

Real-time Respiratory Motion Analysis
Using GPU-accelerated
Range Imaging

Echtzeitanalyse der Atembewegung
mittels GPU-beschleunigter
Tiefenbildgebung

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr.-Ing.

vorgelegt von

Jakob Wasza
aus
München

Als Dissertation genehmigt von der
Technischen Fakultät der
Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	26.09.2019
Vorsitzender des Promotionsorgans:	Prof. Dr.-Ing. R. Lerch
Gutachter:	Prof. Dr.-Ing. J. Hornegger Prof. P. Keall, Ph.D. Prof. Dr.-Ing. J. Denzler

Abstract

Respiratory motion analysis and management are crucial issues for a plurality of medical applications. Of particular scientific concern are methods that allow to analyze the patient's breathing in a non-invasive and real-time manner that does not involve ionizing radiation exposure. For this purpose, range imaging technologies, that allow to dynamically acquire three-dimensional body surface data, have been proposed over the last years. A particular challenge with such methods is a fully automatic investigation and assessment of the body surface data, as well as computation times that comply with real-time constraints. This dissertation is concerned with the application of range imaging principles for real-time automatic respiratory motion analysis. The focus is on the development of efficient methods for data preprocessing and fusion as well as machine learning and surface registration techniques. A particular emphasis of this thesis is the design of the proposed algorithms for GPU architectures to enable real-time computation.

The first part of this thesis covers the general challenges and requirements for respiratory motion analysis in diagnostic and therapeutic applications. Furthermore, the range imaging technologies that are relevant for this thesis are introduced and the suitability of GPU architectures for accelerating the computation of several tasks inherent to range imaging based respiratory motion analysis are investigated.

The second part of this work is concerned with pre-processing and fusion techniques for range data. To account for the low signal-to-noise ratio that is common with range data, this work proposes a processing pipeline that allows to reconstruct the ideal data with an error trueness less than 1.0 mm at run-times of 2 ms. For fusing range image data in a multi-camera setup, as it is required for the simultaneous acquisition of frontal and lateral body surface, this thesis proposes a novel framework that enables the computation of a 180° coverage body surface model consisting of more than 3.0×10^5 points with a computation time of less than 5 ms.

The third part of this work is concerned with patient-specific respiratory motion models. The thesis proposes machine learning techniques to generate a continuous motion model that features the ability to automatically differentiate between thoracic and abdominal breathing as well as to quantitatively analyze the patient's respiration magnitude. By using purposely developed surface registration schemes, these models are then brought in congruence with body surface data acquired by range imaging sensors. This allows for respiratory motion compensated patient positioning that reduces the alignment error observed with conventional approaches by a factor of 3 to less than 4.0 mm. Further, this approach allows to automatically derive a multi-dimensional respiration surrogate that yields a correlation coefficient greater than 0.97 compared to commonly employed invasive or semi-automatic approaches and that can be computed in 20 ms.

The fourth part concludes this thesis with a summary of the presented methods and results, as well as an outlook regarding future research directions and challenges towards clinical translation.

Kurzfassung

Die Analyse und Handhabung von Atembewegungen sind grundlegende Probleme für viele medizinische Anwendungen. Von speziellem wissenschaftlichen Interesse sind hierbei Methoden, die es ermöglichen die Atmung des Patienten nicht-invasiv, ohne den Einsatz ionisierender Strahlung, und in Echtzeit zu analysieren. Für diese Fragestellung wurden kürzlich Technologien der Tiefenbildgebung, die es ermöglichen dreidimensionale Abbildungen der Körperoberfläche dynamisch zu erfassen, vorgeschlagen. Die besonderen Herausforderungen in diesem Umfeld sind eine vollautomatische Untersuchung und Auswertung der Körperoberflächendaten sowie Berechnungszeiten die Echtzeitanforderungen gerecht werden. Diese Dissertation befasst sich mit der Nutzung der Tiefenbildgebung zur echtzeitfähigen automatischen Atembewegungsanalyse. Der Schwerpunkt liegt hierbei auf der Entwicklung von effizienten Methoden zur Datenvorverarbeitung und Sensorfusion sowie Verfahren des maschinellen Lernens und der Oberflächenregistrierung. Ein zentraler Bestandteil ist dabei die Konzipierung der Algorithmen für Grafikprozessoren um eine echtzeitfähige Berechnung zu ermöglichen.

Im ersten Teil der Arbeit werden die generellen Herausforderungen und Anforderungen der Atembewegungsanalyse für diagnostische und therapeutische Verfahren dargelegt. Des Weiteren werden für diese Arbeit relevante Technologien der Tiefenbildgebung beschrieben, sowie die Eignung von Grafikprozessoren zur effizienten Berechnung von Problemen der tiefenbildgebungsbasierten Atembewegungsanalyse untersucht.

Der zweite Teil der Arbeit befasst sich mit der Vorverarbeitung und Fusion von Tiefendaten. Zur Verbesserung des oftmals niedrigen Signal-Rausch-Verhältnisses von Tiefendaten wird eine Verarbeitungskette vorgestellt, die es ermöglicht die idealen Daten mit einer Genauigkeit unter 1.0 mm und einer Laufzeit von 2 ms zu rekonstruieren. Zur Fusionierung von Tiefendaten in einem Mehrkameraaufbau, wie er zur gleichzeitigen Erfassung der vorderen und seitlichen Körperoberfläche notwendig ist, beschreibt diese Arbeit ein neuartiges Verfahren das es ermöglicht ein dreidimensionales 180° Abbild der Körperoberfläche bestehend aus über 3.0×10^5 Punkten und einer Laufzeit von unter 5 ms zu berechnen.

Der Fokus des dritten Teils der Arbeit liegt auf patientenspezifischen Atembewegungsmodellen. Es werden Verfahren des maschinellen Lernens zur Erstellung kontinuierlicher Atembewegungsmodelle beschrieben, welche eine automatische Unterscheidung zwischen Brust- und Bauchatmung sowie eine quantitative Bewertung der Atemmagnitudo ermöglichen. Durch speziell entwickelte Registrierungsverfahren werden diese Modelle mit Oberflächendaten in Deckungsgleichheit gebracht. Dies ermöglicht eine atembewegungskompensierte Patientenpositionierung die, im Vergleich zu herkömmlichen Verfahren, den Positionierungsfehler um den Faktor 3 auf unter 7.0 mm reduziert. Zudem erlaubt dieser Ansatz die vollautomatische Ableitung eines mehrdimensionalen Atemsurrogates das zu konventionellen semi-automatischen oder invasiven Verfahren einen Korrelationskoeffizienten von mehr als 0.97 aufweist, und in 20 ms berechnet werden kann.

Der vierte Teil schließt die Arbeit mit einer Zusammenfassung der vorgestellten Verfahren und Ergebnisse sowie einem Ausblick bezüglich weiterführenden Forschungsrichtungen und Herausforderungen einer klinischen Umsetzung.

Acknowledgment

First and foremost, I like to thank Prof. Dr. Joachim Hornegger for his great lectures that aroused my interest in medical imaging as an undergraduate student and for giving me the opportunity to do my Ph. D. at the Pattern Recognition Lab. I appreciate the great research environment he created at the lab, his encouragement, support, scientific guidance and confidence in me over the last years. I like to express my sincere gratitude to Prof. Paul Keall, Ph. D. , for his valuable input and review of this thesis and Prof. Dr. Andreas Maier for his great support during the final phase of this work. I also like to express my thanks to Prof. Dr. Elmar Nöth for various events that contributed to the great atmosphere at the lab.

I like to thank all my colleges at the lab for the friendly atmosphere, knowledge sharing, scientific discussions and the great time we had at various conferences or social events. In particular, let me thank my colleagues Dr. Sebastian Bauer, Dr. Sven Haase and Peter Fischer for the splendid time in our office and the inspiring scientific discussions and reviews that had a substantial impact on this work. I also like to thank my colleges from the admin group as well as the technical and administrative staff at the lab for keeping things up and running. Here, I especially like to thank Dr. Christian Riess for the great time we had fixing critical hardware outages and our various discussions.

Further, I acknowledge the support by the European Regional Development Fund (ERDF) and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (StMWIVT), in the context of the R&D program IuK Bayern under Grant No. IUK338/001. In this regard I would like to express my thanks to our industrial and academic project partners. At Siemens AG I like to thank Dr. Natalia Anderl, Dr. Annemarie Bakai, Stefan Sattler and Stefan Schuster for the background on clinical workflows regarding radiation therapy and interventional imaging. At Softgate GmbH I like to thank Dr. Florian Höpfel, Sebastian Reichert and Christiane Kupczok for the support in project management, camera calibration and software development. At the chair for Hardware-Software-Co-Design, Friedrich-Alexander-Universität Erlangen, Nürnberg, I like to thank Dr. Frank Hannig and Dr. Moritz Schmid. I would like to express my sincere thanks to Metrilus GmbH and in particular to Dr. Christian Schaller for providing equipment and calibration technology for several experiments.

Last but not least, I like to thank my family and friends for their support over the years. In particular, I like to thank Katrin for her patience and support.

Jakob Wasza

Contents

I	Introduction	1
	Chapter 1 Introduction	3
1.1	Scientific Contributions	4
1.2	Organization of the Thesis	6
	Chapter 2 Thematic Background	7
2.1	Respiratory Motion in Therapy and Diagnostics	7
2.1.1	Image Guided Radiation Therapy	8
2.2	Range Imaging	10
2.2.1	Range Imaging Acquisition Principles	11
2.2.2	RI Systems for Respiratory Motion Analysis	13
2.3	General Purpose Computing on GPUs.	14
2.3.1	GPU Architectures	15
2.3.2	GPU Programming Paradigms	16
II	Real-time Range Imaging	17
	Chapter 3 Foundations of Range Imaging	19
3.1	Acquisition Geometry in Range Imaging	19
3.2	Data Structures and Data Representation	21
3.2.1	Range Measurements and Reconstructed Surface	22
3.2.2	Confidence Maps and Reliability Indicators	23
3.3	Synthetic Range Data By Sensor Simulation	24
3.4	RITK: A Framework for 3-D Range Image Stream Processing	25
	Chapter 4 Range Image Pre-Processing	27
4.1	Motivation	27
4.2	Related Work	28
4.3	Pre-Processing Pipeline Overview	29
4.4	Restoration of Invalid Depth Measurements	30
4.4.1	Normalized Convolution	31
4.4.2	Spectral Deconvolution	32
4.5	Edge-preserving Denoising	35
4.5.1	Bilateral Filtering	35
4.5.2	Guided Filtering	36

4.6	Temporal Denoising	38
4.6.1	Conventional Temporal Averaging	38
4.6.2	Bilateral Temporal Denoising	38
4.7	Experiments and Results	39
4.7.1	Quantitative Results Using a Respiration Phantom	40
4.7.2	Performance Study	49
4.8	Discussion and Conclusion	51

Chapter 5 Multi-View Range Data Fusion 53

5.1	Motivation	53
5.2	Related Work	54
5.3	Implicit Surface Representation Using Signed Distance Transforms	56
5.3.1	Approximated Point-to-Surface Distance	57
5.3.2	Weighting of Approximative Distances	58
5.3.3	Fusion of Multiple Distance Transforms	59
5.4	Explicit Surface Reconstruction by Ray Casting	60
5.4.1	Pinhole Camera Ray Casting	61
5.4.2	Manifold Ray Casting	61
5.5	Experiments and Results	62
5.5.1	Quantitative Results Using a Respiration Phantom	63
5.5.2	Qualitative Results on Real RI Data	68
5.5.3	Performance Study	69
5.6	Discussion and Conclusion	72

III Respiratory Motion Analysis Using 4-D Shape Priors 73

Chapter 6 Foundations of 4-D Shape Priors for Respiratory Motion Analysis 75

6.1	Motivation	75
6.2	Related Work	76
6.3	Principles of 4-D Shape Priors	78
6.3.1	Shape Motion Models Using Displacement Fields	78
6.3.2	Condition of Body Surface Displacement Fields	79
6.4	Shape Correspondences by Deformable Registration	82
6.4.1	Non-rigid RI Surface Registration	82
6.4.2	Non-rigid Registration of Distance Transforms	83
6.4.3	Non-rigid Point Set Registration	84
6.5	Dimensionality Reduction Techniques	85
6.5.1	Linear Methods	85
6.5.2	Non-linear Methods	87
6.6	Experiments and Results	89
6.6.1	Body Surface Displacement Fields	89
6.6.2	Sparse Linear Motion Models	96
6.7	Discussion and Conclusion	98

Chapter 7 Real-time Respiratory Motion Analysis Using 4-D Shape Priors	101
7.1 Motivation	101
7.2 Related Work	102
7.3 4-D Shape Motion Model to RI Surface Alignment	105
7.3.1 Conventional Single-Link Iterative Closest Point	107
7.3.2 Robust Multi-Link Coherent Point Drift	108
7.4 Efficient Corresponding Point Search	112
7.4.1 Random Ball Cover	112
7.4.2 Projective Data Association	113
7.5 Experiments and Results	115
7.5.1 Corresponding Point Search Strategies	115
7.5.2 Continuous Respiratory Motion Monitoring	118
7.5.3 Motion Compensated Patient Alignment	127
7.6 Discussion and Conclusion	132
IV Outlook and Summary	135
Chapter 8 Outlook	137
Chapter 9 Summary	141
V Addendum	145
Appendix	147
List of Symbols	153
List of Figures	157
List of Tables	159
Bibliography	161

Part I
Introduction

Introduction

1.1 Scientific Contributions	4
1.2 Organization of the Thesis	6

Compensating for respiration induced motion is a crucial factor for the success of a plurality of medical procedures. For example, in fractionated radiation therapy the patient must be aligned to pre-procedurally obtained planning data which implies the need to compensate for the difference between the patient's current respiration state and the one that governs the planning data. Further, for accurate dose delivery the patient must be continuously monitored in order to instantaneously initiate measures that compensate for tumor motion caused by respiration.

For both, alignment and continuous motion monitoring, methods based on range imaging (RI) that provides dense body surface data at real-time frame-rates have been proposed in the recent past. Compared to commonly employed imaging techniques such as cone beam computed tomography (CT) or portal imaging for setup verification and pressure belts or reflective vests for motion monitoring, RI systems are marker-less, non-intrusive and do not involve ionizing radiation exposure. Besides a high degree of accuracy and reliability, low latencies are key issues for the success of such RI systems in clinical routine. This poses a challenge as contemporary RI sensors typically stream a huge amount of raw data that must be processed to account for noise and outliers, fused to a unifying representation in a multi-camera setup and ultimately analyzed for the actual motion compensation purposes. Though algorithmic techniques may be used to speed up selected computation steps, dedicated hardware acceleration concepts are in general mandatory.

The thesis at hand is concerned with methods for real-time respiratory motion analysis using Graphics Processing Unit (GPU) accelerated RI. As the main contribution, the thesis proposes a real-time capable framework using prior knowledge on 4-D shape deformations to unify the problems of respiratory motion compensated patient positioning and continuous respiration monitoring. From a methodological perspective, the thesis covers the following aspects. First, pre-processing and multi-sensor fusion methods for RI data as a fundamental prerequisite for subsequent analysis are investigated. Second, the thesis researches machine learning and pattern recognition techniques for patient specific 4-D shape motion models and dedicated surface registration methods to drive real-time RI-based respiratory motion analysis.

1.1 Scientific Contributions

The scientific focus of this thesis is the investigation of existing and the development of novel RI concepts for respiratory motion analysis in computer assisted interventions. Due to the importance of real-time constraints in clinical routine, a focus throughout this work is on hardware acceleration using GPUs. The main contributions along with the corresponding scientific papers that have been published during the course of this thesis are:

- **Real-time RI processing:** High frame-rates and low latencies are the cornerstone for high temporal resolution respiratory motion analysis. This requires a proper software architecture and efficient data pre-processing strategies to enhance the raw data streamed by RI sensors. The thesis proposes GPU-based hardware acceleration techniques that have been published in [Wasz 11a, Wasz 11c, Wasz 11b]. As these publications did not focus on respiratory motion in particular, this thesis provides an additional thorough evaluation of RI processing techniques in a respiratory motion scenario.
- **Real-time fusion of multi-view RI data:** Respiratory motion analysis using a single RI sensor is often not feasible, for example due to occlusions by clinical devices, and further prohibits high coverage body surface acquisition. Thus, multiple sensors are required which implies the need for efficient methods to fuse and consolidate this data. The contribution for this task is a GPU accelerated volumetric fusion approach in conjunction with a novel ray casting technique that is tailored to the human anatomy and that enables an 180° coverage body surface reconstruction at real-time frame-rates. These concepts were firstly published in [Wasz 13, Wasz 16].
- **4-D shape motion models for RI-based respiration analysis:** Existing methods for RI-based respiratory motion analysis are commonly based on heuristics and do not account for the anatomical nature of motion. This thesis proposes pattern recognition methods to set up anatomically founded shape motion models for fully automatic analysis of respiratory motion as published in [Wasz 12b, Wasz 12a]. Elaborating on this concept, this thesis proposes sparse models to enable an unsupervised differentiation between different breathing patterns as published in [Wasz 12a, Wasz 16]. Further, this thesis provides a joint theoretical and anthropological investigation of respiration induced body surface deformation as firstly published in [Wasz 13] to quantify requirements for RI-based respiration analysis.
- **Real-time respiratory motion analysis using 4-D shape priors:** Building upon the developed concepts for RI data processing and 4-D shape motion models, this thesis proposes a real-time framework that unifies the problems of motion compensated patient alignment [Wasz 12b, Wasz 13] and continuous respiration monitoring using external surrogates [Wasz 16]. The main contributions are novel schemes to derive non-redundant external model driven respiration surrogates from dense body surface data and efficient GPU accelerated surface registration methods for real-time computation.

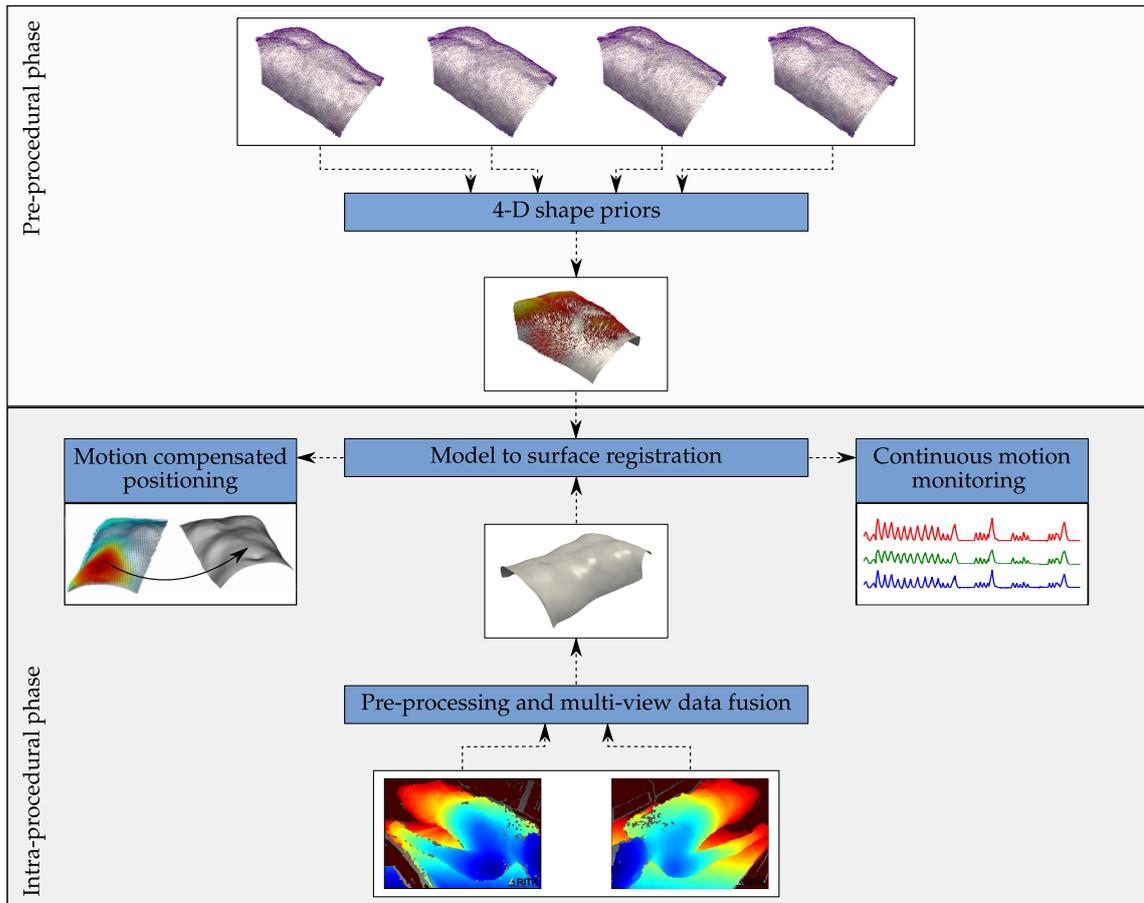


Figure 1.1: The respiratory motion analysis framework proposed in this thesis along with the major scientific contributions highlighted in blue. The framework divides into a pre-procedural stage to set up a 4-D shape motion model that is intra-procedurally registered to multi-view RI body surface data enabling motion compensated patient positioning and continuous respiratory motion monitoring.

Altogether, these contributions provide a unifying framework for real-time RI-based respiratory motion analysis that is illustrated in Fig. 1.1. The key principle of this framework is to divide the problem of respiratory motion analysis into a *pre-procedural* stage to set up a patient specific 4-D motion model capturing prior knowledge on body surface deformations that is registered in the *intra-procedural* phase to RI data for real-time deformation estimation. In particular, this framework enables motion compensated patient positioning outperforming conventional alignment strategies by a factor of about 3 and continuous motion monitoring using a non-intrusive radiation free method that significantly correlates with conventionally employed intrusive or heuristic techniques. Using the principle of GPU computing that is pursued throughout this thesis, real-time run-times of approximately 20 ms or 50 Hz are achieved on off-the-shelf hardware. These run-times satisfy current clinical requirements imposing a maximum time delay of few hundred milliseconds and comply with recommendations on future systems to keep delays as short as possible.

Contributions to other work Several RI concepts that were developed during the course of this thesis contributed to other scientific work. This includes contributions to photometric surface deformation estimation [Baue 12c], multi-modal marker-less coarse patient setup [Baue 11] in fractionated radiation therapy, organ surface registration in an augmented reality scenario [Mull 11], and 3-D endoscopy [Haas 13b, Haas 13a, Kohl 13] for minimally invasive surgery. Besides clinical applications, techniques developed in this thesis have also been used in a computer vision scenario that targets 3-D scene reconstruction [Neum 11, Baue 13b].

1.2 Organization of the Thesis

The thesis is divided into four major parts according to the general issues and methodical challenges, see also Fig. 1.1 for an overview.

Part I The background embedding this thesis and problems that are of relevance w. r. t. real-time respiratory motion analysis using RI technology is covered in this part. Chapter 2 outlines the clinical background, requirements and challenges of RI-based respiratory motion analysis in computer assisted interventions. Further, the basic RI data acquisition principles and sensors used in this work are outlined. As an important aspect, the paradigm of general purpose computing on GPUs (GPGPU) that is pursued in this thesis for real-time computation is introduced.

Part II The focus of the second part is on real-time RI as the key for respiratory motion analysis as investigated in this thesis. Chapter 3 introduces the mathematical foundations on RI data acquisition and the basic notation that is used throughout this thesis. Subsequently, algorithmic methods for range data enhancement and denoising strategies to cope with the low quality of the raw depth measurements obtained from RI sensors are investigated in Chapter 4. The second part of this thesis concludes with Chapter 5 that covers a framework for the fusion of range data obtained from multiple sensors in conjunction with a reconstruction technique that enables high coverage patient body surface models in real-time.

Part III The third part elaborates on pattern recognition and machine learning techniques for RI-based respiratory motion analysis. The general principle of 4-D shape priors is covered in Chapter 6 with a focus on establishing surface correspondences across different respiration states using non-rigid registration methods and dimensionality reduction techniques. Additionally, this chapter details a method for unsupervised decomposition of body surface deformation fields into distinct breathing patterns. The application of 4-D shape priors for motion compensated patient alignment and continuous respiratory motion monitoring is the focus of Chapter 7. Methodologically, this chapter focuses on efficient point-set and surface registration schemes to align motion models to RI data.

Part IV The thesis concludes with an outlook, future research directions and challenges for clinical translation in Chapter 8 and a summary in Chapter 9.

Thematic Background

2.1 Respiratory Motion in Therapy and Diagnostics	7
2.2 Range Imaging.	10
2.3 General Purpose Computing on GPUs.	14

This chapter covers the general background sustaining this thesis. After outlining the importance of respiratory motion analysis and management in therapy and diagnostics, RI technology as means for real-time scene acquisition is presented. Further, this chapter introduces the paradigm of general purpose computing on GPUs which is pursued throughout this thesis for the purpose of complying with real-time constraints.

2.1 Respiratory Motion in Therapy and Diagnostics

Respiratory motion is considered a major challenge in many computer assisted interventions and diagnostic applications such as medical image reconstruction. This is due to the fact that respiration induced internal deformations cause deviations to pre-procedural planning data for interventions and severe motion artifacts with image reconstruction. The simplest option to decrease the effect of respiratory motion is to reduce the respiration magnitude using for example breath-holding, abdominal pressure or active breath-control systems [Hof03], cf. Fig. 2.1a. However, these techniques are inadequate for many solutions. This led to the development of gating or binning techniques where the intervention or acquisition is limited to a certain respiration phase. For CT acquisition for example, a gated reconstruction reduces the image artifacts by up to 50% [Bern15]. Further, sophisticated motion models that continuously correct for the effect of respiratory motion have been proposed. For a comprehensive review of motion models for image guided interventions as well as for motion corrected image acquisition and reconstruction, the reader is referred to the work of McClelland et al. [McCl13].

One crucial aspect for respiratory motion analysis and compensation is whether the target motion is tracked directly or measured via a respiration surrogate. Further, it is of importance whether the system must be real-time capable as motion analysis is performed during the procedure. While for motion compensated reconstruction retrospective correction techniques based on the actual image data, as for example proposed by Wachinger et al. [Wach12], can be used, respiratory motion

(a) Siemens Artiste. ¹(b) BrainLab ExacTrac. ²(c) Accuray Cyberknife. ³

Figure 2.1: Clinically available IGRT systems with different strategies to cope with respiratory motion. The left system reduces the respiration magnitude using abdominal pressure, the platform in the middle performs gated treatment and the right system performs motion tracking. Note that most motion management solutions are realized as extensions to existing IGRT systems [Will 12].

analysis and compensation in many computer assisted interventions typically require real-time capable systems and a direct observation of respiratory motion is not feasible or even prohibitive.

For generating respiration surrogates, RI sensors have gained increasing popularity over the last years as they are non-invasive, real-time capable, and provide dense body surface data. The thesis at hand is concerned with novel concepts for RI-based respiratory motion analysis such as model-based respiration surrogates or motion-compensated patient-positioning techniques. In particular, the investigated methods hold benefit for image guided radiation therapy that inherently requires real-time capability and where a direct observation of the moving target structures might be prohibitive due to additional ionizing radiation exposure or invasiveness. Due to the importance of the topic for this thesis, the basic principles of image guided radiation therapy and the proposed RI solutions for respiratory motion management will be detailed in the following section.

2.1.1 Image Guided Radiation Therapy

Besides chemotherapy and surgery, image guided external beam radiation therapy (IGRT) is a commonly used treatment option for cancer patients. IGRT aims for destroying tumor cells by using ionizing radiation beams delivered by linear particle accelerators (LINACs) that are free to move around the patient. To ensure both a proper tumor irradiation and sparing of healthy tissue, dose delivery is performed using different LINAC positions and beam directions. The actual irradiation protocol is based on a patient-specific pre-procedurally derived image guided treatment plan that contains both the so-called clinical target volumes (CTVs) defining different areas of dose delivery as well as surrounding critical structures and organs where it is crucial to reduce the radiation dose to a low level. Typically, this

¹<http://siemens.com/press/photo/sohim201103-01d>

²http://middleeasthospital.com/wp-content/uploads/2009/09/exactrac_xray_6d.jpg

³http://accuray.com/sites/default/files/vpk_assets/cyberknife_system_product.jpg

treatment plan is derived from CT, however, recent approaches also employ magnetic resonance imaging (MRI) enabling superior soft-tissue visualization without potentially harmful ionizing radiation exposure [Evan 08].

A common approach for dose delivery is fractionated IGRT where the total dose is spread across several sessions. The objective here is to allow healthy cells to recover between fractions and to maximize the probability of irradiating malignant cells in phases of low radio-resistance. Besides the treatment plan the success of fractionated IGRT for both ensuring a proper dose delivery to the target area as well as not harming healthy tissue depends on two critical factors: *inter-fractional* variations and *intra-fractional* changes of tumor location and lung function [Will 12, Kipr 15]. Inter-fractional variations are due to a variety of reasons such as organ movement caused by the digestive and urinary system, differences in the breathing baseline or the patient position itself. Minimizing inter-fractional movement by an accurate patient setup prior to dose delivery is therefore mandatory. Subsequently, during dose delivery, motion mitigation strategies must be used to cope with intra-fractional internal movement that are mainly due to respiratory and cardiac motion. Especially for thoraco-abdominal cancer treatment procedures, intra-fractional tumor movement due to respiration is recognized as a major challenge with motion magnitudes over 10 mm in superior-inferior (SI) direction [Liu 07] that causes significant dosimetric errors [Yu 98, Keal 06, Bert 11]. A comprehensive survey on motion in radiation therapy is beyond the scope of this thesis. The reader is referred to the report of the American Association of Physicists in Medicine (AAPM) task group 76 [Keal 06] or the reviews of Bert and Durante [Bert 11] and Korreman [Korr 12].

One option to decrease the effect of respiration induced tumor movement is to reduce the respiratory motion magnitude using for example active breath-control systems or abdominal pressure [Hof 03]. The latter strategy is depicted in Fig. 2.1a. However, such techniques are limited by the ability of cancer patients to maintain a specific lung capacity and a total elimination of respiration induced motion is impossible [Yan 06]. More advanced approaches are based on motion monitoring techniques to generate a time-resolved respiration signal. This signal can then be used for breathing guidance [Poll 15] or as basis for gating techniques where the radiation beam is only active during specific parts of the breathing cycle as implemented in the ExacTrac system (Brainlab AG, Feldkirchen, Germany) that is depicted in Fig. 2.1b, or tracking solutions that adjust the beam continuously as pursued with the Vero platform (Brainlab AG, Feldkirchen, Germany & Mitsubishi Heavy Industries Ltd., Minato (Tokyo), Japan) or the CyberKnife system (Accuray Inc., Sunnyvale, USA) that is shown in Fig. 2.1c. All these approaches have in common that the respiration signal can either be derived from the tumor motion itself or from a respiration surrogate. In any case, these systems must be real-time capable and the total time delay should be as short as possible and never exceed 0.5 s with typical time delays ranging from 90 ms to 200 ms for clinically available solutions [Keal 06].

The predominant way to generate direct respiration signals is based on X-ray fluoroscopy, cone-beam CT or implanted markers. Though these direct measurements feature a high degree of accuracy and enable advanced motion tracking as

for example proposed by Chung et al. [Chun 16], they are highly invasive or involve additional ionizing radiation, both imposing risks to the patient. Further, X-ray fluoroscopy is often limited to a few frames per minute to minimize the radiation dose and implanted transponders may interfere with imaging modalities or may not work with high energy particle therapy [Bert 11]. In contrast, respiration surrogates can often be derived with zero ionizing radiation using for example air flow analysis systems or even non-invasively by body surface observation techniques based on optical sensors. Besides their non-invasive nature, these techniques feature high frame-rates that allow for high temporal resolution respiratory motion analysis. For a detailed survey on non-radiographic respiration analysis techniques the reader is referred to the report of the AAPM task group 147 [Will 12].

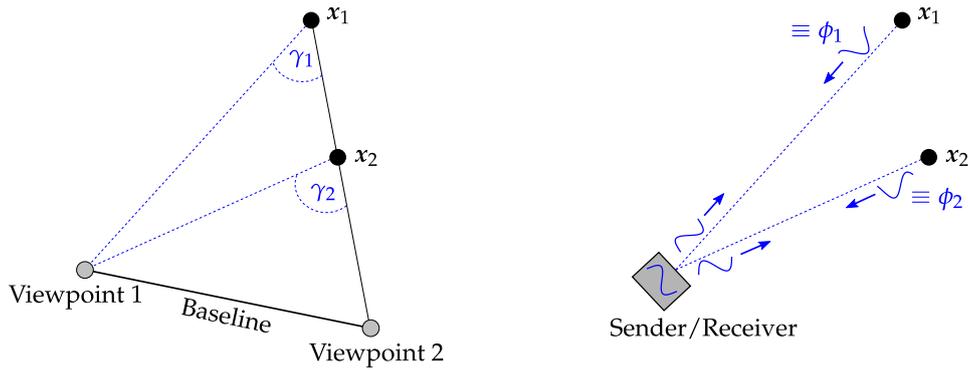
The inherent challenge with respiration surrogates is the consistency of the measured or computed *external* surrogate motion and the actual *internal* target motion. For example, Liu and Koch et al. [Liu 04, Koch 04], Yan et al. [Yan 06] or Fayad et al. [Faya 11] showed for surface based surrogates that the correlation between internal and external motion is strongly influenced by the relative position of the internal target region to the external surrogate and the number of used surrogates. The reported results imply that for predicting internal movement based on observations of the body surface it is advantageous to acquire the surrogates densely and spread across the entire body. These requirements are inherently supported by modern range imaging sensors that feature high resolution with a large field of view at real-time frame-rates. Further, combining multiple sensors allows for a complete body coverage. Thus, RI technology is a promising candidate for generating surface based respiration surrogates in fractionated IGRT.

Combining the advantages of both direct measurements and surrogates allows to eliminate the disadvantages, namely ionizing radiation exposure and low frame-rates with fluoroscopy or low precision with surrogates. This approach is for example pursued in the ExacTrac system depicted in Fig. 2.1b where respiratory motion is continuously monitored using infra-red markers attached to the patient in combination with a periodical verification using a dual X-ray system. However, this requires reproducible and proper marker attachment. This can be superseded by using dense respiration surrogates generated by RI sensors.

2.2 Range Imaging

RI refers to a class of techniques that generate a 2-D image corresponding to the depth representation of the acquired 3-D scene. More precisely, the pixels of an RI hold the distance of 3-D scene points to a specific reference point. In contrast, images acquired by conventional cameras depict the 2-D photometric (e. g. color values or shades of gray) appearance of the 3-D scene. If the RI device is calibrated properly, the distances are given in physical units and the 3-D scene can be reconstructed accurately.

RI is employed in a plurality of application fields as diverse as cultural heritage preservation [Pav107], quality assurance in industry [Sans 09] or human pose recognition in consumer electronics [Shot 13]. Especially the latter gained an incred-



(a) Triangulation-based range computation. Based on the intersection of two viewing directions from different view points and known baseline a triangle with angle γ_i is used to recover depth information.

(b) Indirect run-time range measurement. Based on the phase shift ϕ_i between an emitted modulated signal and the received signal and the speed of light depth information is measured.

Figure 2.2: Basic working principles of triangulation-based depth recovery (left) and coaxial run-time range measurement (right) as popular range imaging techniques. Though relying on fundamentally different sampling principles both techniques are capable of generating dense metric depth maps at real-time frame-rates.

ible increase in popularity with the market launch of Microsoft’s Kinect device in 2010 [Zhan 12]. The Kinect device provides dense range measurements and complementary photometric information at real-time frame-rates (30 Hz) with a mass-market retail price of about \$120. Further, RI also assumes an increasingly important role in healthcare with applications as different as workflow monitoring [Ladi 10], 3-D endoscopy [Maie 13] or radiation therapy [Will 12] as investigated in this thesis. For a comprehensive review on healthcare applications that can benefit from RI technology the reader is referred to the survey of Bauer et al. [Baue 13a].

