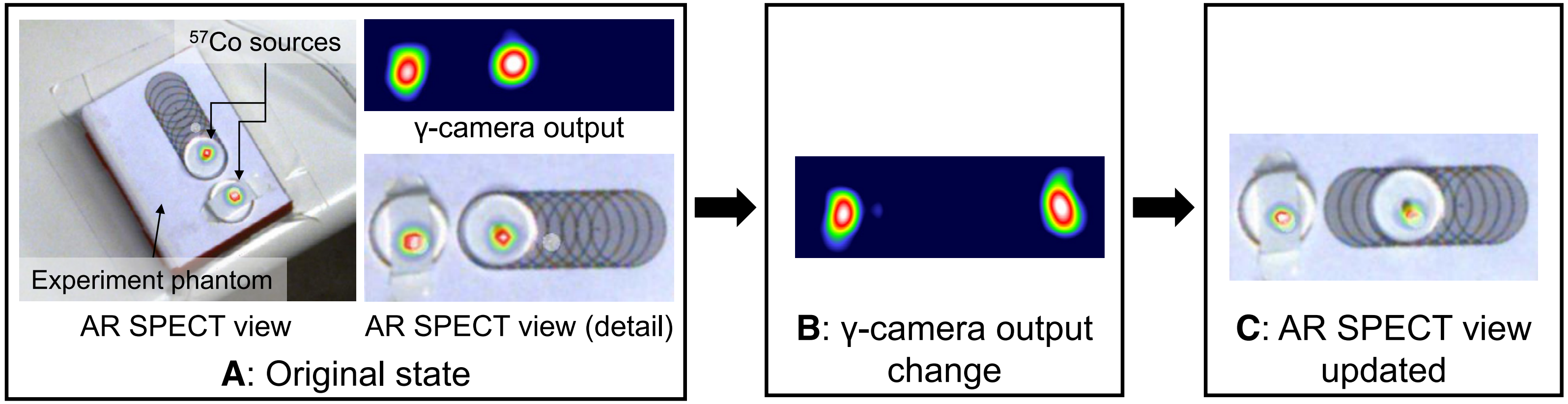

^{57}Co sources
γ-camera output
Experiment phantom
AR SPECT view
AR SPECT view (detail)
A: Original state
B: γ-camera output change
C: AR SPECT view updated

Fig. C.3: Workflow for updating the SPECT AR overlay based on 2D γ-camera acquisitions. We start with original state **A**. When the tracked γ-camera's (off-screen) output changes as in **B**, the overlay is updated as in **C**.

References

[1] Christina Bluemel, Andreas Schnelzer, Alexandra Ehlerding, Klemens Scheidhauer, and Marion Kiechle. “Changing the Intraoperative Nodal Status of a Breast Cancer Patient Using Freehand SPECT for Sentinel Lymph Node Biopsy”. In: *Clinical Nuclear Medicine* 39.5 (2014), e313–e314.

[2] O. R. Brouwer, R. A. Valdes Olmos, L. Vermeeren, C. A. Hoefnagel, O. E. Nieweg, and S. Horenblas. “SPECT/CT and a portable gamma-camera for image-guided laparoscopic sentinel node biopsy in testicular cancer”. In: *J. Nucl. Med.* 52.4 (Apr. 2011), pp. 551–554.

[3] Daan Hellingman, Linda J. de Wit–van der Veen, W. Martin C. Klop, and Renato A. Valdés Olmos. “Detecting Near-the-Injection-Site Sentinel Nodes in Head and Neck Melanomas With a High-Resolution Portable Gamma Camera”. In: *Clinical Nuclear Medicine* 40.1 (Jan. 2015), e11–e16.

[4] Laura L. Holman, Charles F. Levenback, and Michael Frumovitz. “Sentinel Lymph Node Evaluation in Women with Cervical Cancer”. In: *Journal of Minimally Invasive Gynecology* 21.4 (July 2014), pp. 540–545. DOI: `10.1016/j.jmig.2013.12.095`. URL: `http://dx.doi.org/10.1016/j.jmig.2013.12.095`.

[5] David N Krag, Stewart J Anderson, Thomas B Julian, Ann M Brown, Seth P Harlow, Joseph P Costantino, Takamaru Ashikaga, Donald L Weaver, Eleftherios P Mamounas, Lynne M Jalovec, Thomas G Frazier, R Dirk Noyes, Andre Robidoux, Hugh MC Scarth, and Norman Wolmark. “Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial”. In: *The Lancet Oncology* 11.10 (Oct. 2010).

[6] G. H. Lyman, A. E. Giuliano, M. R. Somerfield, A. B. Benson, D. C. Bodurka, H. J. Burstein, A. J. Cochran, H. S. Cody, S. B. Edge, S. Galper, J. A. Hayman, T. Y. Kim, C. L. Perkins, D. A. Podoloff, V. H. Sivasubramaniam, R. R. Turner, R. Wahl, D. L. Weaver, A. C. Wolff, and E. P. Winer. “American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer”. In: *Journal of Clinical Oncology* 23.30 (Oct. 2005), pp. 7703–7720.

[7] Magis Mandapathil, Afshin Teymoortash, Jochen Heinis, Susanne Wiegand, Christian Güldner, Stephan Hoch, Marion Roeßler, and Jochen A Werner. “Freehand SPECT for sentinel lymph node detection in patients with head and neck cancer: first experiences”. In: *Acta Oto-laryngologica* 134.1 (Jan. 2014), pp. 100–104. DOI: `10.3109/00016489.2013.832376`. URL: `http://dx.doi.org/10.3109/00016489.2013.832376`.

[8] Isao Miyashiro, Kentaro Kishi, Masahiko Yano, Koji Tanaka, Masaaki Motoori, Masayuki Ohue, Hiroaki Ohigashi, Akemi Takenaka, Yasuhiko Tomita, and Osamu Ishikawa. “Laparoscopic detection of sentinel node in gastric cancer surgery by indocyanine green fluorescence imaging”. In: *Surg Endosc* 25.5 (Oct. 2010), pp. 1672–1676. DOI: `10.1007/s00464-010-1405-3`. URL: `http://dx.doi.org/10.1007/s00464-010-1405-3`.

[9] Nassir Navab, Tobias Blum, Lejing Wang, Asli Okur, and Thomas Wendler. "First Deployments of Augmented Reality in Operating Rooms". In: *Computer* 99 (2012), pp. 48–55. ISSN: 0018-9162. DOI: `http://doi.ieeecomputersociety.org/10.1109/MC.2012.75`.

[10] A. G. Van der Zee, M. H. Oonk, J. A. De Hullu, A. C. Ansink, I. Vergote, R. H. Verheijen, A. Maggioni, K. N. Gaarenstroom, P. J. Baldwin, E. B. Van Dorst, J. Van der Velden, R. H. Hermans, H. van der Putten, P. Drouin, A. Schneider, and W. J. Sluiter. "Sentinel node dissection is safe in the treatment of early-stage vulvar cancer". In: *Journal of Clinical Oncology* 26.6 (Feb. 2008), pp. 884–889.

[11] L. Vermeeren, R. A. Valdes Olmos, W. Meinhardt, A. Bex, H. G. van der Poel, W. V. Vogel, F. Sivro, C. A. Hoefnagel, and S. Horenblas. "Value of SPECT/CT for Detection and Anatomic Localization of Sentinel Lymph Nodes Before Laparoscopic Sentinel Node Lymphadenectomy in Prostate Carcinoma". In: *Journal of Nuclear Medicine* 50.6 (May 2009), pp. 865–870.

[12] Christoph Vetter, Tobias Lasser, Asli Okur, and Nassir Navab. "1D-3D Registration for Intra-Operative Nuclear Imaging in Radio-Guided Surgery". In: *IEEE Transactions on Medical Imaging* 34.2 (Feb. 2015), pp. 608–617. DOI: `10.1109/tmi.2014.2363551`. URL: `http://dx.doi.org/10.1109/TMI.2014.2363551`.

[13] Sergi Vidal-Sicart, Beatriz Doménech, Blanca Luján, Jaume Pahisa, Aureli Torné, Sergio Martínez-Román, Juan Antonio Lejárcegui, Pere Fusté, Jaume Ordi, Pilar Paredes, and Francesca Pons. "Sentinel node in gynaecological cancers. Our experience". In: *Revista Española de Medicina Nuclear (English Edition)* 28.5 (Sept. 2009), pp. 221–228. DOI: `10.1016/s1578-200x(09)70022-1`. URL: `http://dx.doi.org/10.1016/S1578-200X(09)70022-1`.

[14] Thomas Wendler, Alexander Hartl, Tobias Lasser, Joerg Traub, Farhad Daghighian, Sybille Ziegler, and Nassir Navab. "Towards intra-operative 3D nuclear imaging: reconstruction of 3D radioactive distributions using tracked gamma probes". In: *Medical Image Computing and Computer-Assisted Intervention.* LNCS 4792.2 (Oct. 2007). Ed. by N. Ayache, S. Ourselin, and A. Maeder, pp. 909–917.

[15] Sandra L Wong, Charles M Balch, Patricia Hurley, Sanjiv S Agarwala, Timothy J Akhurst, Alistair Cochran, Janice N Cormier, Mark Gorman, Theodore Y Kim, Kelly M McMasters, et al. "Sentinel lymph node biopsy for melanoma: American Society of Clinical Oncology and Society of Surgical Oncology joint clinical practice guideline". In: *Journal of Clinical Oncology* 30.23 (2012), pp. 2912–2918.

APPENDIX D

First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping (IEEE TMI 2015)

Bernhard Fuerst[1,2], Julian Sprung[1], Francisco Pinto[1,2], Benjamin Frisch[2], Thomas Wendler[2,3], Herve Simon[4], Laurent Mengus[4], Nynke van den Berg[5,6], Henk van der Poel[7], Fijs W.B. van Leeuwen[5,6], and Nassir Navab[1,2]

[1] Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, MD, USA

[2] Computer Aided Medical Procedures, Technical University of Munich, Germany

[3] Department of Nuclear Medicine, Klinikum rechts der Isar, Technische Universität München, 81675 München, Germany

[4] Research and Development Department, Eurorad SA, 67201 Eckbolsheim, France

[5] Interventional Molecular Imaging Laboratory, Leiden University Medical Center, 2300 Leiden, The Netherlands

[6] Department of Urology, Antoni van Leeuwenhoek Hospital, 1000 Amsterdam, The Netherlands

[7] Urology Department, Antoni van Leeuwenhoek Hospital, 1000 Amsterdam, The Netherlands

Copyright © 2015 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/ redistribution requires IEEE permission. See `http://www.ieee.org/publications_standards/publications/rights/index.html` for more information.

Contribution For the very first time, robotic Single-Photon Emission Computed Tomography (SPECT) imaging is presented for minimally invasive robot-assisted surgery. The key technical contributions are the development of a fused optical and kinematic tracking algorithm and the successful reconstruction of radioactive sources using a miniaturized gamma detector. Additionally, this paper presents the path towards an Augmented Reality (AR) visualization for SPECT-guided robot-assisted interventions.

Abstract In this paper we present the usage of a drop-in gamma probe for intra-operative *Single-Photon Emission Computed Tomography* (SPECT) imaging in the scope of minimally invasive robot-assisted interventions. The probe is designed to be inserted and reside inside the abdominal cavity during the intervention. It is grasped during the procedure using a robotic laparoscopic gripper enabling full six degrees of freedom handling by the surgeon. We demonstrate the first deployment of the tracked probe for intra-operative in-patient robotic SPECT enabling augmented-reality image guidance. The hybrid mechanical- and image-based in-patient probe tracking is shown to have an accuracy of 0.2 mm. The overall system performance is evaluated and tested with a phantom for gynecological sentinel lymph node interventions and compared to ground-truth data yielding a mean reconstruction accuracy of 0.67 mm.

Keywords SPECT, Image-guided treatment, Surgical guidance/navigation, Endoscopy/Laparoscopy.

D.1 Introduction

D.1.1 Clinical Background and Motivation

The most important prognostic factor for survival in women with early stage cervical cancer is the involvement of the lymph nodes [10]. In an attempt to evaluate the presence of lymph node metastasis, a lymphadenectomy is performed during which all regional lymph nodes are dissected. However, most of these patients (up to 70%) will not have metastases in lymph nodes [39, 21, 14, 28]. Furthermore, the pelvic lymphadenectomy is associated with a substantial treatment-related morbidity [15, 32].

This has motivated comprehensive research into the validity and effectiveness of the dissection of the lymph nodes most likely reached first by metastasizing tumor cells. This concept is known as *Sentinel Lymph Node* (SLN) biopsy and has become the standard of care for some cancer types, for instance breast cancer [23] and melanoma [43]. It is an accepted and safe procedure for several other cancer types, such as gastric [29], prostate [38, 28, 14], penile [20], vulvar [44, 39] and cervical cancers [15, 32]. In contrast to performing a radical, morbidity loaded and complicated regional lymphadenectomy, SLN biopsy makes it possible to determine the regional lymph node status by excision of the first lymph nodes receiving direct drainage from the primary tumor site and subsequent histopathological examination. In cervical cancer, several studies have shown that a (bilateral) negative SLN biopsy accurately predicts the absence of lymph node metastasis [15, 32].

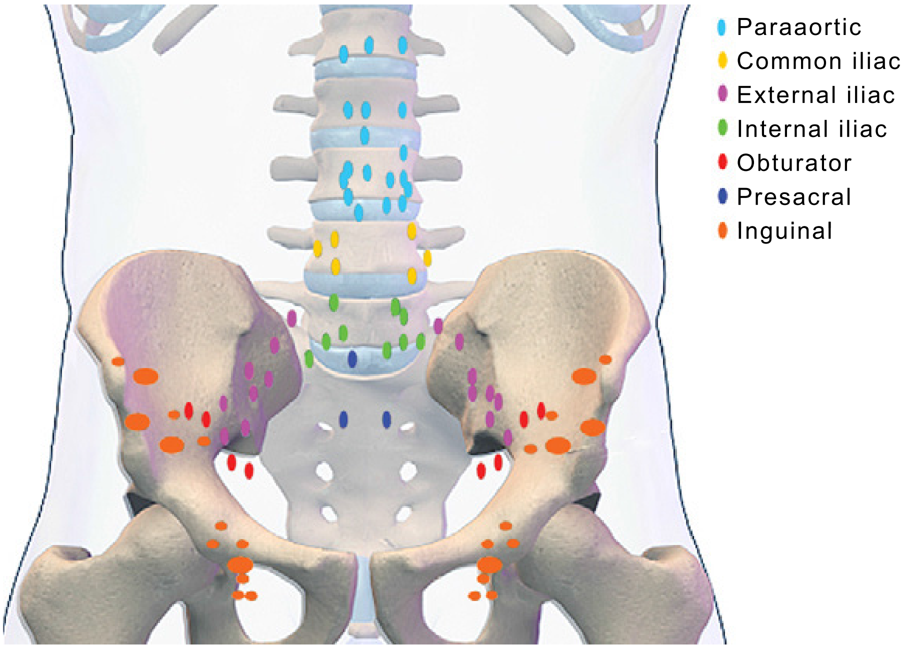


Fig. D.1: Lymph node locations in the pelvic area. Determination of the lymph node(s) to which possible tumor spread has occurred (e.g. identification of the sentinel lymph node(s)), can improve staging of e.g. cervical cancer.

The identification and dissection of SLNs requires the technique of *SLN Mapping* (SLNM), which is incorporated in individually tailored surgery for cases of cervical cancer [32]. Even though the general drainage pathways for lymph nodes in the pelvic area are well documented (see Fig. D.1, [35]), there is still the possibility of inter-patient variance in what pertains to e.g. unexpected drainage pathways [2] which makes SLNM a necessity. Pelvic lymph nodes are often deeply seated [35] and therefore difficult to locate interventionally, requiring pre-operative imaging.

D.1.2 Preoperative Imaging of Sentinel Lymph Nodes

In order to perform preoperative SLNM, a radiotracer has to be injected close to the tumor. The tracer is designed to be only transported by the lymphatic system to the regional lymph nodes. The injection is often done with the patient being awake and can be considered very painful and therefore difficult to perform [32]. Following tracer injection, nuclear imaging is performed to depict the lymph node(s) to which the tracer drains.

Lymphoscintigraphy is a two-dimensional nuclear imaging technique that depicts the lymph nodes that have taken up tracer and sometimes the lymphatic vessels in a two-dimensional lymphoscintigram. Lymphoscintigraphy fails to provide anatomical information.

When dealing with SLNM for cervical cancer, lymphoscintigraphy has only little clinical value, which is also reflected in its poor correspondence between preoperative and interventional mapping results [9]. This motivates the use of other nuclear imaging techniques.

Single-Photon Emission Computed Tomography (SPECT) is a nuclear imaging technique and frequently combined with *X-ray Computed Tomography* (CT). Together, SPECT/CT provides three dimensional images representing both the lymphatic system (SPECT) and anatomical information (CT). This fused imaging method is promising and SLN detection rates of more than 95% have been reported [15, and references therein]. Compared to planar lymphoscintigraphy, SPECT/CT does not only lead to improved SLN detection rates, but also provides better anatomical correlation, and thus enhances the validity and effectiveness of SLN biopsies [19].

The transfer of information visualized on the diagnostic images to the real patient is done by the surgeon, which is a difficult and error prone task due to tissue deformation and varying patient positioning.

It is of relevance to mention that imaging the metastases themselves using functional imaging like *Positron Emission Tomography* (PET) in combination with CT, *Magnetic Resonance Imaging* (MRI), or multiparametric MRI is not a real option. This is mainly due to the fact that clinically relevant lymph node metastases may be smaller than 7 mm in patients with early cervical cancer [19, 7]. In the case of PET, the metastases do not necessarily present with an uptake of the tracer which is high enough relative to the background such that the partial volume effect cannot be ignored.

In contrast to diagnostic imaging, current interventional SLNM is not limited by the size of the lymph node nor the affinity of the tracer to the tumor as the contrast to the background is theoretically infinite. Also, tissue deformation due to patient positioning does not need to be taken into account when performing interventional SLNM.

D.1.3 Interventional Sentinel Lymph Node Mapping

The interventional mapping techniques can be categorized based on the contrast agent, namely either colorimetric or radioisotope. The first group of mapping techniques requires the injection of color dyes (mostly isosulfan blue, methylene blue or similar substances) into either the cervix or near the tumor location (direct injection into the tumor reduces the detection rate of SLN). The dyes are taken up by the lymphatic vessels and transported to the sentinel lymph nodes within 5 to 10 minutes. During the intervention, the SLN and corresponding vessels are visually colored and are identified by visual inspection [2]. However, due to the coloring, the tissue differentiation is more difficult, which may complicate the R0 resection of the main tumor and the delimitation of the SLN. Furthermore, using dyes requires a direct line of sight to the tissue of interest, which makes SLNM with dye inferior to a radioisotope or combined radioisotope and dye technique in terms of sensitivity and detection rate [32].

The typical radioisotope (also referred to as radionuclide) used for SLN detection in cervical cancer is the metastable technetium-99m ($^{99\mathrm{m}}$Tc), which mainly emits its radiation as γ radiation at 0.141 MeV and has a half-life time of approximately 6 hours. Typical doses are in the range of 74 to 148 MBq (2 to 4 mCi) and 7.4 to 37 MBq (0.2 to 1 mCi) for preoperative and interventional mapping respectively[1] [32]. The typical uptake by the SLNs is $0.1 - 5\%$, exceptionally up to 10%, of the injected dose [42]. The image acquisition is typically performed after 15 to 20 minutes [15]. Currently, a laparoscopic or hand-held γ-probe is used to detect SLN that exhibit three- to ten-times more radioactivity than the baseline activity [30]. The difficulty in distinguishing a radioactive-positive SLN from background radioactivity with such one-dimensional probes leads to the use of interventional two- and three-dimensional nuclear imaging, e.g. freehand SPECT. This concept has first been demonstrated for arbitrary radioactive distributions and allows the reconstruction of SPECT images using a tracked γ-probe [41, 12]. The freehand SPECT system is used to image SLN for breast cancer [42, 4], and head and neck cancers [13].

A recently presented clinical case showed that laparoscopic freehand SPECT may be used to generate images of radioactive SLNs located near the injection site (small dots vs. the large signal present in the augmented reality overlay) [3]. At the same time the image in Fig. D.2, [3] illustrates the physical complexity of performing laparoscopic freehand SPECT during robot-assisted surgery: The movements of the laparoscopic γ-probe, needed to acquire the 3D scan, are limited by the robotic arms present in the scanning area. In fact, one robotic arm had to be disconnected from the patient to allow for the generation of this particular image. Optical tracking of the four fiducials on the laparoscopic γ-probe is hindered by the robotic arms (and their sterile draping), but also by the positioning of the optical tracking device in the layout of the operating room. The latter means that the person performing the freehand SPECT acquisition is continuously blocking the optical tracking process. By using a drop-in γ-probe in combination with an alternative mode of tracking, these major limitations can be overcome. This in turn will significantly increase the translational nature of laparoscopic freehand SPECT in a robotic setting.

The interventional use of a mini γ-camera has been demonstrated, for instance, for SLNM in urological cancers [37]. The γ-cameras typically only provide two dimensional views of the radioactivity, and do not provide depth information. Research towards clinical use of SPECT imaging by means of a mini γ-cameras is promising for several, well-accessible operation sites [27, 26].