As diverse as RI application fields is the multitude of different RI principles that have been proposed over the past decades. The following two sections outline range acquisition principles and RI systems that already have found their way into clinical practice or have been proposed for respiratory motion analysis in computer assisted interventions.

2.2.1 Range Imaging Acquisition Principles

Of interest for this thesis are real-time capable RI sensors that provide dense metric range measurements and feature a working distance of roughly 500 mm to 2000 mm. Suitable range acquisition principles that meet these requirements are outlined in the following sections. A survey of state-of-the-art RI technology and sensor specific advantages or limitations is beyond the scientific scope of this thesis and the reader is referred to the work of Blais [Blai 04], Sansoni et al. [Sans 09] or Häusler and Ettl [Haus 11].

Triangulation Methods

The governing principle of methods based on triangulation is that for a 3-D point being observed from two different viewpoints its position in space can be calculated if the relative positions of the viewpoints and their viewing directions are known. More precisely, the baseline between the viewpoints and viewing directions enclose a triangle and the position of the 3-D point is given by the intersection of the viewing directions, see Fig. 2.2a for an illustration. The uncertainty of depth recovery depends on the triangulation angle and thus, the viewpoints must be separated by a certain distance. This entails that there is a lower limit for the size of triangulation sensors and due to the different viewpoints there might occur occlusions and the depth cannot be recovered.

The triangulation principle for depth recovery is commonly employed in stereo vision where a pair of images acquired from two different viewpoints is used and the viewing directions are derived from the projections of the 3-D point in the corresponding image planes. The relative difference in image position of the same 3-D point projected into the two image planes is known as *disparity* and, given the baseline of the two viewpoints, is inversely proportional to the depth of the 3-D point [Hart04]. Naturally, to compute a disparity map given a pair of stereo images it is mandatory to identify corresponding pixels that describe the same 3-D point. This *correspondence problem* is one of the major challenges in stereo vision for both accuracy and run-time.

In general, there are two main approaches on how to address the correspondence problem. In *passive* triangulation the images are acquired by two cameras and correspondences are computed either locally by feature-matching techniques [Bay06] or globally by employing energy minimizing functions with smoothness constraints across the entire image [Boyk01]. Though progress has been made over the last years regarding algorithmic or hardware acceleration strategies and the correspondence problem can be restricted to a single line due to epipolar constraints, computing corresponding points is still a computationally demanding task. Further, in texture-less regions or in the presence of repetitive patterns computing correspondences may fail. This problem can be circumvented by using *active* triangulation where a specific pattern is projected onto the scene using an additional illumination device, thus augmenting texture-less or homogeneous regions. The correspondence problem is now defined between the known pattern as observed in both images. *Active* triangulation can also be realized by replacing one camera with the projector. The correspondence problem is now defined between the observed pattern in the image and the known projected pattern. The simplest form of *active* triangulation uses just one line or a single point and is known as *light sectioning* [Haus88]. This approach substantially simplifies the correspondence problem, however, depth recovery is inherently limited to a single line or a point and requires to successively scan the scene to provide dense range measurements. In contrast, area patterns allow for dense measurements but require the projected pattern to be encoded to account for ambiguities and again imply the need for robust point correspondences. These approaches are known as *structured light* and can be further categorized w. r. t. the pattern encoding including color vs. monochromatic or single-shot vs. temporal encoding [Salv10].

Coaxial Run-time Methods

Run-time methods for range measurements are based on the time it takes for a light signal to travel from a source to an object and back to a sensor. The run-time can either be measured directly or computed indirectly by phase measurements. Run-time methods can be implemented as a coaxial source/detector setup which obviates the need for a baseline as with range imaging principles based on triangulation. Thus, there is no problem with missing data or shadows and the devices can be built very compactly. Further, multi-view calibration steps as with triangulation methods are not needed.

One specific run-time based method is Time-of-Flight (ToF) imaging that has been proposed for both direct measurements using pulsed illumination or indirectly via continuous wave modulation [Kolb 10]. For the latter, the outgoing light is modulated with a carrier signal and the phase shift of the reflected and received carrier signal is used to determine the depth, see Fig. 2.2b for an illustration of this principle. ToF systems feature dense range measurements at very high frame-rates up to 90 Hz with the CamBoard nano (pmdtechnologies GmbH, Siegen, Germany). Further, ToF sensors provide complementary grayscale photometric information without the need of an additional sensor. However, range data acquired by ToF sensors is prone to noise and outliers due to a multitude of error sources. Typically this involves a distance-related error component known as *wiggling*, amplitude related offset or multi-path reflections and sub-surface scattering. Further, ToF devices are known to exhibit a temperature drift [Baue 13a] which restrict their usage for controlled environments, only.

2.2.2 RI Systems for Respiratory Motion Analysis

Though in theory all real-time capable RI acquisition principles outlined in Section 2.2.1 can be used for respiratory motion management there are few systems that have been explicitly proposed for this problem and actually found their way into clinical practice. In fact, the only system that gained widespread usage is the AlignRT system (Vision RT, London, UK) [Bert 05, Scho 07, Kren 09, Peng 10, Scha 12]. The AlignRT system acquires dense range data using the triangulation principle with active illumination and employs multiple sensors for the purpose of high body coverage and occlusion avoidance. The surface model consist of approximately 10 000 3-D points with the highest reported frame-rates of 15 Hz [Scho 07] but practically in the scale of 1 Hz to 3 Hz [Bert 05, Peng 10].

The limited acquisition speed of the AlignRT system led to new developments such as the body surface measurement systems proposed by Price et al. [Pric 12]. This system is based on *fringe projection profilometry* and can thus be categorized as a structured light technique. Reported number of surface points are up to 512×512 with acquisition frame-rates exceeding 20 Hz [Pric 12]. Further, a recent development is the TOPOS system (cyberTECHNOLOGIES GmbH, Ingolstadt, Germany) proposed by Lindl et al. [Lind 13]. This system is based on structured light and features a dual-view surface acquisition with approximately 600 000 3-D points at refresh rates in the scale of 10 Hz.

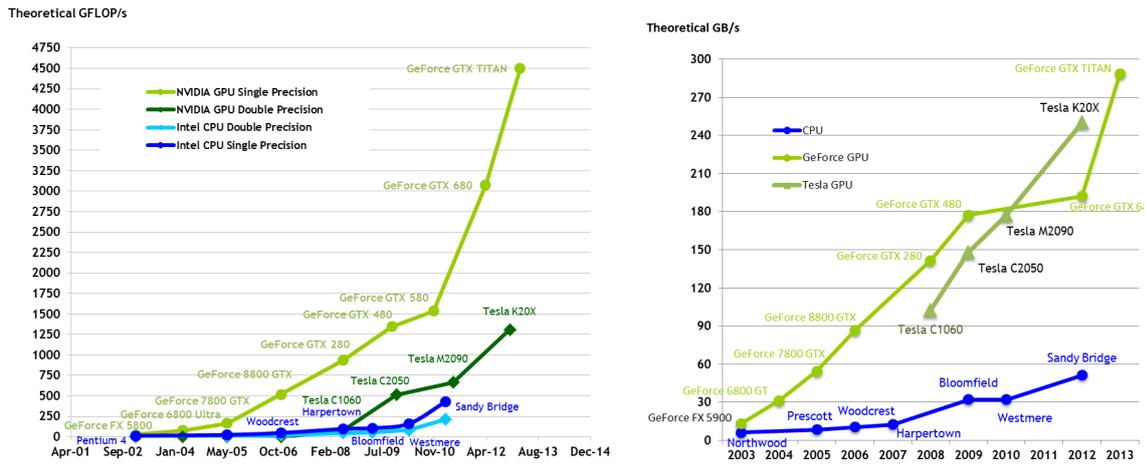


Figure 2.3: Comparison of CPU and GPU FLOPS (left) and memory bandwidth (right). Note the huge discrepancy between CPU and GPU architectures that are due to the fact that GPU devote more resources for computing rather than caching and flow control. Graphics taken from [NVID 13].

Note that sequential surface acquisition techniques as for example used by the Sentinel system (C-RAD AB, Uppsala, Sweden) are of no relevance for respiratory motion analysis due to the long acquisition times that are needed for acquiring the entire torso. Further, though ToF sensors have been proposed over the years for respiratory motion management [Scha 08, Penn 08] such sensors are still limited to research purposes mainly due to the low reliability of the acquired range data.

2.3 General Purpose Computing on GPUs

Traditionally, GPUs have been designed and used for rendering in computer graphics. However, over the last years GPUs have evolved into powerful general purpose processors for a broad variety of applications apart from computer graphics such as CT reconstruction [Sche 07b], astrophysical simulations [Nyla 07] or computational fluid dynamics [Corr 11]. The reason for this increasing popularity is the higher computation power w. r. t. Floating Point Operations Per Second (FLOPS) and memory bandwidth of modern GPUs in comparison to conventional Central Processing Units (CPUs), see Fig. 2.3 for an illustration. GPUs are designed for compute-intensive and massively parallel computation tasks and thus, in contrast to CPUs, can devote more transistors to processing rather than caching or control flow. In particular, problems that exhibit a high level of data parallelism, i. e. the same instructions can be issued on multiple data elements in parallel, are particularly suited for GPUs. As shown in this thesis, respiratory motion analysis based on RI fits this paradigm. This includes conventional 2-D image pre-processing methods from Chapter 4, memory demanding volumetric RI data fusion techniques investigated in Chapter 5, or sophisticated iterative RI surface registration algorithms as covered in Chapter 7. The following paragraphs briefly outline the general GPU hardware architecture and GPU programming paradigms.

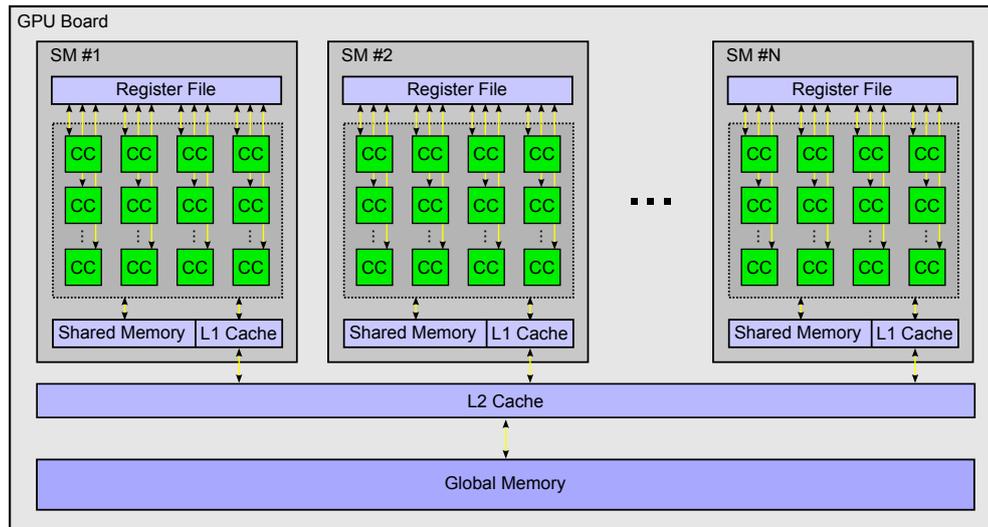


Figure 2.4: *Kepler* GPU architecture. The GPU board consists of an array of multi-threaded SMs each composed of several CCs that handle integer and floating point operations. The memory divides into registers for thread-specific data, fast on-chip memory shared across threads and global memory accessible across SMs.

2.3.1 GPU Architectures

Though the techniques and approaches investigated in this work are generic in the sense that they can be executed on most modern GPUs, the focus is on off-the-shelf consumer GPUs using NVIDIA's *Kepler* architecture and the Compute Unified Device Architecture (CUDA) framework⁴ [NVID13, Wilt13].

The *Kepler* GPU architecture is schematically depicted in Fig. 2.4. The GPU board is built around a scalable array of streaming multiprocessors (SMs) each composed of 192 so-called CUDA cores (CCs) that handle the actual integer and floating point operations. The number of SMs per GPU varies across different models and ranges from 8 SMs for the consumer desktop GPU GeForce GTX680 to 15 SMs with the Tesla K40 that is designed for professional workstation usage. The general design principle of *Kepler* GPUs is similar to older architectures such as the *Fermi* series that features a different number of cores per multi-processor.

Execution Model

The CUDA execution model is to divide a given task into a grid that consists of a problem specific number of *thread blocks* each consisting of a fixed number of *light-weighted* threads that execute the same GPU program called *kernel* in parallel, see Fig. 2.5. The term *light-weighted* means that a context switch between threads can be performed instantaneously without the need to save the thread's current state such as currently used registers as it is common with CPU context switches. The thread blocks are distributed across the multiprocessors and the threads of a block execute concurrently on one SM. More detailed, threads of a block are grouped into so-called *warps* of 32 parallel threads that are scheduled for execution and run

⁴<https://developer.nvidia.com/category/zone/cuda-zone>

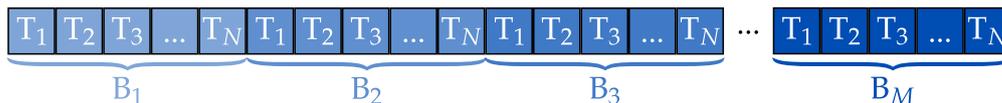


Figure 2.5: The CUDA execution model. A problem of size $N \cdot M$ is divided into a grid of M blocks B_m each consisting of N threads T_n that execute the same GPU program concurrently. Multi-dimensional grids and blocks are supported, too.

in parallel on different CCs on a specific SM. The number of *warps* that can reside simultaneously on one SM is limited by the resource requirements of the kernel so that the *light-weighted* context switch paradigm can be satisfied. The ratio of resident *warps* to the maximum number of *warps* is the occupancy and can have substantial impact on the performance. The reason for this is that latencies, as for example caused by memory transactions, can be hidden by switching to *warps* that are currently not stalled and ready for execution. For details see [NVID 13].

Memory Model

The memory model of the *Kepler* architecture exposes a hierarchy of varying capacity and bandwidth. Each thread has access to its own registers but not to those of other threads. Registers feature a very high bandwidth in the scale of 8000 GB/s but are limited to a maximum of 63 (GK104) or 255 (GK110) registers per thread and a total of 64 K per SM. On-chip memory with a high bandwidth of about 1600 GB/s is provided per SM with a capacity of typically 48 kB. This memory can be used to share data between threads that reside in the same block or as a data cache for individual threads. The former is of particular importance if the threads access the same data. Sharing may help to reduce read operations from off-chip global memory that typically features a rather low bandwidth of < 200 GB/s.

2.3.2 GPU Programming Paradigms

A detailed investigation of general GPU programming paradigms and best-practice techniques is not within the scope of this thesis and the reader is referred to the CUDA programming guide [NVID 13]. However, Appendix A covers a detailed analysis of a problem that is typical for image processing as investigated throughout this thesis. In particular, this case study focuses on a proper design w. r. t. the hierarchical GPU memory model. For an image resolution of 640×480 pixel and an off-the-shelf GTX 680 GPU, the case study shows that a naive implementation has a run-time of 0.6 ms whereas an optimized version features a superior run-time of only 0.06 ms. Though the absolute values are low for both approaches and the difference is of no practical relevance when considering such a processing step in isolation, a speedup of a factor of 10 is of high relevance for filter pipelines in a multi-sensor setup or complex iterative algorithms that require several hundreds of steps. For respiratory motion analysis as investigated in this thesis this also means that latencies are reduced, thus allowing for an instantaneous system response with minimal time lags satisfying the *as fast as possible* recommendation by the AAPM task group 76 for respiratory motion monitoring systems [Keal 06].

Part II

Real-time Range Imaging

Foundations of Range Imaging

3.1 Acquisition Geometry in Range Imaging	19
3.2 Data Structures and Data Representation	21
3.3 Synthetic Range Data By Sensor Simulation	24
3.4 RITK: A Framework for 3-D Range Image Stream Processing	25

This chapter covers the basic concepts and principles of range imaging as investigated in this thesis. In particular, the chapter elaborates on the acquisition geometry relating different sensor coordinate systems, the projective geometry and reconstruction principles that govern RI as well as auxiliary data structures. As a sideline, this chapter introduces the basic nomenclature and RI principles used throughout the thesis. Further, the software framework that was developed in the course of the thesis at hand is outlined.

3.1 Acquisition Geometry in Range Imaging

The concepts and techniques investigated in this thesis rely on multiple *local* acquisition coordinate systems C_k associated with the k -th camera that are embedded in a *global* world coordinate system W . Without loss of generality, this world coordinate system is defined by the canonical origin $\mathbf{o}_W \in \mathbb{R}^3$ and the mutual orthonormal spanning vectors $\{\mathbf{e}_{W,1}, \mathbf{e}_{W,2}, \mathbf{e}_{W,3}\}$, $\mathbf{e}_{W,i} \in \mathbb{R}^3$ defined as:

$$\mathbf{o}_W = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_{W,1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_{W,2} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_{W,3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.1)$$

Similarly, the acquisition coordinate C_k system associated with the k -th camera is defined by an arbitrary origin $\mathbf{o}_k \in \mathbb{R}^3$ and the mutual orthonormal spanning vectors $\{\mathbf{e}_{k,1}, \mathbf{e}_{k,2}, \mathbf{e}_{k,3}\}$, $\mathbf{e}_{k,i} \in \mathbb{R}^3$. The origin \mathbf{o}_k and spanning vectors $\mathbf{e}_{k,i}$ define the position and orientation of the k -th camera in world space as depicted in Fig. 3.1.

Based upon this representation, the transformation T_k from world space W to the k -th camera space C_k can be expressed using homogeneous coordinates as:

$$T_k = \begin{pmatrix} \mathbf{R}_k & \mathbf{t}_k \\ \mathbf{0} & 1 \end{pmatrix}^{-1} \in \mathbb{R}^{4 \times 4}, \quad (3.2)$$

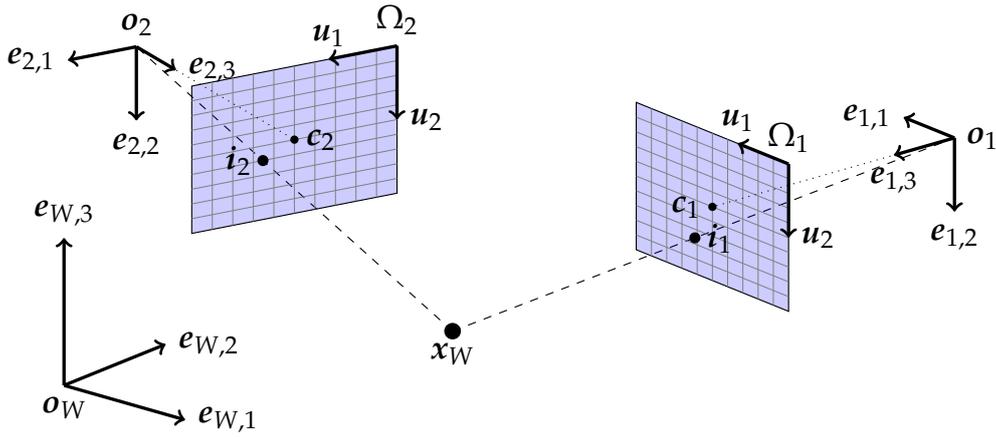


Figure 3.1: The RI acquisition geometry as investigated in this thesis. Two *local* camera coordinates systems C_1 and C_2 are embedded in the *global* world space W . An arbitrary point x_W defined in world space W is then perspectively projected into the image domains Ω_k associated with the k -th camera, yielding the pixel indices $i_1 \in \Omega_1$ and $i_2 \in \Omega_2$, respectively.

where the translation $t_k \in \mathbb{R}^3$ and rotation $R_k \in \text{SO}_3$ are given as:

$$t_k = o_k, \quad R_k = (e_{k,1} \ e_{k,2} \ e_{k,3}). \quad (3.3)$$

Thus, for an arbitrary point $x_W \in \mathbb{R}^3$ defined in world space W , the corresponding representation $x_k \in \mathbb{R}^3$ in the k -th local camera space C_k is given as:

$$\begin{pmatrix} x_k \\ 1 \end{pmatrix} = T_k \begin{pmatrix} x_W \\ 1 \end{pmatrix}. \quad (3.4)$$

Now, let $\tilde{x}_W = (x_W, 1)^\top \in \mathbb{R}^4$ denote the homogeneous representation of x_W in world space and let the perspective projection of \tilde{x}_W into the k -th image domain $\Omega_k \subset \mathbb{R}^2$ be given by the pixel index $i_k \in \Omega_k$. Mathematically, this projection is expressed in homogeneous coordinates as:

$$i_k = \begin{pmatrix} i_{k,1} \\ i_{k,2} \end{pmatrix} = \begin{pmatrix} \tilde{i}_{k,1}/\tilde{i}_{k,3} \\ \tilde{i}_{k,2}/\tilde{i}_{k,3} \end{pmatrix} \curvearrowright \begin{pmatrix} \tilde{i}_{k,1} \\ \tilde{i}_{k,2} \\ \tilde{i}_{k,3} \end{pmatrix} = \tilde{i}_k = C_k (\mathbf{I} \ \mathbf{0}) T_k \tilde{x}_W, \quad (3.5)$$

where \curvearrowright denotes the de-homogenization induced by perspective division. Further, $\mathbf{I} \in \mathbb{R}^{3 \times 3}$ denotes the identity matrix and $\mathbf{0} \in \mathbb{R}^3$ the zero vector. The *camera matrix* $C_k \in \mathbb{R}^{3 \times 3}$ is given as:

$$C_k = \begin{pmatrix} f_{k,1} & 0 & c_{k,1} \\ 0 & f_{k,2} & c_{k,2} \\ 0 & 0 & 1 \end{pmatrix}, \quad (3.6)$$

where $f_{k,i}$ and $c_{k,i}$ denote the camera's focal length and principal point in w. r. t. the i -th image axis u_i , respectively. See Fig. 3.1 for an illustration. Note that C_k is only

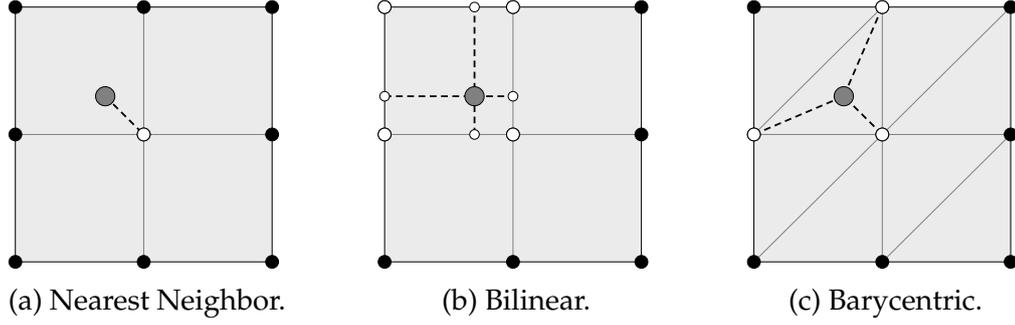


Figure 3.2: Interpolation strategies. The black dots denote depth samples obtained from the RI sensor whereas the gray dot denotes an intermediate pixel index at which the range image needs to be interpolated. White dots denote depth samples that are used to interpolate the range image at the intermediate pixel index.

an approximation as the skew parameters are neglected for simplicity. For convenience, for the remainder of this thesis, the perspective projection of a point x_W in world space into the k -th domain Ω_k as described in Eq. (3.5) or the projection of a point x_k in the k -th camera space into Ω_k will be denoted compactly by projection operators P_k and P_k^T defined as:

$$i_k = P_k(x_k) = P_k^T(x_W). \quad (3.7)$$

Though the mapping between multiple *local* coordinate systems C_k and the projection geometry described in this section is explicitly formulated for camera spaces the same principles and linear transformations can be directly applied for other *local* coordinate systems. This includes coordinate systems associated with medical imaging devices such as C-Arm CTs and LINACs or auxiliary data structures such as signed distance functions that are investigated in Section 5.3.1 for implicit surface representation in a multi-view RI scenario.

3.2 Data Structures and Data Representation

Besides the actual range measurements, modern RI sensors provide complementary information including hardware-based reliability indicators or photometric data. Further, it is usually mandatory to reconstruct a 3-D surface from the observed 1-D range measurements and auxiliary data structures such as *confidence maps* encoding the reliability of measurements must be computed prior to further RI data analysis.

This section covers the data structures and their representation used throughout this thesis. However, note that though often available, photometric information is not investigated further as it assumes no role in this thesis. Due to the Charge-coupled Device (CCD) sensor matrix inherent to the RI sensors investigated in this thesis, all data structures are defined as functions on the k -th RI domain Ω_k sampled on a rectangular grid at $N_1 \times N_2$ indices $i = (i_1, i_2)$. This implies the need to interpolate the functions for continuous indices that do not coincide

with discrete sampling positions. Due to the regular 2-D sampling grid topology of Ω_k , standard interpolation techniques such as nearest neighbor, bi-linear or barycentric interpolation can be used, see Fig. 3.2 for an illustration. Further, this regular grid topology allows to denote $\Omega_{k,i}^r \subset \Omega_k$ as the local quadratic neighborhood with radius r of pixel index $i \in \Omega_k$ w. r. t. the image domain Ω_k according to:

$$\Omega_{k,i}^r = \{i' \in \Omega_k \mid \max(|i'_1 - i_1|, |i'_2 - i_2|) \leq r\}. \quad (3.8)$$

The number of pixels in this quadratic region is given as $|\Omega_{k,i}^r| = (2r + 1)^2$.

3.2.1 Range Measurements and Reconstructed Surface

The RI sensors investigated in this thesis deliver dense 1-D depth or range measurements. The terms range measurements and depth measurements are used interchangeable in the remainder of this thesis. A range image associated with the k -th camera is denoted as the function:

$$\mathcal{R}_k : \Omega_k \rightarrow \mathbb{R}^+, \quad (3.9)$$

where in practice it is often differentiated between *radial* depth values $\mathcal{R}_k^<$ and *orthogonal* measurements \mathcal{R}_k^\perp . For a point x_W in world space the measurements are defined according to:

$$\mathcal{R}_k^<(P_k^T(x_W)) = \mathcal{R}_k^<(P_k(x_k)) = \|x_k\|_2, \quad (3.10)$$

$$\mathcal{R}_k^\perp(P_k^T(x_W)) = \mathcal{R}_k^\perp(P_k(x_k)) = (0, 0, 1)x_k. \quad (3.11)$$

The type of range measurement representation usually depends on the RI sensor, however the two representations can be seamlessly converted into each other. A detailed description is provided in Appendix C.

It is often desirable and for certain applications even mandatory to reconstruct the 3-D scene that corresponds to the range measurements. Similar to the representation of the actual range data in Eq. (3.9), the surface \mathcal{S}_k corresponding to \mathcal{R}_k is denoted as a function or graph:

$$\mathcal{S}_k : \Omega_k \rightarrow \Psi_k, \quad (3.12)$$

where the codomain $\Psi_k \subset \mathbb{R}^3$ holds the 3-D points in camera space as:

$$\mathcal{S}_k(i) = o_{k,i} + \mathcal{R}_k(i) \cdot v_{k,i}. \quad (3.13)$$

Here, $o_{k,i} \in \mathbb{R}^3$ and $v_{k,i} \in \mathbb{R}^3$ denote the origin and direction of the viewing ray associated with index $i \in \Omega_k$, respectively. For a pinhole camera model the viewing ray origins $o_{k,i}$ are defined by the camera's optical center, i. e. $o_{k,i} = o_k, \forall i \in \Omega_k$. However, for virtual camera models as for example investigated in Section 5.4 of this thesis, a generic reconstruction model covering multiple origins is mandatory. The derivations of the viewing rays from the perspective geometry in RI and *radial* or *orthogonal* depth measurements are detailed in Appendix C.

Finally, surface normals $\mathbf{n}_{k,i} \in \mathcal{N}_k : \Omega_k \rightarrow \mathbb{R}^3$ can be computed from \mathcal{S}_k using for example forward differences on the regular grid topology:

$$\mathcal{N}_k(\mathbf{i}) = \left(\mathcal{S}_k(\mathbf{i} + (\Delta, 0)^\top) - \mathcal{S}_k(\mathbf{i}) \right) \times \left(\mathcal{S}_k(\mathbf{i} + (0, \Delta)^\top) - \mathcal{S}_k(\mathbf{i}) \right), \quad (3.14)$$

where $\Delta \in \mathbb{R}^+$ denotes an arbitrary but fixed offset. For completeness, note that surface normals can be computed using backward or central differences, too.

3.2.2 Confidence Maps and Reliability Indicators

Encoding the reliability of range or surface measurements is a crucial issue for RI data analysis. For example this allows to detect outliers or invalid data to be restored or enhanced using pre-processing techniques as discussed in Chapter 4. Further, confidence maps can be used as a prior probability for parameter estimation such as the multi-view data fusion technique detailed in Chapter 5. For notational clarity, the sensor index k is omitted in this section. A *confidence map* is denoted as:

$$\mathcal{C} : \Omega \rightarrow [0, 1], \quad (3.15)$$

where, $\mathcal{C}(\mathbf{i}) = 1$ implies maximum reliability and $\mathcal{C}(\mathbf{i}) = 0$ identifies invalid data w. r. t. the measurement at position \mathbf{i} . The *confidence map* \mathcal{C} as used in this work is generically composed of several reliability indicators $\mathcal{C}_j : \Omega \rightarrow [0, 1]$ as:

$$\mathcal{C}(\mathbf{i}) = \prod_j \mathcal{C}_j(\mathbf{i}). \quad (3.16)$$

In this work, the following four indicators have been identified as most important. These indicators include clues that are directly obtained from a specific RI sensor as well as generic maps that are independent of the particular RI device and are computed from the RI measurements.

Sensor-based Confidence In the simplest form a confidence map can be obtained from the RI sensor directly. This sensor-based confidence is simply defined as:

$$\mathcal{C}_1(\mathbf{i}) = \begin{cases} 1 & \text{if } \mathcal{R}(\mathbf{i}) \text{ is valid w. r. t. the RI device} \\ 0 & \text{else} \end{cases}. \quad (3.17)$$

For example, ToF cameras usually provide an indicator to identify *flying pixels* or saturated and overexposed sensor elements. For the Kinect device a binary validity information can be obtained by checking the range data for zero values.

Acquisition Distance The range measurements itself is a reliability indicator as depth and spatial resolution in range imaging deteriorates with the distance to the captured object. For example, depth maps acquired by the Kinect device are known to exhibit a degradation of depth resolution that is proportionally to the squared distance [Smis 11]. The confidence accounting for the acquisition distance is thus given as:

$$\mathcal{C}_2(\mathbf{i}) = \frac{1}{1 + \sqrt{\mathcal{R}(\mathbf{i})}}, \quad (3.18)$$

where the square root was chosen heuristically.

Acquisition Direction As originally proposed by Curless and Levoy, the angle between the surface normal and the viewing direction is a measure for uncertainty due to degrading illumination for surface regions with normals that are nearly perpendicular to the viewing direction [Curl96]. However, the discrepancy between the surface normal and viewing ray is additionally a direct indicator for the ill-conditioned surface sampling density imposed by the pinhole camera model of RI sensors, i. e. the steeper the edge the lower the spatial sampling. The confidence that accounts for the acquisition direction is given as:

$$C_3(\mathbf{i}) = \left(\mathbf{v}_i^\top \mathcal{N}(\mathbf{i}) \right)^2, \quad (3.19)$$

where \mathbf{v}_i denotes the viewing ray and $\mathcal{N}(\mathbf{i})$ is the normal at index \mathbf{i} , respectively. Squaring the scalar product results in a rapid confidence decay for shallow acquisition angles and was chosen heuristically.

Gradient of Depth Measurements Motivated by the work of Friske and Perry [Fris02], the gradient of the depth measurements $\mathcal{R}(\mathbf{i})$ is incorporated into the composite confidence map. However, instead of using this information for correcting depth measurements as originally proposed, the rationale here is that a high gradient magnitude is a strong indicator for an object boundary. Depth measurements at steep edges or scene discontinuities are often prone to errors such as *flying pixels* in ToF-imaging [Kolb10]. Consequently, the confidence w. r. t. the gradient of depth measurements is given as:

$$C_4(\mathbf{i}) = \frac{1}{1 + \|\nabla \mathcal{R}(\mathbf{i})\|_2}, \quad (3.20)$$

where $\nabla \mathcal{R} : \Omega \rightarrow \mathbb{R}^2$ denotes the gradient of \mathcal{R} . In practice, $\nabla \mathcal{R}$ is computed by convolving \mathcal{R} with a suitable derivative kernel, cf. Chapter 4.

Combining these clues results in high scores for smooth surface regions that are perpendicular to the viewing direction and close to the RI sensor, an illustration using a toy scene is depicted in Fig. 3.3. Note that a binary validity mask $\mathcal{M} : \Omega \rightarrow \{0, 1\}$ can be derived from the confidence map \mathcal{C} by thresholding.

3.3 Synthetic Range Data By Sensor Simulation

For assessing the suitability and validating the correctness of RI processing modules it is often necessary to establish a ground truth reference baseline. In practice this is a difficult task as the ideal 3-D object corresponding to the acquired range measurements is in general not known. However, even if a ground truth baseline of the observed 3-D object is available, for example using high precision tomographic scans or complementary range scans, manual interaction is often required and calibration issues or registration steps to bring both coordinate systems in congruence may introduce additional errors.

A different approach for baseline generation is based on simulated range data obtained from rendering a virtual 3-D scene by using techniques known from computer graphics. The underlying principle here is that the z-buffer representation

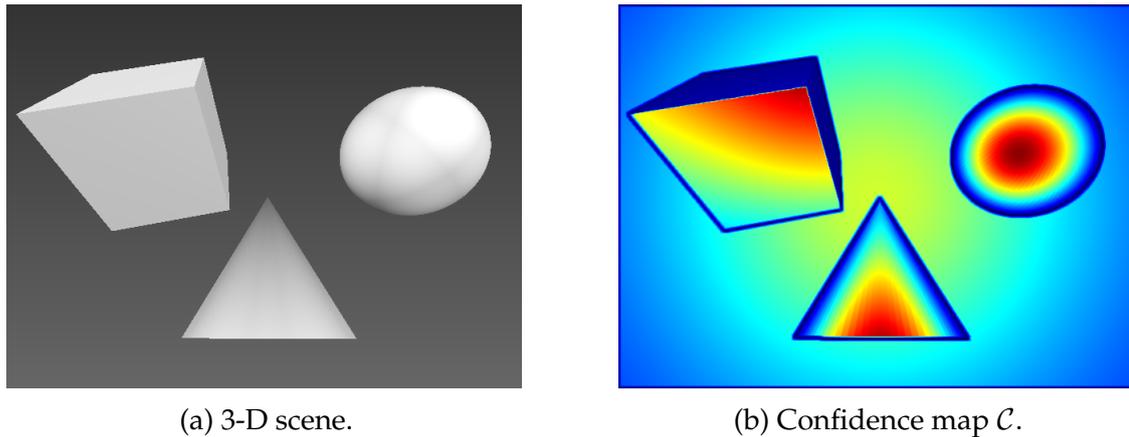


Figure 3.3: Toy scene illustrating a confidence map \mathcal{C} . Three geometric primitives are placed in front of a plane parallel to the RI sensor matrix. Confidence values are mapped to $[0, 1]$ and color coded with blue tones denoting low and red tones representing high certainty. Note the effect of the different terms such as the low confidence at scene discontinuities or the halo-like appearance of the background plane due to the acquisition direction weighted by the distance to the RI sensor.

[Fole 90] of the 3-D scene resembles real-world orthogonal range measurements. Further, the color buffer can be interpreted as complementary photometric data and may be used to simulate the effect of specular reflections or material dependent noise characteristics.

The majority of publications use the standard depth and color buffers including the work of Köhler et al. [Kohl 13] or Keller et al. [Kell 13]. However, Keller et al. also proposed sophisticated device specific simulation programs consisting of multiple render passes [Kell 07, Kell 09].

3.4 RITK: The Range Imaging Toolkit - A Framework for 3-D Range Image Stream Processing

Regardless of the fundamentally different underlying physical sampling principles (cf. Section 2.2), the RI sensors investigated in this thesis provide dense and metric surface information at real-time frame-rates. This poses a challenge in terms of both throughput as well as low latency for subsequent data processing algorithms. For example, off-the-shelf RI devices featuring Video Graphics Array (VGA)-resolution (640×480 px), a typical frame-rate of 30 Hz and single-precision range data representation 32 bit/px generate a bandwidth of

$$640 \text{ px} \times 480 \text{ px} \cdot 32 \text{ bit/px} \cdot 30/\text{s} \sim 300 \text{ Mbit/px}.$$

By using the equivalent 3-D point cloud or surface representation this corresponds to 900 Mbit/s and if complementary data such as photometric information is available the amount of data exceeds 1 Gbit/s. This requires a sophisticated software architecture and data representation in order to comply with real-time constraints

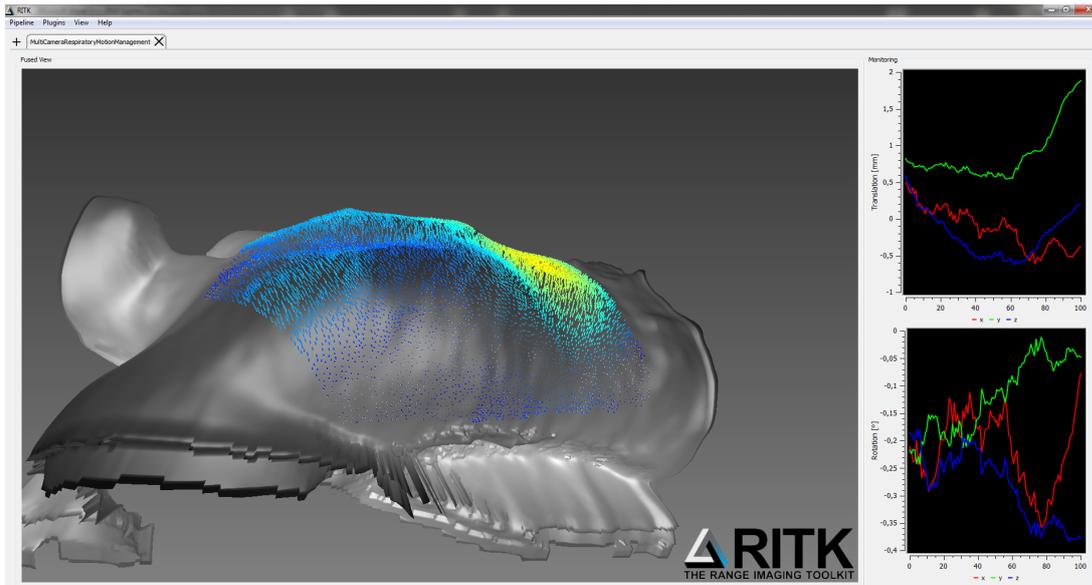


Figure 3.4: A screenshot of RITK used for respiratory motion analysis. RITK is specifically designed for real-time scenarios and supports the integration of dedicated hardware such as GPUs for high performance computing.

in clinical practice. In particular with regard to the integration of dedicated hardware such as GPUs, innovative concepts for data pipeline propagation and memory management are inevitable to minimize the overhead of data transfers.

To cope with the general demands to the underlying software in RI-based respiration analysis the Range Imaging Toolkit (RITK) was developed during the course of this work [Wasz 11a]. RITK is an open source cross-platform and object oriented toolkit written in C/C++ that is explicitly designed for real-time processing of multiple high-bandwidth RI data streams generated by modern RI devices. For this purpose, RITK is conceived with performance in mind, supporting modern multi-core CPU and many-core GPU architectures as well as providing an interface for integrated circuits such as Field-programmable Gate Array (FPGA)

As real-time visualization of dynamic 3-D point clouds and surface data is mandatory for immediate user feedback and scene understanding, RITK employs a resource-efficient dedicated graphics module taking advantage of the interoperability of general purpose computing on the GPU and rendering.

Besides these fundamental concepts, RITK supports both integration of various RI sensors as well as customized pre-processing pipelines at run-time and provides a file format that enables recording and chronologically correct replay of RI data sequences or static snapshots.

A screenshot of RITK deployed in a respiratory motion analysis scenario is given in Fig. 3.4. However, besides respiration analysis, RITK was also successfully used in other range imaging applications including 3-D endoscopy [Haas 13b, Haas 13a, Kohl 13] or 3-D scene reconstruction [Neum 11, Baue 13b].

Range Image Pre-Processing

4.1 Motivation	27
4.2 Related Work	28
4.3 Pre-Processing Pipeline Overview	29
4.4 Restoration of Invalid Depth Measurements	30
4.5 Edge-preserving Denoising	35
4.6 Temporal Denoising	38
4.7 Experiments and Results	39
4.8 Discussion and Conclusion	51

Raw range data directly obtained from RI sensors commonly exhibits a high noise level and is further deteriorated by invalid measurements or systematic errors. This chapter elaborates on RI data pre-processing techniques as a fundamental prerequisite for practical applications. The methods investigated in this chapter are based on preceding work for RI data pre-processing that had a focus on ToF data enhancement in organ surface acquisition [Wasz 11b] and real-time computation using GPUs [Wasz 11c]. To account for the background of the thesis at hand, this chapter explicitly investigates these pre-processing techniques in a dynamic respiratory motion analysis scenario with strict real-time constraints and low latency requirements.

4.1 Motivation

Regardless of the actual sensor technology, raw range data directly obtained from RI sensors commonly exhibits a low signal-to-noise ratio (SNR) and is further deteriorated by invalid measurements or systematic errors. These issues often preclude the raw RI data to be directly used for subsequent computation steps. In particular, this holds true for ToF cameras and off-the-shelf consumer devices such as Microsoft Kinect. With ToF imaging, erroneous range values typically result from overexposed and saturated sensor elements caused by specular reflections or the *flying pixels* at sharp object boundaries [Kolb 10]. For the Kinect sensor difficulties arise due to the ubiquitous discretization steps or when capturing reflective or transparent objects [Smis 11]. Typical artifacts in RI are depicted in Fig. 4.1. With

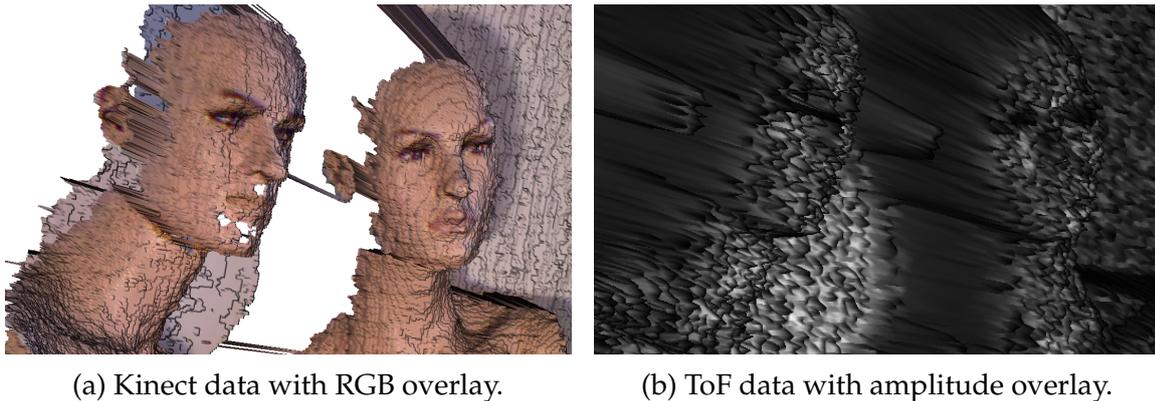


Figure 4.1: Anthropomorphic phantom study illustrating the raw RI surface data as obtained from Microsoft Kinect and a **pmd**technologies CamCube 3.0 ToF camera [Wasz 11c]. For the Kinect note the ubiquitous discretization artifacts and the invalid regions around the male’s chin and nose. For ToF imaging note the low SNR throughout the scene.

regard to the high level of accuracy and reliability required for respiratory motion analysis, RI pre-processing strategies thus constitute a mandatory prerequisite for the applications addressed subsequently.

In general, all image enhancement techniques that have been proposed for image processing or computer vision can be applied to the problem of RI data pre-processing. However, due to real-time constraints in respiratory motion analysis, RI pre-processing techniques must be tractable w. r. t. the computational complexity and have to be eligible for porting on dedicated hardware for acceleration. As a consequence, RI pre-processing in the context of this thesis is always a trade-off between data enhancement and real-time capability.

4.2 Related Work

In image processing, edge-preserving filters for smoothing homogeneous regions while preserving scene discontinuities are of special interest and importance. One of the most popular and established methods is the bilateral filter [Auri, Toma 98]. Beyond its application for a multitude of conventional imaging modalities, it is a common choice for RI data denoising [Lind 10]. The filter is straightforward to implement, but exhibits a poor run-time performance due to its non-linear nature. Recent algorithmic acceleration concepts have attempted to overcome the inherent computational complexity by quantization and approximation techniques, however with the drawback of impairing accuracy [Pari 09, Pori 08, Yang 09, Yosh 10]. In contrast, the concept of guided filtering proposed by He et al. [He 10, He 13] is based on a non-approximative algorithm with a computational complexity that is independent of the filter kernel size. At the same time, it exhibits a comparable degree of edge-preserving smoothing and does not suffer from gradient reversal artifacts. Besides algorithmic acceleration concepts, GPUs assume a prominent role for high-performance data processing. Chen et al. demonstrated real-time frame

rates for high-definition video processing by accelerating the bilateral grid on the GPU [Chen 07]. Furthermore, GPUs were successfully deployed for accelerating denoising and resolution enhancement of RI data, however, without explicitly addressing real-time constraints [Huhl 10].

Prior to denoising the data, the restoration of invalid measurements has to be performed. In contrast to static defect pixels, these invalid measurements occur unpredictable and can affect both an isolated pixel or connected local regions as for example the depth shadows occurring at edge boundaries with the Kinect device [Camp 13]. In computer vision, a plurality of inpainting methods to correct for such artifacts have been proposed. One simple yet effective technique is the normalized convolution proposed by Knutson [Knut 93] that was also applied for range imaging by Frank et al. [Fran 09], however, without explicitly addressing the problem of restoring invalid measurements. More sophisticated methods rely on variational formulations [Getr 12] or spectral domain analysis [Aach 01].

For range imaging, hybrid joint denoising and inpainting techniques have also been proposed. For example, Silberman and Fergus use the cross bilateral filter proposed by Paris and Durand [Pari 06] to compensate for both noise and missing depth values [Silb 11]. Here, the inpainting is guided by the complementary photometric information provided by Kinect sensor. Further hybrid joint denoising techniques have been proposed by Li et al. [Liu 12] and Camplani et al. [Camp 13] for the Kinect and Huhle et al. for ToF data enhancement and super-resolution [Huhl 10]. However, note that these approaches implicitly assume that there exists a correlation between color and depth data which often does not hold true. An example is a flat plane disturbed by glare-lights or that is textured with a checkerboard pattern. In practice, sophisticated techniques are required to account for this issue as for example reported by Köhler et al. for hybrid 3-D endoscopy [Kohl 13].

Further, device and application specific denoising concepts have also been proposed. For example Edeler et al. use prior noise information for ToF data enhancement [Edel 10] and Bauer et al. proposed a framework for joint denoising and registration of RI data [Baue 12b].

4.3 Pre-Processing Pipeline Overview

The proposed pipeline setup is motivated by the observation that independent of the underlying physical reason the raw range data is deteriorated by (i) missing or invalid measurements, (ii) sensor noise or quantization issues and (iii) temporal variations. Note that the correction of systematic errors like the so-called *wiggling* effect for ToF imaging is not part of pre-processing. Consequently, the proposed pipeline restores invalid or missing depth measurements as a first step (Section 4.4). This renders an extra conditioning of invalid data unnecessary in subsequent pre-processing steps. Subsequently, edge-preserving *spatial* filtering is performed (Section 4.5) and as a finalizing step methods to account for *temporal* noise are applied (Section 4.6).

This modular setup and subsequent processing was favored over a joint approach due to performance considerations. In fact, a combined spatial and temporal approach would resemble a conventional filter used in volumetric denoising,

i. e. the first two dimensions are along the image whereas the third dimension accounts for the temporal aspect. Depending on the chosen techniques, these filters exhibit a high computational complexity, especially if they are not fully separable. In this regard, the explicit split between spatial and temporal filtering can be interpreted as enforced separability along the temporal dimension.

The pre-processing techniques investigated in this section are purposely chosen not to include any photometric data such as RGB color information. First, RI devices are not guaranteed to feature complementary color information. Second, the correlation of range data and color information that is inherently required for joint approaches does not hold true in respiratory motion analysis. For a bare-chested scenario, depending on gender and complexion, there potentially exists a strong color gradient around the mamilla whereas the corresponding region in the range data is approximately constant. Further, for the most part, the body surface is rather texture-less which renders the benefit of incorporating additional photometric information questionable.

For notational clarity, the sensor index k will be omitted throughout this chapter. Unless stated otherwise, RI data pre-processing is always performed in the 2-D sampling domain Ω using the 1-D depth measurements \mathcal{R} and not the corresponding 3-D surface data \mathcal{S} . The reason for this is twofold: First, the computational complexity of pre-processing algorithms operating on scalar valued depth data is substantially lower compared to algorithms that use the vector-valued 3-D point data. Second, processing the 3-D point data \mathcal{S} may alter the corresponding indices $i \in \Omega$ and thus implies the need to resample the range image \mathcal{R} . Further, respiratory motion analysis techniques may directly operate on the range data and do not require a closed surface representation of the body as for example pursued by Fayad et al. [Faya 11].

4.4 Restoration of Invalid Depth Measurements

In contrast to the continuous range and surface representation in Section 3.2, conventional interpolation techniques that solely rely on adjacent range values are not an option for restoration of invalid measurements. This is due to the fact that invalid or missing samples occur unpredictable and are often grouped to clusters as previously illustrated in Fig. 4.1a. Invalid or erroneous measurements are identified by a binary mask \mathcal{M} that can be derived from the confidence maps \mathcal{C} introduced in Section 3.2.2. As valid measurements are supposed to remain unchanged the following generic model will be used for the remainder of this section. Let \mathcal{R}_R denote the intermediate output of a restoration filter R . The final output data $\widehat{\mathcal{R}}_R$ is then given according to:

$$\widehat{\mathcal{R}}_R(i) = \mathcal{M}(i) \mathcal{R}(i) + (1 - \mathcal{M}(i)) \mathcal{R}_R(i). \quad (4.1)$$

This model ensures that only invalid or missing range data are replaced by estimated values. For convenience, missing range values will be referred to as invalid measurements for the remainder of this thesis.

4.4.1 Normalized Convolution

One method for restoring invalid depth measurements is to reconstruct an unknown value $\mathcal{R}(\mathbf{i})$ with $\mathcal{M}(\mathbf{i}) = 0$ by a weighted summation of neighboring valid samples $\mathcal{R}(\mathbf{i}')$ forming the set:

$$\{\mathcal{R}(\mathbf{i}') \mid \mathbf{i}' \in \Omega_i^r \wedge \mathcal{M}(\mathbf{i}') \neq 0\}. \quad (4.2)$$

Giving a high priority to nearby pixels, this weighted averaging is closely related to the concept of *Normalized Convolution* (NC) that was originally proposed by Knutsson and Westing [Knut93]. *Normalized Convolution* in the context of the thesis at hand builds upon the observation that invalid measurements are smeared across valid regions when naively convolving the erroneous range image \mathcal{R} with a low-pass kernel. The concept of *Normalized Convolution* accounts for this issue by separating the data into a signal part and a so-called *certainty* part encoding the reliability of a pixel value or measurement. Using the mask image \mathcal{M} as *certainty* function, this is formulated for the intermediate filter output \mathcal{R}_{NC} for Eq. (4.1) as:

$$\mathcal{R}_{\text{NC}}(\mathbf{i}) = \frac{\sum_{\mathbf{i}' \in \Omega_i^r} a(\mathbf{i}, \mathbf{i}') \mathcal{M}(\mathbf{i}') \mathcal{R}(\mathbf{i}')}{\sum_{\mathbf{i}' \in \Omega_i^r} a(\mathbf{i}, \mathbf{i}') \mathcal{M}(\mathbf{i}')}, \quad (4.3)$$

where the so-called *applicability* function $a(\mathbf{i}, \mathbf{i}')$ accounts for the spatial proximity in Ω_i^r . In this work, a Gaussian kernel function is used for the *applicability* function:

$$a(\mathbf{i}, \mathbf{i}') \propto \exp\left(-\frac{\|\mathbf{i} - \mathbf{i}'\|_2^2}{\sigma_s^2}\right), \quad (4.4)$$

where the parameter σ_s controls the spatial similarity or proximity. Regarding the original intention to reconstruct missing values based on neighboring valid measurements, Eq. (4.3) can be interpreted as a convex combination of valid samples with the denominator accounting for a proper normalization.

An important theoretical property of the restoration of invalid measurements employing the concept of *Normalized Convolution* from Eq. (4.3) in conjunction with the Gaussian *applicability* function in Eq. (4.4) lies in the interpretation as a convolution (*) using a linear shift-invariant Gaussian kernel \mathcal{G} :

$$\mathcal{R}_{\text{NC}}(\mathbf{i}) = \frac{\{\mathcal{G} * (\mathcal{M} \circ \mathcal{R})\}(\mathbf{i})}{\{\mathcal{G} * \mathcal{M}\}(\mathbf{i})}, \quad (4.5)$$

where \circ denotes the element-wise multiplication or *Hadamard product*. This property allows to decrease the algorithmic complexity of *Normalized Convolution*, in the naive formulation from Eq. (4.3) being $\mathcal{O}(N_1 N_2 r^2)$, in three ways: First, by exploiting the separability of Gaussian filtering the complexity can be decreased to $\mathcal{O}(N_1 N_2 r)$. The second option of acceleration is to exploit the convolution theorem for linear shift-invariant kernels. This theorem states that the convolution of a signal in the time or spatial domain with a linear shift-invariant kernel such as the Gaussian equals a multiplication in the frequency domain. For the denominator in Eq. (4.5) this can be expressed as:

$$\mathcal{G} * \mathcal{M} = \text{DFT}^{-1}(\text{DFT}(\mathcal{G} * \mathcal{M})) = \text{DFT}^{-1}(\text{DFT}(\mathcal{G}) \circ \text{DFT}(\mathcal{M})), \quad (4.6)$$

where DFT denotes the discrete Fourier transform (DFT) operator. By using this formulation, the complexity of the algorithm is decoupled from the kernel size r and by using the fast Fourier transform (FFT) is $O(N_1 N_2 (\log N_1 + \log N_2))$. Third, by using the recursive definition for Gaussian filtering as proposed by Deriche [Deri 90] the computational complexity of *Normalized Convolution* can be decoupled from the kernel radius r of the local neighborhood Ω_i^r , i. e. the complexity becomes $O(N_1 N_2)$.

4.4.2 Spectral Deconvolution

Though the restoration of invalid measurements using *Normalized Convolution* features an excellent computational complexity, the method has some drawbacks. First, for large defect regions, the assumption that missing values can be properly reconstructed by a weighted summation of surrounding valid measurements does not necessarily hold true. This is due to an over-smoothing implied by a large Gaussian kernel that ultimately results in the loss of salient image features such as edges. Second, the restoration of invalid measurements using *Normalized Convolution* is a spatial method that does not account for local image features such as repetitive texture patterns or intensity gradients. To cope with these issues, an alternative method for the reconstruction of invalid pixels based on spectral deconvolution, originally proposed by Aach and Metzler for defect interpolation in digital radiography [Aach 01], is investigated in this section.

The principle of this spectral deconvolution (SD) method is to model the raw and erroneous image \mathcal{R} as an element-wise multiplication of the ideal sought image \mathcal{R}_{SD} with the binary mask image \mathcal{M} containing the validity information. In the Fourier domain this multiplication equals a convolution and the governing reconstruction model can thus be expressed as:

$$\mathcal{R} = \mathcal{R}_{SD} \circ \mathcal{M} \equiv F_{\mathcal{R}} = \frac{1}{N_1 N_2} F_{\mathcal{R}_{SD}} * F_{\mathcal{M}}, \quad (4.7)$$

where $F_{\mathcal{R}}$, $F_{\mathcal{R}_{SD}}$ and $F_{\mathcal{M}}$ denote the frequency spectra of \mathcal{R} , \mathcal{R}_{SD} and \mathcal{M} , respectively. The effect of the validity mask \mathcal{M} is then removed by deconvolving its spectrum $F_{\mathcal{M}}$ from the spectrum of the corrupted range image $F_{\mathcal{R}}$.

The individual spectral coefficients $F_{\mathcal{R}_{SD}}(\mathbf{I}) \in \mathbb{C}$ with $\mathbf{I} \in \mathbb{N}^2$ denoting an index in the frequency domain are iteratively estimated in a sense that minimizes the error $E^{(t)}$ between the erroneous input $F_{\mathcal{R}}(\mathbf{I})$ and the t -th estimate for the sought spectrum as implied by the convolution model from Eq. (4.7) as:

$$E^{(t)} = \sum_{\mathbf{I}} \left| F_{\mathcal{R}}^{(t)}(\mathbf{I}) \right|^2 = \sum_{\mathbf{I}} \left| F_{\mathcal{R}}^{(t-1)}(\mathbf{I}) - \frac{1}{N_1 N_2} \left\{ \Delta F_{\mathcal{R}_{SD}}^{(t)} * F_{\mathcal{M}} \right\}(\mathbf{I}) \right|^2, \quad (4.8)$$

where $\Delta F_{\mathcal{R}_{SD}}^{(t)}$ denotes the t -th update term and the initial error spectrum $F_{\mathcal{R}}^{(0)}$ is initialized with the spectrum of the erroneous range measurements, i. e. :

$$F_{\mathcal{R}}^{(0)}(\mathbf{I}) = F_{\mathcal{R}}(\mathbf{I}), \quad \forall \mathbf{I}. \quad (4.9)$$

The iterative definition of the error spectrum $F_E^{(t)}$ follows directly from Eq. (4.8). In each iteration t , a single position $\mathbf{I}^{(t)}$ and the corresponding pair of spectral coefficients $F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)})$ and its complex conjugate $F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)})^* = F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{N} - \mathbf{I}^{(t)})$ with $\mathbf{N} = (N_1, N_2) \in \mathbb{N}^2$ are considered. Here, the symmetry of the spectra caused by the real-valued input functions is used. The position $\mathbf{I}^{(t)}$ must be chosen to yield the maximum error reduction w. r. t. the error energy from Eq. (4.8). A simplified approach to save computational expense outlined in [Aach01] is to select $\mathbf{I}^{(t)}$ according to:

$$\mathbf{I}^{(t)} = \underset{\hat{\mathbf{I}}}{\operatorname{argmax}} \left| F_E^{(t-1)}(\hat{\mathbf{I}}) \right|. \quad (4.10)$$

The spectral coefficients of the t -th update term $\Delta F_{\mathcal{R}_{SD}}^{(t)}(\mathbf{I})$ are then given by:

$$\Delta F_{\mathcal{R}_{SD}}^{(t)}(\mathbf{I}) = F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)}) \delta(\mathbf{I} - \mathbf{I}^{(t)}) + F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)})^* \delta(\mathbf{I} - \mathbf{N} + \mathbf{I}^{(t)}), \quad (4.11)$$

where

$$\delta(\mathbf{I}) = \begin{cases} 1 & \text{if } \mathbf{I} = \mathbf{0} \\ 0 & \text{else} \end{cases}. \quad (4.12)$$

The unknown spectral coefficient $F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)})$ is estimated by computing the zero-crossing of the partial derivative of the error energy function from Eq. (4.8) w. r. t. the unknown coefficient $F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)})$ itself and its complex conjugate $F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)})^*$, yielding:

$$F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}^{(t)}) = N_1 N_2 \cdot \frac{F_E^{(t-1)}(\mathbf{I}^{(t)}) F_{\mathcal{M}}(\mathbf{0}) - F_E^{(t-1)}(\mathbf{I}^{(t)})^* F_{\mathcal{M}}(2\mathbf{I}^{(t)})}{|F_{\mathcal{M}}(\mathbf{0})|^2 + |F_{\mathcal{M}}(2\mathbf{I}^{(t)})|^2}. \quad (4.13)$$

Finally, the t -th estimate for the sought deconvolved spectrum is accumulated according to:

$$F_{\mathcal{R}_{SD}}^{(t)}(\mathbf{I}) = F_{\mathcal{R}_{SD}}^{(t-1)}(\mathbf{I}) + \Delta F_{\mathcal{R}_{SD}}^{(t)}(\mathbf{I}), \quad F_{\mathcal{R}_{SD}}^{(0)}(\mathbf{I}) = \mathbf{0}. \quad (4.14)$$

In practice, the algorithm terminates if the number of iterations exceeds a user defined threshold. Clearly, the number of iterations also depends on the scene to reconstruct. Whereas for smooth data few Fourier coefficients are sufficient, the entire spectrum must be estimated for scenes that exhibit strong gradients and complex non-repetitive patterns.

Though this spectral deconvolution method exhibits a strong theoretical foundation its practical usage for a real-time capable restoration of invalid depth measurements in range imaging is very cumbersome. This is due to the inherent computational complexity using imaginary values, its iterative nature and, as the most important aspect, the missing spatial context of the Fourier transform.

Applying the method to the entire image is permitted as this would imply that information throughout the whole image are used for estimating an invalid measurement. Note that this is in contrast to the localized properties of the *Normalized Convolution* method from the previous section. To preserve spatial locality

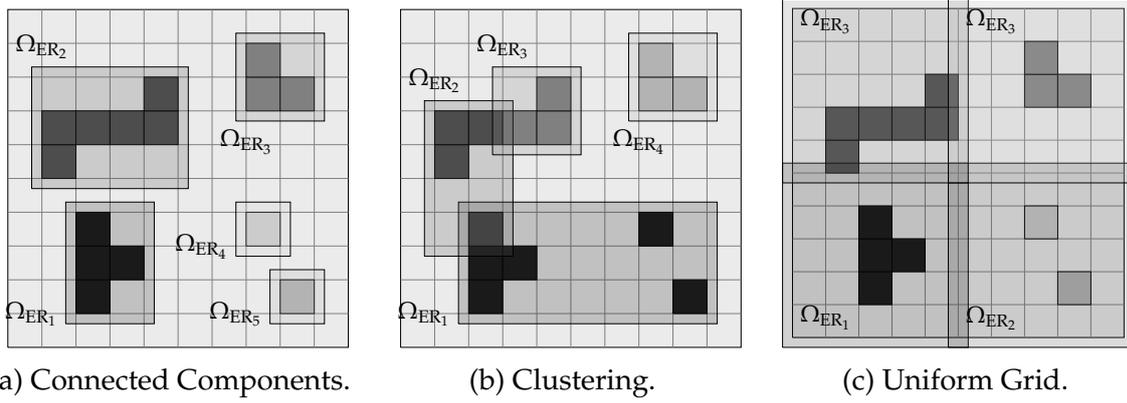


Figure 4.2: Embedding regions for invalid measurements. A strategy based on connected components in general yields a large number of regions. In contrast, clustering approaches decrease the number of regions, however, potentially splitting connected components and merging distant regions. A uniform grid with overlapping patches is a trade-off between spatial locality and number of regions.

for the spectral deconvolution method the restoration technique must be applied to multiple localized and potentially overlapping rectangular regions $\Omega_{ER} \subset \Omega$ embedding the invalid measurements similar to the *Short-time Fourier transform*. For identifying these embedding regions three different strategies can be used, see Fig. 4.2 for an illustration. First, the labels as obtained from a connected component analysis [Suzu 03] applied to the invalid measurements directly define the embedding regions. Second, clustering techniques such as k-means [Duda 00] can be used to obtain the embedding regions. As the third option, a uniform grid covering the whole image can be used to embed invalid measurements. However, spectral deconvolution is only performed for regions that actually contain invalid measurements. Compared to embedding regions derived from connected component analysis and clustering, both being potentially expensive operations in terms of computational complexity, a uniform grid has the advantage that solely a check for invalid measurements inside a grid cell is required. By using so-called *Integral Images* or *Summed Area Tables* [Viol 01, Crow 84] this can be done with a computational complexity that does not depend on the size of the grid cells. Let $\mathcal{I}_{\mathcal{M}} : \Omega \rightarrow \mathbb{N}^+$ denote the *Integral Image* representation of the mask image \mathcal{M} . The number of invalid pixels inside an embedding region Ω_{ER} centered at index $\mathbf{i} = (i_1, i_2)$ and size (s_1, s_2) is then given as:

$$\begin{aligned}
 & |\{\mathbf{i} \mid \mathbf{i} \in \Omega_{ER} \wedge \mathcal{M}(\mathbf{i}) = 0\}| = \\
 & = \mathcal{I}_{\mathcal{M}}(i_1 - 1, i_2 - 1) + \mathcal{I}_{\mathcal{M}}(i_1 + s_1, i_2 + s_2) \\
 & - \mathcal{I}_{\mathcal{M}}(i_1 - 1, i_2 + s_2) - \mathcal{I}_{\mathcal{M}}(i_1 + s_1, i_2 - 1).
 \end{aligned} \tag{4.15}$$

Clearly, this formulation is independent of the size of the embedding region and exhibits a constant computational complexity $\mathcal{O}(1)$.

In fact, a formulation of the spectral deconvolution method based on a *fixed* uniform grid was identified as the most feasible approach regarding run-time constraints. The reasons therefore is that a fixed grid allows for an efficient GPU-based

simultaneous batched computation of the Fourier transform. Further, the computation of each patch can be accelerated using aggressive loop-unrolling strategies.

4.5 Edge-preserving Denoising

Edge-preserving filtering is one of the most important tasks in computer vision and image processing. Such denoising techniques can be interpreted as conventional low-pass filters with an edge-stopping functionality that attenuates the filter kernel weights when the difference between pixel values is large. In particular in range image pre-processing edge-preserving denoising methods are valuable tools as they allow to preserve salient topographic features on surfaces.

4.5.1 Bilateral Filtering

The bilateral filter [Toma 98] is a very popular and established edge preserving spatial denoising filter. The general problem statement for bilateral filtering does not rely on a specific underlying low-pass filter model as, for instance, mean or median filtering, or Gaussian blurring. However, the latter is most commonly used which allows to express the bilateral filter (BF) as:

$$\mathcal{R}_{\text{BF}}(\mathbf{i}) = \frac{\sum_{\mathbf{i}' \in \Omega_i^r} c(\mathbf{i}, \mathbf{i}') s(\mathcal{R}(\mathbf{i}), \mathcal{R}(\mathbf{i}')) \mathcal{R}(\mathbf{i}')}{\sum_{\mathbf{i}' \in \Omega_i^r} c(\mathbf{i}, \mathbf{i}') s(\mathcal{R}(\mathbf{i}), \mathcal{R}(\mathbf{i}'))}. \quad (4.16)$$

The *spatial closeness* $c(\mathbf{i}, \mathbf{i}')$ and the *range similarity* $s(\mathcal{R}(\mathbf{i}), \mathcal{R}(\mathbf{i}'))$ that accounts for the edge-preserving nature of bilateral filtering are given as:

$$c(\mathbf{i}, \mathbf{i}') \propto \exp\left(-\frac{\|\mathbf{i} - \mathbf{i}'\|_2^2}{\sigma_c^2}\right), \quad (4.17)$$

$$s(\mathcal{R}(\mathbf{i}), \mathcal{R}(\mathbf{i}')) \propto \exp\left(-\frac{|\mathcal{R}(\mathbf{i}) - \mathcal{R}(\mathbf{i}')|^2}{\sigma_s^2}\right), \quad (4.18)$$

where σ_c and σ_s control the spatial closeness and range similarity, respectively.

Note that the bilateral filter is related to the concept of normalized convolution from Eq. (4.3) in a sense that the *spatial closeness* $c(\mathbf{i}, \mathbf{i}')$ equals the *applicability* function $a(\mathbf{i}, \mathbf{i}')$ and the *range similarity* $s(\mathcal{R}(\mathbf{i}), \mathcal{R}(\mathbf{i}'))$ can be interpreted as a translation-variant *certainty* function. This translation-variance implies that bilateral filtering is computationally expensive and its complexity is $O(N_1 N_2 r^2)$ where r denotes the radius of the local neighborhood Ω_i^r . This high computation complexity led to the development of edge-preserving denoising techniques that expose a constant complexity w. r. t. the radius of the local region. Here, the guided image filter that will be investigated in the following section is a promising candidate due to its non-approximative algorithm.

4.5.2 Guided Filtering

The concept of guided image filtering was proposed by He et al. [He 10, He 13]. The filter has a non-approximative algorithm for edge-preserving denoising. This is in contrast to recent acceleration strategies for the bilateral filter that rely on quantization methods [Pari 09, Pori 08] and thus may impair accuracy.

The basic idea of guided filtering (GF) is to express the filter output \mathcal{R}_{GF} as a linear transform of a so-called *guidance image* $G : \Omega \rightarrow \mathbb{R}$ in the local neighborhood Ω_i^r of a pixel i :

$$\mathcal{R}_{\text{GF}}(i') = a_{\Omega_i^r} G(i') + b_{\Omega_i^r}, \quad \forall i' \in \Omega_i^r. \quad (4.19)$$

Here, $a_{\Omega_i^r}$ and $b_{\Omega_i^r}$ are linear coefficients that are constant in Ω_i^r . These coefficients are obtained by constraining the filter output \mathcal{R}_{GF} to be similar to the input image \mathcal{R} . With regard to the local linear model from Eq. (4.19) this allows to define a cost function that quantifies the difference between the output \mathcal{R}_{GF} and the input \mathcal{R} in the neighborhood Ω_i^r as:

$$\begin{aligned} \mathcal{J}(a_{\Omega_i^r}, b_{\Omega_i^r}) &= \sum_{i' \in \Omega_i^r} \left(\frac{1}{2} (\mathcal{R}_{\text{GF}}(i') - \mathcal{R}(i'))^2 + \frac{1}{2} \epsilon a_{\Omega_i^r}^2 \right) \\ &= \sum_{i' \in \Omega_i^r} \left(\frac{1}{2} (a_{\Omega_i^r} G(i') + b_{\Omega_i^r} - \mathcal{R}(i'))^2 + \frac{1}{2} \epsilon a_{\Omega_i^r}^2 \right), \end{aligned} \quad (4.20)$$

where ϵ is a regularization parameter that is commonly used in ridge-regression or *Tikhonov regularization*. Consequently, the coefficients $(a_{\Omega_i^r}, b_{\Omega_i^r})$ must minimize the cost function $\mathcal{J}(a_{\Omega_i^r}, b_{\Omega_i^r})$. The coefficients $(a_{\Omega_i^r}, b_{\Omega_i^r})$ can be obtained by equating the corresponding partial derivatives of Eq. (4.20) to zero, yielding:

$$a_{\Omega_i^r} = \frac{\frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} \mathcal{R}(i') G(i') - \frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} \mathcal{R}(i') \frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} G(i')}{\frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} G(i') G(i') - \frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} G(i') \frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} G(i') + \epsilon}, \quad (4.21)$$

$$b_{\Omega_i^r} = \frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} \mathcal{R}(i') - a_{\Omega_i^r} \frac{1}{|\Omega_i^r|} \sum_{i' \in \Omega_i^r} G(i'). \quad (4.22)$$

By interpreting the input \mathcal{R} and guidance image G as uniformly distributed discrete random variables, the coefficients $(a_{\Omega_i^r}, b_{\Omega_i^r})$ can be written in a more compact form as:

$$a_{\Omega_i^r} = \frac{\mathbb{E}_{\Omega_i^r}[\mathcal{R}G] - \mathbb{E}_{\Omega_i^r}[\mathcal{R}] \mathbb{E}_{\Omega_i^r}[G]}{\mathbb{E}_{\Omega_i^r}[GG] - \mathbb{E}_{\Omega_i^r}[G] \mathbb{E}_{\Omega_i^r}[G] + \epsilon} = \frac{\text{Cov}_{\Omega_i^r}[\mathcal{R}, G]}{\text{Var}_{\Omega_i^r}[G] + \epsilon}, \quad (4.23)$$

$$b_{\Omega_i^r} = \mathbb{E}_{\Omega_i^r}[\mathcal{R}] - a_{\Omega_i^r} \mathbb{E}_{\Omega_i^r}[G], \quad (4.24)$$

where $\mathbb{E}_{\Omega_i^r}[\cdot]$, $\text{Var}_{\Omega_i^r}[\cdot]$ and $\text{Cov}_{\Omega_i^r}[\cdot]$ denote the expectation value, variance and covariance of a uniformly distributed discrete random variable in the local neighborhood Ω_i^r , respectively.