D.1.4 Robot-Assisted Lymphadenectomy

Minimally invasive lymphadenectomies for gynecological cancers have been increasingly performed robotically in recent years [24]. On one hand, robot-assisted surgeries exhibit a significantly increased cost and an overall longer duration. Challenges arise in handling, which are partially caused by the bulkiness and complexity of the equipment, combined with the lack of haptic feedback and difficulty of combining preoperative imaging with the interventional view. On the other hand, the outcome of an intervention may benefit from such robot-assisted systems in terms of cosmetic results, postoperative pain and intraoperative blood loss. Additionally, a recent cohort study suggested that the overall recurrence and survival rates are comparable to non-robotic laparoscopic interventions [17], indicating that the use is equivalent with respect to patient outcome. The feasibility of robotic SLN dissection has been demonstrated using several SLNM techniques [11, 40, 34, 36]. When using preoperative imaging for

---

[1] Units: 1 mCi = 1×10^{-3} Curie = 37 MBq = 37×10^{6} Becquerels

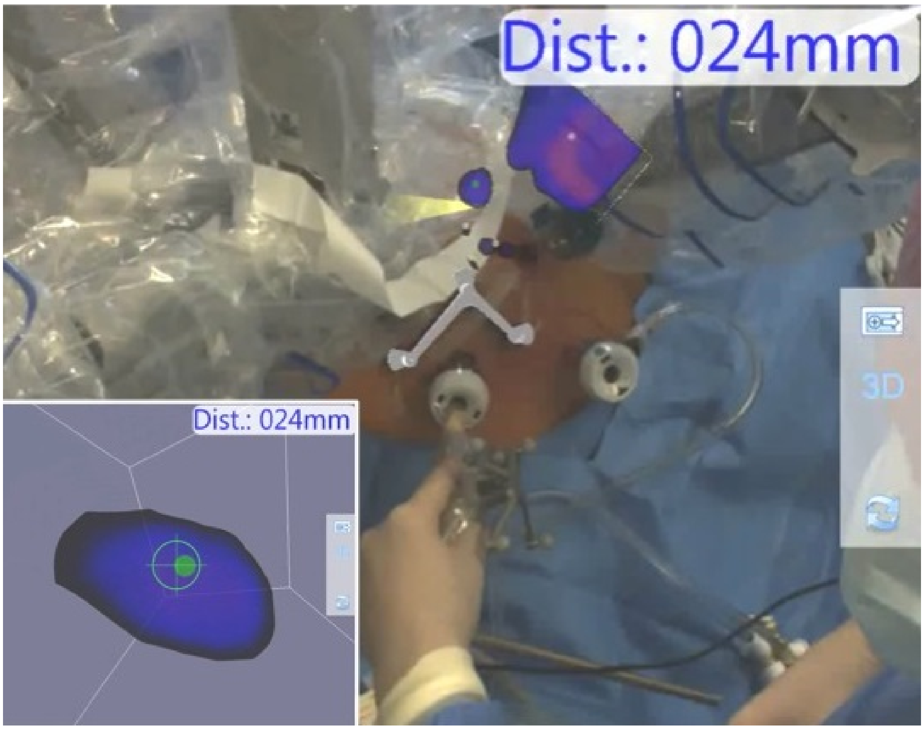


Fig. D.2: Laparoscopic freehand SPECT imaging has been performed during robot-assisted surgery, yielding clinically valuable images. This figure shows the acquisition procedure in which a freehand SPECT volume is obtained using the declipse®SPECT imaging system (SurgicEye GmbH, Munich, Germany) in combination with a laparoscopic γ-probe. To enable freehand SPECT acquisition and navigation, one robotic arm had to be removed, significantly modifying the work-flow.

SLNM (e.g. SPECT/CT), the duration of the intraoperative SLN dissection can significantly be reduced [16].

However, the aforementioned limitations of preoperative imaging may reduce the usefulness during interventions [7]. Therefore, we propose to combine the benefits of both: interventional imaging and robot-assisted lymphadenectomy.

D.1.5 Objective: Minimally Invasive SPECT Imaging

A major gap in the state of art of SLNM appears to be the translation or acquisition of three-dimensional nuclear imaging techniques to these minimally invasive, robot-assisted interventions [5]. To the best of our knowledge, we present the first robot-assisted minimally invasive three-dimensional nuclear imaging concept, designed for sentinel lymph node dissection in the pelvic area. In contrast to existing solutions for intraoperative imaging [41], we use a novel γ-detector designed to be dropped in through a trocar and picked up during a minimally invasive robotic interventions.

D.2 Materials and Methods

D.2.1 System Overview

The proposed interventional SPECT imaging system requires a minimally invasive surgical system, which provides natural dexterity during laparoscopic interventions, a miniaturized γ-probe inserted through the trocar and grabbed using the laparoscopic tool, and a tracking technique to precisely compute the position and pose of the γ-probe within the patient. After the acquisition, the reconstruction of the SPECT image is performed (Sec. D.2.6). Finally, the SPECT image is visualized in the surgeon's console to enable image guided interventions.

To present this novel imaging concept, we utilize a da Vinci® S minimally invasive robotic surgical system (Intuitive Surgical Inc., Sunnyvale, CA, USA; Sec. D.2.3), a drop-in γ-probe (Eurorad, Eckbolsheim, France; Sec. D.2.2), a computer oscilloscope that captures the electrical signal from the γ-probe, and a computer with significant computational power due to a graphics processor unit. The latter two components are part of the commercially available declipse®SPECT imaging system (SurgicEye GmbH, Munich, Germany). Additionally, an infrared tracking system (Polaris Vicra, Northern Digital Inc, Canada) is used to track patient and endoscope movements. The concept of robotic SPECT for minimally invasive SLNM is not restricted to the use of the components mentioned. However, several sections of the presented methodology focus on the combination and use of these components.

D.2.2 Drop-In Gamma Probe

A γ-probe consists in a scintillating crystal that converts the energy of impinging γ photons emitted by a radiation source into light photons, a process called radioluminescence. A photodiode that is optically coupled to the crystal converts the light emitted by the scintillator into an electrical signal which is then amplified and processed. Shielding around the probe and a collimator in front of the crystal at the probe tip limit the field of view of the probe, thus introducing directional information to the γ count rate.

The design of the drop-in γ-probe is based on an existing γ-probe. The modifications allow the insertion of the probe through a trocar, and repeatable handling of the probe for the use of a laparoscopic tool with natural dexterity and six degrees of freedom. The probe has a total diameter of 12 mm, and only 30 mm remain rigid while the rest of the device is placed outside the patient. The distal element is comprised of a Tungsten collimator, and a CsI crystal of 5 mm diameter and 10 mm length, which is coupled to a 25 mm^2 silicon photodiode. The collimator and crystal allow for a 47° field of view. The probe is shown in Fig. D.3a).

The detector is designed to be operated using a standard control unit. In order to be able to be picked-up, a mount was designed such that the ProGrasp™ Forceps of a da Vinci® robot can reproducibly pick it up and keep it fixated. The interlocking mechanism has previously been presented for pick-up ultrasound [33], and is visible on the end of our probe in Fig. D.6. The analog signal of the silicon diode is read after pre-amplification and digitized using an oscilloscope (PicoScope 3204A, Pico Technology, UK) with an adjustable asynchronous trigger. The integral of each pulse is proportional to the energy of the detected photon.

D.2.3 Robotic Surgical System

The da Vinci® S surgical system consists of a surgeon's console, a patient side cart and a vision cart. The system enables a surgeon to perform robot-assisted minimally-invasive (laparoscopic) surgeries by translating the surgeon's manipulator movements to the robotic laparoscopic tools. This allows for precise tool manipulation with six degree of freedom interaction. The system additionally enables the surgeon to perceive a three-dimensional view via the console's stereo viewer, which displays the video stream from the stereo endoscope.

D.2.4 Probe Tracking

Currently, freehand SPECT relies on an infrared tracking system with reflective markers [41]. This tracking technology requires a direct line of sight from the stereo infrared camera system to the markers mounted on the probe and is therefore not applicable for in-patient SPECT acquisition. Work on tracked ultrasound devices has reported the usage of electro magnetic tracking [1, 8]. However, this tracking method is prone to fail when significant amounts of metal are present in the operation site, which is the case in robot-assisted surgery.

The da Vinci® system allows mechanical tracking via computation of the forward kinematics from the reported joint angles and the system's Denavit-Hartenberg parameters. Data retrieval from the system is performed via the research application programming interface (API) which provides the position and orientation of a gripper (e.g. the ProGrasp™ Forceps). Due to the 3.6 m long, 13 degrees-of-freedom kinematic chains, the reported tool tip pose is highly erroneous [31]. This global offset can be coped with using a calibration procedure for relative movements within a small space (less than a 20 mm) around the position of calibration [22]. However, for larger movements, as required for the in-patient SPECT acquisitions, significantly larger errors are expected. Furthermore, a temporal drift of the reported coordinate system (mostly the remote center point) inhibit tracking solely based on the da Vinci® 's kinematics [31].

Therefore, we perform tracking by combining relative displacement information provided by the robotic platform (mechanical tracking) and vision-based tracking using the stereo laparoscopic views. It comprises of following modules:

A The segmentation module labels all pixels in the image as either probe or background based on the color appearance. This includes outlier removal and results in a binary mask for the probe.

B Based on the best fit of several simple polygons, the primary contours are extracted. These contours describe the probe outline and markings along the primary axis.

C By combining the probe outline in the two-dimensional images, the cross-ratio of the markings along the primary axis, the known probe geometry, and the camera intrinsics, the three-dimensional pose is computed.

D A federated Kalman filter is used to compensate possible noise of the vision-based and mechanical tracking, and fuse both tracking streams.

This fused tracking approach enables the surgeon to move the γ-probe freely inside the patient to obtain an optimal nuclear data acquisition. In addition, we integrated the information from the infrared tracking system to validate the endoscope and patient movements.

D.2.5 Robotic In-Patient SPECT Acquisition

The in-patient SPECT acquisitions is performed in a similar manner to that of open surgery freehand SPECT described in [41]. In brief, first the surgeon picks up the γ-probe using the ProGrasp™ Forceps. Then a scanning motion is performed, during which the γ event count rates $\mathbf{A}$ (counts per second) and the corresponding probe positions and orientations relative to the da Vinci® base (${}^{\text{probe}}\mathbf{T}_{\text{base}}$) are recorded.

To enable augmentation of the surgeon's view with the reconstructed volume, it is vital to also obtain the patient-to-probe transformation ${}^{\text{probe}}\mathbf{T}_{\text{patient}}$ as illustrated in Fig. D.4. This transformation can then be used to compensate for endoscope or patient movements. In order to retrieve the desired patient-to-probe transformation ${}^{\text{probe}}\mathbf{T}_{\text{patient}}$, the forceps-to-probe tip transformation ${}^{\text{probe}}\mathbf{T}_{\text{forceps}}$, and infrared tracking-to-base transformation ${}^{\text{base}}\mathbf{T}_{\text{IR}}$ are calibrated before the SPECT acquisition. During the intervention, this transformation is updated by means of an infrared rigid body attached to the endoscope. While the forceps-to-probe tip is based on the probe and tool design, the infrared tracking-to-base is computed using a camera calibration board with an infrared marker. The base-to-probe tip transformation is then computed as:

$$ {}^{\text{probe}}\mathbf{T}_{\text{patient}} = {}^{\text{probe}}\,\mathbf{T}_{\text{forceps}} \cdot {}^{\text{forceps}}\,\mathbf{T}_{\text{base}} \cdot {}^{\text{base}}\,\mathbf{T}_{\text{IR}} \cdot {}^{\text{IR}}\,\mathbf{T}_{\text{patient}} \qquad \text{(D.1)} $$

where the patient-to-infrared tracking and infrared tracking-to-base transformations can be omitted in phantom experiments without movements.

D.2.6 SPECT Reconstruction

The volume of interest is determined interactively with the surgeon. The origin is fixed based the pose of the γ-probe when the recording of poses and probe readings is started. The SPECT reconstruction problem is modeled by defining the vector of measured values $\mathbf{y}$ to be a system operator $\mathbf{S}$ applied to a vector containing the radioactivity $\mathbf{x}$ and noise:

$$\mathbf{y} = \mathbf{S}(\mathbf{x}) + \text{noise}.$$

To approximate $\mathbf{x}$, we model the system operator $\mathbf{S}$ using the system matrix $\mathbf{A}$. The system matrix element A_{ij} consequently represent the probability that an emission of voxel x_i results in an detection in the detector y_j. Note that the index i is used for voxels, while the index j is used for measurements. The number of voxels and therefore the resolution is selected by the user. The problem can now be formulated as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad \text{or} \quad y_j = \sum_{j=1}^{M} A_{ij} x_i, \tag{D.2}$$

In case of SPECT, y_j is known, and the inverse problem $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ needs to be solved. Usually $\mathbf{A}$ can not be inverted. Therefore, we use *Maximum Likelihood Expectation Maximization* (MLEM) to iteratively solve this problem, which is also used in commercial freehand SPECT systems [5, 42]. To determine the system matrix a physical forward model of the radiation propagation and detection process is developed. The model accounts for probe specific properties such as geometry (collimation, field of view), material properties and sensitivity and further for the stochastic nature of radiation and background noise [41].

The model of the γ detection physics of the probe represents the spatial correspondence of a radioactive source's position relative to the probe and the measured count rate. In this work, a look up table that stores the influence coefficients, which is described to be an even equivalent to measuring the physical model in a positioning table experiment [12], is created from an analytical model. The model accounts for the medium attenuation, the solid angle and collimator and detector geometry and dimensions which are extracted from the manufacturer's technical drawings. As the probe's round geometry is axis symmetric around the z-axis a two-dimensional model is sufficient to represent its three-dimensional response.

D.2.7 Augmented Reality Visualization for Surgical Navigation

The stereo endoscopic video is captured from the da Vinci$^{®}$'s vision cart using a *Serial Digital Interface* (SDI) splitter and SDI frame grabber (DeckLink Duo, Blackmagic Design Pty. Ltd.) to allow the primary video source of the da Vinci$^{®}$ to remain original and lag free while computing the image overlay. Due to the transformation between endoscope pose and reconstructed volume being known, the augmentation is straightforward and only requires the camera projection matrix $\mathbf{P}$ of the endoscope. The calibration of the projection matrix $\mathbf{P}$ is performed using a checkerboard pattern. Once the robot-assisted SPECT is acquired, the surgeon manually selects a threshold which is used to augment the laparoscopic camera view. The augmented stereo video is displayed in the surgeon's console on the split screen (TilePro™) in parallel to the original stereo video.

D.3 Experiments and Results

D.3.1 Experimental Setup

We performed a set of experiments using weak ^{57}Co sources at 0.296 and 0.925 MBq (8 and 25 µCi) representing an SLN and the injection site. In two different setups, they are either positioned at a distance of 30 or 44.7 mm using a specially designed source holder. This mounts were additively manufactured (3D printed) with an accuracy of ±0.15 mm and are shown in Figures D.3b), D.6 and D.7b).

The endoscope is positioned to center the sources in the image, while the patient-side manipulator is placed to allow the activity measurements with the γ-probe from at least two sides. The drop-in γ-probe was picked up using the da Vinci® ProGrasp™ Forceps, which was usually successful within 5 seconds. The infrared tracking system is set up to observe the camera motion relative to the patient table, theoretically allowing movements of the patient and camera during the acquisition and augmented reality view.

For reference, SPECT imaging by means of a mini γ-camera by Crystal Photonics, Germany is used. The detector is comprised of a 4×4 cm^2 CdZnTe crystal which has 16×16 pixels. The image acquisition and reconstruction is performed as described in [27], which provides sufficient accuracy for interventional application [6, 25], as mentioned in second last paragraph of Sec. D.1.3.

D.3.2 Data Acquisition

The total duration of the data acquisition is approximately 180 seconds, in which nearly 3000 individual measurements of probe pose and γ activity are recorded. During the acquisition the probe is slowly moved along at least two sides of the volume of interest. Furthermore, varying the orientation of the probe improves the likelihood of a good reconstruction outcome. In general, the reconstruction framework allows the independent acquisition of probe pose and activity, and automatically performs an interpolation of the pose to match the data streams based on a known and constant hardware delay.

D.3.3 Evaluation of Tracking Accuracy

Validation of the fused tracking (mechanical and vision-based) was done for both translational and rotational accuracy. For translational accuracy validation, the probe was mounted at several known points on a specially designed phantom while recording its pose. By comparing the recorded poses with the ground truth, we computed an approximate error of 0.2 mm.

For rotational accuracy validation, the probe was once more placed on an adjustable mount and its pose recorded at 5° adjustment intervals. The ground truth was compared to the actual angles dialed in the mount and the computed error was 0.7° up to 45°. Beyond 45°, the observed area of the probe becomes too small for accurate tracking and the error is in excess of 2°. In a surgical scenario, the surgeon is notified to move the endoscope to provide for a better observation.

D.3.4 SPECT Reconstruction Results

The iterative reconstruction using the drop-in γ-probe held by the ProGrasp™ Forceps is compared to the reconstruction using a mini γ-camera [27, 26], for which it has been shown that the imaging quality is sufficient to support SLNM [6, 25]. The reconstruction algorithm is identical to the one used for the drop-in γ-probe; only the forward models differ.

The reconstruction of the volume based on the γ-probe measurements converges after an average of 20 iterations. Due to the highly optimized and parallelized implementation, the computation is usually performed within 2 seconds. The volumes are then normalized and after manual thresholding, the radioactive sources can clearly be resolved.

The initial lower threshold is set to 10%, which removes artifacts. The second threshold is initially set to 100% and is used to adjust the contrast. As in clinical practice, the threshold is set by medical experts based on the expected radioactivity distribution as seen on preoperative images. Furthermore, the threshold is varied to allow to distinguish between background noise and radioactive tissue or sources. Typical absolute threshold values are not applicable as the reconstructed volume does not represent actual radioactivity counts, but rather the likelihood that a voxel containing radioactive isotopes emitting radioactivity may contribute to the observations.

In general, the comparison of image qualities by a medical expert indicated that the reconstruction based on the scan using the drop-in probe were to support SLNM. Slices through the reconstructed and thresholded volumes are illustrated in Figure D.5. The centroids were computed to allow the evaluation based on the distances between the sources. The deviation between the observed distances and the ground truth obtained by construction are on average 1 mm (3.4%) and 1.1 mm (3.6%) for the reconstruction based on the γ-camera and drop-in probe respectively. Therefore, the localization of a SLN is possible using either techniques.

We evaluated the reconstruction accuracy using the drop-in γ-probe by scanning three sources at known positions. The three-dimensional planar slice through the centroids, and the augmented reality view are shown in Fig. D.7a) and (b) respectively. When comparing the measurements of the distances between the centroids in the reconstructed volume to the known real distances, the average error is 0.67 mm.

The limited access to the sources from only two sides, and tracking inaccuracies beyond 45°, the reconstruction using the drop-in probe tends to exhibit artifacts in sub-volumes which are poorly scanned. This may motivate further research to improve the collimation and tracking algorithm.

D.3.5 Augmented Reality Visualization

After camera calibration using a checkerboard, the norm of the reprojection error in an image with a resolution of 1920 by 1080 pixels is 2.04 pixels. Assuming a realistic average depth, this corresponds to 0.16 mm error. After manual windowing of the reconstructed volume, the live endoscopic video can be augmented with the radioactive sources. Figure D.6 and D.7b) show the semi-transparent overlay of the rendered activity distribution onto the endoscopic view as it is shown to the surgeon.

D.4 Discussion and Conclusion

We introduced a novel concept for robotic SPECT acquisition for minimally invasive robot-assisted interventions, and demonstrated the feasibility and imaging quality using a phantom setup. The technique shows promising performance in the experimental setup. It allows for easy pick up using the ProGrasp™ Forceps and intuitive handling and scanning motion due to its aligned orientation with the gripper tip.

Although current experiments are ex vivo, we are positive that the system will operate as expected in a clinical environment. A recently presented clinical study using a laparoscopic probe for freehand SPECT (see section D.1.3) shows that the reconstruction algorithm produces clinically useful images even in the presence of background radiation or an injection site [3]. In contrast to the laparoscopic freehand SPECT, our approach reduces the limitations on the poses of the detector, which may increase the imaging quality. The use of an external γ-camera for SPECT imaging during a robot-assisted intervention would yield difficulties in terms of accessibility to the patient. Additionally, the typical distance between the γ-camera and the region of interest is less than 15 cm [27], and when exceeding this distance the sensitivity is too low to enable image acquisition.

The current drop-in γ-probe performed well, but only has one pixel and a wide collimation. A theoretical, multiple-pixel revision of the detector would reduce the angle of view (therefore improving the reconstruction accuracy) and consequently reduce the duration of the scan.

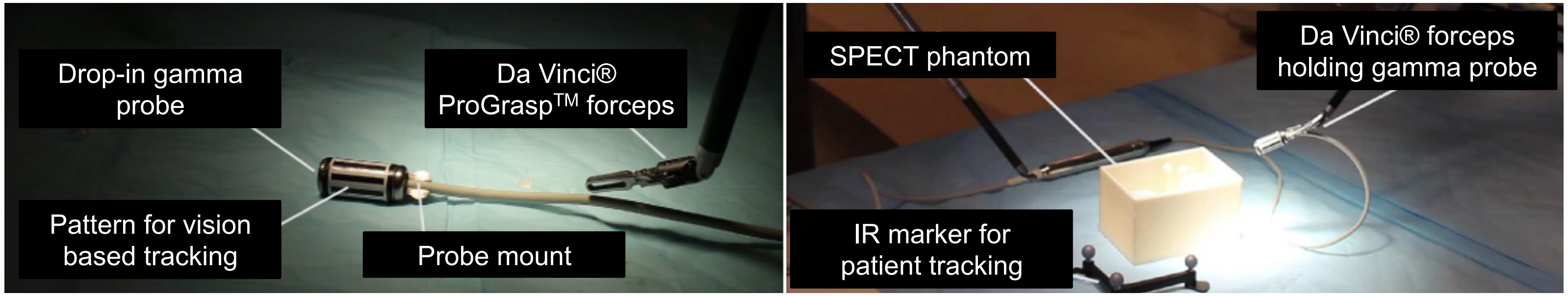
Currently, the tracking of the probe requires a direct line of sight. This clearly works well in phantom experiments, but may be difficult in a real scenario in which blood and smoke are added while the field of view is limited.Further research towards the correction of the offset of the mechanical tracking may reduce the strong dependency on the vision-based tracking. However, the weight of the probe (50 g) causes a bending of the tool, which can currently only be corrected using vision-based tracking. Combining probe and forceps to an optimized tool may hold the opportunity to solve this problem.

Once the resection starts, preoperatively acquired SPECT images do not reflect the deformations made to the tissue anymore. This common problem is shared by all intra-operative SPECT techniques, including the minimally invasive robotic SPECT. However, the probe could remain in the body cavity during the resection. This would allow easy confirmation of SLN removal and rescanning the volume.

The next steps include the integration of the presented system directly into the surgeon's stereo view in order to adapt the system for increased effectiveness in clinical use. Adding intra-operative imaging and navigation to robotic laparoscopic SLNM may reduce the false negative rate (i.e. the rate of missed metastases by the technique that result in a false staging of the patient) and increase sensitivity of intra-operative detection of the SLN. Moreover, a high resolution three-dimensional intra-operative imaging system could make the preoperative injection of the radioactive tracer redundant, as the otherwise needed imaging is moved into the operating room and both the injection and the surgery can be done sequentially during one single anesthesia [18]. The image guidance and the availability of depth information will further enable a fast and intuitive roll-up of the procedure even for less experienced physicians.

Acknowledgment

The authors would like to thank Jonathan Sorger from Intuitive Surgical Inc. for the constructive discussions and access to the application programming interface to the da Vinci®. Furthermore, the authors want to thank Joerg Traub from SurgicEye for his assistance and valuable feedback regarding the reconstruction algorithm.


Drop-in gamma probe
Pattern for vision based tracking
Da Vinci® ProGrasp™ forceps
Probe mount
SPECT phantom
IR marker for patient tracking
Da Vinci® forceps holding gamma probe

(a) Drop-in γ-probe and da Vinci® tool

(b) Experimental setup with SPECT phantom

Fig. D.3: The drop-in γ-probe can be picked up using the da Vinci® ProGrasp™ Forceps (a), and used to scan a region of interest containing radioactive isotopes, such as ^{99m}Tc in a SLN or ^{57}Co in our SPECT phantom (b).