The edge-preserving characteristic of the guided filter can now be understood as follows when considering that the guidance image equals the range image, i. e. $G = \mathcal{R}$, which is the common approach for denoising single channel images.

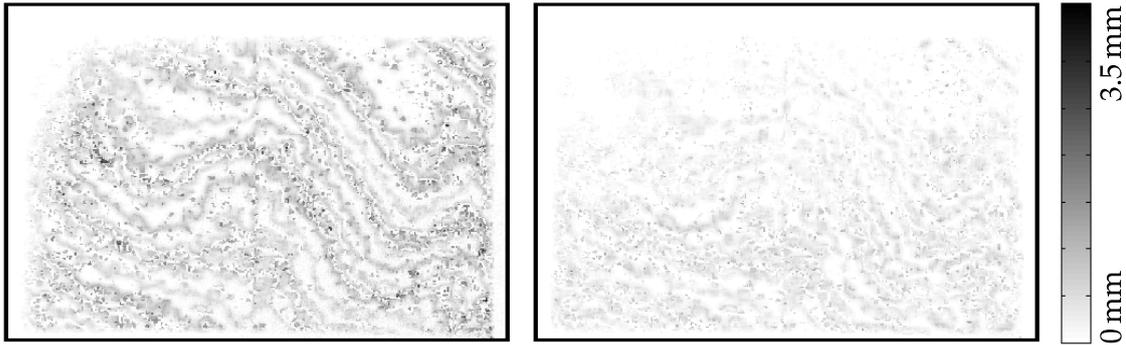


Figure 4.3: Numerical issues (error in [mm]) with mean filtering using single-precision standard 2-D integral images (left) and an implementation using separable 1-D cumulative sums (right) compared to ground truth results computed with double-precision [Wasz 11c]. The test scene shows a plane at a distance of 1 m in front of the camera.

A scene discontinuity such as an edge causes a high variance inside the region Ω_i^r and from Eq. (4.23) it follows that the coefficient $a_{\Omega_i^r}$ tends towards one with the tendency controlled by ϵ . In turn, $b_{\Omega_i^r}$ as derived in Eq. (4.24) evaluates to zero. From the linear model in Eq. (4.19) it follows that no filtering is performed and the edge is eventually preserved. In contrast, smooth regions by definition exhibit a variance of range values close to zero and the coefficient $a_{\Omega_i^r}$ will be close to zero, too. Thus, from Eq. (4.24) it follows that $b_{\Omega_i^r}$ is the expectation value or average of the data inside the neighborhood Ω_i^r and the filter output will be set to these values due to the linear model in Eq. (4.19).

The constant run-time of $O(N_1 N_2)$ independent from the chosen radius is due to the fact that the expectation values from Eqs. (4.23) and (4.24) can be computed using simple mean or box filters for which constant run-time implementations are known. One of the most prominent approaches are based on *Integral Images* or *Summed Area Tables* [Viol01, Crow 84] that were previously used in this thesis for constant-time detection of missing data in arbitrary image regions as outlined Eq. (4.15). For mean filtering in the context of guided filtering, the same theoretical principles apply, however, care has to be taken w. r. t. numerical accuracy when using single-precision floating-point numbers as outlined in prior work [Wasz 11c] and illustrated in Fig. 4.3. Due to the accumulative nature of integral images and the limited accuracy and range inherent to single-precision floating-point numbers errors may be introduced. This effect can be reduced to a certain degree by exploiting the separability of mean filtering, i. e. to perform two successive 1-D convolutions that can be computed efficiently using a parallel prefix sum technique [Harr 07]. However, the best result is obtained when using double-precision floating-point numbers. Unfortunately, this is not always an option for different hardware architectures. For example, certain GPUs may not support double-precision computations at all. Further, the number of floating point operations per second using double-precision arithmetic are considerably smaller than for single-precision and the increased bandwidth load with double-precision may decrease performance as outlined in Section 2.3.

4.6 Temporal Denoising

Temporal denoising (TD) techniques refer to multi-frame methods that generate the filter output by computing a weighted average given a series of T temporally successive frames $\{\mathcal{R}^0, \dots, \mathcal{R}^{T-1}\}$ and associated weights $\{w^0, \dots, w^{T-1}\}$ as:

$$\mathcal{R}_{\text{TD}}(\mathbf{i}) = \sum_{t=0}^{T-1} w_i^t \mathcal{R}^t(\mathbf{i}), \quad \sum_{t=0}^{T-1} w_i^t = 1. \quad (4.25)$$

The superscript denotes the relative temporal shift with 0 denoting the instantaneous and $T - 1$ the 'oldest' considered frame.

4.6.1 Conventional Temporal Averaging

The most simple form of temporal denoising which will be referred to as *conventional temporal averaging* (CTA) assumes a pixel-independent and equal weight w^t for the range image frames \mathcal{R}^t in the generic denoising model from Eq. (4.25), i. e. $w_i^t = 1/T, \forall t \wedge \forall \mathbf{i} \in \Omega$. This translates to computing the pixel-wise arithmetic mean given the set of T successive frames as:

$$\mathcal{R}_{\text{CTA}}(\mathbf{i}) = \frac{1}{T} \sum_{t=0}^{T-1} \mathcal{R}^t(\mathbf{i}). \quad (4.26)$$

A naive implementation of this denoising technique has a computational complexity that, besides the number of pixels, depends on the number of temporal frames, i. e. $O(N_1 N_2 T)$. However, by using a recursive formulation based on a moving average, the computational complexity can be decoupled from the number of temporal frames. Let $\mathcal{R}_{\text{CTA}}^t$ denote the output of conventional temporal denoising at relative temporal shift t . This allows to rewrite Eq. (4.26) in a recursive manner as:

$$\mathcal{R}_{\text{CTA}}^t(\mathbf{i}) = \mathcal{R}_{\text{CTA}}^{t-1}(\mathbf{i}) + \mathcal{R}^t(\mathbf{i}) - \mathcal{R}^{t-T}(\mathbf{i}). \quad (4.27)$$

Thus, the computational complexity is $O(N_1 N_2)$.

Though exhibiting an ideal computational complexity, conventional temporal averaging is of no relevance for this work as it can only be used for static scenes. This problem is addressed in the following section.

4.6.2 Bilateral Temporal Denoising

Conventional temporal denoising has the drawback of introducing blurring artifacts in dynamic scenes. This is due to the fact that it is inherently assumed that range values $\mathcal{R}^t(\mathbf{i})$ along the temporal dimension only differ due to sensor noise and not due to different contents and changing appearance of objects in the captured scene. For example, when capturing the thorax of a subject, respiratory motion induced body deformations will result in changes in the corresponding

depth measurements. By choosing the number T of successive frames inappropriately large, the effect of respiratory motion will be smeared out by conventional temporal averaging.

To cope with this issue, a constrained temporal averaging technique based on the previously introduced spatial bilateral filter from Section 4.5.1 is pursued. The basic idea is that range gradients along the temporal dimension are typical for dynamic scenes and that averaging across these edges falsifies the output. The *bilateral temporal denoising* (BTD) method as referred to as in this thesis computes the weights w_i^t based on the temporal distance as well as the similarity of depth measurements:

$$w_i^t = c(t) s \left(\mathcal{R}^t(\mathbf{i}) - \mathcal{R}^0(\mathbf{i}) \right), \quad (4.28)$$

where \mathcal{R}^0 denotes the instantaneous RI frame. The temporal distance $c(t)$ accounts for the elapsed time between the instantaneous frame and the frame acquired at time t and allows for non-uniform weights in the generic temporal denoising model from Eq. (4.25) that decay according to the 'age' of frames \mathcal{R}^t . In contrast, the similarity of depth measurements $s \left(\mathcal{R}^t(\mathbf{i}) - \mathcal{R}^0(\mathbf{i}) \right)$ does not account for the 'age' of frames but for the distance of corresponding depth measurements $\mathcal{R}^t(\mathbf{i})$. For the denoising model in Eq. (4.25) this allows to penalize or exclude range measurements for the purpose of dynamic scene preservation. In this thesis, both the temporal distance and the similarity of depth measurements are given by a Gaussian function, i. e. :

$$c(t) \propto \exp \left(-\frac{t^2}{\sigma_t^2} \right), \quad s(d) \propto \exp \left(-\frac{d^2}{\sigma_d^2} \right), \quad (4.29)$$

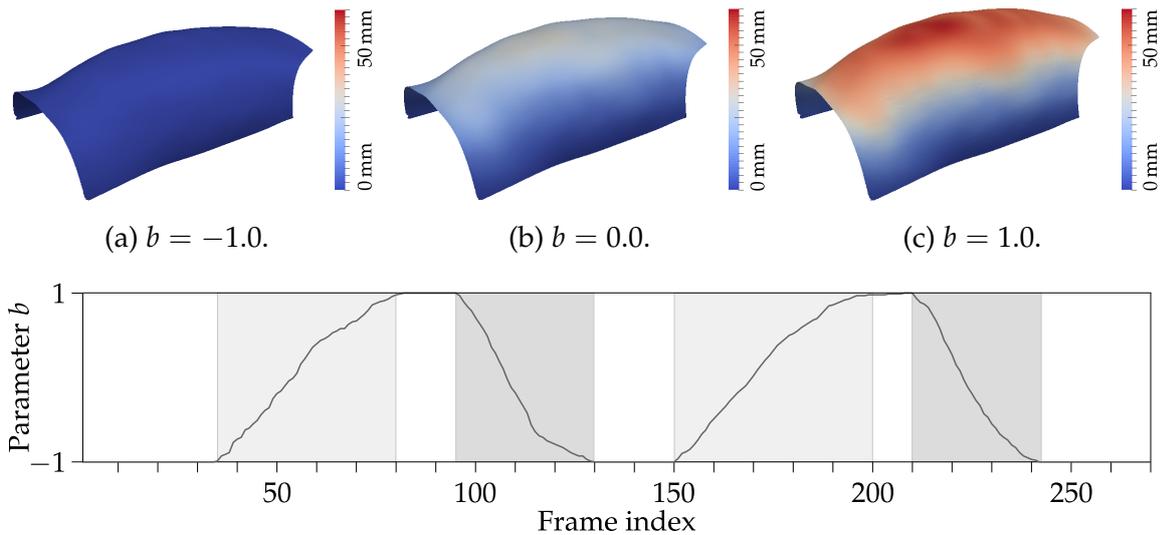
with parameters σ_t and σ_d to control the temporal extent and level of dynamic scene preservation, respectively. The filter definition can thus be stated as:

$$\mathcal{R}_{\text{BTD}}(\mathbf{i}) = \frac{\sum_{t=0}^{T-1} \mathcal{R}^t(\mathbf{i}) c(t) s \left(\mathcal{R}^t(\mathbf{i}) - \mathcal{R}^0(\mathbf{i}) \right)}{\sum_{t=0}^{T-1} c(t) s \left(\mathcal{R}^t(\mathbf{i}) - \mathcal{R}^0(\mathbf{i}) \right)}. \quad (4.30)$$

Due to the fact that the depth measurement similarity $s \left(\mathcal{R}^t(\mathbf{i}) - \mathcal{R}^0(\mathbf{i}) \right)$ is translation variant it follows that the complexity of bilateral temporal averaging is $O(N_1 N_2 T)$.

4.7 Experiments and Results

Experiments for investigating the suitability of the proposed pre-processing techniques divide into a quantitative evaluation using a synthetic respiration phantom and a detailed performance study. Both experiments have a strong focus on assessing the fitness of the investigated pre-processing techniques in a dynamic respiratory motion scenario where both accuracy as well as low run-times constitute crucial issues. Qualitative results for real RI data are purposely omitted in this chapter as they overlap with the evaluation of high coverage body surface models in Section 5.5.2.



(d) Anatomical plausible simulated respiration sequence. The sequence divides into phases of breath hold (white), inhale (light gray) and exhale (dark gray). Note that exhale phases are shorter than inhale phases which is typical for the human respiration system.

Figure 4.4: The virtual test environment for quantitative evaluation. The first row depicts different states of the respiratory motion model based on the NCAT phantom. Color coded is the local respiration magnitude controlled by one global model parameter $b \in [-1, +1]$. The bottom row shows a synthetically generated respiration sequence. Aforementioned surfaces correspond to the first inhale phase at frame indices 30, 55 and 90, respectively.

4.7.1 Quantitative Results Using a Respiration Phantom

The pre-processing techniques discussed in this chapter are first assessed for accuracy in a dynamic respiratory motion scenario based on synthetic RI data generated from a respiration phantom. The purpose of this virtual test environment is to establish the respiration phantom as a reliable ground truth for assessing various pre-processing techniques and their parameterizations. Further, synthetic data allows to study these parameters in the absence of device-specific error sources such as sub-surfaces scattering or multi-path reflections with ToF imaging.

Experimental Setup

The respiration model used for quantitative evaluation is derived from the Nurbs-based Cardiac-torso (NCAT) phantom [Sega 01] and has one free parameter b that basically corresponds to the respiration magnitude. Different respiratory motion states from fully exhale to fully inhale are depicted in Figs. 4.4a to 4.4c. The NCAT motion model consists of approximately 6.4×10^3 vertices that are organized on an unstructured grid and triangulated to form a continuous surface. An investigation on motion models including the NCAT phantom is beyond the scope of this chapter but is detailed in Chapter 6 of this thesis.

Sensor and scene specifications A virtual RI sensor (cf. Section 3.3) with a resolution of 640×480 pixels and vertical field of view of 45° was then placed above the NCAT phantom at an acquisition distance of approximately 800 mm. Respiratory motion was simulated for two inhale/exhale cycles by manually adjusting the model parameter b in an anatomically plausible manner while continuously rendering the motion model at ~ 30 Hz to obtain synthetic ground truth range data \mathcal{R}_{GT} . This setup corresponds to a typical respiratory motion scenario with a single RI sensor. The resulting sequence of motion model parameters consisting of 270 frames and extending approximately 10 s is illustrated in Fig. 4.4d. Note the anatomical plausibility for both the total duration of a sequence consisting of two respiration cycles and smaller duration of exhale phases compared to inhale phases which is typical for the human respiration system.

Noise and data corruption To simulate realistic RI measurements, the ground truth range data \mathcal{R}_{GT} was corrupted on a per-pixel basis to form the test data \mathcal{R} . As the first step, missing measurements were simulated by setting the corresponding confidence values $\mathcal{C}(\mathbf{i})$ and range measurements $\mathcal{R}(\mathbf{i})$ to zero. The decision for a measurement to be invalid or reliable was derived from a manually generated defect probability map defined on the RI domain Ω that is evaluated on a per-frame basis. In the second step the test data was corrupted by additive noise that consists of two components. The first component is modeled by *Perlin noise* which is a coherent noise function commonly used for procedural textures [Perl 85]. The motivation for using *Perlin noise* is that conventionally employed synthetic noise sampled from a zero-centered normal distribution is a simplifying assumption that does not reflect the complex noise characteristics of RI data. For example, pixel-wise corruptions induced by normal distributed noise are not spatially related which is a poor assumption for common RI limitations such as clustered glare light reflections or material properties. In contrast, to a certain degree, *Perlin noise* allows for such clustered or localized disturbances in the range data. For the experiments, the *Perlin noise* was parameterized with a frequency of 64, octave count of 8 and a persistence of 0.5. However, as *Perlin noise* is a static noise function, a second noise component sampled from a normal distribution with zero mean and variances of $\sigma_{\text{N}}^2 = 1$ and $\sigma_{\text{N}}^2 = 2$ is used to simulate temporal variations in the range acquisition process. As a finalizing step, quantization artifacts were introduced by using integer arithmetic according to:

$$\mathcal{R}'(\mathbf{i}) = \lfloor \mathcal{R}(\mathbf{i}) / \Delta_{\text{Q}} + 0.5 \rfloor \cdot \Delta_{\text{Q}}. \quad (4.31)$$

This corresponds to a quantization step size of roughly Δ_{Q} mm. The investigated quantization steps of $\Delta_{\text{Q}} = 2$ and $\Delta_{\text{Q}} = 4$ are suitable approximation regarding a typical RI sensor setup and the Kinect device that exhibits quantization steps of ~ 0.9 mm at 0.6 m and 2.8 mm at 1.0 m acquisition distance, respectively [Smis 11]. An illustration of the different noise and quantization artifacts for the corrupted synthetic RI data is given in Fig. 4.5.

Error assessment Quantitative results are then computed for each frame on a per-pixel error e_i defined as the Euclidean distances between surfaces points of

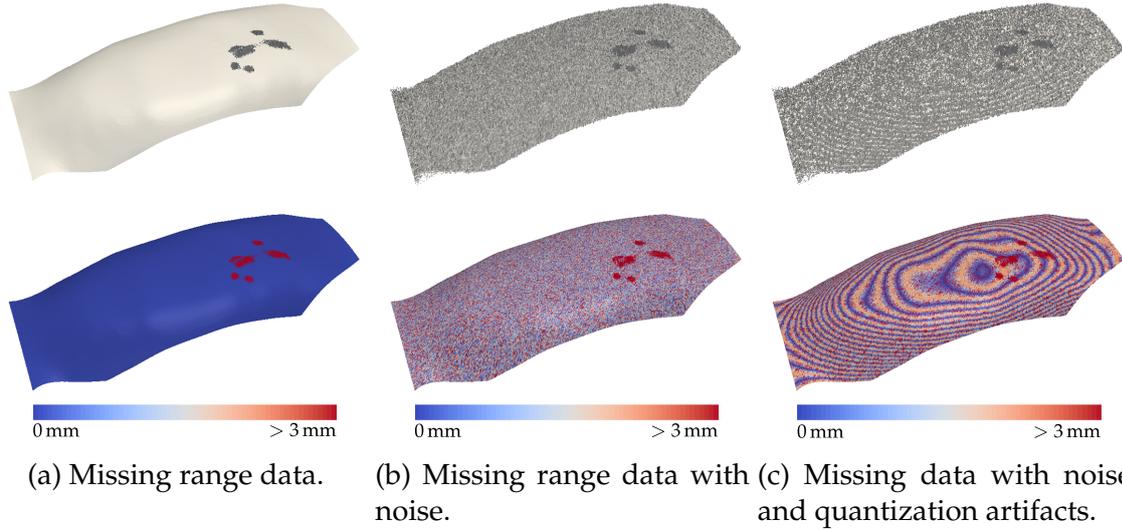


Figure 4.5: Corrupted synthetic RI data on a subdomain $\tilde{\Omega} \subset \Omega$. The top row shows the surface \mathcal{S} corresponding to the corrupted range data \mathcal{R} , the bottom row depicts the color coded error between \mathcal{S} and the ground truth surface \mathcal{S}_{GT} computed from \mathcal{R}_{GT} . Note the missing data in the thorax region (left), complex noise patterns that do not follow a normal distribution (middle) and ubiquitous staircase artifacts due to quantization (right).

the ground truth 3-D data $\mathcal{S}_{\text{GT}}(\mathbf{i})$ corresponding to $\mathcal{R}_{\text{GT}}(\mathbf{i})$ and the corrupted but filtered 3-D data $\mathcal{S}_{\text{F}}(\mathbf{i})$ corresponding to $\mathcal{R}_{\text{F}}(\mathbf{i})$, i. e. :

$$e_i = \|\mathcal{S}_{\text{GT}}(\mathbf{i}) - \mathcal{S}_{\text{F}}(\mathbf{i})\|_2. \quad (4.32)$$

For a compact representation, the first (Q_1), second (Q_2) and third (Q_3) quartiles of the pixel-wise errors e_i are computed over a subdomain $\tilde{\Omega} \subset \Omega$. This domain corresponds to the segmented and cropped NCAT motion model in the RI surface data \mathcal{S}_{GT} , see Fig. 4.5 for an illustration of the subdomain. The subdomain covering the model for evaluation is used to account for a potential bias caused by a static and flat background and unavoidable systematic errors at the sharp boundary between the motion model and the background. To keep computation times manageable, the subdomain was further sub-sampled with a factor of 8 which resulted in approximately 1.5×10^3 uniformly distributed points for which the Q_i errors are reported.

This evaluation scheme of Euclidean distances on the body surface was chosen as the *objective* nature of metric range imaging that actually describes a surface in 3-D is appropriately reflected. In contrast, other popular metrics like the *structural similarity*, though being *objective*, are mainly designed to account for the *subjective* human visual perception in conventional imaging [Wang 04].

Statistical significance To test for statistical significant differences, errors obtained from different approaches or parameterizations are assessed against each other by using a left-tailed *Wilcoxon signed-rank test* [Rams93] across the absolute values of the frame-wise differences in the Q_1 , Q_2 and Q_3 errors. Thus, the null hypothesis H_0 is that the median difference in Q_i errors between two experiments A and B across all frames f equals a threshold δ whereas the alternate hypothesis H_1 is that the median difference is smaller than that threshold, i. e. :

$$H_0 : \text{median}_{\forall f} \left\{ \left| Q_i^{A,f} - Q_i^{B,f} \right| \right\} = \delta \quad (4.33)$$

$$H_1 : \text{median}_{\forall f} \left\{ \left| Q_i^{A,f} - Q_i^{B,f} \right| \right\} \leq \delta. \quad (4.34)$$

The *Wilcoxon signed-rank test* was chosen due to two reasons. First, the individual samples are not statistical independent as they correspond to the same motion model parameter b . Second, a normal distribution of differences in the Q_i reconstruction errors cannot be assumed safely. Clearly, the choice for the difference level is crucial for a sound evaluation. Whereas a threshold close to zero is most accurate in a sense that assesses both approaches for equality, in practice one is usually interested in tolerating a certain error for the sake of faster computation or re-usage of filters. Typical thresholds used for evaluation in this chapter are $\delta = 0.1$ mm or $\delta = 0.01$ mm both being considerably smaller than the synthetic noise level in this chapter or the noise characteristics of real world RI sensors.

Results

The results are divided into different paragraphs in order to investigate the pre-processing techniques w. r. t. their ability to account for the different noise and corruption characteristics introduced in the previous section, i. e. missing data, spatial or temporal noise and quantization artifacts.

Restoration of invalid measurements The first experiment is concerned with the problem of restoring missing range measurements. For normalized convolution a rather large kernel radius of $r = 10$ and $\sigma = 7$ was heuristically chosen. This kernel covers a region of 21×21 pixels and thus embeds the defects used in this experiment and typical invalid regions in real-life measurements. For normalized convolution a fixed grid composed of patches with size 32×32 pixels and covering the domain Ω was chosen. The patches mutually overlap with 8 pixel in each direction and a *Welch* window function [Welc67] is applied to account for the inherent periodicity assumption of the Fourier transform. The number of iterations was set to 50 which was found sufficient to reconstruct a smooth function such as implied by the observed body surface.

The error graph for both methods across the motion model sequence is depicted in Fig. 4.6. Conspicuous is the superior performance of the spectral deconvolution method with a median error of 0.03 mm compared to 0.11 mm with normalized convolution. However, both error scales are below real-life noise levels and thus suitable for restoring missing range values. In particular, this holds

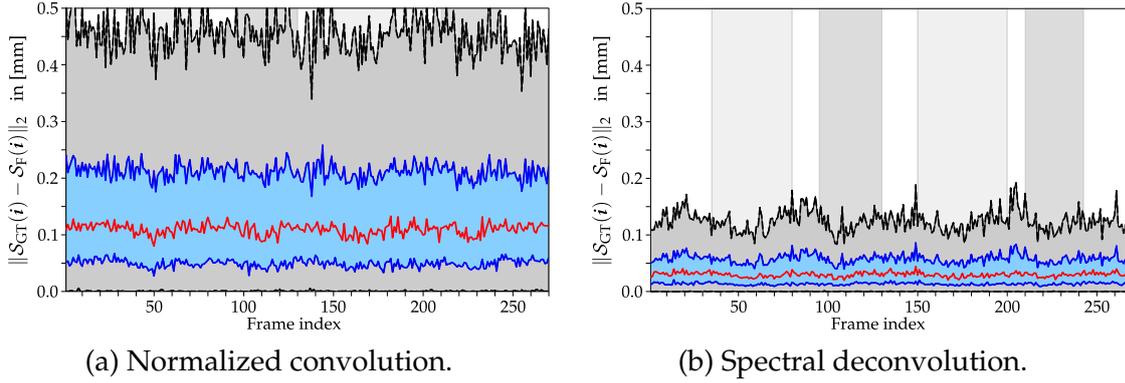


Figure 4.6: Comparison of normalized convolution and spectral deconvolution for restoring missing range measurements. Red denotes the median (Q_2), blue represent the 25th (Q_1) and 75th (Q_3) error percentiles and gray extends to the most extreme distances not considered as outliers.

true when considering that only a very small part of the body surface is affected. Due to its comparable accuracy but superior performance in terms of run-time as shown later in Section 4.7.2, normalized convolution will be used as base-line in the following experiments.

Edge-preserving filtering Results for range data with missing values and that is additionally corrupted by noise are shown in Fig. 4.7. Without any pre-processing, the data differs substantially from the ground truth with a median error of >1.0 mm, cf. Figs. 4.7a and 4.7d. Further, the graphs show no relation between the error and the respiration state. In fact, no significant correlation ($\rho \geq 0.01$) was found between the distances and both the respiration magnitude and the respiration state i. e. inhale/exhale or breath hold. The respiration state was quantified as the absolute value of the derivative of the model parameters w. r. t. the frame index.

For assessing the denoising capabilities of edge-preserving filtering a heuristically determined parameterization was chosen. This resulted in $\sigma_r = 20$, $\sigma_s = 5$ for the bilateral filter and $\epsilon = 400$ for the guided filter. The window size for both filters was chosen as $r = 7$ thus containing $(2 \cdot 7 + 1)^2 = 225$ elements. This configuration also assures that both filters are comparable [He 10]. As expected, applying edge-preserving denoising techniques as depicted in Figs. 4.7b and 4.7e decreases the surface error considerably for both noise levels and yields a median error of ~ 0.2 mm. However, the graphs reveal a dependence of the error on the respiration state. While this effect is hardly visible for bilateral filtering, the dependency is prominent for guided filtering. In fact, the experiments reveal statistical significant ($\rho \leq 0.01$) Pearson correlation coefficients (PCCs) of 0.74, 0.79 and 0.87 between the model parameter b and the Q_1 , Q_2 and Q_3 errors, respectively. This effect is due the fact that the applied edge-preserving filtering techniques are parameterized to give priority to denoising across the body surface and thus slightly smooth across the extreme sharp edge between the range image foreground corresponding to the motion model and the background, see Fig. 4.8 for the error distribution.

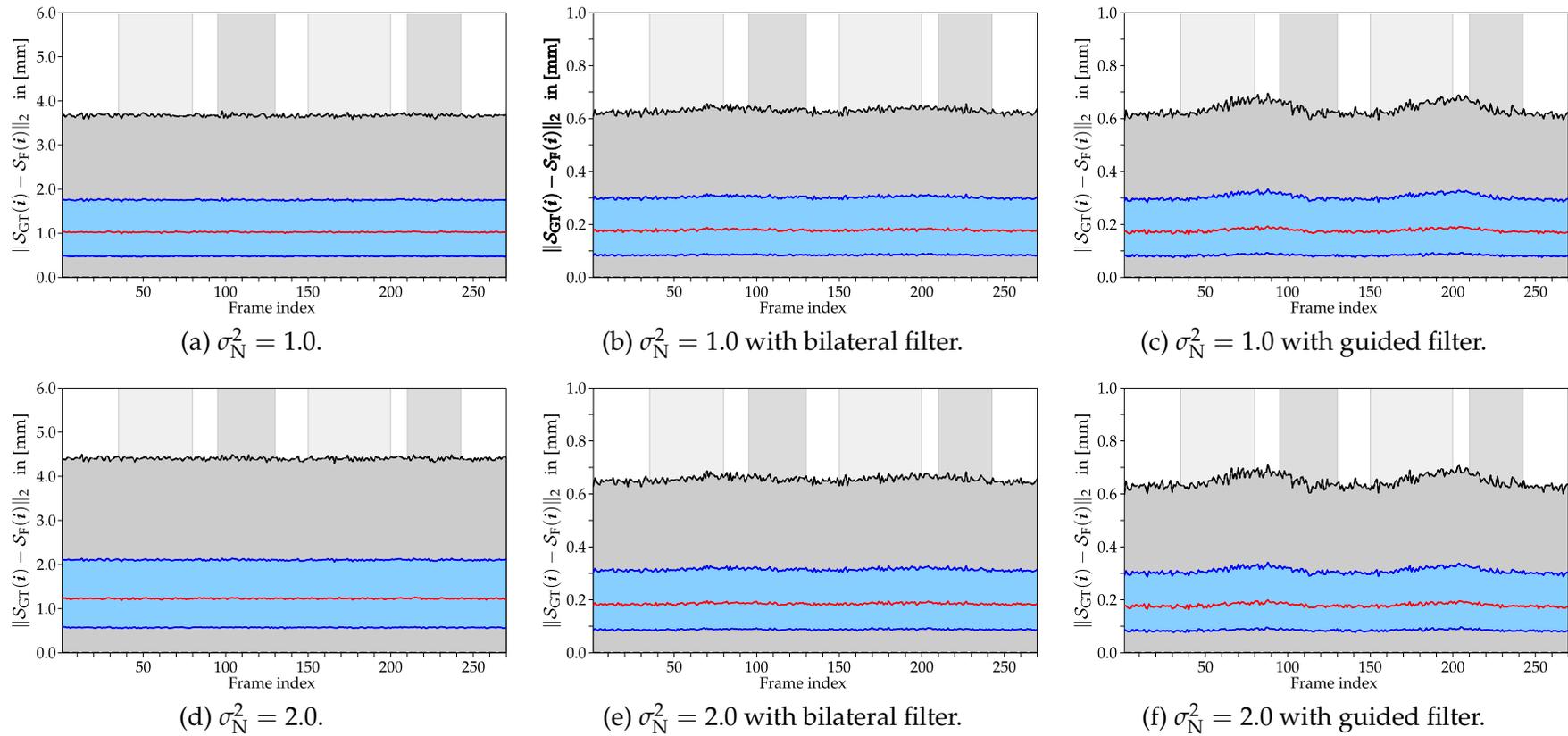


Figure 4.7: Pre-processing results for varying temporal noise levels σ_N^2 and different edge-preserving denoising filters. Note the different error scale of the raw (left column) and denoised (middle and right column) data. Further note the distinct correlation between the error metric and the respiration state for guided filtering (right column).

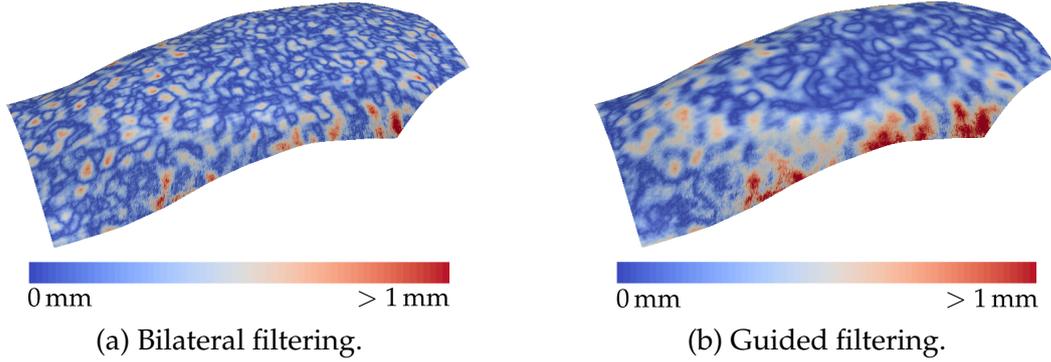


Figure 4.8: Comparison of bilateral filtering (left) and guided filtering (right) for edge-preserving denoising. Color coded is the error metric $\|\mathcal{S}_{GT}(\mathbf{i}) - \mathcal{S}_F(\mathbf{i})\|_2$. Note the larger error of guided filtering at the surface boundaries. The same effect was observed in preceding work [Wasz 11b].

The larger error of RI denoising using guided filtering compared to conventional bilateral filtering at steep edges was also observed in previous work [Wasz 11b]. In this regard, note that only the evaluation is performed on a segmented and cropped model whereas filtering techniques always operate on the entire range data. Due to its slightly more stable performance, the bilateral filtering will be considered as the baseline for evaluating the effect of temporal averaging in the next paragraphs.

Temporal Denoising Assessing the temporal denoising capability is done for the bilateral temporal averaging filter, only. The reasons for this is that the conventional temporal averaging filter can be interpreted as a special case of the bilateral version with $\sigma_d \rightarrow \infty$, cf. Eqs. (4.28) and (4.29). As shown in the experiments, such high values are prohibitive for a respiratory motion scenario, thus obviating the need for an evaluation of the conventional averaging strategy in this chapter. However, temporal averaging is a fundamental feature of the multi-view fusion technique that is proposed in Chapter 5 and the effect of conventional temporal averaging will be investigated in the corresponding evaluation part in Section 5.5.1.