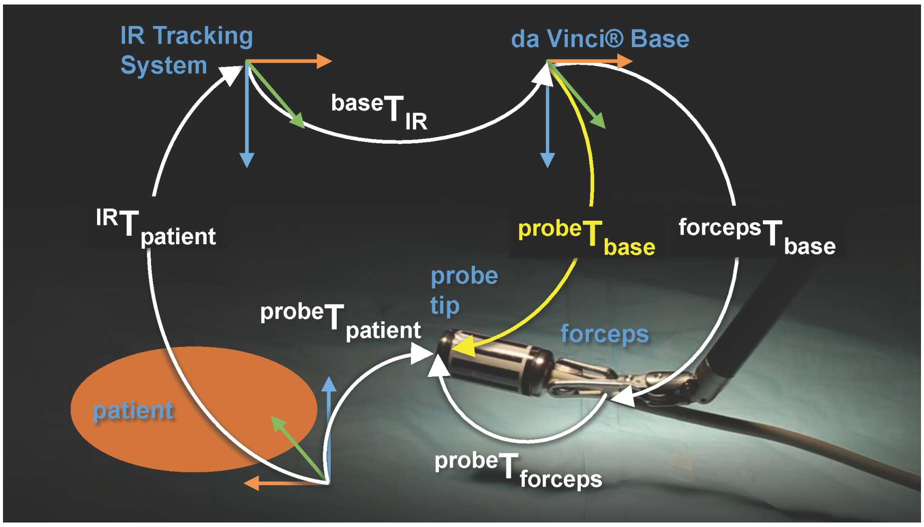
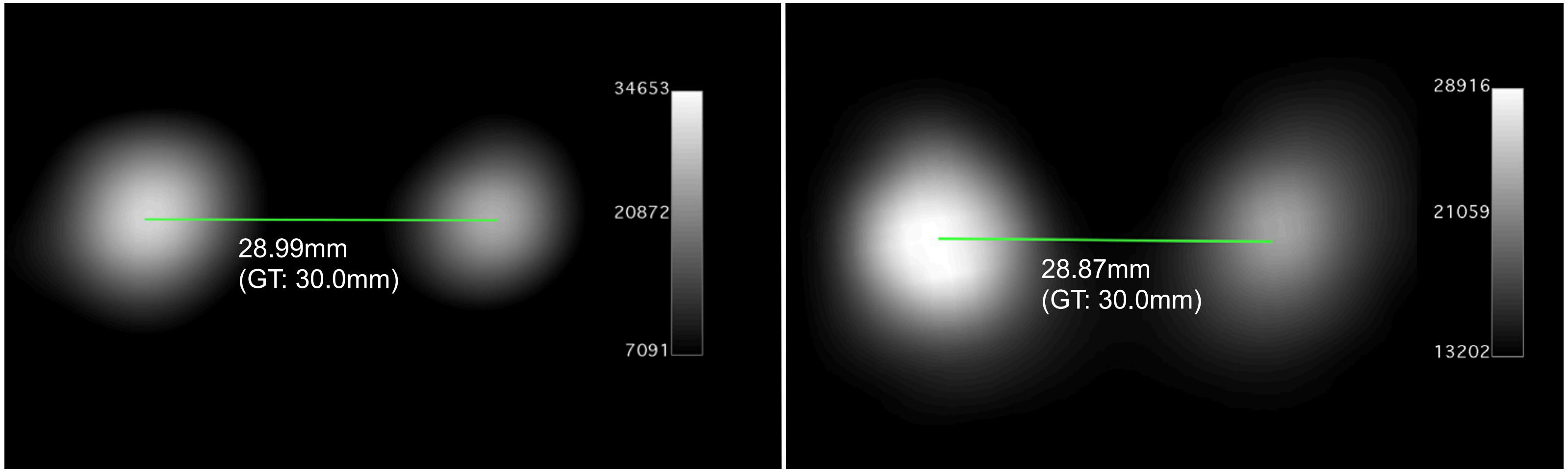

IR Tracking System
da Vinci® Base
baseT IR
IRT patient
probeT base
forcepsT base
probe tip
probeT patient
forceps
patient
probeT forceps

Fig. D.4: The probe tip is tracked in using a federated Kalman filter that combines the vision-based (yellow: ${}^{\text{probe}}\mathbf{T}_{\text{base}}$) and mechanical tracking (${}^{\text{forceps}}\mathbf{T}_{\text{base}}$) streams. The movements of the patient and da Vinci® base are compensated for using an infrared tracking system.



(a) Reconstruction of γ-camera scan

(b) Reconstruction of drop-in γ-probe scan

Fig. D.5: The radioactive sources have been scanned using a freehand γ-camera and the drop-in γ-probe held by the da Vinci® . The reconstructions are visualized in (a) and (b) respectively. Due to the limited angles and limited access from only two sides, the acquisition using the drop-in γ-probe and da Vinci® exhibit some artifacts along the borders of the volume (not illustrated). However, in both cases the reconstruction results clearly indicate two sources and the centroid distances are comparable. The ground truth (GT) distance is 30 mm in both cases.

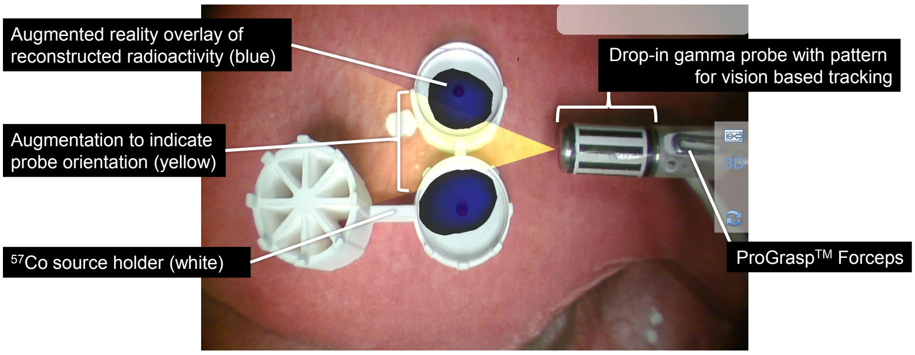
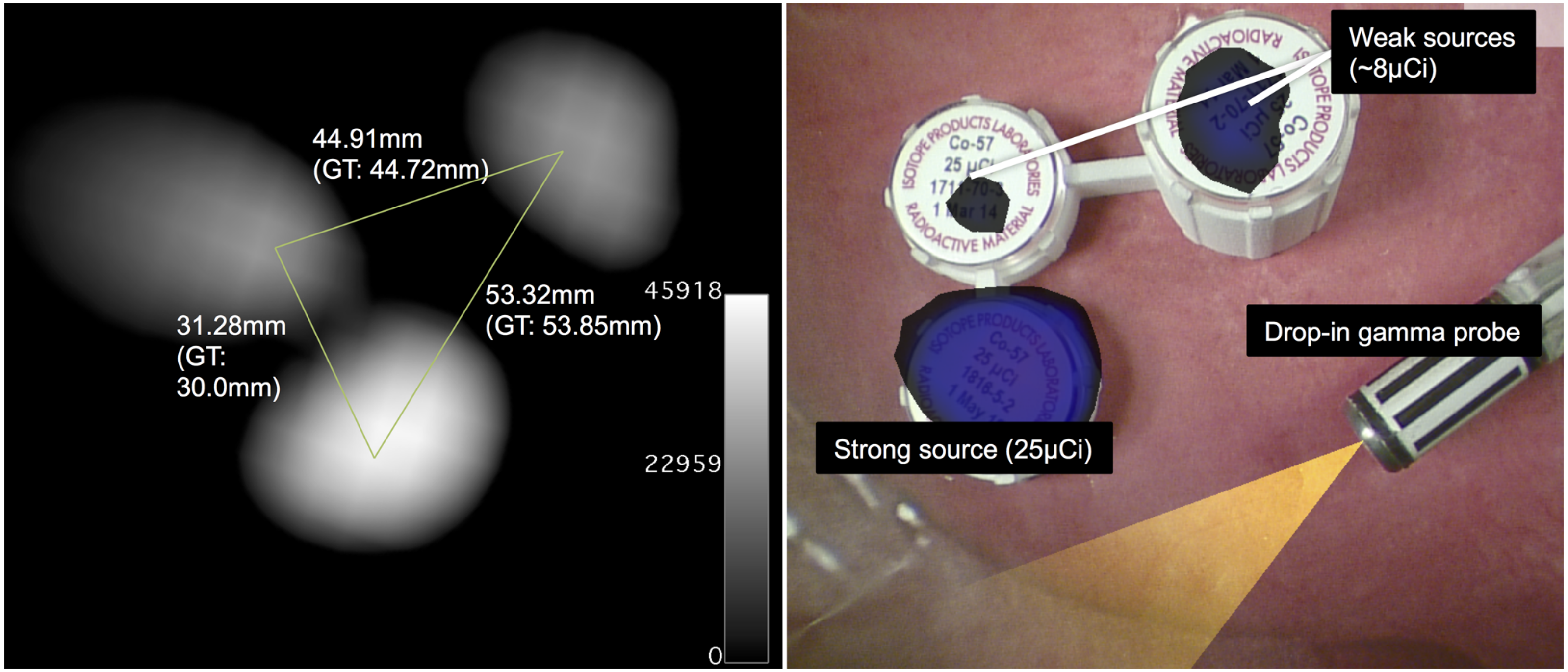

Augmented reality overlay of reconstructed radioactivity (blue)
Augmentation to indicate probe orientation (yellow)
^{57}Co source holder (white)
Drop-in gamma probe with pattern for vision based tracking
ProGrasp^{TM} Forceps

Fig. D.6: This augmented reality view is shown to the surgeon in parallel to the original endoscopic view. It allows the surgeon to see the reconstructed radioactivity distribution (manually windowed SPECT image is visualized in blue) semi-transparently overlaid on the endoscopic view. The stripe pattern on the probe is used for vision-based tracking, which allows for the augmentation of the probe field of view (yellow). During the acquisition, the γ source disks are mounted on a holder (white, 3D printed) to ensure that the ground truth is known. The background is designed to represent the interior of a patient, which contributes to a realistic, ex situ setup.


44.91mm
(GT: 44.72mm)
53.32mm
(GT: 53.85mm)
31.28mm
(GT:
30.0mm)
45918
22959
0
Weak sources
(~8μCi)
Drop-in gamma probe
Strong source (25μCi)

(a) Three-dimensional planar slice through centroids of sources. The measured distances and corresponding ground truth (GT) values are are labeled.

(b) Augmented reality view of 3 source reconstruction

Fig. D.7: To evaluate the accuracy of the reconstruction, three sources with varying intensities have been scanned. The distances between the sources were evaluated on a three-dimensional planar slice through all three centroids (a). The augmentation of the endoscopic view is shown in panel (b).

References

[1] Dean C Barratt, Alun H Davies, Alun D Hughes, Simon A Thom, and Keith N Humphries. "Optimisation and evaluation of an electromagnetic tracking device for high-accuracy three-dimensional ultrasound imaging of the carotid arteries". In: *Ultrasound in Medicine & Biology* 27.7 (2001), pp. 957–968.

[2] Anne-Sophie Bats, Patrice Mathevet, Annie Buenerd, Isabelle Orliaguet, Eliane Mery, Slimane Zerdoud, Marie-Aude Le Frère-Belda, Marc Froissart, Denis Querleu, Alejandra Martinez, et al. "The Sentinel Node Technique Detects Unexpected Drainage Pathways and Allows Nodal Ultrastaging in Early Cervical Cancer: Insights from the Multicenter Prospective SENTICOL Study". In: *Annals of Surgical Oncology* 20.2 (2013), pp. 413–422.

[3] NS van den Berg, GH KleinJan, OR Brouwer, T Wendler, RA Olmos, H van der Poel, and FWB van Leeuwen. "SPECT/CT mixed-reality combined with freehand SPECT augmented-reality for intraoperative navigation in sentinel node biopsy during laparoscopic procedures". In: *European Journal of Nuclear Medicine and Molecular Imaging.* Vol. 40. Springer, USA. 2013, S215–S215.

[4] Christina Bluemel, Andreas Schnelzer, Asli Okur, Alexandra Ehlerding, Stefan Paepke, Klemens Scheidhauer, and Marion Kiechle. "Freehand SPECT for image-guided sentinel lymph node biopsy in breast cancer". In: *European Journal of Nuclear Medicine and Molecular Imaging* 40.11 (2013), pp. 1656–1661.

[5] Oscar R Brouwer, Tessa Buckle, Anton Bunschoten, Joeri Kuil, Alexander L Vahrmeijer, Thomas Wendler, Renato A Valdés Olmos, Henk G van der Poel, and Fijs WB van Leeuwen. "Image navigation as a means to expand the boundaries of fluorescence-guided surgery". In: *Physics in Medicine and Biology* 57.10 (2012), pp. 3123–3136.

[6] Thijs Engelen, Beatrice MF Winkel, Daphne DD Rietbergen, Gijs H KleinJan, Sergi Vidal-Sicart, Renato A Valdés Olmos, Nynke S van den Berg, and Fijs WB van Leeuwen. "The next evolution in radioguided surgery: breast cancer related sentinel node localization using a freehandSPECT-mobile gamma camera combination". In: *American Journal of Nuclear Medicine and Molecular Imaging* 5.3 (2015), pp. 233–245.

[7] Maxime Fastrez, Frédéric Goffin, Ignace Vergote, Jean Vandromme, Philippe Petit, Karin Leunen, and Michel Degueldre. "Multi-center experience of robot-assisted laparoscopic para-aortic lymphadenectomy for staging of locally advanced cervical carcinoma". In: *Acta Obstetricia et Gynecologica Scandinavica* 92.8 (2013), pp. 895–901.

[8] Regina Feurer, C Hennersperger, JB Runyan, CL Seifert, J Pongratz, M Wilhelm, J Pelisek, N Navab, E Bartels, and H Poppert. "Reliability of a freehand three-dimensional ultrasonic device allowing anatomical orientation "at a glance": Study protocol for 3D measurements with Curefab CS". In: *Journal of Biomedical Graphics and Computing* 2.2 (2012), pp. 1–10.

[9] Michael Frumovitz, Robert L Coleman, Isis W Gayed, Pedro T Ramirez, Judith K Wolf, David M Gershenson, and Charles F Levenback. “Usefulness of preoperative lymphoscintigraphy in patients who undergo radical hysterectomy and pelvic lymphadenectomy for cervical cancer”. In: *American Journal of Obstetrics and Gynecology* 194.4 (2006), pp. 1186–1193.

[10] Arlan F Fuller Jr, Nancy Elliott, Cynthia Kosloff, William J Hoskins, and John L Lewis Jr. “Determinants of increased risk for recurrence in patients undergoing radical hysterectomy for stage IB and IIA carcinoma of the cervix”. In: *Gynecologic Oncology* 33.1 (1989), pp. 34–39.

[11] Sébastien Gouy, Aminata Kane, Catherine Uzan, Tristan Gauthier, Jennifer Gilmore, and Philippe Morice. “Single-port laparoscopy and extraperitoneal para-aortic lymphadenectomy: about fourteen consecutive cases”. In: *Gynecologic Oncology* 123.2 (2011), pp. 329–332.

[12] Alexander Hartl, Dzhoshkun I Shakir, Risto Kojchev, Nassir Navab, Sibylle I Ziegler, and Tobias Lasser. “Freehand SPECT reconstructions using look up tables”. In: *SPIE Medical Imaging 2012, Proceedings of.* Vol. 8316. International Society for Optics and Photonics. 2012, 83162H–83162H.

[13] Derrek A Heuveling, Stijn van Weert, K Hakki Karagozoglu, and Remco de Bree. “Evaluation of the use of freehand SPECT for sentinel node biopsy in early stage oral carcinoma”. In: *Oral Oncology* 51 (3 2014), pp. 287–290.

[14] G Holl, R Dorn, H Wengenmair, D Weckermann, and J Sciuk. “Validation of sentinel lymph node dissection in prostate cancer: experience in more than 2,000 patients”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 36.9 (2009), pp. 1377–1382.

[15] Laura L Holman, Charles F Levenback, and Michael Frumovitz. “Sentinel Lymph Node Evaluation in Women with Cervical Cancer”. In: *Journal of Minimally Invasive Gynecology* 21.4 (2014), pp. 540–545.

[16] Jacob P Hoogendam, Monique GG Hobbelink, Wouter B Veldhuis, René HM Verheijen, Paul J van Diest, and Ronald P Zweemer. “Preoperative sentinel node mapping with ^{99m} Tc-nanocolloid SPECT–CT significantly reduces the intraoperative sentinel node retrieval time in robot assisted laparoscopic cervical cancer surgery”. In: *Gynecologic Oncology* 129.2 (2013), pp. 389–394.

[17] JP Hoogendam, RHM Verheijen, I Wegner, and RP Zweemer. “Oncological outcome and long-term complications in robot-assisted radical surgery for early stage cervical cancer: an observational cohort study”. In: *BJOG: An International Journal of Obstetrics & Gynaecology* (2014).

[18] Rene Hornung, Daniele Bolla, Joachim Müller, Thomas Wendler, Alessandra Ray, and Alexander Markus. “One-step laparoscopic sentinel lymph node biopsy (SLN) in endometrium cancer using intraoperative freehand SPECT imaging”. In: *22nd Annual Congress European Society for Gynaecological Endoscopy.* European Society for Gynaecological Endoscopy. 2013.

[19] Ruediger Klapdor, Johanna Mücke, Michael Schneider, Florian Länger, Klaus-Friedrich Gratz, Peter Hillemanns, and Hermann Hertel. "Value and Advantages of Preoperative Sentinel Lymph Node Imaging With SPECT/CT in Cervical Cancer". In: *International Journal of Gynecological Cancer* 24.2 (2014), pp. 295–302.

[20] Joost AP Leijte, Ben Hughes, Niels M Graafland, Bin K Kroon, Renato A Valdés Olmos, Omgo E Nieweg, Cathy Corbishley, Sue Heenan, Nick Watkin, and Simon Horenblas. "Two-center evaluation of dynamic sentinel node biopsy for squamous cell carcinoma of the penis". In: *Journal of Clinical Oncology* 27.20 (2009), pp. 3325–3329.

[21] Charles F Levenback. "How safe is sentinel lymph node biopsy in patients with vulvar cancer?" In: *Journal of Clinical Oncology* 26.6 (2008), pp. 828–829.

[22] Wen P Liu, Sureerat Reaugamornrat, Anton Deguet, Jonathan M Sorger, Jeffrey H Siewerdsen, Jeremy Richmon, and Russell H Taylor. "Toward intraoperative image-guided transoral robotic surgery". In: *Journal of Robotic Surgery* 7.3 (2013), pp. 217–225.

[23] Gary H Lyman, Armando E Giuliano, Mark R Somerfield, Al B Benson, Diane C Bodurka, Harold J Burstein, et al. "American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer". In: *Journal of Clinical Oncology* 23.30 (2005), pp. 7703–7720.

[24] Mohamed Mabrouk, Michael Frumovitz, Marilyn Greer, Sheena Sharma, Kathleen M Schmeler, Pamela T Soliman, and Pedro T Ramirez. "Trends in laparoscopic and robotic surgery among gynecologic oncologists: a survey update". In: *Gynecologic Oncology* 112.3 (2009), pp. 501–505.

[25] P Matthies, J Gardiazabal, A Okur, T Lasser, and N Navab. "Accuracy Evaluation of Interventional Nuclear Tomographic Reconstruction using Mini Gamma Cameras". In: *The Hamlyn Symposium on Medical Robotics.* 2014, p. 31.

[26] Philipp Matthies, José Gardiazabal, Aslı Okur, Jakob Vogel, Tobias Lasser, and Nassir Navab. "Mini gamma cameras for intra-operative nuclear tomographic reconstruction". In: *Medical Image Analysis* 18.8 (2014), pp. 1329–1336.

[27] Philipp Matthies, Kanishka Sharma, Aslı Okur, José Gardiazabal, Jakob Vogel, Tobias Lasser, and Nassir Navab. "First Use of Mini Gamma Cameras for Intra-operative Robotic SPECT Reconstruction". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013.* Springer, 2013, pp. 163–170.

[28] Willem Meinhardt, Renato A Valdés Olmos, Henk G van der Poel, Axel Bex, and Simon Horenblas. "Laparoscopic sentinel node dissection for prostate carcinoma: technical and anatomical observations". In: *BJU international* 102.6 (2008), pp. 714–717.

[29] Isao Miyashiro, Kentaro Kishi, Masahiko Yano, Koji Tanaka, Masaaki Motoori, Masayuki Ohue, Hiroaki Ohigashi, Akemi Takenaka, Yasuhiko Tomita, and Osamu Ishikawa. "Laparoscopic detection of sentinel node in gastric cancer surgery by indocyanine green fluorescence imaging". In: *Surgical Endoscopy* 25.5 (2011), pp. 1672–1676.

[30] Rik Pijpers, Marrije R Buist, Arthur van Lingen, Jan Dijkstra, Paul J van Diest, Gerrit JJ Teule, Peter Kenemans, and René HM Verheijen. "The sentinel node in cervical cancer: scintigraphy and laparoscopic gamma probe-guided biopsy". In: *European Journal of Nuclear Medicine and Molecular Imaging* 31.11 (2004), pp. 1479–1486.

[31] Austin Reiter, Peter K Allen, and Tao Zhao. "Appearance learning for 3D tracking of robotic surgical tools". In: *The International Journal of Robotics Research* (2013), p. 0278364913507796.

[32] Lukas Rob, Helena Robova, Jiri Michael Halaska, Martin Hruda, and Petr Skapa. "Current status of sentinel lymph node mapping in the management of cervical cancer". In: *Expert Review of Anticancer Therapy* 13.7 (2013), pp. 861–870.

[33] Caitlin Schneider, Julian Guerrero, Christopher Nguan, Robert Rohling, and Septimiu Salcudean. "Intra-operative "Pick-Up" Ultrasound for Robot Assisted Surgery with Vessel Extraction and Registration: A Feasibility Study". In: *Information Processing in Computer-Assisted Interventions*. Vol. 6689. Springer, 2011, pp. 122–132.

[34] Gang Shi, D Lu, Zhihong Liu, Dan Liu, and Xiaoyan Zhou. "Robotic assisted surgery for gynaecological cancer". In: *Cochrane Database of Systematic Reviews* (2012).

[35] Blausen.com staff. "Blausen Gallery 2014". In: *Wikiversity Journal of Medicine* (2014).

[36] Ignace Vergote, Bram Pouseele, Toon van Gorp, Bernard van Acker, Karin Leunen, Isabelle Cadron, Patrick Neven, and Frederic Amant. "Robotic retroperitoneal lower para-aortic lymphadenectomy in cervical carcinoma: First report on the technique used in 5 patients". In: *Acta Obstetricia et Gynecologica Scandinavica* 87.7 (2008), pp. 783–787.

[37] L Vermeeren, RA Valdés Olmos, W Meinhardt, A Bex, HG van der Poel, WV Vogel, F Sivro, CA Hoefnagel, and S Horenblas. "Intraoperative radioguidance with a portable gamma camera: a novel technique for laparoscopic sentinel node localisation in urological malignancies". In: *European Journal of Nuclear Medicine and Molecular Imaging* 36.7 (2009), pp. 1029–1036.

[38] Lenka Vermeeren, Renato A Valdés Olmos, Wim Meinhardt, Axel Bex, Henk G van der Poel, Wouter V Vogel, Ferida Sivro, Cees A Hoefnagel, and Simon Horenblas. "Value of SPECT/CT for detection and anatomic localization of sentinel lymph nodes before laparoscopic sentinel node lymphadenectomy in prostate carcinoma". In: *Journal of Nuclear Medicine* 50.6 (2009), pp. 865–870.

[39] Sergi Vidal-Sicart, Beatriz Doménech, Blanca Luján, Jaume Pahisa, Aureli Torné, Sergio Martínez-Román, Juan Antonio Lejárcegui, Pere Fusté, Jaume Ordi, Pilar Paredes, et al. "Sentinel node in gynaecological cancers. Our experience". In: *Revista española de medicina nuclear (English Edition)* 28.5 (2009), pp. 221–228.

[40] Enrico Vizza, Emanuela Mancini, Ermelinda Baiocco, Cristina Vicenzoni, Lodovico Patrizi, Maria Saltari, Monica Cimino, Stefano Sindico, and Giacomo Corrado. "Robotic transperitoneal aortic lymphadenectomy in gynecologic cancer: a new robotic surgical technique and review of the literature". In: *Annals of Surgical Oncology* 19.12 (2012), pp. 3832–3838.

[41] Thomas Wendler, Alexander Hartl, Tobias Lasser, Joerg Traub, Farhad Daghighian, Sibylle Ziegler, and Nassir Navab. "Towards Intra-operative 3D Nuclear Imaging: Reconstruction of 3D Radioactive Distributions Using Tracked Gamma Probes". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*. Vol. 4792. Springer, 2007, pp. 909–917.

[42] Thomas Wendler, Ken Herrmann, Andreas Schnelzer, Tobias Lasser, Joerg Traub, Olivier Kutter, Alexandra Ehlerding, Klemens Scheidhauer, Tibor Schuster, Marion Kiechle, et al. "First demonstration of 3-D lymphatic mapping in breast cancer using freehand SPECT". In: *European Journal of Nuclear Medicine and Molecular Imaging* 37.8 (2010), pp. 1452–1461.

[43] Sandra L Wong, Charles M Balch, Patricia Hurley, Sanjiv S Agarwala, Timothy J Akhurst, Alistair Cochran, Janice N Cormier, Mark Gorman, Theodore Y Kim, Kelly M McMasters, et al. "Sentinel lymph node biopsy for melanoma: American Society of Clinical Oncology and Society of Surgical Oncology joint clinical practice guideline". In: *Journal of Clinical Oncology* 30.23 (2012), pp. 2912–2918.

[44] Ate GJ van der Zee, Maaike H Oonk, Joanne A de Hullu, Anca C Ansink, Ignace Vergote, René H Verheijen, Angelo Maggioni, Katja N Gaarenstroom, Peter J Baldwin, Eleonore B van Dorst, et al. "Sentinel node dissection is safe in the treatment of early-stage vulvar cancer". In: *Journal of Clinical Oncology* 26.6 (2008), pp. 884–889.

APPENDIX E

Vision-based Intraoperative Cone-Beam CT Stitching for Non-overlapping Volumes (MICCAI 2015)

Bernhard Fuerst[1,2] and Javad Fotouhi[1] and Nassir Navab[1,2]

[1] Computer Aided Medical Procedures (CAMP), Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

[2] Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstrae 3, 85748 Garching b. München, Germany

Copyright © Springer International Publishing Switzerland 2015.

Contribution Interventional X-ray imaging is widely in use, and stitching techniques for 2D images have been presented in the past decade. In this paper, we present a system and methodology to align two or more Cone-Beam CT (CBCT) volumes, by only utilizing an optical camera which is rigidly mounted to the C-arm. While moving the C-arm from one position to the next, the algorithm is designed to track features on the patient surface, and recover the scale by identifying the positioning laser which are built in the X-ray machine. This is novel concept allows the alignment of CBCT volumes without the introduction of patient-side reference markers, while the traditional surgical workflow remains intact.

Abstract Cone-Beam Computed Tomography (CBCT) is one of the primary imaging modalities in radiation therapy, dentistry, and orthopedic interventions. While providing crucial intraoperative imaging, CBCT is bounded by its limited imaging volume, motivating the use of image stitching techniques. Current methods rely on overlapping volumes, leading to an excessive amount of radiation exposure, or on external tracking hardware, which may increase the setup complexity. We attach an optical camera to a CBCT enabled C-arm, and co-register the video and X-ray views. Our novel algorithm recovers the

spatial alignment of non-overlapping CBCT volumes based on the observed optical views, as well as the laser projection provided by the X-ray system. First, we estimate the transformation between two volumes by automatic detection and matching of natural surface features during the patient motion. Then, we recover 3D information by reconstructing the projection of the positioning-laser onto an unknown curved surface, which enables the estimation of the unknown scales. We present a full evaluation of the methodology, by comparing vision- and registration-based stitching.

E.1 Introduction

Cone-Beam Computed Tomography (CBCT) enables intraoperative 3D imaging for various applications, for instance orthopedics [3], dentistry [13] or radiation therapy[4]. Consequently, CBCT is aimed at improving localization, structure identification, visualization, and patient positioning. However, the effectiveness of CBCT in orthopedic surgeries is bounded by its limited field of view, resulting in small volumes. Intraoperative surgical planning and verification could benefit from an extended field of view. Long bone fracture surgeries could be facilitated by 3D absolute measurements and multi-axis alignment in the presence of large volumes, assisting the surgeon's mental alignment.

The value of stitched fluoroscopy images for orthopedic surgery was investigated in [8]. Radio-opaque referencing markers attached to the tool were used to perform the stitching. Trajectory visualization and total length measurement were the most frequent features used by the surgeons in the stitched view. The outcome was overall promising for future development, and the usability was counted as good. Similarly, [10, 5] employed X-ray translucent references positioned under the bone for 2D X-ray mosaicing. In [15, 16], optical features acquired from an adjacent camera were used to recover the transformation. The aforementioned methods all benefit from external features for 2D mosaicing, thus do not require large overlaps. However, it remains a challenge to generalize these approaches to perform 3D volume stitching, as illustrated in Fig. E.1.

A validation study on using 3D rotational X-ray over conventional 2D X-rays was conducted for intra-articular fractures of the foot, wrist, elbow, and shoulder [3]. The outcome reported a reduction of indications for revision surgery. A panoramic CBCT is proposed in [4] by stitching overlapping X-rays acquired from all the views around the interest organ. Reconstruction quality is ensured by introducing a sufficient amount of overlapping regions, which in return increases the X-ray dose. Moreover, the reconstructed volume is vulnerable to artifacts introduced by image stitching. An automatic 3D image stitching technique is proposed in [6]. Under the assumption that the orientational misalignment is negligible, and sub-volumes are only translated, the stitching is performed using phase correlation as a global similarity measure, and normalized cross correlation as the local cost. Sufficient overlaps are required to support this method. To reduce the X-ray exposure, [9, 7] incorporate prior knowledge from statistical shape models to perform a 3D reconstruction.

Previous approaches are either limited to the overlap size or the existing prior shape models. Providing large overlaps will significantly increase the exposure. On the other hand, the bone fractures cause large deformation, hence preoperative and postoperative structures of the region of interest are signifi-

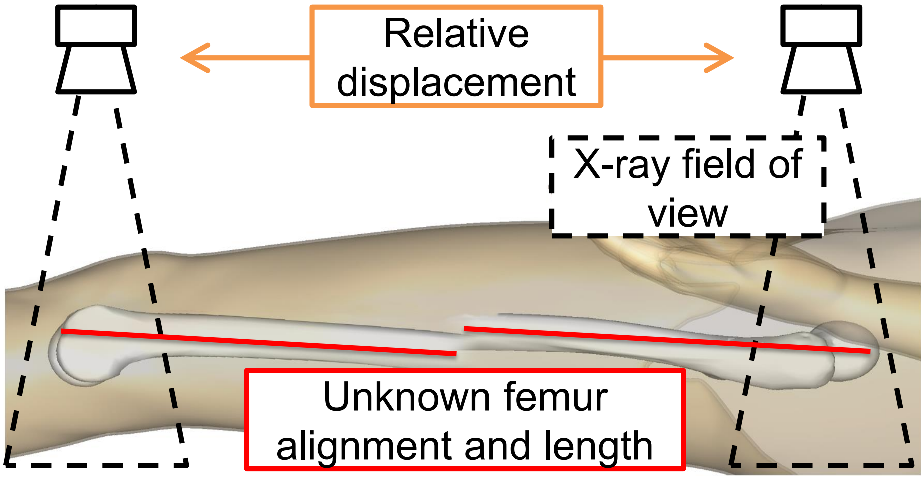


Fig. E.1: The 3D misalignment of bones (red lines) may be difficult to quantify using 2D images. CBCT contributes as a valuable tool for interventions in which the 3D alignment is of importance, for instance in acute fracture treatment or joint replacement. Background image courtesy of BodyParts3D, Center for Life Science, Japan.

cantly different, and one cannot benefit from prior scans for alignment. Lastly, incorporating external trackers leads to an increase in surgical complexity and line of sight problem. In this work, we propose a novel stitching approach, using a co-registered X-ray source with an optical camera attached to the C-arm [11, 12], and a patient positioning-laser to recover the depth scale. Therefore, the system is mobile, self-contained and independent of the OR, and the workflow remains intact. It could be deployed after a single factory calibration. The alignment transformation of volumes is computed based on the video frames, and prior models are not required. We target cases with large gaps between the volumes and focus our approach on spatial alignment of separated regions of interest. Image quality will remain intact, and the radiation dose will be linearly proportional to the size of the individual non-overlapping sub-volumes of interest.

E.2 Materials and Methods

E.2.1 System setup and calibration

The CBCT-enabled motorized C-arm is positioned relative to the patient by utilizing the positioning-lasers, which are built into the image intensifier and C-arm base. To enable the stitching of multiple sub-volumes, the transformation of the patient relative to the C-arm center must be recovered. In contrast to existing techniques we do not require additional hardware setup around the C-arm, but we attach a camera to the C-arm in such manner that it does not obstruct the surgeon's access to the patient. By using one mirror, the camera and the X-ray source centers are optically identical. The system setup is outlined in Fig. E.2.

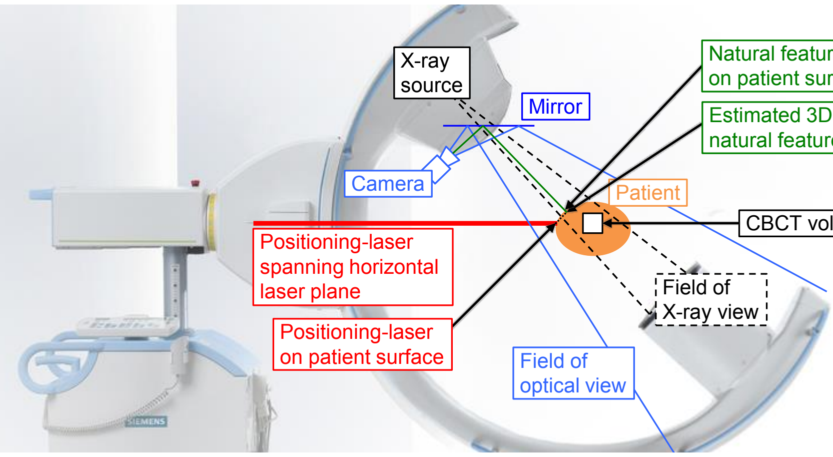

X-ray source
Mirror
Natural feature on patient surface
Estimated 3D natural feature
Camera
Patient
CBCT volume
Positioning-laser spanning horizontal laser plane
Positioning-laser on patient surface
Field of X-ray view
Field of optical view
SIEMENS

Fig. E.2: A mobile C-arm, the positioning-laser (red), and an optical camera (blue) are illustrated. The mirror (purple) aligns the optical camera and X-ray source centers. The patient motion relative to the C-arm is estimated by observing both the positioning-laser and natural features (green) on the patient's surface. The 3D positions of the features are estimated using the depth of the nearest positioning-laser on the patient (black dotted line intersecting green line), of which the depth is based on calibration.

Our system is composed of a mobile C-arm, ARCADIS Orbic 3D, from Siemens Medical Solutions and an optical video camera, Manta G-125C, from Allied Vision Technologies. The C-arm and the camera are both connected via ethernet to the computer with custom software to store the CBCT volumes and video. The X-ray and optical images are calibrated in an offline phase [11, 12].

The positioning-laser in the base of the C-arm spans a plane, which intersects with the unknown patient surface, and can be observed as a curve in the camera image. To determine the exact position of the laser relative to the camera, we perform a camera-to-plane calibration. Multiple checkerboard poses (n) are recorded for which the projection of the positioning-laser intersects with the origin of the checkerboard. Once the camera intrinsics are estimated, the camera-centric 3D checkerboard poses are computed. Under the assumption that the 3D homogeneous checkerboard origins, $\mathbf{x}^{(3)} = \{\mathrm{x}_i \,|\, \mathrm{x}_i = [x, y, z, 1]^\top\}_{i=0}^{n}$ (see footnote [1] for notation), lay on the laser plane, the plane coefficients $A = [a, b, c, d]$ are determined by performing RANdom SAmple Consensus (RANSAC) based plane fitting to the observed checkerboard origins, which attempts to satisfy:

$$\arg\min_{A} \sum_{x \in \Omega} |Ax_j|, \tag{E.1}$$

where Ω is subset of checkerboard origins, which are inliers to the plane fitting.

---

[1] Superscripts $^{(2)}$ and $^{(3)}$ denote 2D and 3D points; $^{(s)}$ denotes points up to a scale.

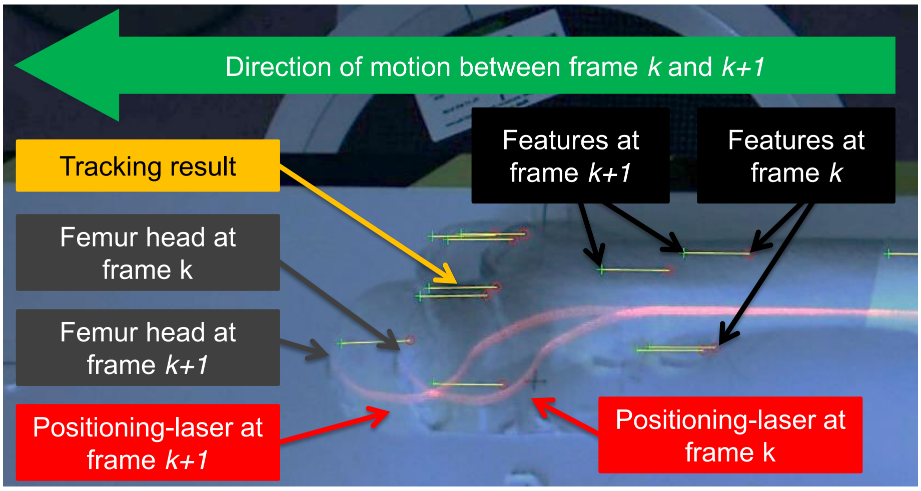

Direction of motion between frame *k* and *k+1*
Tracking result
Features at frame *k+1*
Features at frame *k*
Femur head at frame k
Femur head at frame *k+1*
Positioning-laser at frame *k+1*
Positioning-laser at frame k

Fig. E.3: The figure shows the overlay of two frames to illustrate the feature correspondences to estimate the movement of the patient. From both frames, the positioning-laser (red) and natural surface features are extracted. The tracking results of the matched features in frame k (+) and frame $k+1$ ($\circ$) are illustrated as yellow lines.

E.2.2 CBCT Volume and Video Acquisition

To acquire a CBCT volume, the patient is positioned under guidance of the lasers. Then, the motorized C-arm orbits 190° around the center visualized by the laser lines, and automatically acquires a total of 100 2D X-ray images. The reconstruction is performed using the Feldkamp method, which utilizes filtered back-projection, resulting in a cubic volume with a 256 voxels along each axis and an isometric resolution of 0.5 mm. During the re-arrangement of C-arm and patient for the next CBCT acquisition, the positioning-laser is projected at the patient, and each video frame is recorded. For simplicity, we will assume that in the following the C-arm is static, while the patient is moving. However, as only the relative movement of patient to C-arm is recorded, there are no limitations on allowed motions.

E.2.3 Two-Dimensional Feature Detection and Matching

The transformation describing the relative patient motion observed between two video frames is estimated by detecting and matching a set of natural surface features and the recovery of their scale. For each frame, we automatically detect Speeded Up Robust Features (SURF) as described in [2], which are well suited to track natural shapes and blob-like structures. To match the features in frame k to the features in frame $k+1$, we find the nearest neighbor by exhaustively comparing the features, and removing weak or ambiguous matches. Outliers are removed by estimating the Fundamental Matrix, $\mathbf{F}_k$, using a least trimmed squares formulation and rejecting up to 50% of the features, resulting in a set of n_k features $\mathbf{f}_k^{(2)} = \{\mathrm{f}_{k,j} \,|\, \mathrm{f}_{k,j} = [x, y, 1]^\top\}_{j=1}^{n}$ in frame k (see Fig. E.3). To estimate the 3D transformation, the 3D coordinates of this set of features need to be estimated.

E.2.4 Recovering Three-Dimensional Coordinates

In each frame k, the laser is automatically detected. First the color channel corresponding to the laser's color is thresholded and noise is removed by analyzing connected components. To find the m_k 2D points, $\mathbf{p}_k^{(2)} = \{\mathrm{p}_{k,i} \,|\, \mathrm{p}_{k,i} = [x, y, 1]^\top\}_{i=1}^{m}$, which are most likely on the plane, the resulting binary image is thinned [17]. Each 2D laser point $\mathrm{p}_{k,i}^{(2)}$ is projected back to a point up to a scale $\mathrm{p}_{k,i}^{(s)} = [x_{k,i}^{(s)}, y_{k,i}^{(s)}, 1, 1]^\top$ using the Moore-Penrose pseudo-inverse of the camera projection matrix, $\mathbf{P}$:

$$\mathrm{p}_{k,i}^{(3)} = s_{k,i}\mathrm{p}_{k,i}^{(s)} = s_{k,i}\mathbf{P}^{+}\mathrm{p}_{k,i}^{(2)}, \tag{E.2}$$

where the scale $s_{k,i}$ is recovered by intersecting the point up to a scale $\mathrm{p}_{k,i}^{(s)}$ with the plane:

$$s_{k,i} = \frac{-d}{a x_{k,i}^{(s)} + b y_{k,i}^{(s)} + c}. \tag{E.3}$$

Once the 3D laser points are recovered, the scale for each feature, $\mathrm{f}_{k,j}^{(s)} = s_{k,j}\mathbf{P}^{+}\mathrm{f}_{k,j}^{(2)}$, can be estimated by interpolating the scales of the closest points $\mathrm{p}_{k,i}^{(3)}$.

E.2.5 Estimating 3D Transformation and CBCT Volume Stitching

After the estimation of the 3D coordinates of the matched features, the transformation for the frames k and $k+1$ is computed by solving the least squares fitting for two sets of 3D points [1], obtaining the transformation matrix $\mathbf{T}_k$. Note that, only features in a small neighborhood of the laser line, < 1 cm, are used. Hence, features on other body parts, e.g. the opposite leg, are discarded. To verify the estimated transformation, the Iterative Closest Point (ICP) algorithm is used to perform a redundancy test using the laser points. In other words, ICP is applied after transforming the laser points $\mathrm{p}_i^{(3)}$ from frame k to the next $k+1$ only for verification. Consequently, for long bones, translation along the laser line is not lost. This results in a transformation $\hat{\mathbf{T}}_k$. If $\hat{\mathbf{T}}_k$ is not nearly identity, the frame $k+1$ is rejected and the frames k and $k+2$ are used to compute $\hat{\mathbf{T}}_k$. To obtain the overall transformation $\mathbf{T}_{\mathrm{CBCT}}$, all transformations $\mathbf{T}_k \in \Gamma$ are accumulated, where Γ is the domain of all valid transformations:

$$\mathbf{T}_{\mathrm{CBCT}} = {}^{\mathrm{CBCT}}\mathbf{T}_{\mathrm{camera}} \prod_{\mathbf{T} \in \Gamma} \mathbf{T}_k, \tag{E.4}$$

where ${}^{\mathrm{CBCT}}\mathbf{T}_{\mathrm{camera}}$ is the transformation from camera coordinate system to the CBCT coordinate system obtained during calibration.

E.3 Experiments and Results

The novel laser-guided stitching method is evaluated in two different, but realistic scenarios. For each phantom, we performed vision-based stitching and

Table E.1: Errors are computed by comparing the vision-based stitched CBCT to the real objects. The final row presents the difference to the intensity-based stitching.

Error	X	Y	Z	Norm
Long Bone (Femur) Phantom				
Alignment error (mm)	0.75	0.83	0.37	1.18
Absolute distance error (%)	1.30	1.11	0.10	n/a
Fiducial Phantom				
Alignment error (mm)	0.08	0.10	0.26	0.29
Absolute distance error (%)	0.52	2.00	2.60	n/a
Vision- vs. intensity-based (mm)	0.11	0.08	0.18	0.23

evaluated the quality by measuring 3D distances in the stitched volumes and real object. In addition, the stitching quality was compared to intensity-based mosaicing using overlapping CBCT volumes, indicating the accuracy of the overall 3D transformation $\mathbf{T}_{\mathrm{CBCT}}$.

The result of vision-based stitching is illustrated in Fig. E.4 (a) on the long bone phantom in the absence of overlaps, and in Fig. E.4 (c) on the fiducial phantom with overlaps. The absolute distances are compared to real world measurements which are illustrated in Fig. E.4 (b) and (d). Detailed results are reported in table E.1, which shows the differences of measurements of the vision-based stitched CBCT volumes and real objects. The errors are apportioned according to the coordinate frames illustrated in Fig. E.4, while the *norm* reflects the overall error. In addition, the *absolute distance error* reports the percentage of error with respect to the absolute distances measured. Average errors are in the range of 0.65 ± 0.28 mm and 0.15 ± 0.11 mm for long bone and fiducial phantom stitching, respectively. Lastly, for overlapping volumes, we have compared the vision- and intensity-based stitching by performing rigid registration using normalized cross correlation as similarity measure. The intensity-based stitching deviated from the vision-based stitching by 0.23 mm, indicating an overall good alignment.

E.4 Discussion and Conclusion

The proposed technique is an overlap-independent, low dose, and accurate stitching method for CBCT sub-volumes with intact workflow complexity. We attached an optical camera to a mobile C-arm, and used the positioning-laser to recover the 3D depth scales, and consequently aligned the sub-volumes. As a result of this method, the stitching is performed with low dose radiation, linearly proportional to the size of non-overlapping sub-volumes. We expect this to be applicable to intraoperative planning and validation for long bone fracture or joint replacement interventions, where multi-axis alignment and absolute distances are difficult to visualize and measure from the 2D X-ray views.