Results for bilateral temporal averaging with and without preceding edge preserving spatial denoising are depicted in Fig. 4.9. For this experiment the temporal noise level was fixed to $\sigma_N = 1.0$. Without preceding edge-preserving denoising, the median error is ~ 0.2 mm thus corresponding to the error scale of the data without temporal noise, cf. Fig. 4.7a. In fact, at least a temporal variance of $\sigma_d = 0.5$ for the bilateral temporal averaging filter is necessary to produce any visible effect. For smaller values, there is no temporal averaging at all due to the high noise level that causes corruptions in the range data that are interpreted as gradients along the temporal dimension that need to be preserved. In contrast, the temporal gradients with $\sigma_d = 0.5$ are related to the changing respiration state and not to noise. This is depicted in Fig. 4.9a where the error decreases successively in steady respiration states and immediately changes back to the base level for inhale and exhale phases, respectively. However, the decrease of < 0.05 mm in the median error and

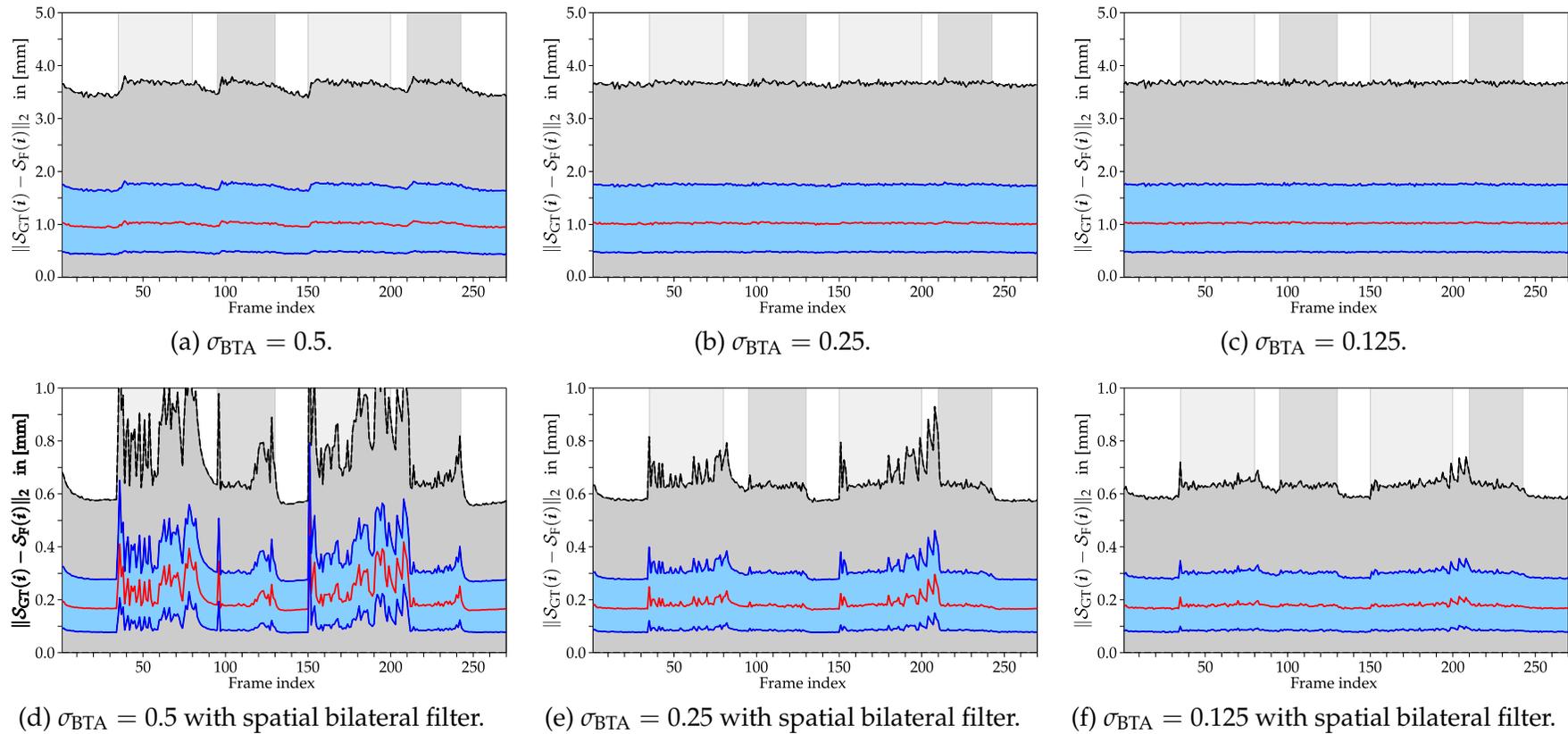


Figure 4.9: Pre-processing results with bilateral temporal averaging (BTA). For all experiments the temporal noise level was fixed to $\sigma_{\text{N}} = 1.0$. The top row shows results for *temporal* bilateral averaging only whereas the bottom row depicts results with preceding edge preserving denoising using a *spatial* bilateral filter with $\sigma_s = 5$ and $\sigma_r = 20$. Note the strong influence of changing respiration state on the Euclidean distance error.

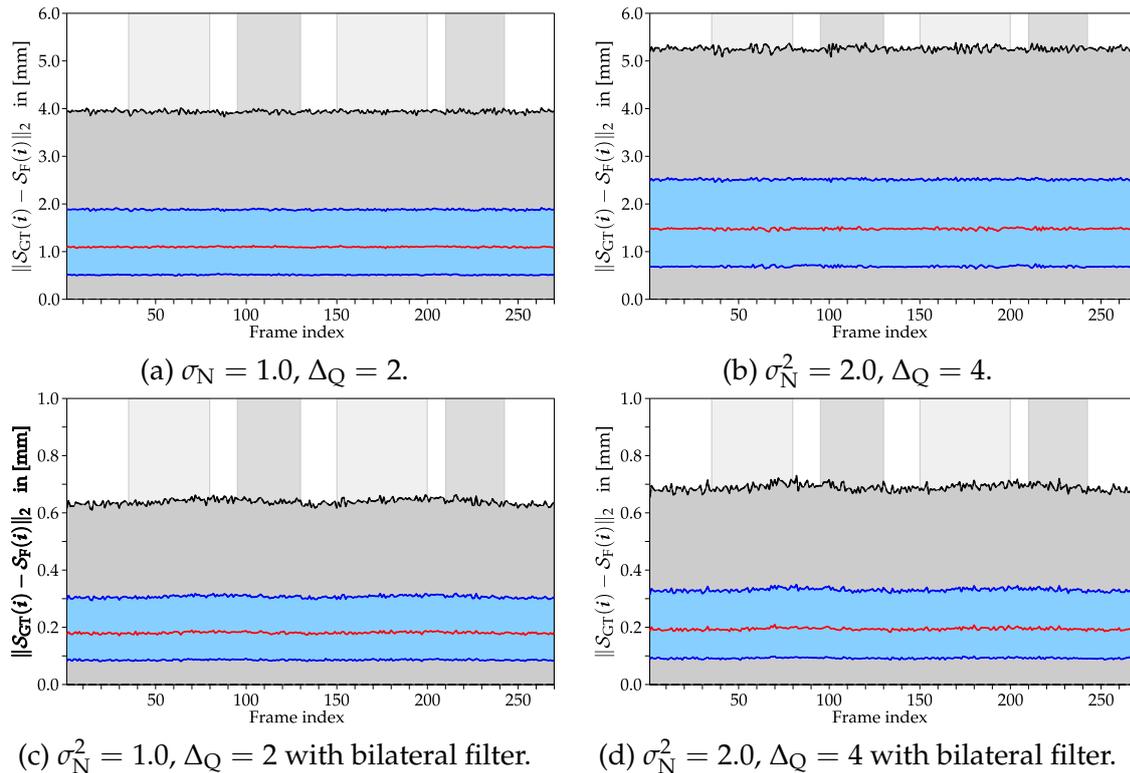


Figure 4.10: Influence of different noise levels and quantization artifacts for spatial bilateral filtering. For heavy corruptions caused by quantization the filter shows a high denoising capability with distance errors comparable to the noise-only case shown in Figs. 4.7b and 4.7e.

steady phases is small compared to the median error drop-off from ~ 1.0 mm to ~ 0.2 mm for spatial denoising.

Applying a bilateral filter as preceding step reveals a crucial issue with temporal averaging, see Figs. 4.9d to 4.9f. Though for steady respiration states the error again successively decreases there are tremendous fluctuations for changing respiration states, especially for larger values of σ_{BTA} as depicted in Fig. 4.9d. The reason for this lies in the fact that after edge-preserving denoising the difference in the local respiration magnitude is not classified as an edge to be preserved anymore. Eventually, this results in an erroneous averaging of distinct respiration states and a high deviation from the ground-truth data. By decreasing the range similarity parameter σ_{BTA}^2 this effect can be lessened to a certain degree as depicted Fig. 4.9f. However, this is naturally accompanied by a loss in denoising capability for steady respiration states.

Quantization artifacts The error graphs for the proposed pre-processing techniques for different noise levels and quantization steps are depicted in Fig. 4.10. For the raw data as shown in Figs. 4.10a and 4.10b a median error of ~ 1.1 mm and ~ 1.5 mm for medium and strong artifacts was determined, respectively. Especially for a quantization step size of $\Delta_Q = 4$ the error is considerably higher compared to the data that was solely corrupted by noise and median error of

~ 1.0 mm, cf. Figs. 4.7a and 4.7d. In contrast, the error of the filtered data as outlined in Figs. 4.10c and 4.10d is very similar to the noise only graphs in Figs. 4.7b and 4.7e with median errors of ~ 0.2 mm in all cases. In fact, independent from the noise level and quantization steps, no statistical significant difference ($\delta = 0.1$ mm, $\rho \leq 0.01$) was found for the $Q_{1,2,3}$ errors between the fully corrupted data and its counterpart that misses quantization artifacts. For a more rigorous threshold of $\delta = 0.01$ mm only the Q_3 error for a noise level of $\sigma_N^2 = 2$ and $\Delta_Q = 4$ showed a significant difference ($\rho \leq 0.01$).

4.7.2 Performance Study

The last part of the experiments conducted in this chapter covers the run-time performance of the investigated pre-processing techniques. In order to provide a fair and comprehensive evaluation of a real-life RI pre-processing pipeline, the performance study is also concerned with the basic data preparation steps including the ubiquitously required transfer of data from the host (CPU) to device (GPU) memory, the 3-D surface reconstruction from Eq. (3.13) and optional confidence computation as described in Section 3.2.2.

Experimental Setup

In order to establish run-times that are independent from the actual scene, restoration of missing data was applied for the entire image. In practice, these methods can be restricted to certain pixels or regions that actually contain invalid measurements, cf. the constant time approach to identify regions with missing values from Eq. (4.15). The reported run-times can thus be understood as an upper boundary for the execution time. Naturally, the remaining pre-processing techniques for edge-preserving and temporal denoising are executed on the entire RI data. For details on the hardware and software setup as well as time measurement principles the reader is referred to Appendix B.

Results

GPU run-times for a typical pre-processing pipeline are given in Table 4.1. As outlined in the experiments, normalized convolution for restoring missing measurements is preferred to spectral deconvolution as it yields a better run-time. In fact, spectral deconvolution is an extremely challenging method for porting on the GPU and a highly optimized implementation yielded run-times in the scale of 14 ms for a uniform grid consisting of 300 patches with a size of 32×32 pixels and 50 iterations, each. Though such run-times seem acceptable at a first glance note that by assuming a typical RI sensor frame-rate of 30 Hz the restoration of a few invalid measurements would account for approximately 50% of the available time between frames. In contrast, normalized convolution shows an excellent run-time performance of 0.28 ms even for a rather large kernel with radius of $r = 10$. This low run-time is due to the separability of convolution reducing the number of operations from a quadratic complexity of $(2 \cdot 10 + 1)^2 = 441$ to $2 \cdot (2 \cdot 10 + 1) = 42$. Further, separable filters can be implemented very efficiently on the GPU taking

Step	GPU time [ms]	Relative time [%]
Host/device transfer	0.10 ± 0.01	4.76
Normalized convolution ($r = 10$)	0.28 ± 0.02	13.3
Bilateral <i>spatial</i> filtering ($r = 7$)	0.91 ± 0.01	43.3
Bilateral <i>temporal</i> denoising ($T = 30$)	0.50 ± 0.02	23.8
Confidence computation	0.23 ± 0.02	11.0
3-D Surface reconstruction	0.08 ± 0.01	3.81
Total	2.10 ± 0.09	100

Table 4.1: GPU run-times for a typical pre-processing pipeline and image size of $\Omega \in \mathbb{R}^{640 \times 480}$ as investigated in this thesis. Edge-preserving and temporal denoising account for the most time, however, inconspicuous but ubiquitous steps such as data transfer from the main memory to the device or the final 3-D surface reconstruction cannot be neglected.

advantage of both the massive parallelism as well as the hierarchical memory architecture, see Appendix A for exemplary implementation details.

In contrast, bilateral *spatial* denoising as a non-separable filtering technique exhibits a higher run-time with a smaller kernel radius compared to the normalized convolution method. Interestingly, bilateral *temporal* denoising yields an even higher run-time w. r. t. the number of elements that are computed for each pixel. The *spatial* bilateral filter has the full quadratic complexity, i. e. $(2 \cdot 7 + 1)^2 = 225$ elements, whereas the *temporal* version only has 30 elements. The reasons for the time discrepancy lies in the GPU memory architecture. Whereas global memory loads can be decreased by sharing data among neighboring pixels as implemented in the *spatial* filter, this is inherently not an option for the *temporal* variant as data is not re-used across neighboring elements. Further, the total amount of data that needs to be loaded for the *temporal* variant is directly related to the number of frames and thus is 30 times higher. This result underlines the importance of a proper implementation of GPU programs not only regarding parallel processing principles but also w. r. t. the memory architecture.

The last part of the performance evaluation is concerned with the run-times of bilateral and guided filtering as two competing strategies for edge-preserving denoising. A plot of run-times against the kernel radii is given in Fig. 4.11. Clearly visible are the quadratic complexity of bilateral filtering and the constant complexity of guided filtering. However, for small to medium kernel radii the bilateral filter provides a much better performance and the break-even-point for guided filtering is rather high with radius $r = 8$. The reasons for this lies in the complex infrastructure and the workflow of guided filtering that, among others, consists of multiples intermediate results and several box-filters (cf. Eqs. (4.21) and (4.22)) that require integral image representations for constant time computation. All this comes along with a non-neglectable overhead caused by GPU kernel invocations and global memory access.

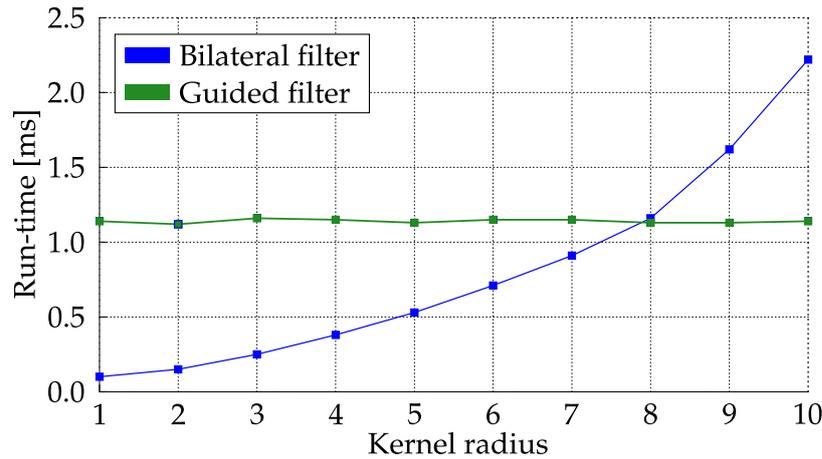


Figure 4.11: Run-time comparison of the bilateral and the guided filter for an image size of 640×480 pixels and varying kernel radius. Note the quadratic complexity for bilateral filtering in contrast to the constant complexity for guided filtering. However, for small radii bilateral filter exhibits a performance as no complex infrastructure in terms of intermediate image representations is required.

4.8 Discussion and Conclusion

This chapter was concerned with real-time capable pre-processing techniques for RI data enhancement in a dynamic respiratory motion scenario. The results confirm that pre-processing is a valuable step to account for noise and artifacts that may be introduced during the range sampling process.

For regions of missing data that are typical with range imaging in respiratory motion monitoring, normalized convolution has shown to be a simple yet effective strategy for restoring measurements, however with a median error of 0.11 mm being inferior to a more complex approach based on spectral deconvolution that yields an error of 0.03 mm. However, these results cannot be transferred directly to other RI applications. This is at least partially due to the simple shape and appearance of the human thorax and in particular of the respiration phantom that does not feature salient topographic landmarks.

In this regard, it must be also questioned if edge-preserving denoising techniques, though definitely not being harming, are necessary at all for enhancing RI-based body surface data. In fact, the parameterizations that were determined heuristically to yield the best noise reduction across the examined surface part comes along with a denoising characteristic for both the bilateral and guided filter that resembles those of conventional filters. Yet, for very steep edges such as the transition of the body to the background an edge-preserving functionality is still present with bilateral filtering being slightly superior to guided filtering. The conclusion of the conducted experiments is that for moderate temporal noise levels with variance in the scale of $\sigma_N = 1.0$ and quantization steps up to 4 mm edge-preserving denoising methods are a valuable tool towards recovering the original data with a residual error in the scale of ~ 0.6 mm. Though exhibiting a slightly

different performance at sharp boundaries, bilateral and guided filtering both are appropriate techniques for edge-preserving denoising.

Whereas the benefit of restoring invalid measurements and edge-preserving *spatial* filtering is invariably underlined by the experiments, the gain of using *temporal* denoising methods is not self-evident. On the one hand temporal denoising yields more stable surface data with a slightly decreased surface error for steady respiration states. On the other hand, serious artifacts are introduced for changing respiration states. Incorporating a functionality to detect changing respiration states as pursued with bilateral temporal averaging only partially accounts for this issue. Especially, for higher noise levels and slightly changing respiration magnitude, there is no means to separate per-pixel variations in range measurements from respiratory motion and temporal noise. In either case, the parameter must be chosen carefully and applying a temporal denoising technique must be considered as a trade-off between a stable and steady surface and an accurate reconstruction.

Regarding real-time constraints that are imposed by a respiratory motion scenario GPU architectures have proven to be suitable. A typical pre-processing pipeline as proposed in this chapter executes in approximately 2 ms on off-the-shelf consumer hardware, thus easily coping with the frame-rates of modern RI sensors. An interesting result here is that constant time methods for edge-preserving denoising such as *guided filtering* do not necessarily exhibit a lower run-time compared to the conventional *bilateral filter* when porting to GPU architectures. This is due to the complex nature of constant time methods in contrast to the brute-force computation principle for *bilateral filtering*. Further, the reconstruction and performance results for restoring missing range information using normalized convolution or spectral deconvolution demonstrate that in general a trade-off between accuracy on the one hand and run-time on the other hand has to be made. In this regard it must be mentioned that, though often assumed in the image-processing community, GPU implementations per se do not necessarily yield run-times to satisfy time constraints as imposed by practical demands.

Multi-View Range Data Fusion

5.1 Motivation	53
5.2 Related Work	54
5.3 Implicit Surface Representation Using Signed Distance Transforms	56
5.4 Explicit Surface Reconstruction by Ray Casting	60
5.5 Experiments and Results	62
5.6 Discussion and Conclusion	72

Patient monitoring using one single RI sensor is often not possible in image guided interventions due to a variety of reasons such as occlusion by clinical devices, staff or the acquisition geometry itself. This chapter elaborates on a multi-sensor setup and investigates a volumetric method for fusing RI streams from different sensors in conjunction with a ray-casting technique for patient body surface reconstruction. The investigated methods were first proposed for respiratory motion analysis in preceding work [[Wasz 13](#), [Wasz 16](#)].

5.1 Motivation

For respiratory motion analysis in image guided interventions a camera setup using one single RI sensor is often not feasible. First and foremost, a single static camera is not capable to provide a complete body coverage. This is mandatory in fractionated radiation therapy for arbitrary beam directions and different patients. Further, the incomplete body coverage implied by a single sensor may render a model-based respiratory motion analysis ill-conditioned [[Wasz 13](#)] and contradicts the principle that surface-based respiration surrogates should be acquired densely and spread across the entire body, cf. Section 2.1.1. Second, due to self-occlusions of the patient by the body surface itself, staff and clinical devices such as C-Arm CT systems or the LINAC with potentially attached multi-leaf collimators, a continuous coverage of the target position cannot be guaranteed. Yet, even if there are no blocking objects, small reconstruction frustums, shallow acquisition angles and a large working distance may deteriorate the quality of RI data. Multi-camera setups help to remedy these issues by providing data from multiple viewpoints which allows to address occlusions or body coverage issues and to enable super-resolution techniques for data enhancement.

Dedicated multi-sensor systems have been proposed in fractionated radiation therapy by Lindl et al. [Lind 13] or Price et al. [Pric 12] where the latter system features fast acquisition rates of > 20 Hz, thus allowing for high temporal resolution respiratory motion analysis. A commercially available system with multi-camera functionality is the AlignRT system that is used in several clinical facilities throughout the world. See Section 2.2.2 for a more detailed description of these systems. However, regardless of the underlying range imaging principle, these existing multi-sensor systems yet do not provide a unifying fusion of RI data streams from multiple sensors. This is a crucial drawback for patient alignment and respiratory motion analysis algorithms. For example, the marker-less patient setup systems proposed by Bauer et al. [Baue 11] or Placht et al. [Plac 12] rely on surface features that encode the local topography of the body surface. For non-fused RI surfaces this would imply to reconstruct the local neighborhood from the disjunct or overlapping multiple surfaces at border regions. Further, data redundancy also requires to explicitly select the most reliable measurement. As an important practical issue, explicit handling of junctions and data redundancy also hampers the deployment of high-performance algorithms. The reason for this is that 3-D points being close to each other in world space may be stored in different memory regions due to the separate data structures obtained from the individual RI sensors. For GPUs this hinders an effective texture cache usage or coalesced memory access to exploit the maximum available bandwidth for data transfer between off-chip and on-chip memory, cf. the GPU architecture introduction in Section 2.3.1.

In this chapter a real-time and modality-independent framework for high coverage body surface reconstruction from multiple fused RI sensors is investigated. The proposed method employs a volumetric range data fusion approach encoding the patient body surface in an implicit manner. Based upon this implicit representation, an explicit surface reconstruction technique that utilizes a novel manifold ray casting technique tailored to the human anatomy is described. As a unique characteristic compared to existing approaches, the presented method inherently features a high body surface coverage and a surface topology that allows for a very efficient computation of local neighborhoods. These features facilitate the usage of algorithms designed to cope with real-time constraints.

5.2 Related Work

The fusion of point cloud data, such as delivered by RI sensors, that is acquired from multiple view-points is a widely investigated problem in geometric modeling. Generally, the problem can be divided into either a pure fusion and *consolidation* of the multiple point cloud data [Paul 02, Rusi 02, Lind 13, Pric 12], or a unified *surface reconstruction* from this 3-D data [Hopp 92, Curl 96, Kazh 06]. A pure consolidation of the point data is sufficient for geometry processing algorithms that solely require a discrete set of 3-D point data, whereas surface reconstruction techniques are mandatory if the algorithm is based on a continuous surface representation.

The standard strategy for simple merging of point cloud data is based on clustering techniques [Paul 02]. The governing principle here is to subdivide the bound-

ing box that contains all different scans into grid cells. All 3-D points that map to the same cell are then replaced by a common representative. The representative is typically derived by simple averaging all points within the cell [Rusi 02, Lind 13]. The subdivision itself is commonly performed using a fixed grid with equal size [Rusi 02, Lind 13], however, due to memory constraints space partitioning strategies like *kd-tress* have also been proposed [Pric 12]. Note that this consolidation strategy only yields a discrete set of 3-D points and no continuous surface representation. However, this is sufficient for algorithms that solely require a discrete set of 3-D point data. For example, the RI based patient alignment systems proposed by Price et al. and Lindl et al. [Pric 12, Lind 13] only use a subset of the available point data to compute the transformation of the subject to the reference position. In this context, a consolidation strategy using grids in conjunction with a thinning or a space partitioning strategy can also be interpreted as acceleration strategies to reduce the computational complexity.

For surface reconstruction techniques, as an intermediate step, an implicit surface representation is required. The most prominent representations here are based on signed distance functions (SDFs) that encode the signed distance of a 3-D point to the surface [Hopp 92, Curl 96, Kazh 06] or *occupancy grids* in which each grid cell holds the probability of being occupied [Pirk 11]. Here, it is worth noting that it is neither required to compute the exact SDF nor it is required to compute the SDF over the whole domain embedding the 3-D point clouds. For example, Hoppe et al. [Hopp 92] proposed to compute the SDF only for a small region near the original discrete set of point data. As the exact surface is not known but only measured by a discrete set of 3-D samples, the SDF is approximated by the distance of a point to a tangent plane computed from local samples. Another example is based on approximated projective distance transforms as proposed by Curless and Levoy [Curl 96]. Here, the SDF is computed by approximating the point to surface distance using a fast projective data association scheme. Further, the SDF is truncated and clamped to avoid inference issues with potentially opposing acquisition directions that would falsify the implicit surface representation. To address the problem of high memory requirements of implicit representations based on regular grids, recent approaches employ space partitioning techniques based on *octrees* [Zeng 12] or point-based fusion frameworks [Kell 13].

Regardless of the actual implicit surface representation, these reconstruction techniques perform a level set based surface extraction as the finalizing step. For SDFs the sought level set is given by the zero crossing whereas for probabilistic occupancy grid one seeks for the iso-level 0.5 identifying the maximum posterior probability [Pirk 11] which, by using the concept of log odds corresponds to the zero level of an SDF in the interval $[-1, +1]$. For iso-level extraction the well known *marching cubes* algorithm [Lore 87] or *marching tetrahedra* techniques [Tree 99] are popular. Though efficient GPU-based implementations have been introduced for these techniques, reconstruction times of ~ 150 ms for a volume size of $512 \times 512 \times 70$ as reported in [Cizn 12] are beyond real-time constraints imposed by a respiratory motion analysis scenario. A different approach for extracting the sought surface is based on ray casting techniques [Park 98]. In this context a noteworthy approach is the *Kinect Fusion* framework [Newc 11, Izad 11] which

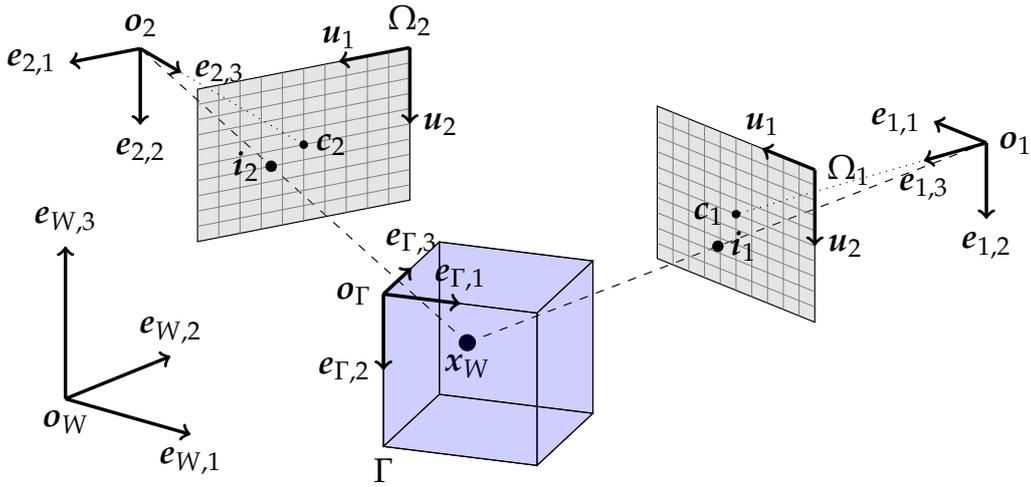


Figure 5.1: The embedding domain Γ of a signed distance function \mathcal{D} as an implicit surface representation in world space W , cf. the multi-view acquisition geometry depicted in Fig. 3.1.

aims at live scene reconstruction using a hand-held device. By utilizing sophisticated GPU implementations for SDF data integration as well as explicit surface reconstruction from ray casting the SDF, run-times of < 30 ms for a volume size of 512^3 using current off-the-shelf consumer hardware have been reported.

5.3 Implicit Surface Representation Using Signed Distance Transforms

As the first step in a multi-sensor scenario, the simultaneously acquired range data from the individual RI sensors must be consolidated and fused to a unifying representation. Due to real-time constraints in RI based patient respiratory motion monitoring, an efficient fusion method is mandatory. For this purpose, a variant of approximative SDFs originally proposed by Curless and Levoy [Curl96] is investigated in this section. To cope with the special demand w. r. t. simultaneously acquired RI data from multiple sensors, this section elaborates on joint *simultaneous* multi-view fusion and *sequential* temporal SDF integration.

An implicit surface representations encodes the surface \mathcal{S}_k with the zero level set of its SDF $\mathcal{D}_{\mathcal{S}_k}$ and an associated weighting function $\mathcal{W}_{\mathcal{S}_k}$:

$$\mathcal{D}_{\mathcal{S}_k}(\mathbf{x}_\Gamma) : \Gamma \rightarrow [-1, +1], \quad (5.1)$$

$$\mathcal{W}_{\mathcal{S}_k}(\mathbf{x}_\Gamma) : \Gamma \rightarrow [0, 1]. \quad (5.2)$$

Here, $\Gamma \subset \mathbb{R}^3$ describes the domain embedding the SDF and weighting function. Similar to the *local* RI camera coordinate systems C_k introduced in Eq. (3.4) the SDF $\mathcal{D}_{\mathcal{S}_k}$ has an associated transformation $T_\Gamma \in \mathbb{R}^{4 \times 4}$ that describes the mapping from

a point $x_W \in \mathbb{R}^3$ given in *global* world space W to a point $x_\Gamma \in \Gamma$ given in the *local* SDF coordinate system as:

$$\begin{pmatrix} x_\Gamma \\ 1 \end{pmatrix} = T_\Gamma \begin{pmatrix} x_W \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_\Gamma & \mathbf{t}_\Gamma \\ \mathbf{0} & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_W \\ 1 \end{pmatrix}, \quad (5.3)$$

with rotation $\mathbf{R}_\Gamma \in \text{SO}_3$ and translation $\mathbf{t}_\Gamma \in \mathbb{R}^3$ similarly to the camera space from Eq. (3.3) defined as $\mathbf{t}_\Gamma = \mathbf{o}_\Gamma$ and $\mathbf{R}_\Gamma = (e_\Gamma^1 \ e_\Gamma^2 \ e_\Gamma^3)$. The origin \mathbf{o}_Γ and spanning vectors e_Γ^i describe the actual embedding of Γ in the world space W , see Fig. 5.1 for an illustration.

5.3.1 Approximated Point-to-Surface Distance

Computing the exact signed distance transform of a surface is a computationally expensive task. Instead, conventional SDFs are approximated by using a projective data association scheme. Exploiting the projective acquisition geometry of RI sensors, the signed distance of a point $x_\Gamma \in \Gamma$ to the surface \mathcal{S}_k is computed from the range measurements \mathcal{R} as:

$$d(x_\Gamma, \mathcal{S}_k) = \|\mathcal{S}_k(P_k(x_k))\|_2 - \|x_k\|_2 = \mathcal{R}_k^<(P_k(x_k)) - \|x_k\|_2. \quad (5.4)$$

By using a further approximation step this distance can be expressed in terms of orthogonal range measurements in the proximity of the surface \mathcal{S}_k as:

$$d(x_\Gamma, \mathcal{S}_k) \approx (0, 0, 1) \mathcal{S}_k(P_k(x_k)) - (0, 0, 1) x_k = \mathcal{R}_k^\perp(P_k(x_k)) - (0, 0, 1) x_k. \quad (5.5)$$

In both versions the projection operator $P_k : \mathbb{R}^3 \rightarrow \Omega_k$ is associated with the k -th RI sensor, cf. Eq. (3.7), and the transformation from x_Γ defined in the SDF embedding domain Γ to a point x_k defined in the k -th camera space is given as:

$$\begin{pmatrix} x_k \\ 1 \end{pmatrix} = T_k T_\Gamma^{-1} \begin{pmatrix} x_\Gamma \\ 1 \end{pmatrix}. \quad (5.6)$$

The projective distances $d(x_\Gamma, \mathcal{S}_k)$ from Eqs. (5.4) and (5.5) can be interpreted as computing the difference between the point x_k as seen from the sensor and the actual measured surface point $\mathcal{S}_k(P_k(x_k))$ along the viewing ray implied by the projection $P_k(x_k)$. Note that positions x_k that are located between the sensor and the surface $\mathcal{S}_k(P_k(x_k))$ are assigned a positive distance whereas positions behind the surface obtain a negative value. A support region controlled by ϵ is defined around the surface \mathcal{S}_k to account for the approximative nature of $d(x_\Gamma, \mathcal{S}_k)$ and uncertainty of range measurements w. r. t. the noise characteristics of the RI sensor. For points x outside the support region, i. e. $d(x_\Gamma, \mathcal{S}_k) > +\epsilon \vee d(x_\Gamma, \mathcal{S}_k) < -\epsilon$, the corresponding distances are clamped according to:

$$\eta(d) = \min(\epsilon, |d|) \text{sign}(d). \quad (5.7)$$

The effect of clamping can be understood as assigning a rough information for a point x being outside the support interval somewhere in front or behind the

surface \mathcal{S}_k as seen from the k -th sensor. In contrast, points x inside the support region are assumed to provide a reliable distance. From Eqs. (5.5) and (5.7) the ϵ -normalized approximated SDF $\mathcal{D}_{\mathcal{S}_k}$ w. r. t. the surface \mathcal{S}_k is given as:

$$\mathcal{D}_{\mathcal{S}_k}(x_\Gamma) = \frac{1}{\epsilon} \eta(d(x_\Gamma, \mathcal{S}_k)). \quad (5.8)$$

5.3.2 Weighting of Approximative Distances

Attached to $\mathcal{D}_{\mathcal{S}_k}$ is the weighting function $\mathcal{W}_{\mathcal{S}_k}$ that encodes the applicability of approximative point-to-surface distances as derived in Eq. (5.8). Following the original work by Curless and Levoy [Curl96], the weighting component $\mathcal{W}(x_\Gamma)$ comprises two concepts: First, the weight reflects the certainty of surface measurements $x_i \in \Psi_k$. This information is encoded in the confidence map \mathcal{C}_k from Eq. (3.16). Recall that the confidence map \mathcal{C}_k yields high scores for smooth surfaces that are perpendicular to the viewing direction and close to the k -th RI sensor. This is of great importance for multi-view surface acquisitions of one single object as the scans in general partially overlap. Thus, an individual surface point is in general captured from different distances and viewing angles.

Now, instead of naively fusing the multiple acquisitions into the common implicit model from Section 5.3.1 these acquisition parameters are taken from the confidence maps which eventually allows for a weighted fusion as later outlined in Eq. (5.16) and Eq. (5.17) giving priority to reliable surface measurements. As the second component, the weighting term incorporates a distance term $w(d)$ that penalizes high distances $d(x_\Gamma, \mathcal{S}_k)$ for points that are not in proximity to the surface \mathcal{S}_k . In this work, the distance related weighting term performs a smoothly decaying weighting in front and a sharp cut-off behind the surface w. r. t. the support region ϵ according to:

$$w(d) = \begin{cases} \frac{1}{1+|\frac{d}{\epsilon}|^2} & \text{if } d \geq -\epsilon \\ 0 & \text{else} \end{cases}. \quad (5.9)$$

The rationale behind the distance component $w(d)$ is to prevent interference when fusing approximative SDFs that have been computed from substantially different view points. The original work by Curless and Levoy [Curl96] used a linear attenuation of weights followed by a sharp truncation when the distance exceeds a certain threshold. However, in a dynamic scenario such as RI-based respiratory motion analysis a sharp truncation in front of the surface is obstructive. This is due to the fact that in a dynamic scenario a position in space may be occupied in one frame (e. g. fully inhale) but empty in a subsequent frame (e. g. fully exhale). Fusing the corresponding SDFs with a rigorous weight cut off yields an implicit representation that, due to discarding of free space information, erroneously contains the old surface that is not present anymore as well as the new correct surface. In contrast, smoothly attenuating weighting coefficients in front of the surface allow to incorporate the information of non-occupancy into the SDF fusion.

Finally, the combined weighting term $\mathcal{W}_{\mathcal{S}_k}(\mathbf{x}_\Gamma)$ is given by the product of the confidence data and the distance term as:

$$\mathcal{W}_{\mathcal{S}_k}(\mathbf{x}_\Gamma) = \mathcal{C}_k \left(P_k^T(\mathbf{x}_\Gamma) \right) w(d(\mathbf{x}_\Gamma, \mathcal{S}_k)). \quad (5.10)$$

5.3.3 Fusion of Multiple Distance Transforms

Without loss of generality, let $\{(\mathcal{D}^{(t)}, \mathcal{W}^{(t)})\}$ denote a set of distance and weighting functions that were acquired both at different time and from varying view points. As outlined in [Curl96], under certain assumptions, the fused representation $\widehat{\mathcal{D}}$ is given in a least squares sense according to:

$$\widehat{\mathcal{D}}(\mathbf{x}_\Gamma) = \frac{\sum_{t=1}^T \mathcal{W}^{(t)}(\mathbf{x}_\Gamma) \mathcal{D}^{(t)}(\mathbf{x}_\Gamma)}{\sum_{t=1}^T \mathcal{W}^{(t)}(\mathbf{x}_\Gamma)}. \quad (5.11)$$

Practically, it is not feasible to keep track of all SDFs and a memory efficient yet equivalent approach for the least squares estimate $\widehat{\mathcal{D}}(\mathbf{x}_\Gamma)$ is given by a recursive formulation of Eq. (5.11):

$$\widehat{\mathcal{D}}^{(t)}(\mathbf{x}_\Gamma) = \frac{\widehat{\mathcal{D}}^{(t-1)}(\mathbf{x}_\Gamma) \widehat{\mathcal{W}}^{(t-1)}(\mathbf{x}_\Gamma) + \mathcal{W}^{(t)}(\mathbf{x}_\Gamma) \mathcal{D}^{(t)}(\mathbf{x}_\Gamma)}{\widehat{\mathcal{W}}^{(t)}(\mathbf{x}_\Gamma)}, \quad (5.12)$$

$$\widehat{\mathcal{W}}^{(t)}(\mathbf{x}_\Gamma) = \widehat{\mathcal{W}}^{(t-1)}(\mathbf{x}_\Gamma) + \mathcal{W}^{(t)}(\mathbf{x}_\Gamma), \quad (5.13)$$

with the initial estimates given by $\widehat{\mathcal{D}}^{(0)} = \widehat{\mathcal{W}}^{(0)} = 0$. Indeed, this is a valid formulation regarding a *static* model reconstruction problem targeted in the original work with a limited number of RI scans [Curl96]. However, this approach is not feasible for a *dynamic* scenario with rapidly streaming RI data. The reason for this is that the weighted average in Eq. (5.12) and Eq. (5.13) is solely data driven and does not account for the data acquisition protocol. Eventually, this mixes up data being captured *simultaneously* in a multi-sensor setup and data that were acquired *sequentially* in a dynamic setting.