Our approach does not limit the working space, nor does it require any additional hardware besides a simple camera. The C-arm remains mobile and independent of the OR. One requirement is that the C-arm does not move during the CBCT acquisition, but we believe that the use of external markers could solve this problem and may yield a higher accuracy. However, in our scenario we intentionally did not rely on markers, as they would increase complexity and alter the surgical workflow. Our approach uses frame-to-frame tracking, which can cause drift. In fact, the ICP verification helps us to detect such drifts as it is based on points which were not used for motion estimation. Therefore, if the estimated motion from ICP increases over time, we can detect the drift and use ICP to correct if necessary. Alternatively, the transformations could be refined using bundle adjustments [14]. Further studies on the effectiveness during interventions are underway. Also, the reconstruction of the patient surface during the CBCT acquisition may assist during the tracking of the patient motion.

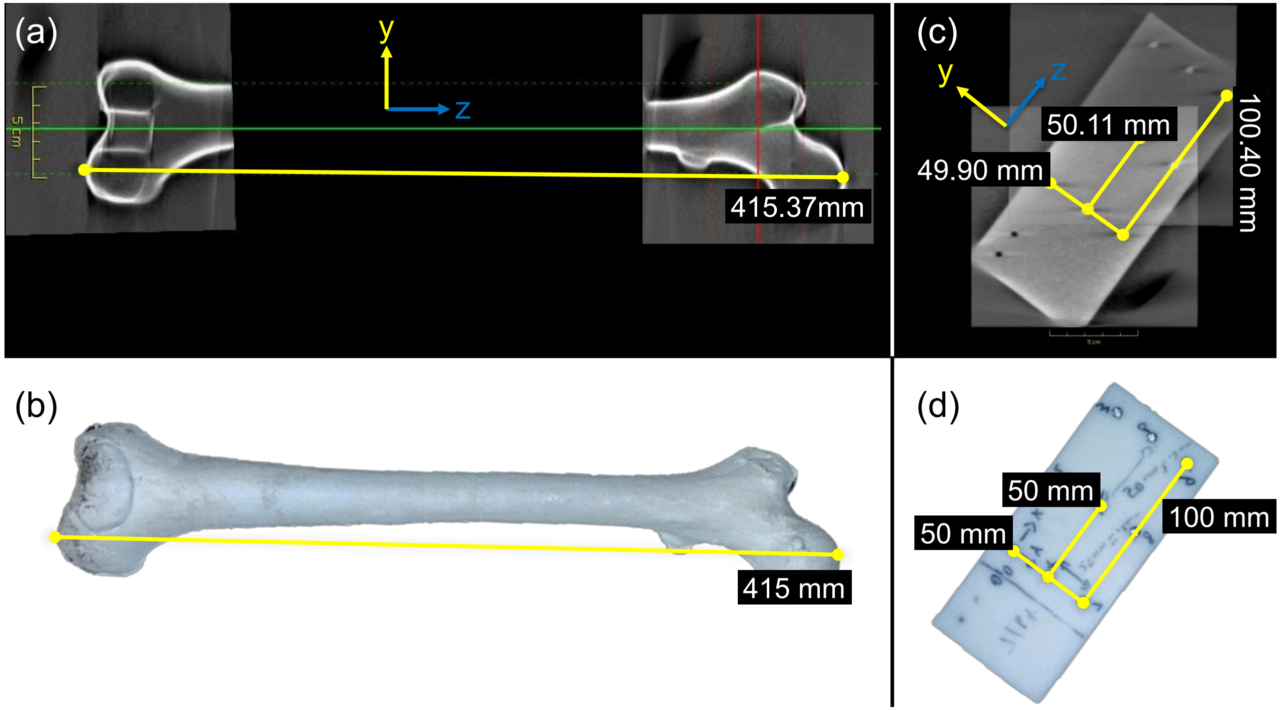

(a)
y
z
5 cm
415.37mm
(c)
y
z
50.11 mm
49.90 mm
100.40 mm
(b)
415 mm
(d)
50 mm
50 mm
100 mm

Fig. E.4: Absolute distance of the aligned sub-volumes in (a) is measured (415.37 mm), and compared to the real world measurements (415 mm) of the femur phantom in (b). Similarly, a fiducial phantom was scanned and the vision-based stitching (c) compared to the real world object (d). For visualization purposes and volumetric appearance in (a) and (c), multiple parallel slices are averaged.

References

[1] K Somani Arun, Thomas S Huang, and Steven D Blostein. "Least-squares fitting of two 3-D point sets". In: *PAMI, IEEE Transactions on* 5 (1987), pp. 698–700.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "SURF: Speeded Up Robust Features". In: *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[3] B. Carelsen, R. Haverlag, D.Th. Ubbink, J.S.K. Luitse, and J.C. Goslings. "Does intraoperative fluoroscopic 3D imaging provide extra information for fracture surgery?" English. In: *Archives of Orthopaedic and Trauma Surgery* 128.12 (2008), pp. 1419–1424. ISSN: 0936-8051.

[4] Jenghwa Chang, Lili Zhou, Song Wang, and K. S. Clifford Chao. "Panoramic cone beam computed tomography". In: *Medical Physics* 39.5 (2012), pp. 2930–2946.

[5] Cheng Chen, Risto Kojcev, Daniel Haschtmann, Tamas Fekete, Lutz Nolte, and Guoyan Zheng. "Ruler Based Automatic C-Arm Image Stitching Without Overlapping Constraint". English. In: *Journal of Digital Imaging* (2015), pp. 1–7. ISSN: 0897-1889.

[6] Mario Emmenlauer, Olaf Ronneberger, Aaron Ponti, Patrick Schwarb, Alessandra Griffa, Alida Filippi, Roland Nitschke, Wolfgang Driever, and Hans Burkhardt. "XuvTools: free, fast and reliable stitching of large 3D datasets". In: *Journal of microscopy* 233.1 (2009), pp. 42–60.

[7] M. Fleute and S. Lavallée. "Nonrigid 3-D/2-D Registration of Images Using Statistical Models". English. In: *Medical Image Computing and Computer-Assisted Intervention*. Ed. by Chris Taylor and Alain Colchester. Vol. 1679. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1999, pp. 138–147. ISBN: 978-3-540-66503-8. DOI: `10.1007/10704282_15`.

[8] Michael Kraus, Stephanie von dem Berge, Hendrik Schoell, Gert Krischak, and Florian Gebhard. "Integration of fluoroscopy-based guidance in orthopaedic trauma surgery: A prospective cohort study". In: *Injury* 44.11 (2013), pp. 1486–1492.

[9] H. Lamecker, T.H. Wenckebach, and H.-C. Hege. "Atlas-based 3D-Shape Reconstruction from X-Ray Images". In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 1. 2006, pp. 371–374.

[10] Peter Messmer, Felix Matthews, Christoph Wullschleger, Rolf Hügli, Pietro Regazzoni, and Augustinus. Jacob. "Image Fusion for Intraoperative Control of Axis in Long Bone Fracture Treatment". English. In: *European Journal of Trauma* 32.6 (2006), pp. 555–561. ISSN: 1439-0590. DOI: `10.1007/s00068-006-5159-5`.

[11] N. Navab, S.-M. Heining, and J. Traub. "Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications". In: *Medical Imaging, IEEE Transactions on* 29.7 (July 2010), pp. 1412–1423. ISSN: 0278-0062.

[12] S Nicolau, PY Lee, HS Wu, M Huang, R Lukang, L Soler, and J Marescaux. "Fusion of C-arm X-ray image on video view to reduce radiation exposure and improve orthopedic surgery planning: first in-vivo evaluation". In: *Proceedings of Computer Assisted Radiology and Surgery*. Vol. 6. 2011, pp. 115–116.

[13] Ruben Pauwels, Kazuyuki Araki, JH Siewerdsen, and Saowapak S Thongvigitmanee. "Technical aspects of dental CBCT: state of the art". In: *Dentomaxillofacial Radiology* 44.1 (2015).

[14] Bill Triggs, PhilipF. McLauchlan, RichardI. Hartley, and AndrewW. Fitzgibbon. "Bundle Adjustment - A Modern Synthesis". English. In: *Vision Algorithms: Theory and Practice*. Ed. by Bill Triggs, Andrew Zisserman, and Richard Szeliski. Vol. 1883. Springer, 2000, pp. 298–372.

[15] Lejing Wang, Joerg Traub, SandroMichael Heining, Selim Benhimane, Ekkehard Euler, Rainer Graumann, and Nassir Navab. "Long Bone X-Ray Image Stitching Using Camera Augmented Mobile C-Arm". English. In: *Medical Image Computing and Computer-Assisted Intervention*. Vol. 5242. Springer Berlin Heidelberg, 2008, pp. 578–586. ISBN: 978-3-540-85989-5.

[16] Lejing Wang, Joerg Traub, Simon Weidert, Sandro Michael Heining, Ekkehard Euler, and Nassir Navab. "Parallax-free intra-operative X-ray image stitching". In: *Medical Image Analysis* 14.5 (2010).

[17] TY Zhang and Ching Y. Suen. "A fast parallel algorithm for thinning digital patterns". In: *Communications of the ACM* 27.3 (1984), pp. 236–239.

APPENDIX F

Pre-Clinical Usability Study of Multiple Augmented Reality Concepts for K-Wire Placement (IJCARS 2016)

Marius Fischer[1,2] * and Bernhard Fuerst[2,3] *, Sing Chun Lee[2], Javad Fotouhi[2], Severine Habert[3], Simon Weidert[1], Ekkehard Euler[1], Greg Osgood[4] and Nassir Navab[2,3]

* M. Fischer and B. Fuerst contributed equally to the work and should be regarded as joint first authors.

[1] Trauma Surgery Department, Klinikum Innenstadt, LMU Munich, Germany

[2] Computer Aided Medical Procedures (CAMP), Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

[3] Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstrae 3, 85748 Garching b. München, Germany

[4] Orthopaedic Trauma, Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, MD, US

Copyright © Springer International Publishing Switzerland 2016.

Contribution This manuscript presents the methods, results and conclusions of a pre-clinical usability study. The technical components are in regard to the system calibration and setup (see Appendix G), and the execution and statistical analysis of the usability study, while the key scientific contributions are the identification of potential benefits of the 2D/2D and 3D/3D augmented reality systems. The manuscript present a clear indication that the 3D/3D system provides superior support during guide wire placements in bone phantoms. Therefore, further research should investigate the potential impact on other medical applications using C-arm imaging, and complete the system in order to move towards clinical studies.

PRE-CLINICAL USABILITY STUDY OF MULTIPLE AUGMENTED REALITY CONCEPTS FOR K-WIRE PLACEMENT (IJCARS 2016)

Abstract *Purpose:* In many orthopedic surgeries there is a demand for correctly placing medical instruments (e.g. K-wire or drill) to perform bone fracture repairs. The main challenge is the mental alignment of X-ray images acquired using a C-arm, the medical instruments and the patient, which dramatically increases in complexity during pelvic surgeries. Current solutions include the continuous acquisition of many intra-operative X-ray images from various views, which will result in high radiation exposure, long surgical durations, and significant effort and frustration for the surgical staff. This work conducts a pre-clinical usability study to test and evaluate mixed reality visualization techniques using intra-operative X-ray, optical, and RGBD imaging to augment the surgeon's view to assist accurate placement of tools.
Method: We design and perform a usability study to compare the performance of surgeons and their task load using three different mixed reality systems during K-wire placements. The three systems are interventional X-ray imaging, X-ray augmentation on 2D video, and 3D surface reconstruction augmented by digitally reconstructed radiographs and live tool visualization.
Results: The evaluation criteria include duration, number of X-ray images acquired, placement accuracy and the surgical task load, which are observed during 21 clinically relevant interventions performed by surgeons on phantoms. Finally, we test for statistically significant improvements, and show that the mixed reality visualization leads to a significantly improved efficiency.
Conclusion: The 3D visualization of patient, tool and DRR shows clear advantages over the conventional X-ray imaging and provides intuitive feedback to place the medical tools correctly and efficiently.

Keywords Interventional Imaging, Usability Study, Orthopaedic and Trauma Surgery

F.1 Introduction

A continuous and rapid evolution of technology has changed the face of trauma and orthopaedic surgeries in the past decades. Especially, minimally invasive techniques are widely accepted for treatment of bone fractures in spine and pelvis, thanks to the development of modern imaging technology and computer aided navigation systems. The benefits of minimally invasive orthopeadic surgeries are the reduction of blood loss, collateral tissue damage and overall operating duration [5]. However, these techniques usually yield a higher X-ray exposure for both patient and clinical staff, and may increase fatigue and frustration due to the difficulty in continuous re-positioning of the mobile X-ray machine (C-arm) [1, 32].

The main challenge during percutaneous K-wire placement and screw fixation is the mental alignment of patient, medical instruments and the intra-operative X-ray images [27, 30], which also requires the frequent re-positioning of the C-arm [31]. For instance, in pelvic acetabulum fractures, the surgeon needs to find the correct trajectory of the K-wire through a small bony structure, namely the superior pubic ramus. The misplacement of the K-wire could cause severe damage to the external iliac artery and vein, obturator nerve, or to structures such as the inguinal canal and intra-articular hip joint [8]. It

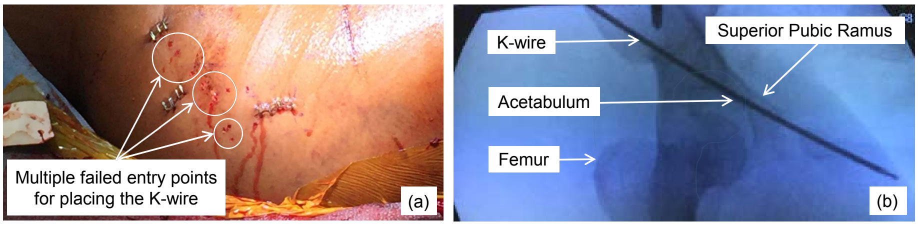


Fig. F.1: a) Lateral view of the hip in pelvic surgery: several skin punctures demonstrate the number of failed attempts to place the K-wire. b) Anteroposterior X-ray image of hip region in pelvic surgery. The narrow superior pubic ramus requires a precise placement of the K-wire, especially considering that a misplacement could cause severe damage to external iliac artery or vein.

is not unusual that a single K-wire placement for one screw takes up to ten minutes [29].

The standard treatment procedure for undisplaced superior pubic ramus fractures requires several K-wire placements and subsequent screw insertions. For each K-wire, the surgeon first locates the entry point location and performs a skin incision at the lateral side of the hip, which requires several intra-operative X-ray images from various perspectives to confirm the exact tool orientation. It is common to correct the K-wire placement, as displayed in Fig. F.1a). While advancing the K-wire through soft tissue and into the bone, X-ray images from various perspectives are acquired to constantly validate the trajectory. Fig. F.1b) shows the narrow path through the superior pubic ramus. After the K-wire is placed, the procedure concludes by drilling and placing a cannulated screw.

Computer aided surgical navigation systems have been introduced to assist the placement of K-wires and screws. Current solutions use pre-operative Computed Tomography (CT) volumes, external optical tracking systems, and tracked markers as reference on medical instruments, the patient and the C-arm. Navigation systems then provide intra-operative information on the spatial relation of surgical instruments and medical images. The validation of the K-wire placement is performed using conventional X-ray imaging.

The benefits of navigation systems are controversial. Some publications indicate a reduction of the radiation dose and an increase in accuracy [7, 6], while a more recent study shows no clear advantage of using navigation systems in some procedures [14]. A major drawback of navigation systems is the high cost, which limits the availability of such systems to major hospitals and research facilities [12, 7]. The cost is driven by external hardware, which constitutes a logistical problem due to the bulkiness and consumption of space in the OR. Beyond hardware requirements, the systems also impose a change of the surgical workflow [31]. In summary, after two decades of surgical navigation systems, expert surgeons are starting to realize that these systems have failed to provide the advantages promised. They do not reduce the required OR time, show no systematic, significant influence on the patient outcome and do not reduce the frustration of the surgeon and staff. The additional efforts required to use modern surgical navigation systems often outweigh the benefits in many scenarios. Therefore, interventions are frequently performed without surgical navigation systems even though navigation would be available and theoretically present a benefit [13], which has been especially researched for spine surgery [11].

An alternative solution, which is comparatively inexpensive, contained in existing equipment and intuitive, has been proposed in [18]. This solution adds a mirror and video camera to a C-arm, such that the X-ray and optical views align. After a single calibration and warping step, the video view can be augmented with the X-ray images, which provides an intuitive optical feedback to the surgeon. In cadaver studies, this system leads to reduced radiation dose and increase in surgical efficiency in terms of duration and accuracy [16, 33]. During orthopaedic and trauma procedures, the use of a camera augmented intra-operative X-ray system resulted in improved incisions, reduced radiation exposure of the surgeon, and simplified instrument tool alignment [19, 3].

However, the mirror construction reduces the free moving space of the surgeon, which can be overcome in mounting the camera next to the X-ray source [22]. That setup will only be able to augment the video view with warped X-ray images, which are clinically less relevant. Both approaches require the X-ray source to be positioned on the top rather than below the surgical table, which is an unusual setup and may increase the exposure of the surgeon to scatter radiation.

In continuation to [18], in [9] an RGBD (Red Green Blue Depth) camera was mounted to a C-arm instead of a video camera. Similarly to an RGB camera, an RGBD camera provides a 2D color image, and additionally provides a depth value for every pixel which represents the distance between the observed object and the camera origin. This allows to reconstruct the 3D surfaces of an object. The system using the RGBD camera rather than the RGB camera, enables an offline 3D/2D mixed reality visualization of X-ray on the reconstructed patient surface. The main limitation of this work is due to 2D projective nature of the X-ray image. As soon as the display viewpoint of the surface is different than the X-ray source optical axis, the visualization is physically wrong. Using CBCT may allow to overcome this issue, since a new simulated X-ray (DRR) corresponding to the viewpoint can be generated dynamically. In [10], two RGBD sensors were mounted on a mobile C-arm in order to synthesize the video as seen from the X-ray source viewpoint without the need of a mirror construction.

The integration of a stereo camera near the X-ray detector enables tool tracking within the working space of the C-arm. If CT images are transferred to the inter-operative setup, a Digitally Reconstructed Radiograph (DRR) can be computed and augmented onto the one camera view [25]. This system has been presented as a good combination of augmented reality visualization and surgical navigation systems, but requires markers on the patient and tools. The change of the augmented view requires the movement of the entire system, and may introduce errors of the alignment of CT and optical view in case the patient marker is occluded.

Systems with augmented video may benefit of the use of RGBD cameras, which allows the positioning of the virtual cameras and renderings of the patient surface from arbitrary perspectives [9]. RGBD information can also be used to improve the understanding of the environment and enhance the augmentation [24].

In this paper, we present a pre-clinical usability study to provide a more comprehensive understanding whether enhanced C-arm systems provide a clinically relevant benefit. We will compare K-wire placement using *(i)* conventional X-ray imaging, *(ii)* 2D RGB video augmented with X-ray images, and *(iii)* a novel 3D RGBD video augmented with DRRs generated from Cone-Beam CT (CBCT). The later system allows the surgeon virtually rotate the entire scene (DRR, patient surface and tools) and simultaneously view the scene from different perspectives. A total of 21 K-wire placements are performed by 7 surgeons, ranging from residents to attending physicians. We compare the system usabilities in terms of surgical efficiency, which is defined by the number of X-ray images, duration, accuracy and surgical task load.

F.2 Method

In this section we first describe the imaging systems to be compared. These include conventional intra-operative X-ray imaging, X-ray image augmented 2D video, and a novel 3D RGBD view augmented with DRR. Finally, we present the questionnaires and statistical methods to perform the usability study.

F.2.1 Imaging Systems

To evaluate the usability of mixed-reality visualization techniques, we acquire a baseline using conventional intraoperative X-ray imaging. Examples of the three visualizations are illustrated in Fig. F.2.

Conventional intra-operative X-ray Imaging This imaging method using a standard C-arm provides the baseline performance as it is the most commonly used system to perform image-guided K-wire placement. The images are obtained in the Digital Radiography (DR) mode. This allows for a single, brief exposure at higher than normal mA to capture a higher quality single image. For reasons of comparability between subjects, we limit the functionality of the C-arm to this mode.

2D RGB Video and X-Ray Visualization To achieve a fused RGB and X-ray visualization we attached a camera near the X-ray source. Using a mirror construction, the X-ray source and optical camera centers are virtually aligned as described in [18]. To be able to observe the surgical site using the RGB camera, the X-ray source and camera are positioned above the patient.

The X-ray images are obtained using the standard C-arm in DR mode. After camera calibration [34], the alignment registration of optical and X-ray images is performed using a single plane phantom with radiopaque markers that are also visible in the optical view [20].

Finally, this first augmented reality system allows the simultaneous display of live RGB video overlaid with DR images obtained at the user's discretion. Additionally, we provide the user with the option to control the alpha blending to change the transparency to be able to focus on the X-ray image or video background.

3D RGBD and DRR via CBCT Visualization The previous system requires the re-positioning of the C-arm in order to change the optical and X-ray view. To overcome this limitation we introduce a novel system using an RGBD

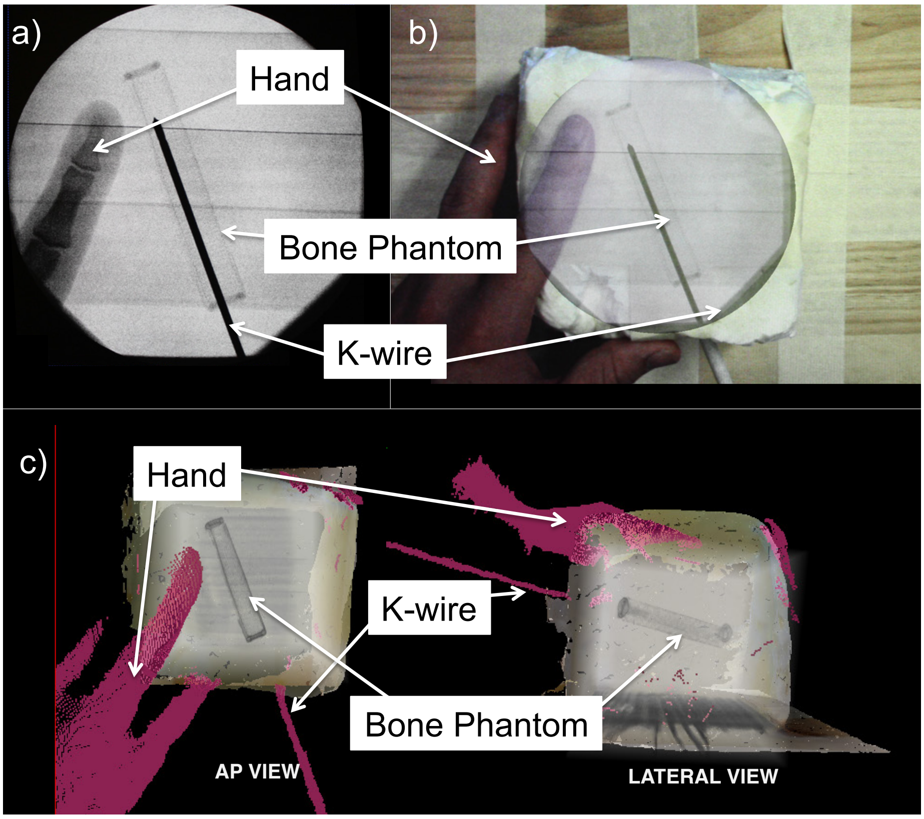


Fig. F.2: The same stage in the K-wire placement has been recreated using the different image-guidance systems. In a), the K-wire is placed under conventional C-arm guidance, which requires frequent imaging and may result in a higher radiation dose for the surgeon. b) The X-ray image is augmented onto a live video stream, and the surgeon can update the X-ray image at his discretion. c) The use of an RGBD camera and DRR computed from a CBCT allows for the simultaneous visualization of the patient from different views. The surgeon can choose which views should be displayed, which will be updated using live RGBD information.

camera and Cone-Beam CT (CBCT) volumes, which allows the simultaneous visualization of the patient and medical data from multiple arbitrary views. As the RGBD camera is rigidly mounted to the X-ray detector, the X-ray source can be positioned under the surgical table as done during conventional image-guided surgery.

To calibrate the RGBD information and the CBCT volume, we simultaneously acquire the CBCT and the surface information using the RGBD camera of an arbitrary object. We extract the surface from the CBCT by simple thresholding, and reconstruct the surface observed by the RGBD camera as described in [21], resulting in a smooth and precise surface mesh. The calibration is obtained by means of surface matching [15].

After the calibration is obtained, the CBCT and patient's surface scan are acquired. This data is fused into a mixed reality scene, in which the patient's surface, DRR from CBCT, and live RGBD data (e.g. hand or tool) are visualized. The surgeon can now define multiple arbitrary views of the fused DRR and RGBD data. The system allows perspectives that are usually not possible using conventional X-ray imaging, as the free moving space is limited by the patient, surgical table or OR setup. The live RGBD data provides an intuitive understanding of the relation of CBCT volume, patient's surface, surgeon's hand, and medical tools.

F.2.2 Evaluation Method

During the usability study, we evaluate the performance achieved using each system. Our hypothesis states that the mixed reality visualizations improve the surgical efficiency. Our data cannot be assumed to be of normal distribution, but are ordinal. Using Friedman's ANOVA [4], we test if the differences in observations are coincidental or statistically significant. Additionally, we need to test whether the individual systems yield a significant difference in term of the surgical efficiency. As a normal distribution of our data cannot be assumed, these post-hoc tests are performed using the Wilcoxon signed-rank tests with Bonferroni correction [35].

Surgical Efficiency Measure Together with our clinical partners we identified following measures to express the surgical efficiency. First, the duration of each K-wire placement is of importance. During hip surgeries, this process is often the most time consuming and is followed by a relatively quick drilling step and screw placement. Surgical navigation systems often do not yield the advantage of reducing the overall OR time. Next, the number or X-ray images and cumulative area dose product is of importance both to the patient and surgeon. During conventional C-arm guided placement, a large number of X-ray images is acquired during the planning and propagation of the K-wire. One of our systems will acquire a pre-incision CBCT, for which we will include the dose into our statistics. Finally, the error is defined by the medical need of the K-wire remaining in the superior pubic ramus. We will compute the average distance between the ideal path, which is the center line of bone phantom, and the placement of it. However, as all study participants are trained surgeons, we do not expect that any significant improvement will be possible.

Surgical Task Load The workload is measured using a standardized questionnaire, namely the Surgical Task Load Index (SURG-TLX) [36]. This test is designed to evaluate the mental demands, physical demands, temporal demands, task complexity, situational stress, and distractions during surgical interventions. It is specifically designed and validated to analyze implications for categorizing the difficulty of certain procedures, and the implementation of new technology in the operating room.

F.3 Experiments

The K-wire placement through the superior pubic ramus (acetabulum arc) is a complex and cumbersome procedure, which is performed frequently and in case of an undislocated fracture usually minimally invasive [17]. In our experiments we mimicked this scenario by designing adequate radiopaque phantoms. The surgeons each performed three K-wire placements using the image-guidance systems in a randomized order.

Phantom Design The superior pubic ramus is a thin tubular bone with an diameter around 10 mm. In case of an undislocated fracture, a 2.8 mm thin K-wire needs to be placed through a narrow safe zone [26, 2]. Later, a 7.3 mm cannulated screw is inserted [28]. Our phantom was created out of methylene bisphenyl diisocyanate (MDI) foam, which is stiff, lightweight and not radiopaque. The bone phantom was created out of an thin aluminium mesh filled with MDI. The begin and end of the bone were marked with a rubber radiopaque ring. Therefore, the bone phantom is very similar to the superior pubic ramus in terms of haptic feedback during K-wire placement, as the K-wire will easily exit the bone without significant resistance. The orientation of the bone within the phantom was randomized and phantoms were not reused for other experiments.

Experimental Setup and Design In all our experiments, we use a CBCT enabled C-arm, SIEMENS ARCADIS Orbic 3D from Siemens Healthcare GmbH, which automatically computes the cumulative area dose for the patient for each imaging session. The second and third system use an optical video camera, Manta G-125C, from Allied Vision Technologies, or an RGBD camera, Intel RealSense Camera (F200), Intel Corporation, respectively.

Each surgeon was asked to perform three independent K-wire placements using the different imaging modalities. The order of the modalities was randomized, but for simplicity we will refer to the first (S1), second (S2) and third system (S3) in the order presented in Sec. F.2.1. Fig. F.3 shows the experimental setup for the 3D RGB and DRR visualization system (S3). Using a 2.8 mm K-wire, the surgeons identified the entry point on the phantom's surface, drilled towards the begin of the bone phantom, and passed through the tubular bone structure. When finished, the K-wire was removed from the drill and a CBCT was acquired to measure the error of the K-wire placement post-operatively.

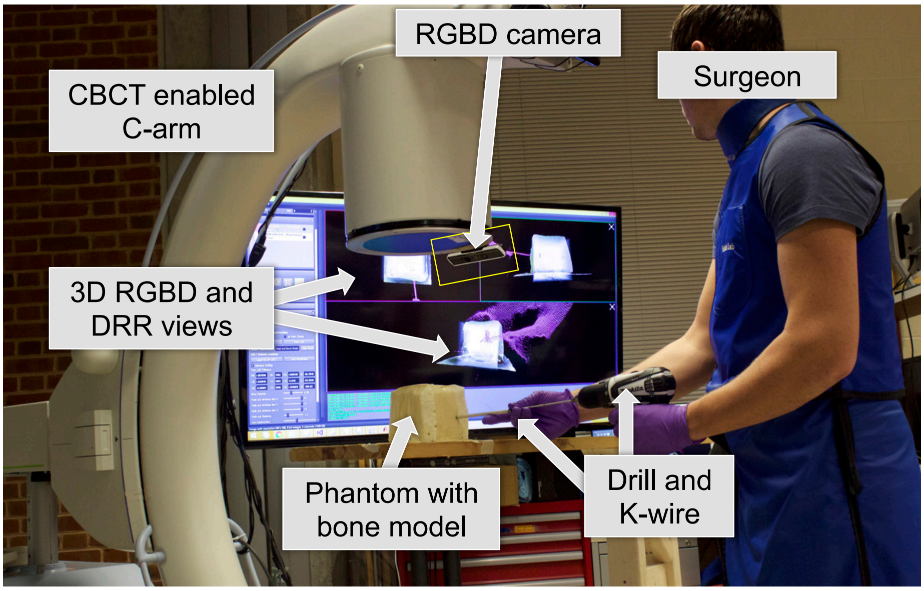

RGBD camera
Surgeon
CBCT enabled C-arm
3D RGBD and DRR views
Phantom with bone model
Drill and K-wire

Fig. F.3: During the experiments the surgeons drilled a K-wire into a phantom. This figure shows the experimental setup during a procedure guided by the 3D RGBD and DRR visualization (S3). The RGBD camera is mounted on the C-arm X-ray detector, and the surgeon is watching the live 3D RGBD and DRR views on the monitor while drilling into the phantom containing the bone model.

Table F.1: This table presents all observed values for each study participant and system used. For the RGBD and DRR visualization a CBCT was acquired, which is included in the dose measurement, but not in the number of X-ray images acquired.

	Participants						
	1	2	3	4	5	6	7
System 1: Conventional C-Arm							
Time (sec)	937	686	617	464	636	388	432
X-ray images	80	47	44	33	32	21	29
Dose (cGycm^2)	7.68	1.73	3.54	4.38	5.62	2.69	5.38
Error (mm)	3.08	7.88	11.43	3.01	1.87	2.27	2.72
Task Load	76	25.67	41.67	17.67	53.33	19.33	70.67
System 2: RGB and X-Ray Visualization							
Time (sec)	360	431	521	295	436	691	768
X-ray images	19	13	20	13	18	20	30
Dose (cGycm^2)	3.07	1.3	1.57	1.92	1.42	2.38	5.56
Error (mm)	7.92	2.69	3.85	4.23	4.88	3.44	1.74
Task Load	60.33	10	20	21.67	26	22.33	62.33
System 3: RGBD and DRR Visualization							
Time (sec)	182	180	380	181	190	254	339
X-ray images	1	2	2	2	2	3	3
Dose (cGycm^2)	1.76	1.9	1.48	1.44	1.55	1.47	1.59
Error (mm)	7.38	6.39	8.45	6.53	1.39	2.31	3.48
Task Load	20.33	5	24.33	23	11.33	8.67	30.33

F.4 Results

We observed a total of 21 minimally invasive K-wire placements using different image-guidance systems. Table F.1 present the observed time in seconds, number of acquired X-ray images, cumulative area dose product (dose) in cGycm^2, error relative to the ideal path in mm, and surgical task load index for each participant and system used. Note that the task load is a accumulative scale, for which the score of 5 and 100 represents the lowest and highest possible load, respectively.

The aggregated observations are presented in table F.2. When comparing the use of a conventional C-arm to the use of a mixed reality system, a clear tendency towards a decreased operation time, lower number of X-ray images acquired, reduced dose and reduced task load can be observed, as illustrated in Fig. F.4. The measure of the dose includes the acquisition of the CBCT volume required for the RGBD and DRR visualization (system 3). The accuracy does not improve, as it is already in an acceptable range.

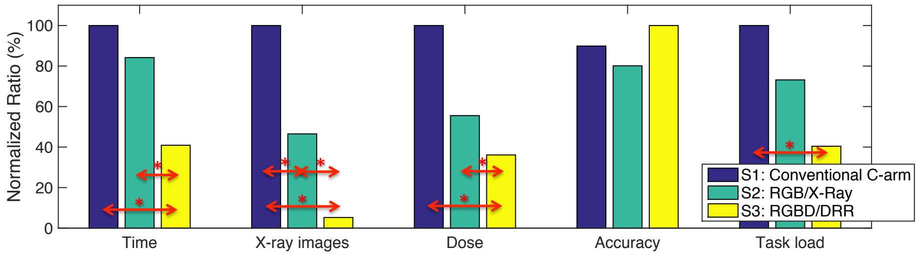


Fig. F.4: This plot illustrates duration of the intervention, number of X-ray images taken, radiation dose, K-wire placement error and surgical task load, where each bar shows the accumulated values using one of the systems (conventional X-ray, RGB/X-ray fusion, or RBGD/DRR). Each measure is normalized relative to the maximum value observed. The $\star$ symbols indicate significant differences.

Table F.2: The accumulated values and standard deviations for the observations (table F.1) are presented in this table.

	S1: C-Arm	S2: RGB/X-ray	S3: RGBD/DRR
Time (sec)	594 ±188	500 ±172	243 ±84
X-ray images	40.86 ±19.38	19.00 ±5.72	2.14 ±0.69
Dose ($cGycm^2$)	4.43 ±2.00	2.46 ±1.50	1.60 ±0.17
Error (mm)	4.61 ±3.62	4.11 ±1.97	5.13 ±2.72
Task Load	43.48 ±24.03	31.81 ±20.76	17.57 ±9.33

F.4.1 Statistical Evaluation

Statistical tests were performed to study the changes of the surgical efficiency measures. Significance is achieved for p-values lower than 0.05, indicating that the chance of the change being coincidentally observed is less than 5% [23]. A Friedman test was calculated to compare each measure as a normal distribution of the data could not be assumed. We found a significant difference in time ($\chi^2(3) = 11.14$, $p < 0.01$), number of X-ray images ($\chi^2(3) = 12.29$, $p < 0.01$), and radiation dose ($\chi^2(3) = 6.00$, $p < 0.05$) depending on the kind of assistance that was provided to the subjects. The post-hoc tests were computed using the Wilcoxon signed-rank tests with Bonferroni correction.

Time: The tests show significant differences between the first system (S1: Conventional C-Arm) and the third system (S3: RGBD and DRR Visualization) ($Z = -2.366$, $p < 0.05$), and significant differences between second system (S2: RGB and X-ray Visualization) and S3 ($Z = -2.366$, $p < 0.05$). This indicates that the 3D placement of the K-wire is best supported with a multi-view 3D visualization.

X-ray Images: All combinations of S1, S2 and S3 show a significant reduction of the number of X-ray images acquired: S1 to S2 ($Z = -2.117$, $p < 0.05$), S2 to S3 ($Z = -2.375$, $p < 0.05$), and S1 to S3 ($Z = -2.366$, $p < 0.05$).

Radiation Dose: Although we have included the dose caused by the CBCT in S3, the tests show that the intervention using the conventional C-arm causes a significantly higher cumulative area dose product: S1 to S2 ($Z = -2.197$, $p < 0.05$), S1 to S3 ($Z = -2.197$, $p < 0.05$). However, the dose difference between S2 and S3 is not significant.

Error: No significant difference in error between based on the use of different systems can be observed. Therefore, the reported changes in accuracies are most likely coincidental.

Surgical Task Load Index: Similarly to the changes of the duration of the intervention, the reduction of task load evaluated using the SURG-TLX is only significant between S1 and S3 ($Z = -2.197$, $p < 0.05$).

In conclusion, S3 yields better results in terms of all observed surgical efficiency measures except for the accuracy, for which the difference is not statistically significant. Even though S3 is not a fully developed product, our usability study indicates that there are clear advantages over the conventional C-arm system when guiding K-wire placement.

F.5 Discussion and Conclusion

In this paper we presented a usability study using three different mixed reality visualization systems to perform K-wire placement into the superior pubic ramus. This procedure was chosen because of the high clinical relevance, frequent prevalence, and the especially challenging minimal invasive surgical technique.

Our attention was focused on the usability and clinical impact of the three different visualization systems. For that reason we were not only interested in the quality of a procedure (e.g. accuracy), but also in the workload and frustration that the surgeons experienced while using the different systems. We observed the 21 interventions performed by 7 surgeons, and used the Surgical TLX to evaluate the task load.

Our results show that the 3D visualization yields the most benefit in terms of surgical duration, number of X-ray images taken, overall radiation dose and surgical workload. This is despite the fact, that the mixed reality visualizations currently do not provide an augmentation of a tracked tool. The conventional C-arm constitute the system yielding the poorest results, indicating a high potential for improvements to the currently used image-guidance systems. In all scenarios, the surgeons placed the K-wire within clinically relevant tolerance. The change in accuracy of the placed K-wire is not significant, which shows that all three systems provide sufficient support in terms of placement quality.

This study also showed the clear necessity to continue research and development of the mixed reality systems. For instance, movement of the C-arm or surgical table may lead to loss of tracking, which results in an outdated mixed reality visualization. However, in a clinical scenario, the failure of the mixed reality system is immediately visible and the surgeon can continue using the conventional X-ray imaging capabilities.

In our evaluation we have not take the learning curve under consideration as we frequently observed that surgeons unfamiliar to the mixed reality system adopted very quickly. Perhaps an initial training phase would further emphasize the advantages of the augmentations.

Future studies will include other complex K-wire placement procedures, such as performed in case of an ilio-sacral fracture, or pedicle screw placement. We will attempt to include even more surgeons during the next studies, which will allow for a more detailed statistically analysis. Additionally, we will investigate the usability for other procedures, such as Jamshidi needle placement or general needle biopsies.

Our usability study showed that mixed reality systems have great potential to increase surgical efficiency, and should be in the focus of research on computer assisted interventions. The integration of better visualization techniques, or tool tracking and identification may yield more opportunities in assisting surgeons to conduct interventions more efficiently and reduce the task load.

Acknowledgments

The authors want to thank Wolfgang Wein and his team from ImFusion GmbH, Munich for the opportunity of using the ImFusion Suite, and Gerhard Kleinzig and Sebastian Vogt from SIEMENS for their support and making a SIEMENS ARCADIS Orbic 3D available for this research. The authors want to especially thank all surgeons for their participation and valuable feedback.

Conflict of interest: The authors declare that they have no conflict of interest.
Ethical approval: For this type of study formal consent is not required.
Informed consent: This articles does not contain patient data.

References

[1] Bronek M Boszczyk, Michael Bierschneider, Stephanie Panzer, Werner Panzer, Roger Harstall, Katharina Schmid, and Hans Jaksche. “Fluoroscopic radiation exposure of the kyphoplasty patient”. In: *European Spine Journal* 15.3 (2006), pp. 347–355.

[2] ML Chip Routt, Peter T Simonian, and Leslie Grujic. “The retrograde medullary superior public ramus screw for the treatment of anterior pelvic ring disruptions: a new technique”. In: *Journal of orthopaedic trauma* 9.1 (1995), pp. 35–44.

[3] Benoit Diotte, Pascal Fallavollita, Lejing Wang, Simon Weidert, Peter-Helmut Thaller, Ekkehard Euler, and Nassir Navab. “Radiation-free drill guidance in interlocking of intramedullary nails”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 18–25.

[4] Milton Friedman. “A comparison of alternative tests of significance for the problem of m rankings”. In: *The Annals of Mathematical Statistics* 11.1 (1940), pp. 86–92.

[5] Spencer B Gay, Christopher Sistrom, Gwo-Jaw Wang, David A Kahler, Thomas Boman, Nancy McHugh, and Henry T Goitz. “Percutaneous screw fixation of acetabular fractures with CT guidance: preliminary results of a new technique.” In: *AJR. American journal of roentgenology* 158.4 (1992), pp. 819–822.

[6] Florian T Gebhard, Michael D Kraus, Eugen Schneider, Ulrich C Liener, Lothar Kinzl, and Markus Arand. “Does computer-assisted spine surgery reduce intraoperative radiation doses?” In: *Spine* 31.17 (2006), pp. 2024–2027.

[7] Florian Gras, Ivan Marintschev, Kajetan Klos, Thomas Mückley, Gunther O Hofmann, and David M Kahler. “Screw placement for acetabular fractures: which navigation modality (2-dimensional vs. 3-dimensional) should be used? An experimental study”. In: *Journal of orthopaedic trauma* 26.8 (2012), pp. 466–473.

[8] Pierre Guy, Mohammad Al-Otaibi, Edward J Harvey, and Nader Helmy. “The ”safe zone” for extra-articular screw placement during intra-pelvic acetabular surgery”. In: *Journal of orthopaedic trauma* 24.5 (2010), pp. 279–283.

[9] Severine Habert, Jose Gardiazabal, Pascal Fallavollita, and Nassir Navab. “RGBDX: first design and experimental validation of a mirror-based RGBD Xray imaging system”. In: *International Symposium on Mixed and Augmented Reality (ISMAR)*. Fukuoka, Japan, Oct. 2015.

[10] Severine Habert, Ma Meng, Wadim Kehl, Xiang Wang, Federico Tombari, Pascal Fallavollita, and Nassir Navab. “Augmenting mobile C-arm fluoroscopes via Stereo-RGBD sensors for multimodal visualization”. In: *International Symposium on Mixed and Augmented Reality (ISMAR)*. Fukuoka, Japan, Oct. 2015.

[11] Roger Härtl, Khai Sing Lam, Jeffrey Wang, Andreas Korge, Frank Kandziora, and Laurent Audigé. "Worldwide survey on the use of navigation in spine surgery". In: *World neurosurgery* 79.1 (2013), pp. 162–172.

[12] T Hüfner, T Stübig, T Gösling, D Kendoff, J Geerling, and C Krettek. "Cost-benefit analysis of intraoperative 3D imaging". In: *Der Unfallchirurg* 110.1 (2007), pp. 14–21.

[13] Michael Kraus, Stephanie von dem Berge, Hendrik Schoell, Gert Krischak, and Florian Gebhard. "Integration of fluoroscopy-based guidance in orthopaedic trauma surgery: A prospective cohort study". In: *Injury* 44.11 (2013), pp. 1486–1492.

[14] Michael Kraus, Julia Weiskopf, Jens Dreyhaupt, Gert Krischak, and Florian Gebhard. "Computer-Aided Surgery Does Not Increase the Accuracy of Dorsal Pedicle Screw Placement in the Thoracic and Lumbar Spine: A Retrospective Analysis of 2,003 Pedicle Screws in a Level I Trauma Center". In: *Global spine journal* 5.2 (2015), p. 93.

[15] Sing Chun Lee, Bernhard Fuerst, Javad Fotouhi, Marius Fischer, Greg Osgood, and Nassir Navab. "Calibration of RGBD Camera and Cone-Beam CT for 3D Intra-operative Mixed Reality Visualization". Manuscript submitted for publication. Draft attached for reviewers. 2016.

[16] Roberto Londei, Marco Esposito, Benoit Diotte, Simon Weidert, Ekkehard Euler, Peter Thaller, Nassir Navab, and Pascal Fallavollita. "Intraoperative augmented reality in distal locking". In: *International journal of computer assisted radiology and surgery* (2015), pp. 1–9.

[17] Dana C Mears. "Surgical treatment of acetabular fractures in elderly patients with osteoporotic bone". In: *Journal of the American Academy of Orthopaedic Surgeons* 7.2 (1999), pp. 128–141.

[18] N. Navab, S.-M. Heining, and J. Traub. "Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications". In: *Medical Imaging, IEEE Transactions on* 29.7 (July 2010), pp. 1412–1423. ISSN: 0278-0062.

[19] Nassir Navab, Tobias Blum, Lejing Wang, Aslı Okur, and Thomas Wendler. "First deployments of augmented reality in operating rooms". In: *Computer* 7 (2012), pp. 48–55.

[20] Nassir Navab, Matthias Mitschke, and Oliver Schütz. "Camera-augmented mobile C-arm (CAMC) application: 3D reconstruction using a low-cost mobile C-arm". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI'99*. Springer. 1999, pp. 688–697.

[21] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. "KinectFusion: Real-time dense surface mapping and tracking". In: *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. Oct. 2011, pp. 127–136. DOI: `10.1109/ISMAR.2011.6092378`.