In this work, the temporal nature of fusion is explicitly addressed by introducing an additional weighting factor $\alpha \in [0, 1]$ that controls the averaging of successive frames and thus accounts for the level of dynamic scene preservation. Extending Eq. (5.12) and Eq. (5.13) by the temporal weighting factor α consequently reads as:

$$\widehat{\mathcal{D}}^{(t)}(\mathbf{x}_\Gamma) = \frac{(1 - \alpha) \widehat{\mathcal{W}}^{(t-1)}(\mathbf{x}_\Gamma) \widehat{\mathcal{D}}^{(t-1)}(\mathbf{x}_\Gamma) + \alpha \mathcal{W}^{(t)}(\mathbf{x}_\Gamma) \mathcal{D}^{(t)}(\mathbf{x}_\Gamma)}{\widehat{\mathcal{W}}^{(t)}(\mathbf{x}_\Gamma)}, \quad (5.14)$$

$$\widehat{\mathcal{W}}^{(t)}(\mathbf{x}_\Gamma) = (1 - \alpha) \widehat{\mathcal{W}}^{(t-1)}(\mathbf{x}_\Gamma) + \alpha \mathcal{W}^{(t)}(\mathbf{x}_\Gamma). \quad (5.15)$$

For *static* scenes one is usually interested in a smooth reconstruction by averaging multiple subsequent frames. This is controlled by small α values. In contrast, for dynamic scenes priority should be given to the instantaneous measurements.

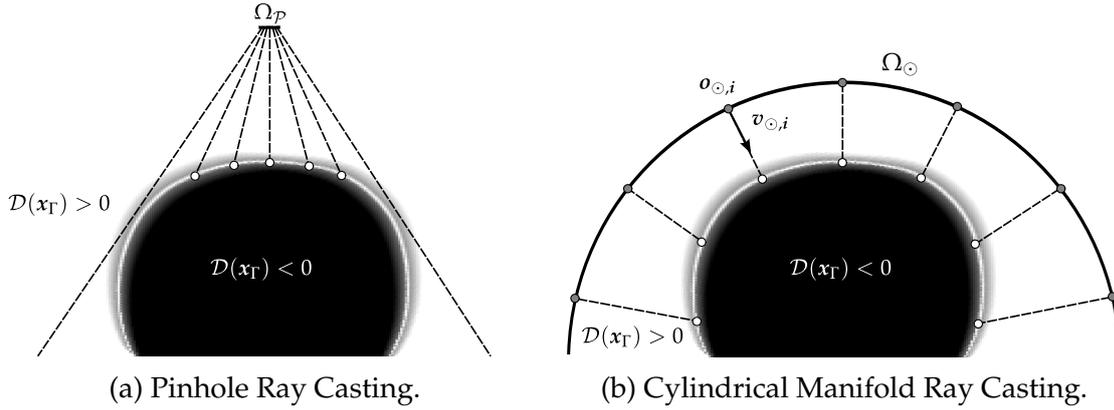


Figure 5.2: Explicit surface reconstruction by ray casting an SDF \mathcal{D} [Wasz 13]. The left figure depicts an approach based on a pinhole camera model and the right figure illustrates manifold ray casting based on a cylindrically shaped virtual sensor plane. Note that with manifold ray casting a 180° coverage patient surface model can be reconstructed.

Thus, depending on the application, α must be chosen close to 1. As the final step, the instantaneous SDF composed of $\mathcal{D}^{(t)}(\mathbf{x}_\Gamma)$ and $\mathcal{W}^{(t)}(\mathbf{x}_\Gamma)$ at time t is separately modeled to account for the *simultaneous* acquisition of RI surfaces in a multi-sensor scenario.

Given, the approximated point-to-surface distance from Eq. (5.8) and the definition of the total weight from Eq. (5.10), the instantaneous fused SDF is modeled by a convex combination of the instantaneous multi-view measurements:

$$\mathcal{D}^{(t)}(\mathbf{x}_\Gamma) = \frac{1}{\mathcal{W}^{(t)}(\mathbf{x}_\Gamma)} \sum_{k=1}^K \mathcal{W}_{S_k}(\mathbf{x}_\Gamma) \mathcal{D}_{S_k}(\mathbf{x}_\Gamma), \quad (5.16)$$

$$\mathcal{W}^{(t)}(\mathbf{x}_\Gamma) = \sum_{k=1}^K \mathcal{W}_{S_k}(\mathbf{x}_\Gamma). \quad (5.17)$$

Clearly, the convex coefficients are solely data driven and thus this fusion approach effectively accounts for the *simultaneous* acquisition property present in a multi-sensor setup.

5.4 Explicit Surface Reconstruction by Ray Casting

The distance transform \mathcal{D} encodes the body surface in an implicit manner via the zero level set of approximated point to surface distances. However, the majority of surface processing or analysis techniques require an explicit representation, i. e. vertices and edges. Though explicit surface reconstruction techniques such as *marching cubes* algorithm [Lore 87] can be used for zero level extraction, real-time run-times are hard to achieve. A different approach for extracting the sought surface is based on ray casting techniques [Park 98] that, given run-times of ~ 5 ms for a volume size of 512^3 voxels [Newc 11], have proven to satisfy real-time constraints. The basic idea pursued in surface reconstruction using ray casting is to

simulate a certain camera model and the range sampling process. In the following sections the most popular ray casting technique based on a pinhole camera model and a dedicated approach tailored to respiratory motion analysis are investigated.

5.4.1 Pinhole Camera Ray Casting

Analogous to the RI acquisition geometry and 3-D surface reconstruction for real-life sensors detailed in Section 3.1 and Section 3.2, respectively, let $\Omega_{<}$ denote the virtual sampling domain for pinhole ray casting. Consequently, the virtual range image is denoted by $\mathcal{R}_{<} : \Omega_{<} \rightarrow \mathbb{R}^+$ and the corresponding surface is denoted by $\mathcal{S}_{<} : \Omega_{<} \rightarrow \Psi_{<}$ with $\Psi_{<} \subset \mathbb{R}^3$. Per pixel viewing rays $v_{<,i}$ with $i \in \Omega_{<}$ are computed from the camera matrix, cf. Eq. (9.6) in Appendix C. Further, without loss of generality, the optical center is set to the canonical origin, i. e. $\mathbf{o} = (0, 0, 0)^\top$. Now, the problem of explicit surface reconstruction from the distance transform is to compute the virtual range image $\mathcal{R}_{<}$ and the surface $\mathcal{S}_{<}$ such that for a 1-D depth measurement the corresponding reconstructed 3-D point coincides with the zero level of the distance transform \mathcal{D} :

$$\mathcal{D}(\mathcal{S}_{<}(i)) = \mathcal{D}(\mathbf{o} + \mathcal{R}_{<}(i) \cdot v_{<,i}) \stackrel{!}{=} 0, \forall i \in \Omega_{<}. \quad (5.18)$$

See also Eq. (3.13) for the relationship to real range measurements.

Pinhole ray casting has a severe drawback in the multi-sensor scenario investigated in this chapter. Though the fusion of multiple RI data allows to address issues that are due to occlusion or partial visibility there is inherently no possibility to reconstruct a high coverage model of the body surface. This issue is depicted in Fig. 5.2a and requires more sophisticated surface reconstruction methods such as the *Manifold Ray Casting* technique that is described in the next section.

5.4.2 Manifold Ray Casting

The basic idea of *Manifold Ray Casting* is to have the viewing rays to emanate from a 2-manifold that surrounds the patient instead of one single optical center. A sketch of this idea is given in Fig. 5.2b. As a distinguishing feature compared to a conventional technique based on a pinhole camera model is the ability to reconstruct a 180° body surface model with a 2-D parameterization imposed by the manifold sampling domain $\Omega_{\odot} \in \mathbb{R}^2$. Of course this requires both a suitable choice for the manifold and a parametric representation of the manifold to specify the topology of Ω_{\odot} .

In this work the focus is on a manifold based on a half cylinder enclosing the distance transform \mathcal{D} and thus the patient body. The reasons for this choice is threefold. First, a half cylinder is a suitable approximation of the human body and thus provides prior knowledge of the body shape. Second, the 2-D parameterization to define the topology of Ω_{\odot} can be deduced from cylinder coordinates in a straight-forward manner. Third, the projection of a point $x \in \mathbb{R}^3$ onto the cylinder manifold \odot is given by a closed-form solution which is a very important aspect towards run-time performance and projective data association schemes as used in Chapter 7 for efficient nearest neighbor computation.

Assuming a discretization of Ω_{\odot} with $N_1 \times N_2$ sample points $\mathbf{i} = (i_1, i_2) \in \Omega_{\odot}$ the 2-D parametric cylinder coordinates (ρ_{i_1}, ϕ_{i_2}) are given as:

$$\begin{aligned}\rho_{i_1} &= i_1 \cdot \frac{h}{N_1 - 1} - \frac{h}{2}, \\ \phi_{i_2} &= i_2 \cdot \frac{\Phi}{N_2 - 1} - \frac{\Phi}{2}.\end{aligned}\tag{5.19}$$

Here, $h \in \mathbb{R}^+$ is the physical length of the cylinder and $\Phi \in [0, 2\pi]$ denotes the radial coverage. Now, given the radius r of the cylinder, the ray origins $\mathbf{o}_{\odot, i}$ are given by the corresponding 3-D cylinder coordinates as:

$$\mathbf{o}_{\odot, i} = (r \cos(\phi_{i_2}), r \sin(\phi_{i_2}), \rho_{i_1})^{\top}.\tag{5.20}$$

The viewing rays $\mathbf{v}_{\odot, i}$ with $\|\mathbf{v}_{\odot, i}\|_2 = 1$ are defined by the cylinder normals as:

$$\mathbf{v}_{\odot, i} = (\cos(\phi_{i_2}), \sin(\phi_{i_2}), 0)^{\top}.\tag{5.21}$$

Surface reconstruction using ray casting is then similar to the pinhole model from Eq. (5.18) with the difference that there are now multiple optical centers.

For reconstructed range measurements $\mathcal{R}_{\odot}(\mathbf{i})$ that are smaller than the cylinder radius r the original cylinder manifold and the reconstructed surface \mathcal{S}_{\odot} are homotopic, i. e. they can be continuously deformed into each other without breaking the topology imposed by Ω_{\odot} . For surfaces \mathcal{S}_{\odot} that do not exhibit strong discontinuities this also implies that mutual distances of points in a local neighborhood of \mathcal{S}_{\odot} can be modeled by mutual distances of the corresponding index neighborhood in Ω_{\odot} . Note that a similar idea governs the *Isomap* manifold learning algorithm [Tene00] where one aims to find a distance-preserving mapping for non-linear dimensionality reduction. For RI this has the consequence that the 2-D domain Ω_{\odot} can be used for efficient image or mesh enhancement techniques on \mathcal{R}_{\odot} or \mathcal{S}_{\odot} , respectively. Naturally, this enables to use the pre-processing techniques investigated in Chapter 4 as post-processing methods on \mathcal{R}_{\odot} after ray-casting prior to reconstructing \mathcal{S}_{\odot} . Further, as investigated later in Chapter 7, a nearest neighbor problem in \mathcal{S}_{\odot} can be approximated very efficiently via Ω_{\odot} using projective data association techniques.

5.5 Experiments and Results

The experiments for the proposed framework for multi-view RI data fusion and high-coverage surface reconstruction are divided into three different categories. First, a study on synthetic RI data using a respiration phantom is performed for quantitatively assessing the suitability of the proposed approach in a dynamic scenario. Next, qualitative results using real RI data of human subjects are presented. The section concludes with a comprehensive performance study.

5.5.1 Quantitative Results Using a Respiration Phantom

Quantitative results for the suitability of the proposed fusion and reconstruction framework in a dynamic respiratory motion scenario are based on simulated RI data and a synthetic respiration phantom. The purpose of this virtual test environment is to assess the intrinsic parameters of the fusion framework in the absence of extrinsic error sources such as physical limitations in the diverse range measurement principles, device specific acquisition frame-rates and multi-camera synchronization issues as well as sensor calibration or manufacturing inaccuracies.

Experimental Setup

Evaluation is performed by using the respiratory motion sequence generated from the NCAT phantom as outlined in Section 4.7.1. However, instead of one single virtual camera, a triple RI sensor setup that enables a 180° body coverage was used. For this purpose, two additional virtual cameras were placed lateral to the torso at opposing viewpoints to capture the motion model with an overlap at the anterior thorax region.

Sensor characteristics For evaluation, the motion model sequence was sampled both at full $f_s = 30$ Hz and half $f_s = 15$ Hz sampling rate. These two sampling rates are introduced to assess the effects of long computation or processing times that hinder a real-time system to cope with the frame rate of the RI sensor which ultimately results in the loss of some patient surface data. For some experiments, to assess the fusion framework in a more realistic manner, the synthetic range data $\mathcal{R}_{1,2,3}$ from the three cameras was corrupted with the same noise characteristics as in Section 4.7.1, i. e. missing data based on sensor specific defect probability maps, static *Perlin* noise and temporal variations that are sampled from a normal distribution ($\sigma_N = \sqrt{2}$) and as a last step quantization artifacts ($\Delta_Q = 4$). Optionally, for corrupted data, post-processing using techniques outlined in Chapter 4 was performed on the reconstructed RI data \mathcal{R}_\odot prior to computing \mathcal{S}_\odot .

Fusion and reconstruction configuration The physical extent of the embedding domain Γ was chosen to cover the respiration phantom and corresponds to a rectangular cuboid with an edge length of approximately $664 \text{ mm} \times 257 \text{ mm} \times 422 \text{ mm}$ in SI, anterior-posterior (AP) and medio-lateral (ML) direction, respectively. Unless stated otherwise, the reported results were obtained by using a discretization of the distance transform according to $\Gamma \in \mathbb{R}^{512 \times 512 \times 512}$. The support region for approximated distances was empirically chosen as $\epsilon = 20$ mm, the temporal integration parameter α is varied between experiments to assess its influence. For surface reconstruction using manifold ray casting a half cylinder along the SI direction was chosen with suitable radius $r = 255$ mm close to the AP extent and a height equal to the SI size, i. e. $h = 664$ mm. For all experiments the virtual RI domain for ray-casting was kept fixed with an arbitrarily chosen resolution according to $\Omega_\odot \in \mathbb{R}^{640 \times 480}$.

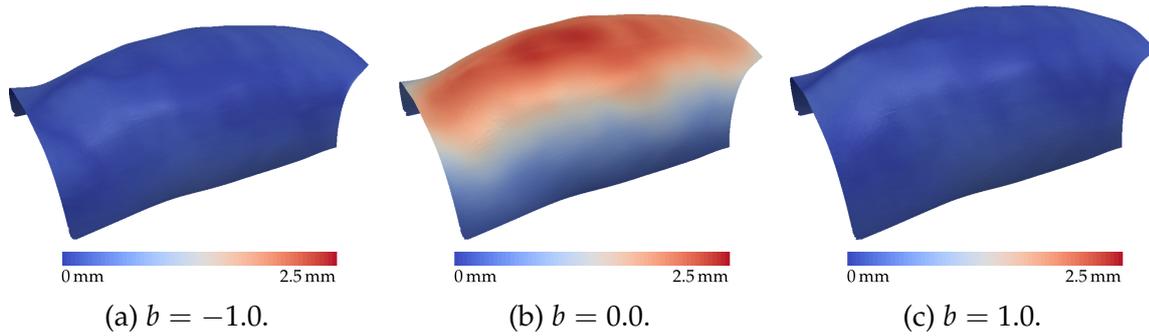


Figure 5.3: Color-coded point-wise distances from the respiration phantom to the reconstructed surface. The left and right figures approximately correspond to a steady exhale and inhale phase, respectively, whereas the center figure depicts an intermediate state between exhale and inhale. The reconstructions were obtained from successive temporal data integration with $\alpha = 0.4$.

Error assessment The actual performance of the fusion framework is assessed by the reconstruction error that is expressed as the point-to-surface distance from the ground truth respiration phantom to the corresponding reconstructed surface. A point-to-surface distance was favored over a point-to-point metric to cope with systematic errors that would occur due to the different topology of the reconstructed surface defined on a dense regular grid and the motion model that is governed by a sparse unstructured grid. In contrast to the preceding section, the evaluation is not restricted to a specific sub region but is performed across the entire motion model surface. This restriction could be relaxed as a direct consequence of the high coverage reconstruction ability of the proposed framework. For compactification, the first (Q_1), second (Q_2) and third (Q_3) quartiles of the errors across the entire surface are reported.

To assess the difference between different approaches or parameterizations a left-tailed *Wilcoxon signed-rank test* was performed. The null hypothesis is that the median difference in Q_i errors between two experiments across all frames equals a threshold δ whereas the alternate hypothesis is that the median difference is smaller than that threshold. For details see the description in Section 4.7.1.

Results

Experimental results are reported for the general accuracy of the proposed fusion and reconstruction framework, the ability to account for noise and artifacts and, last, for the discretization of the SDF embedding domain Γ .

General error and temporal integration Exemplary distributions of point-to-surface distances for different motion model states are depicted in Fig. 5.3 and a temporal integration parameter $\alpha = 0.4$. Conspicuous is the discrepancy of reconstruction errors between the steady respiration states in Figs. 5.3a and 5.3c and the intermediate state visualized in Fig. 5.3b. Whereas the former errors do not exceed 0.5 mm the reconstructed intermediate respiration state features a distinct

error clustering in surface regions that correspond to respiration motion in AP direction and exhibit an error scale of ~ 2.5 mm.

This effect is detailed in Fig. 5.4 for varying temporal integration parameters α along with different sampling rates f_s across the entire respiration sequence. Similar to the errors corresponding to the intermediate state in Fig. 5.3b there are distinct spikes in the error graphs that coincide with a changing respiration state or magnitude. In fact, for a high temporal averaging and low sampling rate as outlined in Fig. 5.4a a statistical significant PCC of 0.85 ($p \leq 0.01$) was found between the frame-wise median reconstruction errors and the absolute values of the derivative of the model parameters w. r. t. the frame index. This correlation coefficient drops down to 0.15 ($p \leq 0.01$) for $\alpha = 0.9$ and a sampling rate of $f_s = 30$ Hz. The reason for this effect is that for high temporal averaging, i. e. small values for α , the reconstructed surface is by the definition of data fusion from Eq. (5.14) composed of several instantaneous body surfaces that were captured at different time points. Fusing body surfaces acquired at different respiration states thus inevitable results in an erroneous surface reconstruction. This effect is heavily influenced by the surface sampling frequency as outlined in the left column of Fig. 5.4 for the half sampling rate of $f_s = 15$ Hz. These errors are similar to pre-processing results as investigated in Chapter 4 in the context of temporal denoising. For noise-free data, performing no temporal integration at all ($\alpha = 0$) yields the lowest reconstruction error with a median error in the scale of 0.1 mm. However, there is no significant difference ($\delta = 0.1, p \leq 0.01$) between $\alpha = 1.0$ and $\alpha = 0.9$ for the Q_1, Q_2 and Q_3 surface reconstruction errors and the full sampling rate $f_s = 30$ Hz.

In this regard, the reconstruction error corresponding to exclusive instantaneous data constitutes the lower error bound for the fusion and reconstruction framework proposed in this thesis. This lower error bound is a superposition of various effects including voxel spacing, approximated distances and surface reconstruction using ray-casting. The effect of voxel spacing in the embedding domain Γ will be detailed later as it is directly linked to the run-time performance of the fusion framework. However, the other effects are not investigated further as the overall reconstruction error is considerably smaller than the noise level of the synthetic and real-life RI data as investigated in this thesis.

Noise and artifacts Surface reconstruction errors for RI data that was corrupted by noise are illustrated in the left column of Fig. 5.5 across the entire motion sequence. Again, distinctive spikes corresponding to changing respiration states are noticeable. Though minor noise reductions are observable for steady respiration states and a high temporal integration parameter of $\alpha = 0.4$ as depicted in the first 10 frames in Fig. 5.5a the error scale during phases of changing respiration state is very high and not acceptable in practice. Further, when performing a subsequent post-processing step as shown in the right column of Fig. 5.5, the effect of temporal denoising in steady respiration states is hardly visible and performing no temporal integration yields the best result with a median error of ~ 0.2 mm. Interestingly, though the exact same parameterization for noise and artifacts as in the quantitative pre-processing evaluation in Section 4.7.1 was used, the fusion reconstruction error for the raw data in the scale of 0.4 mm is considerably smaller compared to a

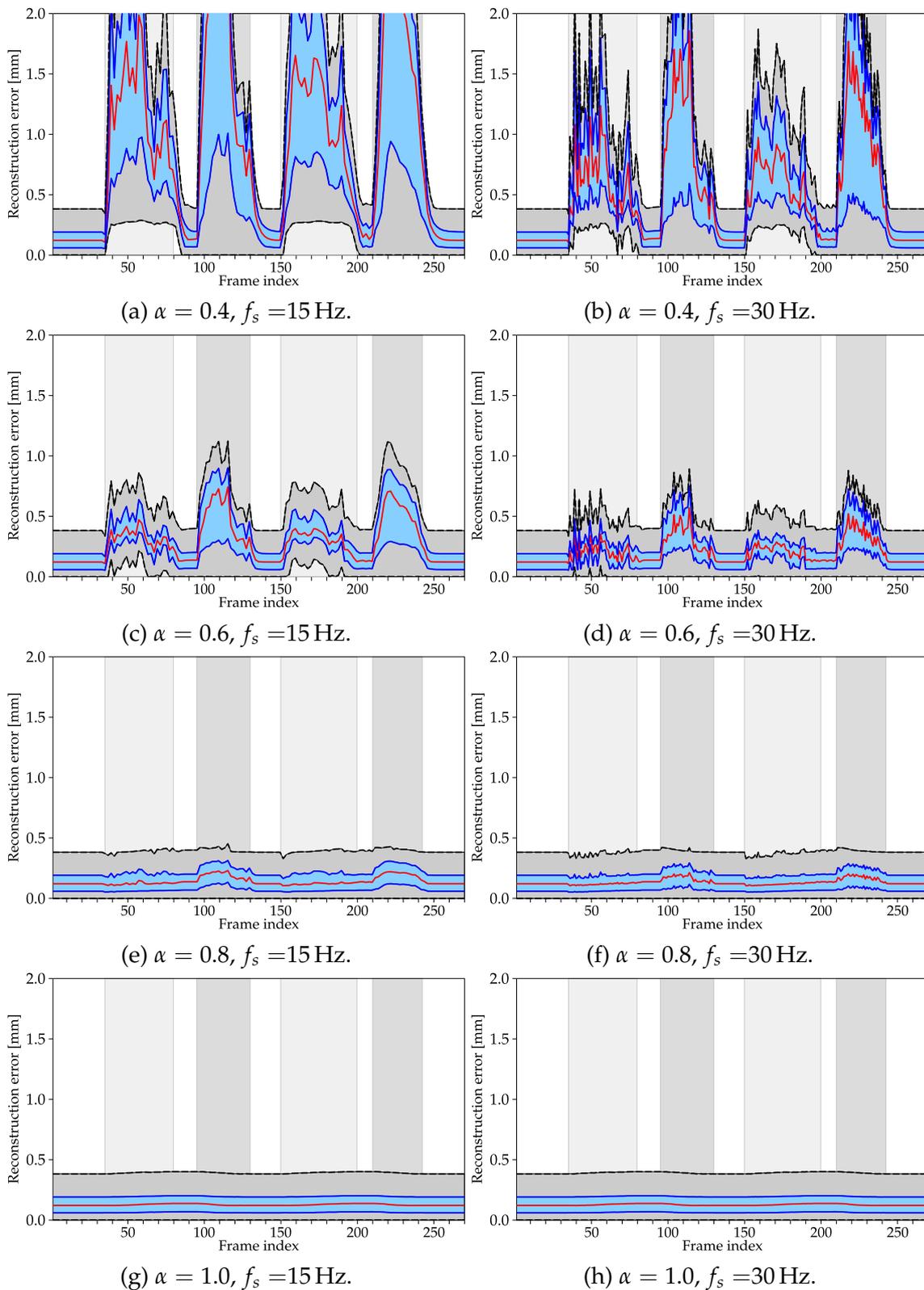


Figure 5.4: Reconstruction error expressed as the point-to-surface distance between the ground-truth respiration phantom and the reconstructed body shape for varying temporal integration parameters α and sampling rates f_s .

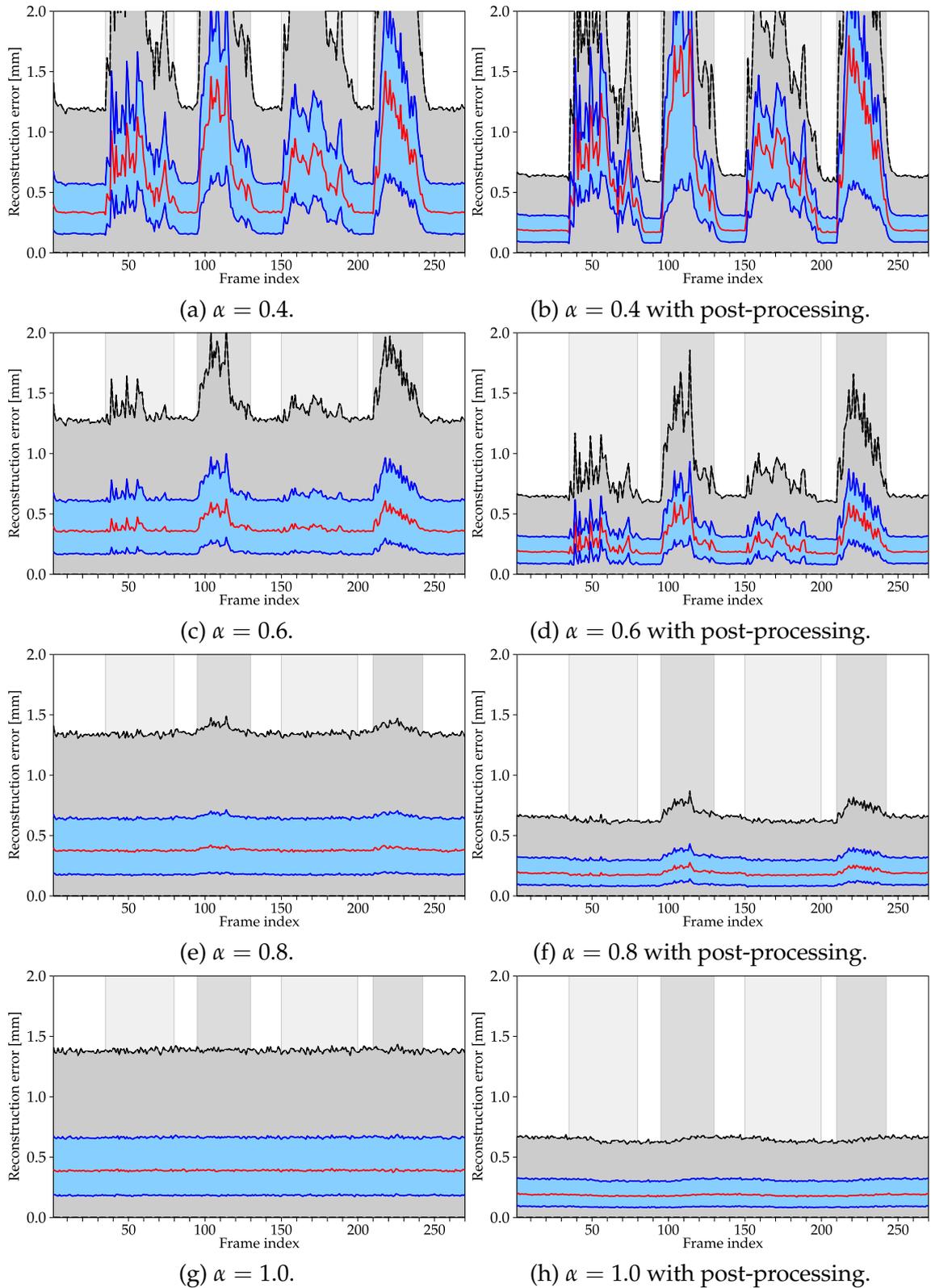


Figure 5.5: Surface distance between the ground-truth respiration phantom and the reconstructed body shape for varying temporal integration parameter α in the presence of noise and artifacts. Optionally, as shown on the right column, post-processing was performed on the reconstructed range data \mathcal{R}_\odot using normalized convolution and bilateral filtering, cf. Chapter 4.

Size of Γ	Ideal data			Corrupted data		
	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
128^3	0.06	0.12	0.19	0.10	0.21	0.35
256^3	0.06	0.13	0.19	0.09	0.19	0.32
512^3	0.06	0.13	0.19	0.09	0.19	0.31

Table 5.1: Reconstruction errors for varying discretizations of the SDF embedding domain Γ . For corrupted data post-processing is performed. Reported are the mean Q_i errors in [mm] across the entire motion model sequence. The standard deviation did not exceed 0.01 and thus is omitted. Note that the resolution of Γ has no substantial impact on the reconstruction accuracy.

median error of ~ 1.5 mm that was observed for corrupted data in Chapter 4. See Fig. 5.5g and Fig. 4.10b for a direct comparison. The reason for this discrepancy is partially due to the overlapping field of views in the triple sensor setup that allows for a super-resolved surface reconstruction especially when considering quantization artifacts. Similar effects were reported for scene reconstruction in the Kinect fusion framework [Newc 11], however, using a single but moving camera.

Discretization of Γ The last part of this evaluation is concerned with the impact of the discretization of the SDF embedding domain Γ to the reconstruction accuracy. The errors across the entire motion sequence are reported in Table 5.1. The results show that the error induced by the proposed fusion framework is mostly independent from the chosen discretization of the SDF embedding domain Γ . For example there is no significant difference ($p \leq 0.01$) for the ideal Q_1 , Q_2 and Q_3 reconstruction errors between a discretization with 128^3 and 512^3 voxels. This effect becomes clear by recalling that for reconstructing the fused representation \mathcal{R}_\odot a ray casting technique is applied, cf. Section 5.4. The ray casting method is implemented to perform a trilinear interpolation of the SDF and thus the sensitivity w. r. t. the spacing of the SDF is decreased.

5.5.2 Qualitative Results on Real RI Data

The qualitative assessment of the proposed fusion and reconstruction framework on real data is performed for four healthy male subjects $S_1 - S_4$.

Experimental Setup

The subjects' body surface was acquired using two Kinect RI sensors that were positioned and oriented to provide a high coverage with a slightly overlapping field-of-view in the thorax region. The Kinects were calibrated using an external tool based on the corner-detection framework in the OpenCV software library [Brad 08]. Time-synchronized multi-view RI data consisting of 10 successive frames for each sensor were then fused with a temporal integration weight of $\alpha = 0.6$ that was chosen heuristically for the Kinect RI devices to provide a trade-off between denoising and dynamic scene preservation. The distance transform \mathcal{D} and the



Figure 5.6: Qualitative results for RI data fusion and surface reconstruction for subject S_1 . The subject was captured using two Kinect RI sensors. The proposed fusion and reconstruction framework enables a high coverage body surface model, however, the mesh reflects the noise characteristics of the RI device and exhibits artifacts due to sensor interference.

weighting function \mathcal{W} were discretized on a regular grid covering Γ with 256^3 voxels using a subject specific spacing that was adjusted on-the-fly. The body surface was then reconstructed using manifold ray casting with a cylindrically shaped sensor domain that was discretized with 640×480 pixels.

Results

Qualitative surface reconstruction results using this configuration are depicted in Fig. 5.6. This figure demonstrates that high coverage body surface models are enabled by the proposed multi-view RI fusion and reconstruction framework. However, though 10 successive frames were used for temporal data denoising, the reconstructed mesh still shows some rough parts that resemble the discretization artifacts known from the Kinect sensor, cf. Chapter 4. Note that a super-resolved scene as computed by Newcombe et al. [Newc 11] is unfeasible as the RI sensors are mounted statically and observe a body surface that is either not moving or deforming non-rigidly. Though, in theory, a non-rigid surface registration could help to set up point-correspondences for a super resolution approach this is not feasible in practice due to run-times that do not satisfy real-time constraints. Besides the rough surface, the reconstructed mesh exhibits discontinuities in the umbilicus region that are due to sensor interference.

Therefore, prior to computing the actual 3-D mesh, the range data \mathcal{R}_\odot as computed from ray-casting the distance transform is enhanced using normalized convolution to account for missing values and, as a finalizing processing step, guided filtering for the purpose of edge-preserving denoising, cf. Sections 4.4.1 and 4.5.2. The reconstruction results using this post-processing steps are depicted in Fig. 5.7. For all subjects post-processing yields a smooth surface representation with clearly visible retained salient landmarks such as the umbilicus and the ribcage.

5.5.3 Performance Study

The performance study is conducted on both the synthetic respiration phantom using a triple-sensor setup as well as the real life subjects with a dual-sensor setup.

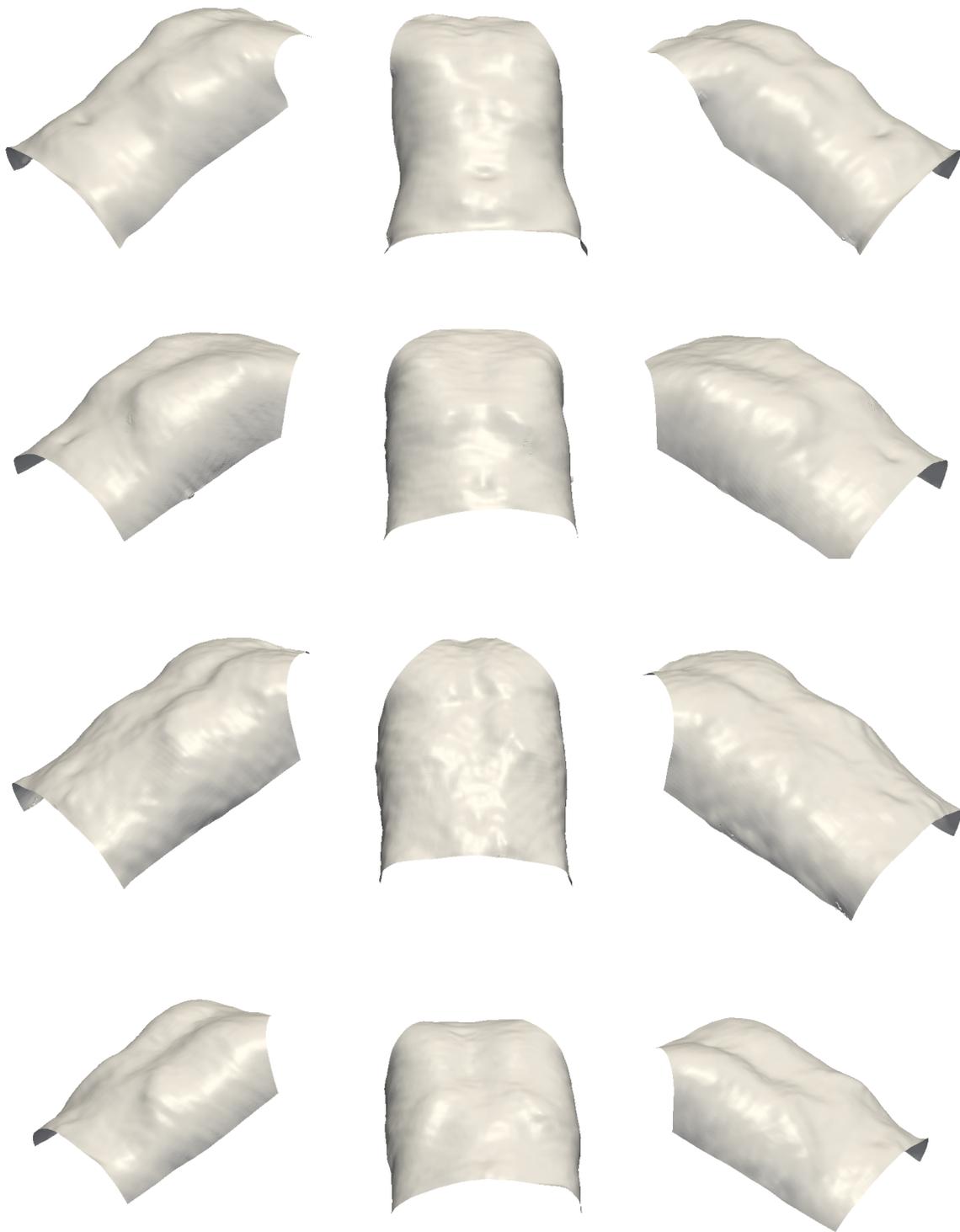


Figure 5.7: Qualitative results for RI data fusion and surface reconstruction for four male subjects. The body surfaces were captured using two Kinect RI sensors. Note the wide field of view coverage for all subjects and salient anatomical landmarks. For visualization purposes the surface meshes are cropped to a region of interest covering the upper body excluding arms and the neck.

Size of Γ	Step	P ($K = 3$)	S_1 ($K = 2$)	S_2 ($K = 2$)	S_3 ($K = 2$)	S_4 ($K = 2$)
128^3	T_P	3.16 ± 0.05	1.41 ± 0.18	1.42 ± 0.19	1.41 ± 0.18	1.42 ± 0.17
	T_F	1.86 ± 0.03	0.18 ± 0.01	0.19 ± 0.01	0.18 ± 0.01	0.21 ± 0.02
	T_R	0.72 ± 0.04	0.66 ± 0.06	0.64 ± 0.06	0.70 ± 0.07	0.66 ± 0.05
	T_{PP}	1.89 ± 0.04	2.13 ± 0.16	2.14 ± 0.14	2.14 ± 0.15	2.15 ± 0.15
	T	7.63 ± 0.16	4.38 ± 0.41	4.39 ± 0.40	4.43 ± 0.41	4.44 ± 0.39
256^3	T_P	3.21 ± 0.17	1.43 ± 0.18	1.47 ± 0.24	1.41 ± 0.19	1.45 ± 0.20
	T_F	2.33 ± 0.02	1.03 ± 0.02	1.02 ± 0.02	1.13 ± 0.03	1.07 ± 0.02
	T_R	0.87 ± 0.05	0.78 ± 0.07	0.75 ± 0.06	0.83 ± 0.06	0.79 ± 0.07
	T_{PP}	1.88 ± 0.06	2.15 ± 0.16	2.16 ± 0.15	2.18 ± 0.20	2.17 ± 0.16
	T	8.29 ± 0.30	5.09 ± 0.43	5.40 ± 0.47	5.55 ± 0.48	5.48 ± 0.45
512^3	T_P	3.18 ± 0.15	1.45 ± 0.19	1.40 ± 0.18	1.41 ± 0.19	1.44 ± 0.21
	T_F	11.3 ± 0.19	7.34 ± 0.15	7.21 ± 0.08	8.07 ± 0.06	7.59 ± 0.10
	T_R	2.40 ± 0.16	1.95 ± 0.08	1.80 ± 0.08	2.32 ± 0.07	2.00 ± 0.09
	T_{PP}	1.89 ± 0.05	2.20 ± 0.17	2.15 ± 0.16	2.14 ± 0.16	2.17 ± 0.20
	T	18.8 ± 0.55	12.9 ± 0.59	12.6 ± 0.50	13.9 ± 0.48	13.2 ± 0.60

Table 5.2: Average run-times in [ms] for the multi-view RI framework. Reported are data preparation T_P (host-device RI data transfer, confidence computation), multi-view SDF fusion T_F , surface reconstruction using manifold ray-casting T_R , post-processing T_{PP} and the total time T . For the respiration phantom P a triple-sensor setup ($K = 3$) and for subjects S_1 - S_4 a dual-sensor setup ($K = 2$) was used.

Experimental Setup

For better interpretation, the total run-time was subdivided into four distinct components consisting of (1) data preparation time T_P which includes the host-device RI data transfer and the computation of the confidence maps, (2) SDF integration or fusion time T_F , (3) reconstruction time using manifold ray casting T_R and (4) post-processing time T_{PP} . Due to its importance regarding the trade-off between spatial resolution on the one hand and memory requirements or computational complexity on the other hand, run-times were investigated for different resolutions of the embedding domain Γ . For details on the hardware configuration and time measurements the reader is referred to Appendix B.

Results

The fusion and reconstruction run-times for the respiration phantom and subjects S_1 - S_4 are given in Table 5.2. For large resolutions, the overall run-times are dominated by the SDF integration or fusion and run-times are directly proportional to the SDF resolution, i. e. $T_F \propto |\Gamma|$. However, increasing the resolution also results in higher reconstruction times T_R which is among others due to the fact that ray casting a larger volume also increases the number of texture cache misses.

In general, run-times vary only slightly across the different respiration states. This is not self-evident as the number of arithmetic operations and memory transactions to compute \mathcal{D} and \mathcal{W} are scene dependent. For example, for small weights

in the current estimate for \mathcal{W} close to zero there is no need to update the new estimate and points that are outside the viewing frustum of the cameras, i. e. they are not projected into the sensor planes Ω , are discarded completely. As expected, the run-times for data preparation T_P and post-processing T_{PP} do not depend on the volume configuration and constitute an overhead that must not be neglected, especially for small resolutions.

5.6 Discussion and Conclusion

This chapter proposed and investigated a real-time capable framework that enables unifying high coverage body surface models computed from multi-view RI. The results show that a synthetic respiration phantom can be reconstructed with a typical reconstruction error in the scale of 0.2 mm for corrupted data. For real-life RI data, distinct anatomical landmarks are perceptible throughout the reconstructed surface. Regarding real-time constraints, a GPU implementation has proven to be very efficient. As the reconstruction accuracy is not influenced substantially by the chosen discretization of the implicit surface representation, an entire pipeline consisting of dual sensor multi-view data fusion, explicit surface reconstruction and subsequent post-processing can be realized in 4.4 ms. This copes with the frame-rate of current RI sensors and enables low latencies while simultaneously saving resources for subsequent data processing and analysis steps.

The key issue with the proposed method is the temporal integration of RI data in a respiratory motion scenario. Though temporal denoising is an often used technique, the quantitative evaluation show that there exists a strong correlation between the surface reconstruction error and changing respiration states for aggressive up to moderate temporal averaging strategies. This is similar to the results for temporal denoising that were investigated in Chapter 4 of this thesis. At a first glance this favors to perform no temporal integration at all. However, the spatial distribution of errors induced by temporal averaging shows a predominant cluster that corresponds to regions of high respiration magnitude at the anterior thorax whereas corruptions by noise are evenly spread across the whole surface. In certain applications, particularly for patient positioning, it might be beneficial to give priority to stationary body surface regions that do not move with respiratory motion. Clearly, these regions can benefit from a temporal averaging strategy.

With regard to future generations of RI devices and the steady trend towards higher frame-rates in range imaging the partially negative effect of temporal integration can potentially be lessened. This is underlined by the experiments using different sampling rates of 15 Hz and 30 Hz. However, also given frame rates of 80 Hz as available with ToF devices, the temporal integration parameter must always be chosen carefully and temporal integration must be considered as a trade-off between a stable and steady surface and an accurate reconstruction.

In particular with regard to temporal integration issues, the experiments conducted in this chapter underline the fact that RI data processing in a highly dynamic scenario such as respiratory motion analysis is fundamentally different from conventional static scenarios.

Part III

Respiratory Motion Analysis Using 4-D Shape Priors

Foundations of 4-D Shape Priors for Respiratory Motion Analysis

6.1 Motivation	75
6.2 Related Work	76
6.3 Principles of 4-D Shape Priors	78
6.4 Shape Correspondences by Deformable Registration	82
6.5 Dimensionality Reduction Techniques	85
6.6 Experiments and Results	89
6.7 Discussion and Conclusion	98

This chapter covers the fundamental concepts of 4-D shape priors for respiration analysis. In particular, this includes the basic techniques needed for training 4-D motion models as pursued in preceding work [[Wasz 12a](#), [Wasz 12b](#), [Wasz 13](#), [Wasz 16](#)]. Further, a joint theoretical investigation of respiration induced deformations for unambiguous respiratory motion analysis that was first proposed in preceding work [[Wasz 13](#)] is detailed. As a major part of this chapter, sparse representation schemes for 4-D respiratory motion models enabling an unsupervised classification of thoracic and abdominal breathing patterns as published in [[Wasz 12a](#), [Wasz 16](#)] are investigated.

6.1 Motivation

Pre-procedurally trained motion models that encode prior knowledge of patient specific distinct breathing patterns and respiration induced body surface deformations hold many benefits for intra-procedural respiratory motion analysis. Such models that are referred to as 4-D shape priors or motion models in this thesis can be used to train specific breathing patterns for breath-hold motion mitigation strategies, detect deviations of the instantaneous breathing pattern from the planning phase or to enable an automatic differentiation between thoracic and abdominal breathing. Further, prior knowledge on the 4-D appearance of the patient's body surface can be used for motion compensated patient positioning techniques and surrogate driven continuous respiration monitoring. These clinical applications are subsequently detailed in Chapter 7. As another application field, motion

models can be used as a dynamic ground truth for algorithmic verification as used in Chapters 4 and 5 of this thesis.

6.2 Related Work

From a methodological point of view, the body surface motion model as used in this thesis can be interpreted as a statistical shape model (SSM) with pioneering work by Cootes et al. [Coot95] for active shape models. SSMs are widely used in medical image segmentation and for a comprehensive overview on this application field the reader is referred to the review by Heimann and Meinzer [Heim09]. It is worth noticing that, in contrast to most SSMs used for segmentation, the motion models proposed in this thesis are patient-specific. Besides segmentation in medical imaging, statistical shape models have also been proposed for different applications such as growth modeling of human mandibles [Hilg03], synthesis of 3-D faces [Blan99] or for predicting aging trajectories of faces [Sche07a]. Of particular importance for SSMs and for this chapter are the problems of establishing point correspondences across different shapes using non-rigid registration techniques and subsequent modeling of shape variations using dimensionality reduction techniques. However, note that different approaches for computing correspondences apart from non-rigid registration schemes have also been proposed. An example is the minimum description length criterion that reformulates the correspondence problem as a spherical mapping optimization problem [Davi02, Thod03].

As statistical respiratory motion models of the external body surface are not yet described in literature, the related work covers a more broad range of application fields, however, sharing the same basic principles and methodology. For a review of general motion models for respiration analysis the reader is referred to the work of McClelland et al. [McCl13].

Non-rigid surface registration Setting up point correspondences across different shapes is a fundamental part for building SSMs and non-rigid deformation techniques are popular means for this task. For example, Liu et al. use an intensity based non-rigid deformation technique to construct a shape model of the lung [Liu10]. The point correspondences are extracted from the elastic dense deformation fields that map intensity images from different respiration states to a properly chosen reference in the respiration cycle. For registration, a purposely developed technique for lung motion estimation was used [Fosk05]. In this regard, a more recent dedicated registration scheme for lung motion estimation was proposed by Papiez et al. [Papi13]. This approach extends conventional registration schemes with diffusion regularization such as the well known Demons algorithm [Thir98] with an adaptive regularized bilateral filter of the deformation field. Similar to conventional bilateral filtering or edge-preserving denoising in general (cf. Section 4.5), this prevents the deformation field to be smoothed across object boundaries. A dense intensity based registration scheme has also been proposed by Ehrhardt et al. for constructing a 4-D lung motion model [Ehrh11]. Be-

sides intensity based registration schemes for lung motion estimation, different representations have also been proposed. For example Klinder et al. [Klin 10] construct a statistical motion model of the lung using point correspondences that are directly but sparsely established on organ surfaces that have been extracted from segmented and binarized image data. Besides lung motion analysis, binary images have also been used for dense volumetric registration of organs in the SSM-based kidney segmentation framework proposed by Spiegel et al. [Spie 09]. Here, a dedicated curvature similarity metric is incorporated to account for shape properties. However, instead of using shapes or binary data directly, Huang et al. [Huan 06] proposed to perform non-rigid registration on the distance transform or implicit shape representations.

So far these methods had a strong focus on registration of internal shapes and only recently dedicated methods for registering body surface data have been proposed. Among the first was the work by Schaerer et al. that use a non-rigid Iterative Closest Point (ICP) variant to generate high dimensional respiration surrogates [Scha 12]. This non-rigid ICP was originally proposed by Amberg et al. as a general purpose point-set registration technique [Ambe 07]. In this context, alternative generic point-set registration schemes include the work of Myronenko and Song [Myro 10] or Jian and Vemuri [Jian 11] where the registration problem is interpreted as a density estimation problem for a Gaussian mixture model. In contrast to these general purpose mesh or point-set registration schemes, Bauer et al. proposed dedicated variational surface deformation estimation techniques that are explicitly designed to exploit the governing representation of intra-procedural RI and pre-procedural planning data [Baue 12b, Baue 12a]. This includes both the expected RI noise level and data reliability as well as RI sampling density.

Dimensionality reduction Modeling shape variations using dimensionality reduction techniques is an ubiquitous part for SSMs. The fundamental principle here is that a set of corresponding shapes, each typically consisting of thousands of individual surface points, is actually governed by only a few intrinsic parameters that have to be recovered from this training set. Any unseen test shape is then approximated by estimating the parameters such that the corresponding high dimensional model resembles the test shape. Typically, this step also includes plausibility enforcement by placing boundaries for parameter values.

One key characteristic of dimensionality reduction techniques is if varying one low-dimensional parameter impacts the entire shape on a global scale or if the effect is limited to a locally clustered set of points. Conventional techniques like Principal Component Analysis (PCA) are global methods [Coot 95] and different approaches for enforcing locality or sparsity have been proposed. For example, Stegman et al. used orthomax rotations to construct sparse modes of variations based on a conventional preceding PCA [Steg 06]. In contrast, Sjöstrand et al. use the sparse PCA proposed by Zou et al. [Zou 06] to directly generate a sparse representation for modeling shape variability [Sjos 07]. Other popular sparse methods used for shape modeling include *independent component analysis* [Uzum 03] or *maximum autocorrelation factors* [Hilg 03].

A different aspect of dimensionality reduction techniques is concerned with linear and non-linear decomposition schemes of the training data. The original linear methods were soon relaxed by using regression techniques [Sozo 94], mixture models [Coot 99] or kernel PCA (KPCA) [Twin 01]. However, in contrast to linear decomposition schemes like PCA and derived sparse variants, non-linear methods such as KPCA or manifold-learning techniques including Locally Linear Embedding (LLE) [Rowe 00] and Laplacian Eigenmaps (LEM) [Belk 03] do not provide direct solutions for the so-called *out-of-sample* and the *pre-image* problems. The *out-of-sample* problem refers to assigning a label or a low-dimensional value to data that was not seen in the training stage. This can also be interpreted in terms of *inductive* opposed to *transductive* reasoning [Zhu 08, Chap 06]. Typical solutions to this problem are the re-computation of the model or incremental learning techniques [Law 06, Fisc 14]. In contrast, the *pre-image* problem is concerned with the mapping from the low dimensional space to the original high dimensional space [Hone 11]. These problems usually require a re-computation of the model or iterative optimization techniques as proposed for non-linear SSMs by Twining and Taylor [Twin 01] or Kirschner et al. [Kirs 11]. For specific kernels dedicated non-iterative methods have been proposed [Kwok 04], however, these approaches are ill-conditioned thus requiring appropriate penalizing strategies [Zhen 10]. In general, non-linear methods for shape modeling are not widespread, and for example in medical image segmentation the majority of approaches rely on linear methods [Heim 09].

6.3 Principles of 4-D Shape Priors

The basic principle of 4-D shape priors as pursued in this thesis is to *pre-procedurally* acquire the patient's body surface at different respiration states and then perform dimensionality reduction techniques on the aligned training data to generate a 4-D motion model that encodes prior knowledge of patient specific breathing patterns. Here, the term 'aligned' refers to registered shapes with established point or landmark correspondences. *Intra-procedurally*, this motion model is then registered to RI surface data which eventually provides the basis for motion compensated patient alignment as well as continuous respiration monitoring. These clinical applications and their unification in a mathematical framework are subsequently detailed in Chapter 7.

6.3.1 Shape Motion Models Using Displacement Fields

One crucial challenge for respiratory motion models is the pre-procedural representation of the patient body surface at different respiration states. In contrast to the previous sections where a continuous RI body surface was represented as a set of points organized on a regular grid, a well-defined topology and uniform sampling of surface points is not guaranteed. This is due to the fact that pre-procedural body surfaces do not necessarily have to be acquired by RI cameras but instead could be extracted from 4-D volumetric planning data as obtained from CT or MRI. Further, laser scanners or markers attached to the body do not provide a

continuous surface representation at all. Thus, in the context of motion models as pursued in this thesis, the patient's body is represented as a *point-cloud* or *point-set* rather than a continuous surface.

Following the point set paradigm, let $\mathcal{P}^s = \{\mathbf{x}_1^s, \dots, \mathbf{x}_N^s\}$, $\mathbf{x}_i^s \in \mathbb{R}^3$ denote the set of N points that describe the body surface at respiration state s . For the motion model generation it is of utmost importance that the point sets are aligned properly such that individual 3-D points from different respiration states describe corresponding landmarks and that differences are solely due to respiratory motion and not due to artifacts caused by data representation or sampling. This property can be enforced by modeling the surface points \mathcal{P}^s for respiration state s through a set of points $\mathcal{P}^0 = \{\mathbf{x}_1^0, \dots, \mathbf{x}_N^0\}$, $\mathbf{x}_i^0 \in \mathbb{R}^3$ from a reference respiration state 0, e. g. fully exhaled, that are warped with an elastic deformation. This deformation is represented as a displacement field $\mathcal{U}^s = \{\mathbf{u}_1^s, \dots, \mathbf{u}_N^s\}$, $\mathbf{u}_i^s \in \mathbb{R}^3$ encoding the point-wise displacements induced by respiratory motion as:

$$\mathbf{x}_i^s = \mathbf{x}_i^0 + \mathbf{u}_i^s, \forall \mathbf{x}_i^s \in \mathcal{P}^s. \quad (6.1)$$

The displacement field \mathcal{U}^s that eventually accounts for valid point correspondences between different respiration states can be computed using various non-rigid surface registration techniques that are outlined in Section 6.4. Note that this formulation does not cover non-respiratory motion as a proper patient alignment is assumed. See Chapter 7 for details on such methods.

For a subsequent analysis of different respiration states a formulation in a high-dimensional vector space $\mathcal{H} \subset \mathbb{R}^{3N}$ is often preferable compared to point sets or displacement vectors. Let \mathcal{L} be defined as an operator that concatenates a set of 3-D vectors or points $\mathcal{Z} = \{\mathbf{z}_i\}$, $\mathbf{z}_i \in \mathbb{R}^3$ to a single vector $\mathbf{z} \in \mathcal{H}$ as:

$$\mathbf{z} = \mathcal{L}(\mathcal{Z}) = \left(\mathbf{z}_1^\top, \dots, \mathbf{z}_N^\top \right)^\top. \quad (6.2)$$

Conversely, \mathcal{L}^{-1} is defined as the 'inverse' transformation operator that restores the initial representation, i. e. $\mathcal{L}^{-1}(\mathcal{L}(\mathcal{Z})) = \mathcal{Z}$. Without loss of generality, the point-wise surface displacement formulation from Eq. (6.1) can thus be equivalently expressed in \mathcal{H} using a linear model according to:

$$\mathcal{L}(\mathcal{P}^s) = \mathcal{L}(\mathcal{P}^0) + \mathcal{L}(\mathcal{U}^s). \quad (6.3)$$

Generating the motion model now consists of recovering the intrinsic structure that governs the high-dimensional representations of the training data $\mathcal{L}(\mathcal{P}^s)$ with $s = 1, \dots, S$. This is done by using dimensionality reduction techniques that, for the intra-procedural model application, must also be capable to perform well on unseen point-sets that are not contained in the training data. These issues are investigated in Section 6.5.

6.3.2 Condition of Body Surface Displacement Fields

Independent from non-rigid registration and dimensionality reduction techniques the formulation of shape correspondences using point-wise displacements from

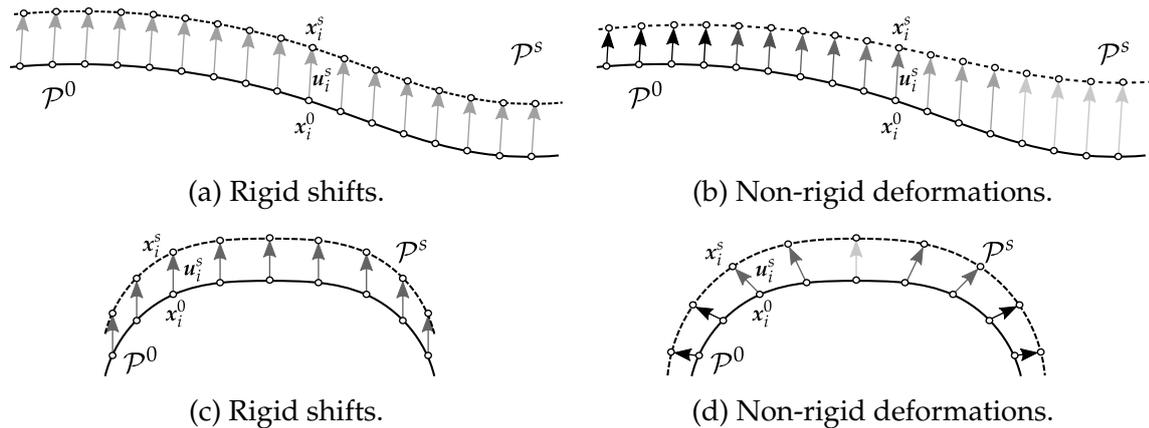


Figure 6.1: Surface displacement fields encoding patient movement (left) and respiratory motion (right) as observed for sagittal (top) and transverse (bottom) views of the torso. Patient movement is characterized by piece-wise global rigid shifts whereas respiratory motion is governed by an elastic deformation. On a local scale displacements corresponding to non-rigid deformations appear as rigid shifts, cf. grayscales of the respiration magnitude. On a global scale, displacements in the sagittal plane exhibit a translational component which is typically not the case for transverse plane.

Eq. (6.1) or the equivalent formulation in \mathcal{H} from Eq. (6.3) allow to deduce a theoretical concept for assessing the condition or fitness of body surface displacement fields for respiratory motion analysis purposes. The general problem that has to be faced is to automatically and quantitatively assess if a given body surface displacement field appropriately reflects non-rigid deformations induced by respiratory motion or if the displacements for example encode rigid shifts that are caused by patient or treatment table movement. This issue is of particular importance if a differentiation between these motion patterns is essential as for example with respiratory motion compensated patient positioning strategies that are investigated in Chapter 7.

A sketch of the general problem is given in Fig. 6.1. Self-evident is the non-rigid nature of respiration induced body surface deformations. However, this only holds true on a *global* scale, that is by observing the entire body surface. On a *local* scale, i. e. when considering only a small region on the body surface, the displacements may appear as rigid shifts that could be erroneously classified as a patient alignment problem and not a respiratory motion mitigation issue. Thus, displacement fields for respiratory motion analysis ideally should not contain any similarity transforms, i. e. translations, rotations and scalings, to account for potential ambiguity errors. This need for robust displacement fields also arises from the fact that in statistical shape modeling it is common to perform a so-called Procrustes Alignment (PA) of the training data [Heim 09]. This removes similarity transforms from the training samples such that only pure shape deformations remain. Consequently, an ideal displacement field for shape modeling does not exhibit any similarity transforms. However, for RI-based respiratory motion analysis, performing PA is counterproductive. As mentioned above, the displacement field depicted

in Fig. 6.1b exhibits a linear component that would be removed by PA. However, if the two surfaces are to be aligned again this would now involve a false additional translation or erroneous alignment step that exactly corresponds to the one removed in the preceding PA step.

For quantitative assessment of body surface displacement fields there are several options. In preceding work [Wasz 13], the assessment problem was formulated in the high-dimensional space \mathcal{H} using the formulation from Eq. (6.3). Inspired by pattern recognition principles where one aims to maximize the distance between different classes [Duda 00] the idea is that the high-dimensional representation of the displacement field $\mathbf{u}^s = \mathcal{L}(\mathcal{U}^s)$ constitutes a direction vector in \mathcal{H} and one wants to maximize the angle to directions $\mathbf{e}_{\{x,y,z\}} \in \mathcal{H}$ that correspond to rigid shifts in x -, y - and z -direction, respectively. Translations in \mathcal{H} are defined as linear combinations of the canonical forms:

$$\begin{aligned} \mathbf{e}_x &= (1, 0, 0, 1, 0, 0, \dots, 1, 0, 0, 1, 0, 0)^\top, \\ \mathbf{e}_y &= (0, 1, 0, 0, 1, 0, \dots, 0, 1, 0, 0, 1, 0)^\top, \\ \mathbf{e}_z &= (0, 0, 1, 0, 0, 1, \dots, 0, 0, 1, 0, 0, 1)^\top. \end{aligned} \quad (6.4)$$

Thus, a metric \mathcal{K}_{SP} to assess and quantify the condition of a body surface displacement field \mathbf{u}^s can be deduced according to the scalar product (SP) as:

$$\mathcal{K}_{\text{SP}}(\mathbf{u}^s) = 1 - \left(\frac{1}{3} \sum_{i=\{x,y,z\}} \frac{|\langle \mathbf{u}^s, \mathbf{e}_i \rangle|}{\|\mathbf{u}^s\|_2 \|\mathbf{e}_i\|_2} \right) \in [0, 1]. \quad (6.5)$$

If \mathbf{u}^s corresponds to a rigid shift, e.g. in x -direction and thus parallel to \mathbf{e}_x , then at least one addend of Eq. (6.5) equates to one. Such rigid shifts are illustrated in Figs. 6.1a and 6.1c. In contrast, if \mathbf{u}^s corresponds to a 'maximally' non-rigid displacement field then all addends will equate to zero. Such a 'maximally' non-rigid displacement field may for example be characterized in the form of $\mathbf{u}^s = (1, 0, 0, 1, 0, 0, \dots, -1, 0, 0, -1, 0, 0) \in \mathcal{H}$ being orthogonal to all rigid shifts $\mathbf{e}_{x,y,z} \in \mathcal{H}$. In RI based respiratory motion analysis this corresponds to distinct body surface regions that move in opposing directions as can be observed in the transverse plane of the torso as depicted in Fig. 6.1d. In contrast, observing solely the anterior regions of the thorax as depicted in Fig. 6.1b is prone to errors as basically all displacement vectors share the same direction and linear dependency is solely broken due to slightly different displacement magnitudes.

A direct consequence of these theoretical and anthropological considerations is that motion models that provide a high body surface coverage are beneficial and of importance in RI based respiratory motion analysis. This is enabled by the multi-view fusion framework proposed in preceding work [Wasz 13] and detailed in Chapter 5 of this thesis.

Though the theoretical basis of the condition metric from Eq. (6.5) allows to derive direct requirements for RI based respiratory motion analysis it does not reflect the whole spectrum of affine transformations. While shearing, reflections and scaling are of no practical relevance in respiratory motion this does not necessarily hold true for rotations. A different formulation for assessing the condition of

displacement fields that incorporates rotations is based on the conventional point-wise displacement formulation from Eq. (6.1). Here, the reconstruction error of displacement fields w. r. t. an optimal rigid body transformation similar to PA are analyzed. Let $(\widehat{\mathbf{R}}, \widehat{\mathbf{t}})$ denote the rotation $\widehat{\mathbf{R}} \in \text{SO}_3$ and translation $\widehat{\mathbf{t}} \in \mathbb{R}^3$ that approximate the displacement field \mathcal{U}^s in a least squares sense according to:

$$(\widehat{\mathbf{R}}, \widehat{\mathbf{t}}) = \underset{\mathbf{R}, \mathbf{t}}{\operatorname{argmin}} \sum_{i=1}^N \|(\mathbf{x}_i^0 + \mathbf{u}_i^s) - (\mathbf{R}\mathbf{x}_i^0 + \mathbf{t})\|_2^2. \quad (6.6)$$

The condition \mathcal{K}_{RE} of \mathcal{U}^s can thus be derived from the residual error (RE) of Eq. (6.6) using a formulation similar to root mean squared errors according to:

$$\mathcal{K}_{\text{RE}}(\mathcal{U}^s) = \sqrt{\frac{\sum_{i=1}^N \|(\mathbf{x}_i^0 + \mathbf{u}_i^s) - (\widehat{\mathbf{R}}\mathbf{x}_i^0 + \widehat{\mathbf{t}})\|_2^2}{\sum_{i=1}^N \|\mathbf{u}_i^s\|_2^2}} \in [0, 1]. \quad (6.7)$$

For a displacement field \mathcal{U}^s that can be perfectly modeled by rotations and translations there is no reconstruction error at all and the nominator of Eq. (6.7) consequently equates to zero. In contrast, in the presence of a 'maximally' non-rigid displacement field that cannot be described by rotations and translations the least squares estimator for Eq. (6.7) necessarily must yield the identity transform, i. e. $\mathbf{R} = \text{diag}(1, 1, 1)$ and $\mathbf{t} = (0, 0, 0)^\top$. In this case the nominator equals the denominator and the condition metric equates to one.

6.4 Shape Correspondences by Deformable Registration

As outlined in the related work in Section 6.2 there exists several non-rigid registration strategies for computing point correspondences across shapes. This section summarizes the registration techniques that have been used in preceding work and are employed in this thesis. The presented approaches can be directly used for the generic model from Eq. (6.1) where one is interested in estimating the displacement field \mathcal{U}^s that describes the elastic deformation ϕ^s matching a patient reference surface and the patient shape acquired at respiration state s .

6.4.1 Non-rigid RI Surface Registration

This method was purposely developed for RI surface registration and the underlying variational formulation was originally proposed by Bauer et al. [Baue 12b] for joint ToF data denoising and registration. However, the method was also successfully employed in preceding work [Wasz 12b] for motion model generation.

For this specific registration scheme, the fixed and moving point sets $\mathcal{P}^0, \mathcal{P}^s$ are represented by the RI surface data $\mathcal{S}^0, \mathcal{S}^s$ that form graphs on the domain Ω , see Eq. (3.12) for the function or graph definition of RI surfaces. Further, the elastic deformation ϕ^s is represented by a displacement field $\tilde{\mathcal{U}}^s : \Omega \rightarrow \mathbb{R}^3$ defined as:

$$\phi^s(\mathcal{S}^0(\mathbf{i})) = \mathcal{S}^0(\mathbf{i}) + \tilde{\mathcal{U}}^s(\mathbf{i}). \quad (6.8)$$

Now, the goal is to estimate ϕ^s in a sense that $\phi^s(\mathcal{S}^0) \approx \mathcal{S}^s$. For this purpose, \mathcal{S}^s is represented by its signed distance function $\mathcal{D}^s : \mathbb{R}^3 \rightarrow \mathbb{R}$ where the sign is positive outside the body and negative inside. This is similar to the volumetric fusion scheme investigated in Chapter 5, however in this chapter the distance transform is neither approximative nor truncated or clamped. By definition of signed distance functions, $\nabla \mathcal{D}^s(x)$ is an outward pointing normal on \mathcal{S}^s and $|\nabla \mathcal{D}^s| = 1$. Based on the distance transform, the projection of a point $x \in \mathbb{R}^3$ onto the closest point on \mathcal{S}^s can be derived according to:

$$P^s(x) := x - \mathcal{D}^s(x) \cdot \nabla \mathcal{D}^s(x). \quad (6.9)$$

This allows to quantify the closeness of a displaced fixed point $\phi^s(\mathcal{S}^0(\mathbf{i}))$ to the moving surface \mathcal{S}^s as:

$$|P^s(\phi^s(\mathcal{S}^0(\mathbf{i}))) - \phi^s(\mathcal{S}^0(\mathbf{i}))| = |\mathcal{D}^s(\phi^s(\mathcal{S}^0(\mathbf{i})))| \quad (6.10)$$

Based on this closeness measure, a variational formulation is used for estimating $\tilde{\mathcal{U}}^s$ as a minimizer of the functional:

$$\begin{aligned} \mathcal{E}[\mathcal{S}^0, \mathcal{D}^s, \tilde{\mathcal{U}}^s] &= \mathcal{E}_M[\mathcal{S}^0, \mathcal{D}^s, \tilde{\mathcal{U}}^s] + \alpha \mathcal{E}_R[\tilde{\mathcal{U}}^s] \\ &= \int_{\Omega} \left(\mathcal{D}^s(\mathcal{S}^0(\mathbf{i}) + \tilde{\mathcal{U}}^s(\mathbf{i}))^2 + \alpha \|J_{\tilde{\mathcal{U}}^s}(\mathbf{i})\|_F^2 \right) d\mathbf{i}, \end{aligned} \quad (6.11)$$

where $\|J_{\tilde{\mathcal{U}}^s}(\mathbf{i})\|_F^2$ denotes the Frobenius norm of the Jacobian of the deformation field $\tilde{\mathcal{U}}^s$ and α is a regularization weight. The matching term \mathcal{E}_M ensures that $\phi^s(\mathcal{S}^0) \approx \mathcal{S}^s$ whereas the regularization term \mathcal{E}_R is a smoothness prior.

6.4.2 Non-rigid Registration of Distance Transforms

In the previous section, one body shape was transformed to an implicit representation via its corresponding distance transform. This allowed to derive a metric to quantify the distance between two shapes, however the governing principle is still a point-to-surface distance measure. A different approach is to represent both shapes with their distance transforms and to apply intensity based volumetric registration techniques. This can be formulated with a generic model according to:

$$\mathcal{E}[\mathcal{D}^0, \mathcal{D}^s, \tilde{\mathcal{U}}^s] = \mathcal{E}_M[\mathcal{D}^0, \mathcal{D}^s, \tilde{\mathcal{U}}^s] + \alpha \mathcal{E}_R[\tilde{\mathcal{U}}^s]. \quad (6.12)$$

Again, \mathcal{E}_M is a similarity metric and \mathcal{E}_R is a regularization term that accounts for a smooth deformation field. There have been numerous approaches published for both the similarity metric as well as the regularization term and, in general, each have their advantages and drawbacks or may be limited to certain applications. For a review the reader is referred to the preceding related work section and the overview by Zitova and Flusser [Zito 03] or Sotiras et al. [Soti 13].

For the registration of body surface shapes represented by their corresponding distance transforms, the level-set motion method proposed by Vemuri et al. [C 03] has been found to be suitable in this thesis. This method does not provide the

displacement field directly but instead performs a level-set evolution to compute a mapping between intensities in both images. The geometric warp and smoothness of the deformation fields is eventually derived during numerical computation of the corresponding partial differential equation.