[22] S Nicolau, PY Lee, HS Wu, M Huang, R Lukang, L Soler, and J Marescaux. "Fusion of C-arm X-ray image on video view to reduce radiation exposure and improve orthopedic surgery planning: first in-vivo evaluation". In: *Proceedings of Computer Assisted Radiology and Surgery.* Vol. 6. 2011, pp. 115–116.

[23] Regina Nuzzo. "Statistical errors". In: *Nature* 506.7487 (2014), pp. 150–152.

[24] Olivier Pauly, Benoit Diotte, Pascal Fallavollita, Simon Weidert, Ekkehard Euler, and Nassir Navab. "Machine learning-based augmented reality for improved surgical scene understanding". In: *Computerized Medical Imaging and Graphics* 41 (2015), pp. 55–60.

[25] S Reaungamornrat, Yoshito Otake, Ali Uneri, Sebastian Schafer, DJ Mirota, Sajendra Nithiananthan, J Webster Stayman, Gerhard Kleinszig, A Jay Khanna, Russell H Taylor, and Jeffrey H Siewerdsen. "An on-board surgical tracking and video augmentation system for C-arm image guidance". In: *International journal of computer assisted radiology and surgery* 7.5 (2012), pp. 647–665.

[26] ML Chip Routt Jr, Sean E Nork, and William J Mills. "Percutaneous fixation of pelvic ring disruptions." In: *Clinical orthopaedics and related research* 375 (2000), pp. 15–29.

[27] ML Chip Routt Jr, Peter T Simonian, and William J Mills. "Iliosacral screw fixation: early complications of the percutaneous technique". In: *Journal of orthopaedic trauma* 11.8 (1997), pp. 584–589.

[28] Daniel Schweitzer, Alejandro Zylberberg, Marcelo Córdova, and Jaime Gonzalez. "Closed reduction and iliosacral percutaneous fixation of unstable pelvic ring fractures". In: *Injury* 39.8 (2008), pp. 869–874.

[29] Adam J Starr, Charles M Reinert, and Alan L Jones. "Percutaneous fixation of the columns of the acetabulum: a new technique". In: *Journal of orthopaedic trauma* 12.1 (1998), pp. 51–58.

[30] AJ Starr, AL Jones, CM Reinert, and DS Borer. "Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the actabulum". In: *Injury* 32 (2001), pp. 45–50.

[31] Ulrich Stöckle, Klaus Schaser, and Benjamin König. "Image guidance in pelvic and acetabular surgery-expectations, success and limitations". In: *Injury* 38.4 (2007), pp. 450–462.

[32] Michael Synowitz and Juergen Kiwit. "Surgeon's radiation exposure during percutaneous vertebroplasty". In: *Journal of Neurosurgery: Spine* 4.2 (2006), pp. 106–109.

[33] Joerg Traub, Tim Hauke Heibel, Philipp Dressel, Sandro Michael Heining, Rainer Graumann, and Nassir Navab. "A multi-view opto-Xray imaging system". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007.* Springer, 2007, pp. 18–25.

[34] Roger Y Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses". In: *Robotics and Automation, IEEE Journal of* 3.4 (1987), pp. 323–344.

[35] Frank Wilcoxon. “Individual comparisons by ranking methods”. In: *Biometrics bulletin* (1945), pp. 80–83.

[36] Mark R Wilson, Jamie M Poolton, Neha Malhotra, Karen Ngo, Elizabeth Bright, and Rich SW Masters. “Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX)”. In: *World journal of surgery* 35.9 (2011), pp. 1961–1969.

APPENDIX G

Calibration of RGBD Camera and Cone-Beam CT for 3D Intraoperative Mixed Reality Visualization (IJCARS 2016)

Sing Chun Lee[1] * and Bernhard Fuerst[1,2] *, Javad Fotouhi[1], Marius Fischer[1,3], Greg Osgood[4] and Nassir Navab[1,2]

* S.C. Lee and B. Fuerst contributed equally to the work and should be regarded as joint first authors.

[1] Computer Aided Medical Procedures (CAMP), Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

[2] Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstrae 3, 85748 Garching b. München, Germany

[3] Trauma Surgery Department, Klinikum Innenstadt, LMU Munich, Germany

[4] Orthopaedic Trauma, Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, MD, US

Copyright © Springer International Publishing Switzerland 2015.

Contribution The calibration of Cone-Beam CT (CBCT) volumes and 3D optical imaging (RGBD) is the key prerequisite to enable interventional Augmented Reality (AR) for medical applications. This paper presents a clear methodology to perform the calibration, namely the extraction of surface information from both modalities, the initialization of the calibration using Fast Point Feature Histograms (FPFH), and the refinement of the data alignment using the Iterative Closest Point (ICP) algorithm. Finally, we present the calibration accuracy, show the repeatability, and provide application examples. A usability study using a fully calibrated system is presented in Appendix F.

CALIBRATION OF RGBD CAMERA AND CONE-BEAM CT FOR 3D INTRA-OPERATIVE MIXED REALITY VISUALIZATION (IJCARS 2016)

Abstract *Purpose:* This work proposes a novel algorithm to register Cone-Beam Computed Tomography (CBCT) volumes and 3D optical (RGBD) camera views. The co-registered real-time RGBD camera and CBCT imaging enable a novel augmented reality solution for orthopedic surgeries, which allows arbitrary views using Digitally Reconstructed Radiagraphs (DRRs) overlaid on the reconstructed patient's surface without the need to move the C-arm.
Methods: An RGBD camera is rigidly mounted on the C-arm near the detector. We introduce a calibration method based on the simultaneous reconstruction of the surface and the CBCT scan of an object. The transformation between the two coordinate spaces is recovered using Fast Point Feature Histograms (FPFH) descriptors and the Iterative Closest Point (ICP) algorithm.
Results: Several experiments are performed to assess the robustness and the accuracy of this method. Target Registration Error (TRE) is measured on multiple visual and radio-opaque landmarks to evaluate the accuracy of the registration. Mixed reality visualizations from arbitrary angles are also presented for simulated orthopedic surgeries.
Conclusion: To the best of our knowledge, this is the first calibration method which uses only tomographic and RGBD reconstructions. This means that the method does not impose a particular shape of the phantom. We demonstrate a marker-less calibration of CBCT volumes and 3D depth cameras, achieving reasonable registration accuracy. This design requires a one-time factory calibration, is self-contained, and could be integrated into existing mobile C-arms to provide real-time augmented reality views from arbitrary angles.

Keywords Augmented Reality, Cone-beam CT, C-arm, Intra-operative imaging, 3D-3D Calibration

G.1 Introduction

X-ray imaging is an important tool for percutaneous ilio-sacral and pedicle screw placements in spine surgeries. To avoid potential damages to soft tissues and the nervous system near the vertebra, and reduce muscle retraction, significant amount of fluoroscopic/X-ray images are acquired from multiple views during these interventions. Foreign body removal surgeries also require a high number of X-ray image acquisitions, as there are significant risks of inadequately performing the wound debridement [15]. Multiple attempts to remove them could lead to larger incisions, additional trauma, delay in healing, and worsened outcomes.

To place or remove a rigid object during minimally invasive image-guided orthopedic operations, the surgeon first locates the point of entry on the skin by acquiring multiple X-ray images from different views while having a tool for reference in the scene. The reference tool (e.g. needle, K-wire, drill, etc.) is used during the intervention to assist the surgeons with the mental alignment. An example of the entry point localization is illustrated in Fig. G.1.

An exemplary workflow involves the collection of a set of anteroposterior X-ray images in which the target anatomy and the drill are visible. Next, the direction of the medical instrument is corrected in corresponding lateral and oblique views, which may introduce small displacements in the anteroposterior

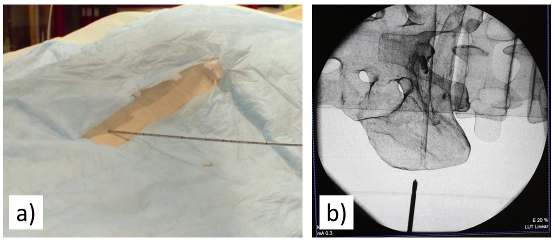


Fig. G.1: The first step during a minimally invasive orthopedic surgery is the entry point localization. The surgical site, containing patient and tools, is shown in (a). Panel (b) shows an X-ray image during entry point localization.

side. To ensure the accurate placement of the medical instrument, this procedure is repeated several times, and during each iteration the drill is traversed further through the tissue until the target is reached. Lastly, verification images are acquired, and depending on the performed placement the procedure may need to be repeated.

Delicate operations of these types have long surgical durations as they require the acquisition of many intra-operative images which implies frequent C-arm repositionings. This leads to a high surgical effort and major frustration for the surgical staff, long procedures, and high radiation exposure. This is in contrast to the need of shortening anesthesia durations for elderly patients due to post-operative complications. Thus, it is a challenge for the surgeon to operate accurately in a limited time and minimizing collateral damage to surrounding tissue.

External surgical navigation systems are used to provide the spatial relation among the anatomy in medical images, the patient's body in the operation room (OR), and the surgical tool. This information is used to avoid potential damage to surrounding tissue. Alternatively, additional sensors such as cameras are directly attached to the C-arm to perform tracking of the surgical tools, or aim at providing intuitive visualizations of the patient surface and the medical data.

Optical-based image-guided navigation systems were used to recover the spatial transformation between surgical tools and a 3D rotational X-ray enabled C-arm with sub-millimeter accuracy [8]. Significant reduction in radiation exposure was achieved by navigating the surgical tool together with a tracked C-arm with markers attached to the detector plane [9]. Navigation-assisted fluoroscopy in minimally invasive spine surgery with optical tracker for placing pedicle screws was evaluated in [10, 7]. Both publications reported a reduction in radiation exposure. However, no statistically significant changes in the time of surgery was found. There are two main problems associated with these systems: First, they increase the complexity of the surgery, require additional hardware, occupy significant amount of space and require line-of-sight between patient and hardware. Second, the surgeon performs the surgery by only observing a display, which does not provide visual feedback of actual patient and current deformation of tissue. Good outcomes highly depend on the surgeon's experience, confidence, and surgical skills.

Alternative systems are fully integrated within the absolutely necessary C-arm and range from systems with video cameras [11] to solutions with integrated tracking devices [16]. A video camera attached to a mobile C-arm was co-registered with the X-ray source in [11] to provide real-time augmentation of the surgical site. X-ray and optical centers are virtually aligned using an X-ray transparent mirror construction and calibrated by utilizing X-ray and optically visible markers. Next, a homography is estimated to warp the optical images to enable the augmentation with the undistorted X-ray image. This system was used in over 40 orthopedics and trauma procedures [12, 2], and the results indicated improvements in terms of optimal incision, avoidance of direct radiation exposure, and instrument axis alignment. The evaluation through assisted interlocking procedures showed that as long as depth information is not required, the system can assist the surgeon in accurately placing the tools inside the patient.

Similarly, [14] incorporated a video camera and performed an online calibration by placing a metal rod with known shape close to patient's bone. Cadaver studies for screw placement and femoral fractures on real patients reported 26-30% decrease in radiation exposure, but no statistically significant change in the procedure time were reported. In contrast to [11, 14], a color and depth camera was mounted on the mobile C-arm to replace the video camera in [5]. The 3D/2D mixed reality visualization is demonstrated using a X-ray image augmented on a patient surface offline. The main limitation of this work is due to 2D projective nature of the X-ray image, resulting in a physically wrong visualization as soon as the viewpoint is different from the X-ray source. In [6], the optical view from the viewpoint of the X-ray source has been synthesized using two RGBD cameras, enabling a system like [11], but without mirror construction.

A vision-based tracking system using natural features observed in the view of an optical camera attached to a mobile C-arm was suggested to enable the extension of the field of view of CBCT volumes with minimum radiation exposure [4]. Frame-to-frame registration results acquired from the optical camera were applied to CBCT sub-volumes by calibrating CBCT volumes with the optical camera in advance.

An intuitive visualization based on Digitally Reconstructed Radiographs (DRR) was proposed in [16], and addresses the limitations of conventional navigation systems by providing an augmented reality view of DRR, video and tracked tools. A vision-based navigation is performed by mounting stereo cameras on the C-arm near the detector. Patient and C-arm are then registered using a mixed set of visual and radio-opaque markers on a single calibration phantom. This system reached sub-millimeter tracking accuracy, and has the benefit of not using any external tracking system, thus remaining self-contained. The main complexity arises when visualization from different angles during augmentation requires moving and rotating the C-arm. To put this into perspective, even though the number of X-rays is decreased and the line-of-sight issue is minimized, the additional work load of re-positioning the C-arm is not eliminated.

The main problem associated with RGB cameras is the unknown depth in arbitrary scenes. RGBD cameras are sensing systems capable of acquiring RGB images and co-registered depth information, thus providing the means to a 3D visualization or marker-less tracking. A calibration of an RGBD camera to 2D X-ray images of C-arm was proposed in [20]. Registration is performed by computing the projection matrix between a 3D point cloud and corresponding 2D points on the X-ray image plane using a visual and radio-opaque planar

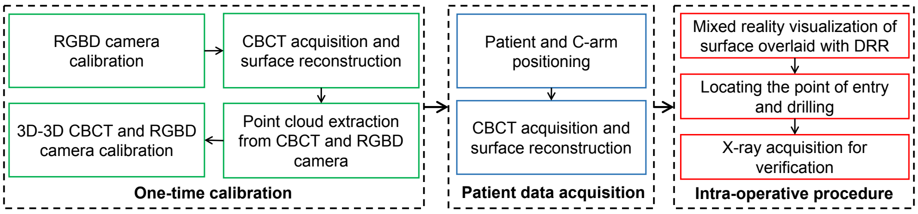

RGBD camera calibration
CBCT acquisition and surface reconstruction
3D-3D CBCT and RGBD camera calibration
Point cloud extraction from CBCT and RGBD camera
One-time calibration
Patient and C-arm positioning
CBCT acquisition and surface reconstruction
Patient data acquisition
Mixed reality visualization of surface overlaid with DRR
Locating the point of entry and drilling
X-ray acquisition for verification
Intra-operative procedure

Fig. G.2: The offline calibration of RGBD camera to CBCT origin is performed by introducing an arbitrary object into the common views of both devices. Before an intervention begins, CBCT and surface scans of the patient are acquired simultaneously. During the intervention, the fused visualization of patient's surface, surgeon's hands and tools, together with simulated X-ray images (DRR) are displayed to assist the surgeon.

phantom. This method reached a calibration error of 0.54 mm (RMS).

This work introduces a calibration technique for CBCT volumes and RGBD camera, and enables an intuitive 3D visualization which overlays both physical and anatomical information from arbitrary views. In contrast to the aforementioned technique in [16], where a 2D video is augmented with 2D DRRs, this technique takes the next step by proposing a full 3D-3D registration and enables the augmentation of a 3D optical view and simulated X-ray images from any arbitrary view. This system is capable of providing views which may be impossible to capture due to a limited free moving space of the C-arm, for instance, intra-operative transversal images. The proposed marker-less vision-based technique requires a one-time factory calibration as the depth camera and the X-ray source are rigidly mounted together, and achieves a calibration accuracy of 2.58 mm.

G.2 Method

The proposed technique uses an RGBD camera mounted on a mobile C-arm, and recovers a 3D rigid-body transformation from the RGBD surface point clouds to CBCT. The transformation is recovered using Iterative Closest Point (ICP) with a Fast Point Feature Histogram (FPFH) [18] for initialization. The general workflow is illustrated in Fig. G.2 and is comprised of an offline calibration, patient data acquisition and processing, and intra-operative 3D augmented reality visualization. The following sections describe the system setup (Sec. 2.1), calibration phantom characteristics (Sec. 2.2), transformation estimation (Sec. 2.3), and the augmented reality overlay (Sec. 2.4).

G.2.1 System setup

The system comprises a mobile C-arm, the SIEMENS ARCADIS Orbic 3D from Siemens Healthcare GmbH, and a close-range structured-light Intel RealSense F200 RGBD camera from Intel Corporation which better minimizes the light-power interference and ensures accuracy in shorter ranges compared to time-of-flight or stereo cameras. A structured-light RGBD camera provides reliable

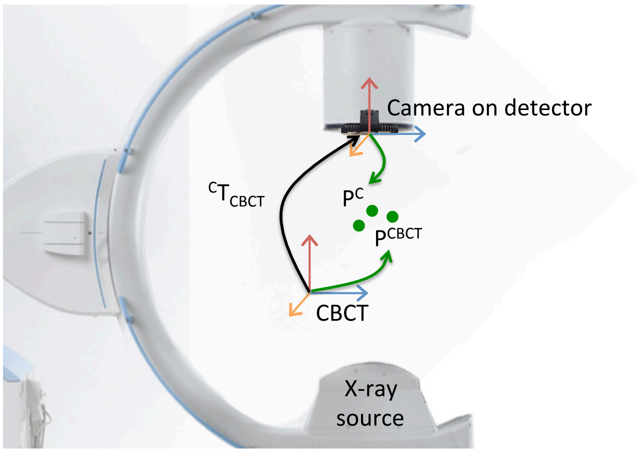


Fig. G.3: System setup: A depth camera is rigidly mounted on the detector, so that the field of view, and depth of view cover the CBCT volume. The transformation $^{c}\mathbf{T}_{CBCT}$ is recovered by registering the point clouds $\mathbf{P}^{c}$ and $\mathbf{P}^{CBCT}$ of the calibration phantom.

depth information by projecting patterned infrared lights onto the surface, and computes the depth information based on the pattern deformations. A typical time-of-flight camera, such as the Microsoft Kinect One (v2), requires additional warm up time of up to 20 minutes and depth distortion correction. In addition, the depth values highly depend on the color and shininess of the scene objects. On the other hand, conventional stereo cameras require textured surfaces for reliable triangulation, which are not suitable in this application [3].

The C-arm is connected via Ethernet to the computer for CBCT data transfer, and the RGBD camera is connected via powered USB 3.0 for real-time frame capturing. The RGBD camera is mounted rigidly near the detector, and its spatial position remains fixed with respect to CBCT's origin. After a one-time calibration, the patient is positioned on the surgical table under the C-arm guidance using the laser aiming guide attached to the C-arm. Thereafter, CBCT is acquired, and the surface is scanned using the RGBD camera simultaneously. The system setup is outlined in Fig. G.3.

G.2.2 Calibration phantom design, point cloud extraction and pre-processing

A planar checkerboard pattern is used to recover intrinsic parameters of the RGB and depth camera, and their spatial relation [21]. Depth camera intrinsics are used to reconstruct the surface in depth camera coordinates, and the intrinsics of the RGB camera together with their spatial transformation are used for reprojecting the color information onto the surface. For simplicity, we will refer to the calibrated RGB and depth camera as the RGBD camera.

A calibration phantom shown in Fig. G.4-a is introduced into the common view of the CBCT and the RGBD camera. Surface point clouds are then computed from both imaging modalities, and are used to estimate a 3D-3D rigid body transformation. The phantom is composed of three pipes and a cylindrical

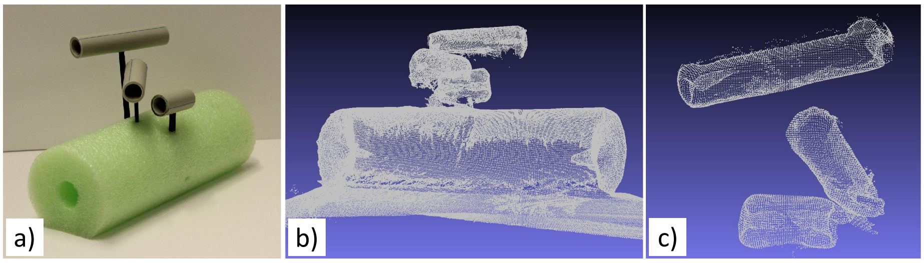


Fig. G.4: (a) The calibration phantom comprised 3 pipes in different heights, lengths, and orientations which represent an unambiguous 3D object. (b) Raw surface point cloud reconstruction from RGBD data, and (c) filtered surface points after the cylinder fitting.

foam base. Each pipe has a different length, and is positioned at diverse height and orientation to provide a unique rigid 3D-3D mapping between two coordinate spaces. Furthermore, the pipes have higher radiation absorption than the foam base, which allows a simple thresholding for point cloud segmentation. In contrast to shape angles or corners, the round surface of the phantom provides a more stable depth information with lower corner reflection effect.

After positioning the calibration phantom at the center of the C-arm, CBCT data is acquired. While the C-arm is rotating, KinectFusion [13] is used to compute the surface reconstruction in RGBD camera space[1]. Raw point clouds $\mathbf{P}^{cr}$ in Fig. G.4-b are subjected to least square cylinder fitting for the tubes α, β, and γ with known radius $r = \{r_\alpha, r_\beta, r_\gamma\}$ and height $h = \{h_\alpha, h_\beta, h_\gamma\}$. Cylinder fitting is performed by minimizing the model function $F(.)$ using M-estimator SAmple and Consensus (MSAC) [19]:

$$E(\mathbf{c},\mathbf{u}) = \min_{\substack{\mathbf{c},\mathbf{u}\in\mathbb{R} \\ |\mathbf{u}\cdot(\mathbf{p}-\mathbf{c})|\le h/2}} \sum_{i\in S} F(\mathbf{p}_i^{cr},\mathbf{c},\mathbf{u})^2 \tag{G.1}$$

where $\mathbf{c} \in \mathbb{R}^3$ is the center, $\mathbf{u} \in \mathbb{R}^3$ is the orientation of the principle axis of the cylinder, S_j is the sampling set at the j-th MSAC iterations, and $F(\mathbf{p}_i^{cr},\mathbf{c},\mathbf{u})$ is the residual defined as below where I is the identity matrix:

$$F(\mathbf{p}_i^{cr},\mathbf{c},\mathbf{u})^2 = (\mathbf{p}_i^{cr} - \mathbf{c})^\intercal (I - \mathbf{u}\mathbf{u}^\intercal)(\mathbf{p}_i^{cr} - \mathbf{c}) - r^2 \tag{G.2}$$

Next, the pre-processed surface points are computed as the inliers to the parametric model $E(\mathbf{c},\mathbf{u})$ with respect to a distance threshold d:

$$\mathbf{P}^c = \{\mathbf{P}^{cr}(\mu) : \|\mathbf{P}^{cr}(\mu) - E(\mathbf{c},\mathbf{u})\| < d\} \tag{G.3}$$

The filtered surface points are shown in Fig. G.4-c.

Filtering the CBCT data is performed in 4 steps. First, due to different absorption coefficients of the foam base and the pipes, the intensities are thresholded manually in CBCT data to filter out the foam (Fig. G.5-a). The remaining points are transformed into mesh grids (Fig. G.5-b) using fast greedy triangulation algorithm [1], and an ambient occlusion value is assigned to each vertex

---

[1] open source code available at `https://github.com/Nerei/kinfu_remake`

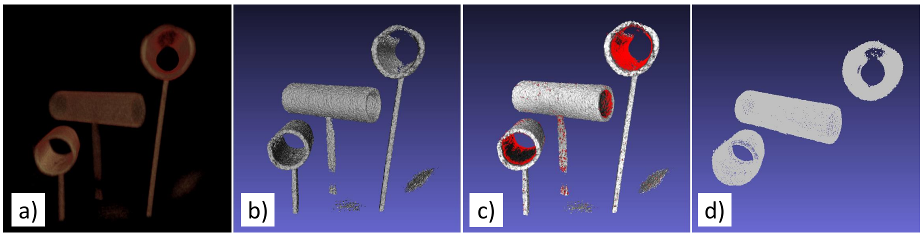


Fig. G.5: The steps to acquire the smooth surface point cloud from CBCT for calibration start with the CBCT volume acquisition (a). The CBCT surface segmentation is performed by thresholding (b). The ambient quality of CBCT points is assigned to each point (c), where white is a higher score (outer surface), and red is a lower scores (inner surface). Based on the ambient score, the outer CBCT surface is defined (d).

(Fig. G.5-c). This score defines the amount which each point in the scene is exposed to the ambient light. Higher values are assigned to outer surfaces, and lower values are assigned to interior of the tubes. Lastly, the outer surface is segmented by thresholding the scores of the vertices (Fig. G.5-d). The two point clouds $\mathbf{P}^c$ and $\mathbf{P}^{CBCT}$ are used in Sec. 2.3 for calibration of CBCT and RGBD data.

G.2.3 Calibration of C-arm and the RGBD camera

The RGBD camera is mounted rigidly on the detector of the C-arm as shown in Fig. G.3, therefore the transformation between them remains fixed and could be modeled as a rigid transformation ${}^{c}\mathbf{T}_{CBCT} \in SE(3)$, where $SE(3)$ is the special Euclidean group.

To register $\mathbf{P}^c$ and $\mathbf{P}^{CBCT}$, ICP is used with an initial guess acquired from a SAmple Consensus Initial Alignment (SAC-IA) with FPFH [18]. FPFH provides a fast and reliable initialization for the two point clouds. To compute the feature histograms (implemented in the Point Cloud Library [17]), the normal $\mathbf{n}_i^S$ is estimated for each point in the CBCT and camera space $\mathbf{P}^S$, where $S \in \{CBCT, c\}$. Next, for each point, a neighborhood Ω_i is defined with respect to a given radius. For every point pair $\{\mathbf{p}_j^S, \mathbf{p}_k^S\}$ inside Ω_i, a point feature histogram (PFH) is computed as following:

$$PFH_i^S = H\left(\{\phi_{jk}^S\}\right) \tag{G.4}$$

where $\boldsymbol{\phi}_{jk}^S$ is the set of angular variations and $H(.)$ is the histogram of these features in Ω_i [18]. FPFH is then computed as a weighted PFH in Ω_i, where w_j is the Euclidean distance between point pairs $\{\mathbf{p}_i^S, \mathbf{p}_j^S\}$:

$$FPFH_i^S = PFH_i^S + \frac{1}{k}\sum_{j\in\Omega}\frac{1}{w_j}PFH_j^S \tag{G.5}$$

The acquired feature histograms $\mathbf{FPFH}^S = \{FPFH_i^S\}_{i=1}^n$ are used in SAC-IA to register the two point clouds iteratively. The transformation $\mathbf{T}^{(i)}$ acquired from registration is used as an automatic initialization, and ICP is used to

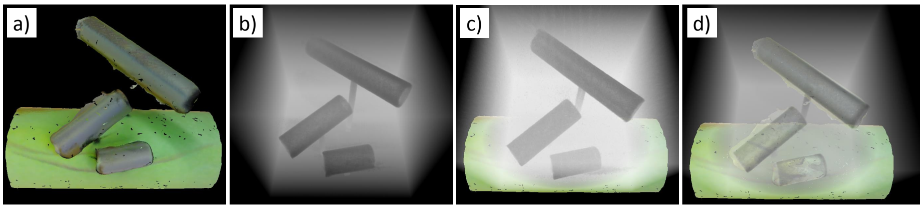


Fig. G.6: The surface reconstruction (a) and the DRR rendering of the calibration phantom (b) are overlaid using the opacities 0.4 and 0.7 in (c) and (d), respectively. This fused view may assist surgeons during orthopedic interventions.

further refine the registration by minimizing the following cost function:

$$^{c}\mathbf{T}_{CBCT} = \min_{\mathbf{T}} \sum_{i,j} \|\mathbf{p}_i^c - \mathbf{T}\mathbf{p}_j^{CBCT}\|_2^2 \tag{G.6}$$

G.2.4 Mixed reality visualization of DRRs overlaid on the patient's surface

Using the calibration results from Sec. 2.3, an augmented reality overlay comprised of DRRs rendered from CBCT volumes and the patient's surface is provided. Moreover, by changing the virtual camera pose after the augmentation, the system provides arbitrary perspectives. Finally, background subtraction in the RGBD view enables a real-time visualization of moving point clouds (surgeon's hands and tools). The output is a mixed reality visualization of anatomical data, patient's surface, and surgeon's hands and tools that are useful for orthopedic interventions. The surface reconstruction of the calibration phantom, DRR, and overlays with different opacities are shown in Fig. G.6.

G.3 Experimental Validation and Results

The robustness of the technique is first assessed by repeatedly performing the calibration using the phantom in Fig. G.6. For each test, the calibration phantom is placed differently such that all pipes are visible in the RGBD view to ensure a full surface reconstruction. The surface reconstruction using an Intel RealSense RGBD camera on the detector is compared with a Microsoft Kinect 360 (v1) camera mounted on the gantry (due to the depth range limitation, the Kinect camera needs to be placed at least 80 cm away from the object). The standard deviation (SD) of the calibration results in x, y, z directions, and rotation Euler angles α, β, γ are shown in table G.1. Results show low deviations, and no statistically significant difference was found between the two cameras. Arbitrary objects such as a stone or the spine phantom in Fig. G.8-a are also used directly as the calibration phantom, and the estimated transformation parameters varied between 0.7% to 2% compared to the original calibration phantom.

Point clouds acquired from the RGBD camera and CBCT are subjected to downsampling using voxel grid filter with different grid sizes. The SD values

Table G.1: The results of the repeated calibration (5 tests) in term of SD of Euler angles α, β, γ of ${}^{c}\mathbf{R}_{CBCT}$, the x, y, z components of ${}^{c}\mathbf{t}_{CBCT}$ and $\|{}^{c}\mathbf{t}_{CBCT}\|_2$, where ${}^{c}\mathbf{R}_{CBCT}$ and ${}^{c}\mathbf{t}_{CBCT}$ are the rotational and translational components of the estimated transformation ${}^{c}\mathbf{T}_{CBCT}$.

Error	α (rad)	β (rad)	γ (rad)	x (mm)	y (mm)	z (mm)	$\|{}^{c}\mathbf{t}_{CBCT}\|_2$
F200	0.0045	0.0059	0.016	0.89	0.29	0.81	0.84
Kinect	0.0047	0.0027	0.0079	0.31	0.74	0.35	0.46

Table G.2: The original data acquired from the RGBD camera and CBCT contains 25226 and 94547 points, respectively. The Euclidean distances of the transformation parameters estimated from the downsampled and original data are compared. Grid size defines the size of the voxel grid for downsampling the point clouds.

Grid Size (mm)	# of pts (CBCT)	# of pts (Depth)	α (rad)	β (rad)	γ (rad)	x (mm)	y (mm)	z (mm)
0.5	51831	21014	0.020	0.0025	0.014	0.22	0.23	0.56
1.0	18684	16221	0.0079	0.0018	0.0027	0.47	0.55	0.29
1.5	9016	8536	0.0018	0.0015	0.0041	0.33	0.25	0.57
2.0	5238	5183	0.0014	0.0025	0.0038	0.48	0.18	0.16

for the transformation parameters are reported in Table G.2. Larger number of points representing the data results in better initialization from the FPFH. The ICP estimation shows small variations in transformation parameters using the downsampled data.

Bilateral filtering is used to remove the noise during the surface reconstruction. Moreover, FPFH and ICP are both tolerant to outliers, thus small amount of noise are discarded during the transformation estimation. Due to the significant difference of the attenuation coefficient of the calibration phantom and the background noise, thresholding the CBCT data eliminates the background noise. Therefore, the calibration method is robust to small of amount of noise and outliers.

To evaluate the accuracy of the calibration, the Target Registration Error (TRE) is computed using the phantom shown in Fig. G.7-a. The phantom contains visual and radio-opaque landmarks and each landmark is selected manually. TRE is computed as the euclidean distance between a visual landmark after applying the transformation and the corresponding radio-opaque landmark. Since the landmarks are not co-linear nor co-planar, small orientational errors are also reflected in TRE. The accuracy test is repeated three times using eight landmarks. The resulting errors are shown in Table G.3. The main misalignment arises from the error in the direction perpendicular to the RGBD

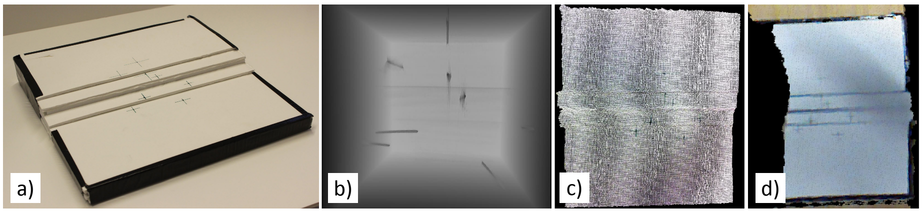


Fig. G.7: The calibration accuracy is evaluated using a phantom (a) with radiopaque markers which are visible in the RGBD view. (b) Test phantom in CBCT (DRR rendering), (c) point clouds of the phantom obtained using the Intel RealSense F200 camera, and (d) point clouds of the phantom from the Kinect camera.

Table G.3: The results of the repeated accuracy tests are shown as TRE, where δx, δy, δz, and $\|\delta\|_2$ are the Euclidean distances. Values are reported as mean ± SD.

TRE 1	δx	δy	δz	$\|\delta\|_2$
F200	1.26 ± 0.73	1.94 ± 1.50	0.98 ± 0.95	2.91 ± 1.10
Kinect	0.58 ± 0.56	2.87 ± 1.97	5.61 ± 1.72	6.54 ± 2.04

TRE 2	δx	δy	δz	$\|\delta\|_2$
F200	0.72 ± 0.78	2.46 ± 1.12	1.12 ± 0.87	2.91 ± 1.37
Kinect	1.09 ± 0.49	3.19 ± 0.83	7.30 ± 1.19	8.11 ± 1.02

TRE 3	δx	δy	δz	$\|\delta\|_2$
F200	0.83 ± 0.57	0.72 ± 0.52	1.40 ± 1.00	1.92 ± 0.98
Kinect	1.40 ± 0.65	1.97 ± 1.33	7.11 ± 1.33	7.60 ± 1.55

camera (due to poor depth quality). In each test the phantom is placed at a different pose by 90 degrees rotation. Different accuracy among the three tests are mainly due to changes of the overall distance of the landmarks to the camera resulting from their non-uniform distribution.

The Intel RealSense camera achieves an average TRE of 2.58 mm, where the Microsoft Kinect 360 (v1) achieves 7.42 mm. Due to poor depth quality, the errors along the z axis are more significant than the other two directions in both cameras.

G.3.1 Mixed reality visualization

An example for inserting a guide wire into a spine phantom is shown in Fig. G.8. This system could also be used for fast foreign body (shrapnel) removal. The simulated scenario for this application is shown in Fig. G.9.

G.4 Discussion and Conclusion

This paper proposes a novel methodology to calibrate an RGBD camera rigidly mounted on a C-arm and a CBCT volume. This combination enables intuitive

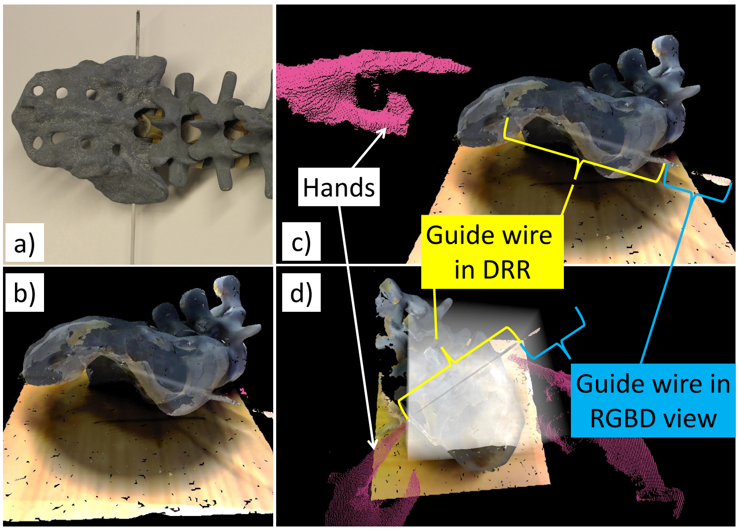

a)
b)
c)
d)
Hands
Guide wire in DRR
Guide wire in RGBD view

Fig. G.8: (a) The spine phantom for visualization experiment, (b) the overlay of DRR and phantom surface, (c) the overlay with a point cloud of a hand, and (d) the overlay from a different view. Depending on the structure of interest, the user can change the windowing (level and width of intensity range) and transparency of the DRR dynamically. This is demonstrated by the different renderings in (b) and (d).

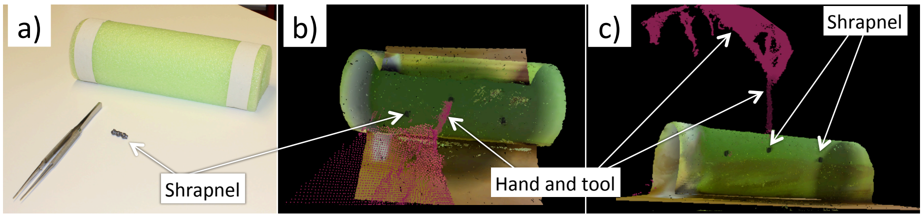

a)
b)
c)
Shrapnel
Hand and tool
Shrapnel

Fig. G.9: A shrapnel phantom (a) is designed to perform foreign body removal under mixed reality visualization, shown in (b) and (c).

intra-operative augmented reality visualization. The accuracy and robustness of the algorithm is evaluated using several tests. Although the spatial resolution of the RGBD cameras in depth is poor (approximately ±5% of the depth), a reasonable registration accuracy of 2.58 mm is achieved. This paper has presented two applications with high clinical impact. First, image-guided drilling for cannulated sacral screw placement was demonstrated. Finally, the experiments are concluded with a simulated foreign body removal using shrapnel models.

To achieve the fused RGBD and DRR view, multiple steps are required. First, the CBCT and the patient's surface scans are acquired. The FPFH matching for fast initialization of ICP yields a robust and efficient calibration of data extracted from CBCT and RGBD. This enables the data overlay, resulting in an augmented reality scene. The calibration accuracy is strongly dependent on the quality of the depth information acquired from the RGBD camera.

In contrast to other calibration techniques, this method does not require a pre-defined marker or known 3D structure. Theoretically, the calibration technique functions with any arbitrary object for which the surface is visible in the CBCT volume and yields enough structural features. In a clinical scenario, a system constructed as our design would require a one-time calibration or at the discretion of the user.

The fusion of CBCT and RGBD into one common coordinate space enables several new concepts. First, any arbitrary view can be visualized as the spatial restrictions in terms of C-arm placement no longer apply. For instance, a view along the spine can be visualized while placing a Jamshidi needle. Secondly, the augmented reality scene can be viewed from different view point simultaneously. This enables surgeons to align tools in all dimensions at the same time, possibly saving significant OR time.

Currently, the changes in the environment is not tracked. For instance, moving the surgical table or RGBD camera may result in the loss of proper image alignment, which motivates further development of the CBCT and RGBD system. Beyond the aforementioned possibilities, the fusion of RGBD and CBCT could facilitate intra-operative surgical navigation as the RGBD camera could be used for tool or activity tracking. Understanding the activity would enable the automatic adjustment of the view in order to provide the most optimal view during interventions.

The proposed technique contributes to a novel calibration for RGBD and CBCT data and achieves an accuracy of 2.58 mm. By acquiring more reliable depth information, this system could be later used for image-guided interventions to assist surgeons to perform more efficient procedures. The mixed reality visualization could enable an entire new field of novel applications for computer aided orthopedic interventions.

Acknowledgments

The authors want to thank Wolfgang Wein and his team from ImFusion GmbH, Munich for the opportunity of using the ImFusion Suite, and Gerhard Kleinzig and Sebastian Vogt from SIEMENS for their support and making a SIEMENS ARCADIS Orbic 3D available for this research.

Conflict of interest: The authors declare that they have no conflict of inter-

est.
Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Informed consent: This articles does not contain patient data.

References

[1] Matthew T. Dickerson, Robert L. Scot Drysdale, Scott A. McElfresh, and Emo Welzl. “Fast Greedy Triangulation Algorithms”. In: *Proceedings of the Tenth Annual Symposium on Computational Geometry.* SCG ’94. Stony Brook, New York, USA: ACM, 1994, pp. 211–220. ISBN: 0-89791-648-4. DOI: `10.1145/177424.177649`. URL: `http://doi.acm.org/10.1145/177424.177649`.

[2] Benoit Diotte, Pascal Fallavollita, Lejing Wang, Simon Weidert, Peter-Helmut Thaller, Ekkehard Euler, and Nassir Navab. “Radiation-free drill guidance in interlocking of intramedullary nails”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012.* Springer, 2012, pp. 18–25.

[3] S. Foix, G. Alenya, and C. Torras. “Lock-in Time-of-Flight (ToF) Cameras: A Survey”. In: *Sensors Journal, IEEE* 11.9 (Sept. 2011), pp. 1917–1926. ISSN: 1530-437X. DOI: `10.1109/JSEN.2010.2101060`.

[4] Bernhard Fuerst, Javad Fotouhi, and Nassir Navab. “Vision-Based Intraoperative Cone-Beam CT Stitching for Non-overlapping Volumes”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015.* Springer, 2015, pp. 387–395.

[5] Severine Habert, Jose Gardiazabal, Pascal Fallavollita, and Nassir Navab. “RGBDX: first design and experimental validation of a mirror-based RGBD Xray imaging system”. In: *International Symposium on Mixed and Augmented Reality (ISMAR).* Fukuoka, Japan, Oct. 2015.

[6] Severine Habert, Ma Meng, Wadim Kehl, Xiang Wang, Federico Tombari, Pascal Fallavollita, and Nassir Navab. “Augmenting mobile C-arm fluoroscopes via Stereo-RGBD sensors for multimodal visualization”. In: *International Symposium on Mixed and Augmented Reality (ISMAR).* Fukuoka, Japan, Oct. 2015.

[7] Choll W Kim, Yu-Po Lee, William Taylor, Ahmet Oygar, and Woo Kyung Kim. “Use of navigation-assisted fluoroscopy to decrease radiation exposure during minimally invasive spine surgery”. In: *The Spine Journal* 8.4 (2008), pp. 584–590.

[8] Everine B van de Kraats, Theo van Walsum, Lance Kendrick, Niels J Noordhoek, and Wiro J Niessen. “Accuracy evaluation of direct navigation with an isocentric 3D rotational X-ray system”. In: *Medical image analysis* 10.2 (2006), pp. 113–124.

[9] Felix Matthews, Dominik J Hoigne, Manfred Weiser, Guido A Wanner, Pietro Regazzoni, Norbert Suhm, and Peter Messmer. “Navigating the fluoroscope’s C-arm back into position: an accurate and practicable solution to cut radiation and optimize intraoperative workflow”. In: *Journal of orthopaedic trauma* 21.10 (2007), pp. 687–692.

[10] Philippe Merloz, Jocelyne Troccaz, Hervé Vouaillat, Christian Vasile, Jérôme Tonetti, Ahmad Eid, and Stéphane Plaweski. “Fluoroscopy-based navigation system in spine surgery”. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 221.7 (2007), pp. 813–820.

[11] N. Navab, S.-M. Heining, and J. Traub. "Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications". In: *Medical Imaging, IEEE Transactions on* 29.7 (July 2010), pp. 1412–1423. ISSN: 0278-0062.

[12] Nassir Navab, Tobias Blum, Lejing Wang, Aslı Okur, and Thomas Wendler. "First deployments of augmented reality in operating rooms". In: *Computer* 7 (2012), pp. 48–55.

[13] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. "KinectFusion: Real-time dense surface mapping and tracking". In: *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on.* Oct. 2011, pp. 127–136. DOI: `10.1109/ISMAR.2011.6092378`.

[14] S Nicolau, PY Lee, HS Wu, M Huang, R Lukang, L Soler, and J Marescaux. "Fusion of C-arm X-ray image on video view to reduce radiation exposure and improve orthopedic surgery planning: first in-vivo evaluation". In: *Proceedings of Computer Assisted Radiology and Surgery.* Vol. 6. 2011, pp. 115–116.

[15] James A Pfaff and Gregory P Moore. "Reducing risk in emergency department wound management". In: *Emergency medicine clinics of North America* 25.1 (2007), pp. 189–201.

[16] S Reaungamornrat, Yoshito Otake, Ali Uneri, Sebastian Schafer, DJ Mirota, Sajendra Nithiananthan, J Webster Stayman, Gerhard Kleinszig, A Jay Khanna, Russell H Taylor, and Jeffrey H Siewerdsen. "An on-board surgical tracking and video augmentation system for C-arm image guidance". In: *International journal of computer assisted radiology and surgery* 7.5 (2012), pp. 647–665.

[17] Radu Bogdan Rusu and Steve Cousins. "3d is here: Point cloud library (pcl)". In: *Robotics and Automation (ICRA), 2011 IEEE International Conference on.* IEEE. 2011, pp. 1–4.

[18] R.B. Rusu, N. Blodow, and M. Beetz. "Fast Point Feature Histograms (FPFH) for 3D registration". In: *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on.* May 2009, pp. 3212–3217. DOI: `10.1109/ROBOT.2009.5152473`.

[19] P.H.S. Torr and A. Zisserman. "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry". In: *Computer Vision and Image Understanding* 78.1 (2000), pp. 138–156. ISSN: 1077-3142. DOI: `http://dx.doi.org/10.1006/cviu.1999.0832`. URL: `http://www.sciencedirect.com/science/article/pii/S1077314299908329`.

[20] Xinag Wang, Severine Habert, Ma Meng, Xiang Wang, Chun-Hao Huang, Pascal Fallavollita, and Nassir Navab. "RGB-D/C-arm Calibration and Application in Medical Augmented Reality". In: *International Symposium on Mixed and Augmented Reality (ISMAR).* Fukuoka, Japan, Oct. 2015.

[21] Z. Zhang. "A Flexible New Technique for Camera Calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (Dec. 2000), pp. 1330–1334. URL: `http://research.microsoft.com/apps/pubs/default.aspx?id=64198`.