This registration scheme is potentially beneficial for RI surface deformation estimation as the image subject to registration evolves approximately along the normal of the level-set that is the gradient of the distance transform, i. e. for the t -th optimization step:

$$\mathcal{D}^0(\mathbf{x}, t+1) = \left(\mathcal{D}^s(\mathbf{x}) - \mathcal{D}^0(\mathbf{x}, t) \right) \|\nabla \mathcal{D}^0(\mathbf{x}, t)\|. \quad (6.13)$$

Note that this registration scheme favors deformation components along surface normals as they are equivalent to gradients in the distance transform.

6.4.3 Non-rigid Point Set Registration

Though being fundamentally different from a methodological point of view, the preceding approaches are eventually surface registration schemes, i. e. they rely on a continuous representation of the body. However, as outlined in Section 6.3, this cannot be taken for granted and point sets are a more generic representation. As outlined in Section 5.2 for range data fusion, there exist methods for approximating distance transforms from point sets [Hopp 92], however, such methods require the surface to be sampled appropriately. A different approach is to register the discrete set of points directly thus superseding any intermediate data representations. Recent methods that have been proposed for this task represent the point sets as Gaussian mixture models (GMMs) and the registration is formulated to minimize a distance or divergence criterion between mixture densities. For a taxonomy of these approaches the reader is referred to the work of Jian and Vemuri [Jian 11]. In this thesis the focus is on the Coherent Point Drift (CPD) framework [Myro 10] where only one point set is represented as a GMM. This framework will be detailed in Chapter 7 for motion model registration and this section thus only provides a very brief description for non-rigid point set registration.

The general idea is that if the fixed point set \mathcal{P}^0 is represented as a GMM then the probability p of a moving point $\mathbf{x}_i^s \in \mathcal{P}^s$ w. r. t. the GMM is given as:

$$p(\mathbf{x}_i^s) = \sum_{\mathbf{x}_i^0} p(\phi^s(\mathbf{x}_i^0)) \exp(-\sigma^{-2} \|\mathbf{x}_i^s - \phi^s(\mathbf{x}_i^0)\|_2^2), \quad (6.14)$$

where the fixed points \mathbf{x}_i^0 form the GMM centroids locations with σ^2 denoting the variance of the Gaussian normal distribution. The sought transformation $\phi^s(\cdot)$ can be estimated by minimizing the negative log-likelihood function:

$$\mathcal{E}_M[\mathcal{P}^0, \mathcal{P}^s, \phi^s, \sigma] = - \sum_{\mathbf{x}_i^s} \log \sum_{\mathbf{x}_i^0} p(\phi^s(\mathbf{x}_i^0)) \exp(-\sigma^{-2} \|\mathbf{x}_i^s - \phi^s(\mathbf{x}_i^0)\|_2^2). \quad (6.15)$$

For enforcing a smooth deformation this matching term is incorporated in a regularized framework according to:

$$\mathcal{E}[\mathcal{P}^0, \mathcal{P}^s, \phi^s, \sigma] = \mathcal{E}_M[\mathcal{P}^0, \mathcal{P}^s, \phi^s, \sigma] + \alpha \mathcal{E}_R[\phi^s], \quad (6.16)$$

where \mathcal{E}_R is a weighted regularization term that regularizes the deformation ϕ^s by minimizing scaled coefficients of its corresponding Fourier transform [Myro 10], i. e. high frequencies that correspond to discontinuities in the deformation are penalized.

6.5 Dimensionality Reduction Techniques

Recovering the intrinsic dimensionality of the training data is the fundamental step for generating motion models. The dimensionality reduction techniques used for this purpose can either be applied to the deformed surface points $\mathbf{p}^s = \mathcal{L}(\mathcal{P}^s)$ or the corresponding displacement field $\mathbf{u}^s = \mathcal{L}(\mathcal{U}^s)$ with $s = 1, \dots, S$. If the data is mean-centered then both formulations coincide. Dimensionality reduction techniques can be grouped in linear and non-linear methods. The following sections outline and discuss dimensionality reduction strategies that have been identified as promising for respiratory motion analysis.

6.5.1 Linear Methods

The most well known methods for linear dimensionality reduction are PCA also known as the *Karhunen-Loève Transform* and Linear Discriminant Analysis (LDA) that is originally known as *Fisher mapping* [Duda 00]. LDA aims for maximizing inter-class distances and thus is a potential candidate for classification and separation of distinct respiratory motion patterns such as thoracic and abdominal breathing. However, LDA is a supervised method for a discrete set of classes and thus not investigated further. In fact, as shown later, extensions to conventional PCA allow for an automatic and unsupervised classification or separation of these breathing patterns. Thus the focus of linear methods for dimensionality reduction techniques in this thesis is on PCA.

PCA One interpretation of PCA is finding a set of mutually orthonormal vectors $\mathbf{e}_i \in \mathcal{H}$ that allow to reconstruct the set of training samples $\{\mathbf{p}^1, \dots, \mathbf{p}^S\}$, $\mathbf{p}^s \in \mathcal{H}$ in a least squares sense. This is done by an eigen-decomposition of the covariance matrix \mathbf{C} of the mean-centered training data defined as:

$$\mathbf{C} = \sum_{s=1}^S (\mathbf{p}^s - \bar{\mathbf{p}})(\mathbf{p}^s - \bar{\mathbf{p}})^\top, \quad \bar{\mathbf{p}} = \frac{1}{S} \sum_{s=1}^S \mathbf{p}^s. \quad (6.17)$$

The eigen-vectors \mathbf{e}_l often referred to as *modes of variation* that correspond to the L largest eigen-values λ_l of \mathbf{C} now define the model basis as:

$$\Phi = [\mathbf{e}_1, \dots, \mathbf{e}_L] \in \mathbb{R}^{3N \times L}. \quad (6.18)$$

Defining the number of modes L smaller than the number of samples S is the dimensionality reduction step. Commonly, L is chosen such that Φ yields an appropriate reconstruction error w. r. t. the training data. This is equivalent to retain a portion of variance in the input data, i. e. $\sum_{l=1}^L \lambda_l \geq \delta \sum_{l=1}^S \lambda_l$ with $\delta \in [0, 1]$ close to 1 for maximal variance.

Now, given a point set $\mathbf{p} \in \mathcal{H}$ that must not necessarily be contained within the training data, its approximation w. r. t. the PCA basis Φ is traditionally computed according to:

$$\mathbf{p} \approx \bar{\mathbf{p}} + \Phi \left(\Phi^\top (\mathbf{p} - \bar{\mathbf{p}}) \right) = \bar{\mathbf{p}} + \Phi \mathbf{b}, \quad (6.19)$$

where $\mathbf{b} \in \mathbb{R}^L$ denotes the low-dimensional representation of \mathbf{p} and will be referred to as model parameter in this thesis. Essentially, Eq. (6.19) can be derived as the least squares estimate regarding point-wise Euclidean distances between corresponding points in the model basis and the test point-set. Alternative strategies that employ different distance metrics and robust estimators dedicated to RI data are investigated in Section 7.3.

To permit valid shapes only, the individual components b_l of the model parameter \mathbf{b} are commonly restricted based on the corresponding variances λ_l as:

$$b_l \in [-3\sqrt{\lambda_l}, +3\sqrt{\lambda_l}], \quad (6.20)$$

i. e. $\sim 99\%$ of all shapes can be synthesized by the model when interpreting PCA as a multivariate normal distribution estimator with standard deviations $\sqrt{\lambda_l}$. Though this property does not necessarily hold true for respiration samples, these bounds were found to be reasonable in this thesis and in preceding work [Wasz 16].

For unseen data that is not contained within the training set PCA thus allows for a straightforward computation of the low-dimensional representation without having to re-compute the mapping. Further, the reconstruction of a high-dimensional point in the original space \mathcal{H} can be performed using a closed-form solution. These two aspects are commonly referred to as the *out-of-sample* and *pre-image* problem, respectively. As discussed later in Section 6.5.2 these problems constitute serious challenges for non-linear dimensionality reduction techniques and often only approximations are possible.

Sparse Motion Modes One drawback of conventional PCA is that the model basis Φ eventually denotes an orthogonal transformation of the original canonical basis in a way that minimizes the reconstruction error w. r. t. the training samples \mathbf{p}^s . Thus, the individual components of the modes \mathbf{e}_l that are also known as *loadings* do not necessarily exhibit sparsity and in fact represent global abstract deformations that do not allow for an intuitive interpretation or distinct anatomically plausible breathing patterns.

One solution to this problem that was published in preceding work [Wasz 12a, Wasz 16] is to perform an additional rotation $\mathbf{R} \in \text{SO}_L$ according to $\Phi' = \Phi \mathbf{R}$ such that the new spanning vectors stored column wise in Φ' exhibit sparsity.

From factor analysis it is known that well-defined rotations of a PCA model basis are means to generate a more simple structure [Kais 58]. The class of so-called *Orthomax Rotations* aims for generating such a more simple structure. A special case are Varimax rotations (VRs) [Steg 06] that estimate the sought rotation \mathbf{R}^{VR} to maximize:

$$\mathbf{R}^{\text{VR}} = \underset{\mathbf{R}}{\operatorname{argmax}} \sum_{l=1}^L \left(\sum_{q=1}^{3N} [\Phi \mathbf{R}]_{l,q}^4 - \left(\sum_{q=1}^{3N} [\Phi \mathbf{R}]_{l,q}^2 \right)^2 \right). \quad (6.21)$$

Essentially, this corresponds to a maximization of the variances of *squared* loadings across the modes of variation e_l' contained column-wise in $\Phi' = \Phi \mathbf{R}^{\text{VR}}$. Due to the orthonormality constraints on \mathbf{R} the only possibility to maximize Eq. (6.21) is to find a rotation \mathbf{R} that brings several squared loadings $[\Phi \mathbf{R}]_{l,q}^2$ close to zero and make others grow large [Steg 06]. Naturally, this favors sparse modes.

For respiratory motion analysis one issue arises from the fact that the rotation optimization problem from Eq. (6.21) rigorously enforces sparsity exclusively w. r. t. the scalar-valued components of the original modes e_l . As by definition all modes e_l have unit length, i. e. $\|e_l\|_2^2 = 1$, this scheme neglects the original variances that are given by the eigenvalues λ_l . Thus, important information about the magnitude of respiration induced displacements is not incorporated into the sparse motion mode. This is of particular importance if the number of modes is chosen in a way that several modes with minor importance are obtained.

As demonstrated in preceding work [Wasz 16], one option to cope with this problem is to incorporate an additional weighting matrix $\Lambda \in \mathbb{R}^{L \times L}$ in the VR optimization process to yield a weighted Varimax rotation (WVR) basis as:

$$\mathbf{R}^{\text{WVR}} = \underset{\mathbf{R}}{\operatorname{argmax}} \sum_{l=1}^L \left(\sum_{q=1}^{3N} [\Phi \Lambda \mathbf{R}]_{l,q}^4 - \left(\sum_{q=1}^{3N} [\Phi \Lambda \mathbf{R}]_{l,q}^2 \right)^2 \right). \quad (6.22)$$

For $\Lambda = \operatorname{diag}(1, \dots, 1)$, Eq. (6.22) corresponds to the original VR principle from Eq. (6.21). Setting $\Lambda = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_L})$, Eq. (6.22) maximizes the variance of the squared loadings weighted by mode-specific standard deviations defined by the eigenvalues λ_l . The rationale behind this scheme is that the eigenvalues are directly related to the amount of shape variation in the training samples. By design, the samples correspond to distinct respiration states with varying degrees of external surface deformation and breathing magnitude. Thus, sparsity w. r. t. the respiration magnitude is injected in the WVR model.

For optimization of Eqs. (6.21) and (6.22), an iterative scheme employing singular value decompositions can be used [Steg 06]. One issue here is that the rotated basis $\Phi^{\text{WVR}} = \Phi \Lambda \mathbf{R}^{\text{WVR}}$ is in general not orthogonal. This can be addressed by computing the WVR using the weighting matrix according to Eq. (6.22) and compute the WVR basis *without* the weighting matrix as $\Phi^{\text{WVR}} = \Phi \mathbf{R}^{\text{WVR}}$ [Wasz 16].

6.5.2 Non-linear Methods

One problem with PCA and related techniques is that they are often not able to recover the true intrinsic dimensionality or the manifold that governs the training data. This follows directly from their linear mathematical foundations. In contrast, non-linear methods are often capable to analyze such arbitrarily structured data in a meaningful way. One differentiation criterion for non-linear methods is if they aim to preserve the global properties of the input data or if they solely aim for preserving local properties. The inherent problem with local methods such as LLE [Rowe 00] or Laplacian Eigenmaps [Belk 03] in the respiratory motion scenario at hand is the limited number of training samples. For example, if the surfaces are extracted from volumetric planning data such as MRI, the number of available

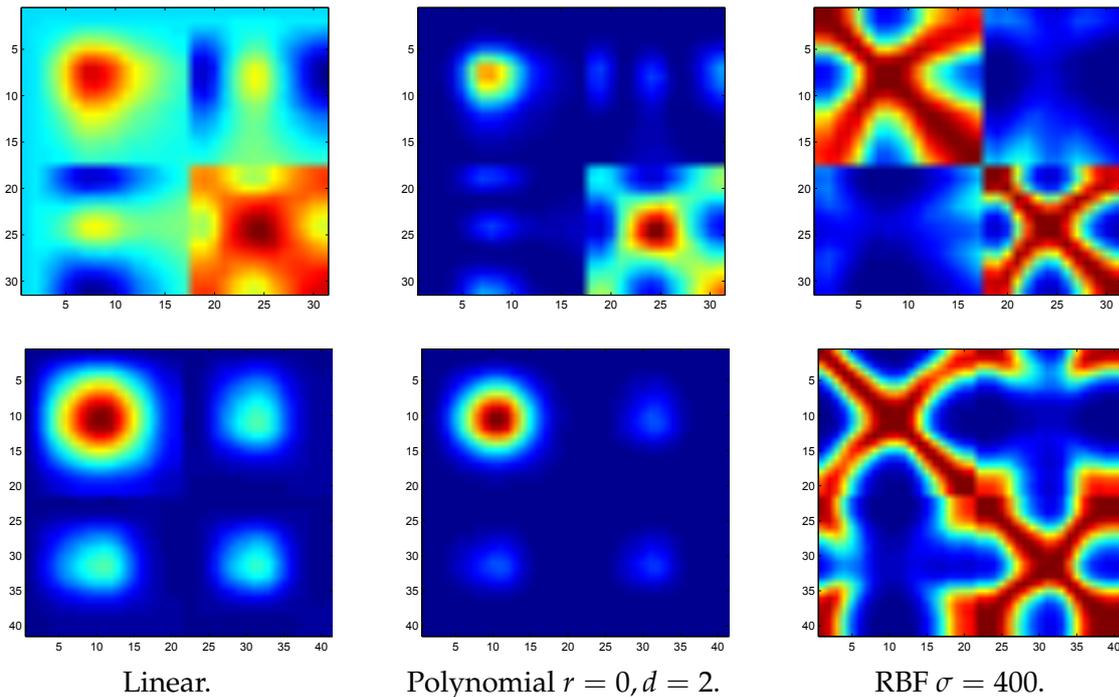


Figure 6.2: Kernel matrices K for different kernels (left to right) and two human subjects (rows). Linear $\kappa(\mathbf{p}^i, \mathbf{p}^j) = \langle \mathbf{p}^i, \mathbf{p}^j \rangle$, polynomial $\kappa(\mathbf{p}^i, \mathbf{p}^j) = (\langle \mathbf{p}^i, \mathbf{p}^j \rangle + r)^d$ and RBF kernels $\kappa(\mathbf{p}^i, \mathbf{p}^j) = \exp(-\sigma^{-2} \|\mathbf{p}^i - \mathbf{p}^j\|_2^2)$ are shown. Note the two distinct breathing patterns that appear as diagonal blocks from top-left to bottom-right in K .

samples is typically not larger than 15 [Grim 15]. Thus, there is no meaningful differentiation between global and local properties. In fact, from a technical point of view, local methods require nearest-neighbor graphs as with LEMs or are based on reconstructing samples based on a linear-combination of neighboring data points as with LLE. For a small number of training samples, the expressiveness of such concepts is rather limited and the resulting optimization problems are often ill-conditioned or similar to that of local methods. Thus, this thesis only discusses KPCA as a global method for non-linear dimensionality reduction.

Kernel PCA The governing idea with kernel-based methods is that problems such as data separation or dimensionality reduction that cannot be tackled with linear methods in the original data space can be handled well with linear techniques in a high-dimensional feature space. Consequently, the principle of KPCA [Scho 98] is to perform PCA not directly in the input space but instead in a high-dimensional possibly infinite-dimensional feature space.

Omitting details that for example can be found in [Mika 98], by using the so-called *kernel-trick*, there is no need to compute the mapping to the high dimensional space *explicitly* and KPCA can be formulated based on the eigen-decomposition

of the *Gramian Matrix* $\mathbf{G} = (g_{i,j}) \in \mathbb{R}^{S \times S}$ that holds the *implicitly* computed inner products of mean centered data points in feature space as:

$$g_{i,j} = k_{i,j} - \frac{1}{S} \sum_{s=1}^S k_{i,s} - \frac{1}{S} \sum_{s=1}^S k_{s,j} + \frac{1}{S^2} \sum_{s=1}^S \sum_{s'=1}^S k_{s,s'}, \quad (6.23)$$

where the elements $k_{i,j}$ of the *Kernel Matrix* $\mathbf{K} \in \mathbb{R}^{S \times S}$ are defined as:

$$k_{i,j} = \kappa(\mathbf{p}^i, \mathbf{p}^j), \quad (6.24)$$

and $\kappa(\cdot, \cdot)$ is a suitable kernel function [Scho 02]. Typical kernels are the linear kernel for which KPCA equals conventional PCA, polynomial kernels and radial basis function (RBF) kernels. Exemplary kernel matrices are depicted in Fig. 6.2. Linear and polynomial kernels encode the angles between displacement fields similar to the condition metric discussed in Section 6.3.2. RBF kernels correspond to a smoothed sum of squared distances across corresponding surface points. As with traditional PCA only the L largest eigen-values λ_l with corresponding eigenvectors $\mathbf{e}_l \in \mathbb{R}^S$ of \mathbf{G} are used for dimensionality reduction.

Given a surface that is not necessarily contained in the training set and that is represented as vectorized point set $\mathbf{p} \in \mathcal{H}$, its low dimensional representation $\mathbf{b} \in \mathbb{R}^L$ can be computed via the *Nyström extension* [Will 00, Beng 03, Aria 07] as:

$$b_l = \frac{1}{\sqrt{\lambda_l}} \sum_{s=1}^S \kappa(\mathbf{p}^s, \mathbf{p}) [\mathbf{e}_l]_s. \quad (6.25)$$

KPCA is thus capable to handle the *out-of-sample* problem, however, compared to traditional PCA the computation complexity is substantially increased.

In contrast, solving the pre-image problem with KPCA is not directly possible and common methods employ non-linear optimization [Mika 98], distance constraints in a multi-dimension-scaling framework [Kwok 04], or the *Nyström extension* [Aria 07]. These techniques have in common that they are computationally expensive and thus hard to handle in a real-time scenario. Thus, non-linear methods will not be investigated further in this thesis.

6.6 Experiments and Results

The experiments are divided into two parts. First, the non-rigid surface registration techniques from Section 6.4 for matching different respiration states are discussed and the condition for displacement fields as introduced in Section 6.3.2 is investigated. The second part covers the proposed motion models with a focus on the sparse linear WVR approach from Section 6.5.1.

6.6.1 Body Surface Displacement Fields

The first part of this evaluation is concerned with the assessment of body surface deformation fields to form the training data that is used to generate 4-D motion models. The evaluation is performed for surface data obtained from real-life RI data of healthy subjects and synthetic data from the NCAT phantom.

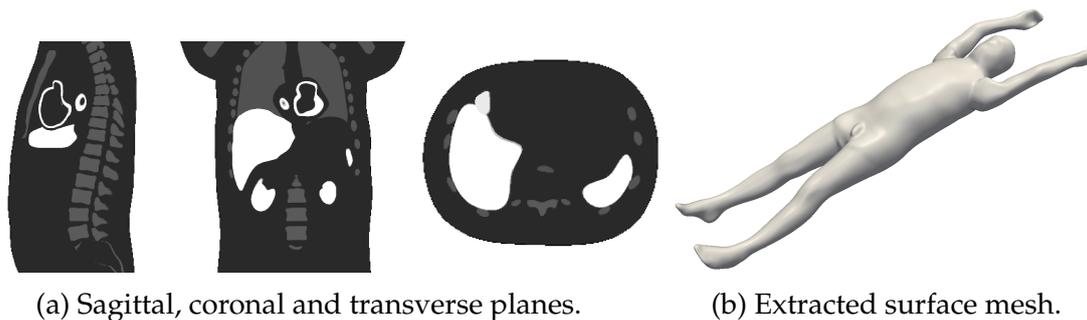


Figure 6.3: The 4-D NCAT phantom. For synthetic RI data a body surface mesh (right) is extracted from the volumetric data (left) by segmenting air and applying the marching cubes algorithm. Respiratory motion is simulated by generating different instances of the phantom.

Experimental Setup

Real life RI data for this experiment was acquired using the multi-view RI framework from Chapter 5. For the sensor characteristics and the general setup with the four healthy subjects, see Section 5.5.2. The subjects were asked to perform one cycle of thoracic and abdominal breathing, respectively. Each cycle was then sampled with 1 Hz which corresponded to 5–7 different respiration states for the individual sequences, see Table D1 for a detailed listing. This sparse temporal sampling was purposely used to stress the fact that the motion models as used in this thesis are generic in a sense that no specific modality to capture surface data from different respiration states is required. For example, 4-D tomographic planning data commonly provides a limited number of different respiration states due to binned reconstruction schemes, radiation exposure with CT or acquisition time with MRI.

The raw surfaces reconstructed with the multi-view RI approach consist of approximately 3.6×10^5 3-D points. The surfaces were cropped to a region of interest, subsequently resampled using a quadric edge collapse strategy and finally processed using Laplacian smoothing. Again, the motivation here is multi-modality which requires the employed algorithms to be robust w. r. t. different spatial sampling, mesh topology and partial matching issues. This resulted in meshes that consist of approximately 1.0×10^4 3-D points to form the training data.

For the virtual test environment surface data extracted from the NCAT phantom [Sega 01] was used. The NCAT phantom utilizes 3-D non-uniform rational B-Spline models of torso organs and the body to simulate cardiac and respiratory motion. The phantom yields a volumetric representation similar to a CT scan, see Fig. 6.3a for an illustration. The body surface was then extracted by first segmenting external air and subsequently applying the marching cubes algorithm [Lore 87] to generate an explicit surface representation, see Fig. 6.3b. Similar to the real-life data, the meshes were cropped to a region of interest that covers the torso and subsequently resampled using a quadric edge collapse strategy. The final meshes for the NCAT phantom consist of approximately 4.0×10^3 3-D points.

Due to the high coverage surface data that are inherent to both the NCAT model as well as the data reconstructed with the multi-view RI framework, the dedicated RI surface registration scheme from Section 6.4.1 to compute displacement fields \mathcal{U}^s matching different respiration states was not applicable. This is due to the fact that the method requires a standard pinhole camera model that is not available with high coverage surface data, cf. Section 5.4. Instead, the distance transform registration (DTR) using the level-set motion registration scheme from Section 6.4.2 for shapes represented as distance-transforms and the CPD point-set registration method outlined in Section 6.4.3 were used. The state of exhaled was chosen as reference.

Results

First, results are reported for the different surface matching techniques. This includes a visual inspection and plausibility analysis as well as a quantitative analysis of the registration errors. Next, the results for the condition of surface deformations are reported.

Elastic surface deformations Qualitative results for typical real-life body surface displacement fields are given in Fig. 6.4. The graphic further illustrates the difference between the DTR and the CPD registration schemes. Though the same respiration states and surface data were used for both methods, the resulting displacement fields are substantially different. In fact, the underlying matching forces as well as the regularization terms that govern the two registration approaches are fundamentally different and constitute one reason for the different displacement fields. However, common deformation patterns such as locally stationary regions or similar motion directions and coinciding deformation magnitudes are visible. From a visual inspection the displacement fields obtained by the DTR method appear more reasonable compared to the CPD approach. This is due to the fact that the CPD displacement fields exhibit distinct vortices or opposed motion directions (cf. lateral region Fig. 6.4e) and unrealistic attracting points (cf. abdominal region Fig. 6.4f). In contrast, there are much less visible artifacts with the deformation fields obtained by the DTR method. However, this is no guarantee for anatomical plausibility.

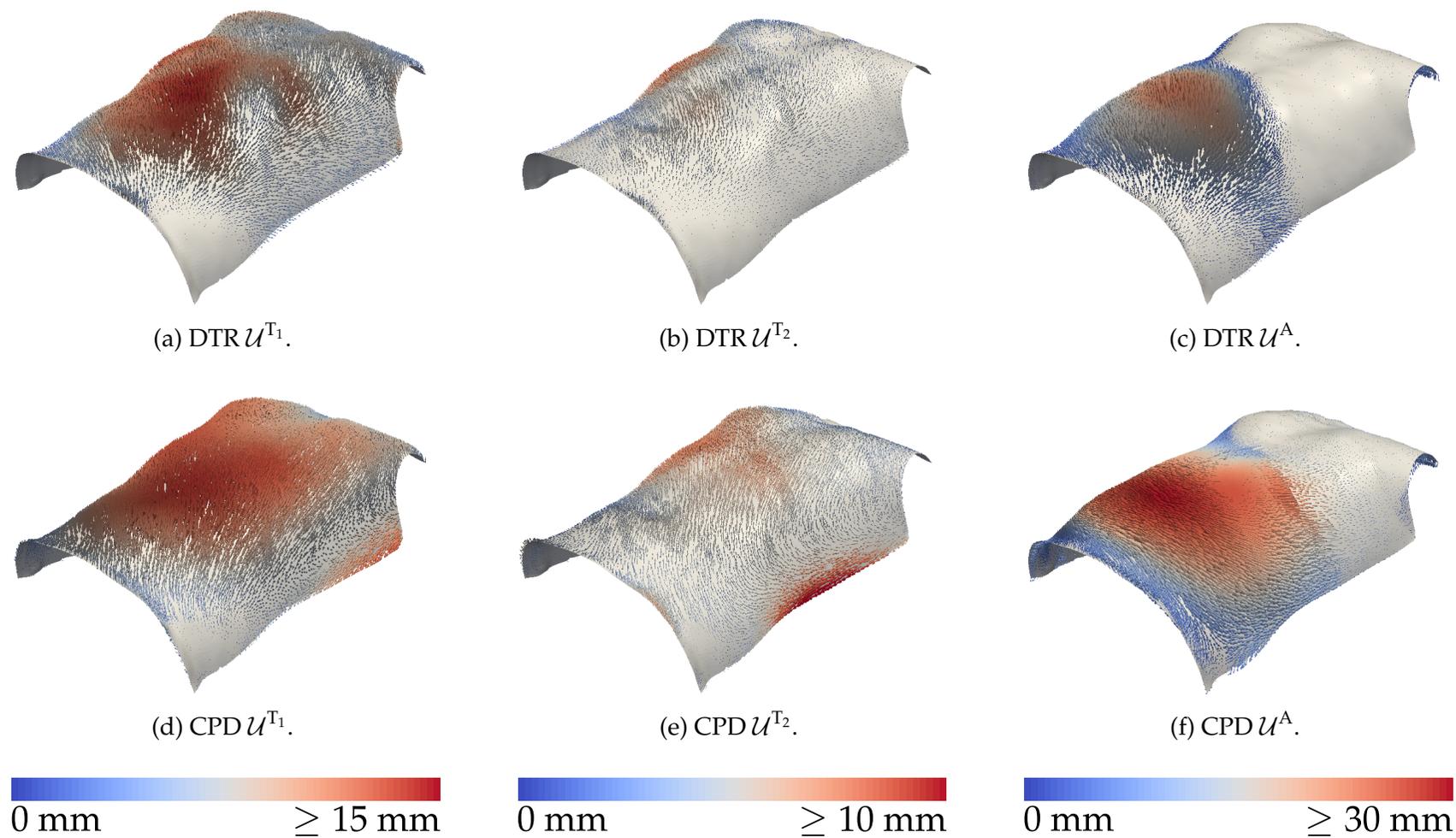


Figure 6.4: Exemplary depiction of body surface deformation fields U using DTR shown in the top row and non-rigid CPD shown in the bottom row. Depicted are the displacements for two thoracic (T_1, T_2) and one abdominal (A) respiration state for one subject. The displacement magnitude in [mm] is color coded. For visualization, the displacement vectors are magnified by a factor of two.

		Displacement Magnitude [mm]			Residual Distance [mm]		
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
P	DTR	1.04 ± 0.63	3.40 ± 2.39	5.17 ± 3.71	0.01 ± 0.01	0.02 ± 0.01	0.04 ± 0.02
	CPD	3.16 ± 1.47	5.23 ± 2.56	7.09 ± 3.56	0.07 ± 0.03	0.14 ± 0.06	0.26 ± 0.11
S_1	DTR	2.15 ± 1.48	3.78 ± 2.80	6.43 ± 3.40	0.11 ± 0.08	0.24 ± 0.18	0.45 ± 0.33
	CPD	4.10 ± 2.29	6.28 ± 3.51	9.26 ± 3.39	0.28 ± 0.16	0.60 ± 0.34	1.04 ± 0.56
S_2	DTR	2.52 ± 2.69	4.36 ± 4.00	6.46 ± 5.29	0.06 ± 0.04	0.14 ± 0.08	0.26 ± 0.15
	CPD	5.47 ± 4.05	8.20 ± 5.75	10.4 ± 6.51	0.18 ± 0.08	0.39 ± 0.17	0.70 ± 0.29
S_3	DTR	1.02 ± 0.67	2.24 ± 1.71	4.53 ± 3.04	0.06 ± 0.03	0.14 ± 0.06	0.27 ± 0.12
	CPD	3.11 ± 1.85	4.73 ± 3.09	6.35 ± 3.66	0.18 ± 0.07	0.39 ± 0.16	0.71 ± 0.29
S_4	DTR	3.93 ± 3.23	6.25 ± 5.26	7.94 ± 6.28	0.07 ± 0.04	0.15 ± 0.09	0.27 ± 0.17
	CPD	5.10 ± 3.68	7.80 ± 5.90	9.60 ± 6.52	0.19 ± 0.10	0.41 ± 0.22	0.72 ± 0.38

Table 6.1: Non-rigid surface registration statistics for the DTR and the CPD methods. Reported is the residual error as the point-wise distance of the warped template points to the reference surface. The magnitude of the estimated displacements are provided for comparison. The first three quartiles are computed across the body surface and averaged over the training samples for four male subjects S_1 - S_4 and the NCAT respiration phantom (P).

To quantify the difference of the two investigated registration approaches and to assess their general capability to match surfaces from different respiration states the residual error defined as the point-to-surface distance between the warped template point-set to the fixed reference surface is analyzed. The results are outlined in Table 6.1 along with the corresponding average displacement magnitudes for comparison. Again, a substantial difference between the DTR approach and the CPD method is noticeable. This applies for both the residual error as well as the estimated displacement magnitude. For all experiments, the CPD method yields a higher residual error compared to the registration based on distance transform. At a first glance, similar to the visual inspection of the deformation fields depicted in Fig. 6.4, the higher residual error with the CPD scheme indicates inferior performance. However, this is a numerical evaluation that does not account for anatomical plausibility of the deformation fields. Further, the residual error as evaluated in this experiment is closely related to the error function that is minimized with the DTR approach. Thus, there is a potential bias towards accepting the DTR method.

Recalling the noise level and quantization steps in the scale of ~ 1 mm for the Kinect RI device (cf. Section 4.7), both registration approaches provide an appropriate match with an average Q_2 residual error < 1 mm. For the NCAT phantom the residual error is considerably smaller. This is due to the fact that the surfaces extracted from the NCAT phantom are a coarse approximation of the human torso that do not feature salient landmarks, see Fig. 6.3b. Further, the synthetic data does not exhibit noise, quantization steps or artifacts that are typical with RI. Thus, a smooth deformation that closely matches the surfaces is easy to obtain. In con-

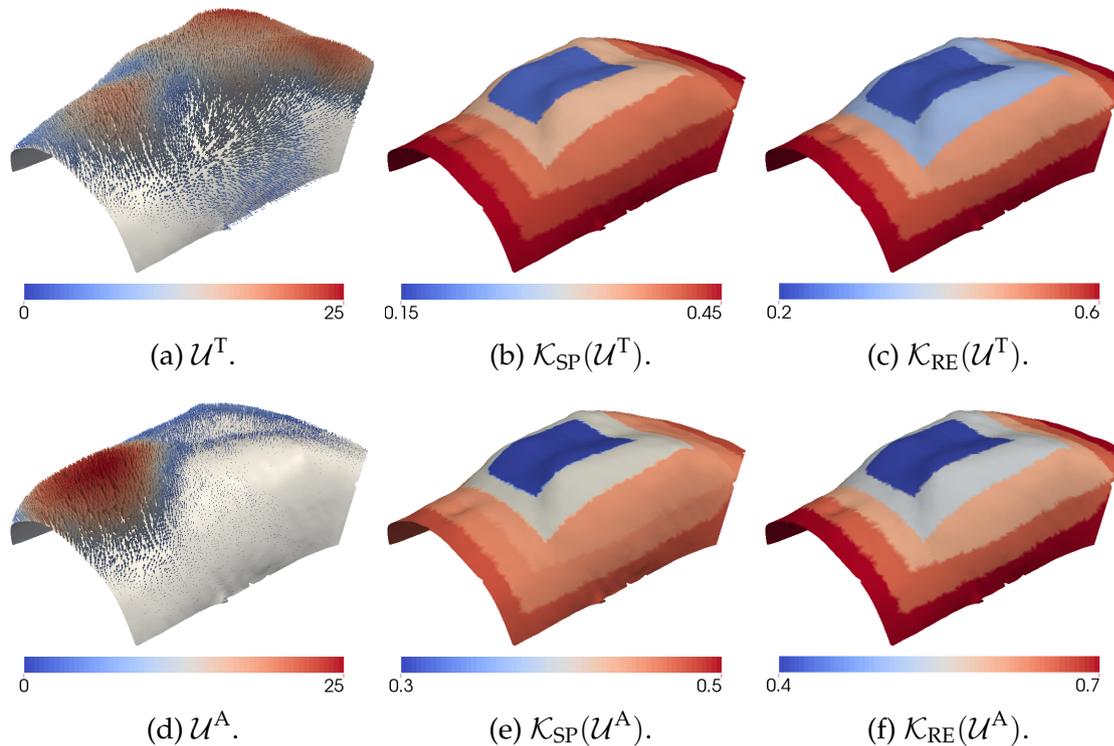


Figure 6.5: Illustration of body surface deformation fields for thoracic (\mathcal{U}^T , top) and abdominal (\mathcal{U}^A , bottom) breathing along with the corresponding condition metrics \mathcal{K}_{SP} and \mathcal{K}_{RE} evaluated for five different surface regions $R_1 \subset R_2 \subset R_3 \subset R_4 \subset R_5$. These regions correspond to varying degrees of body coverage that are typical with different field of views in range imaging. The respiration magnitude in [mm] is color coded, the displacement vectors are magnified by a factor of two for visualization.

trast, a close match for real-life data is commonly not possible due to smoothness constraints on the deformation that is enforced by the regularization term.

Due to more reasonable deformation fields and better surface match the remainder of this chapter is concerned with the displacement fields that have been computed with the DTR method.

Condition of surface deformations The surface deformations obtained from the DTR method are now quantitatively assessed by using the condition metrics \mathcal{K}_{SP} and \mathcal{K}_{RE} as proposed in Section 6.3.2 of this chapter. For this purpose, the body surfaces are divided into five different regions with varying degrees of body coverage. In practice, these regions correspond to different field of views of RI sensors with the largest coverage requiring a multi-sensor setup, cf. Chapter 5.

An exemplary depiction of thoracic and abdominal displacements along with the corresponding condition metrics is given in Fig. 6.5. For both respiration patterns and metrics a higher body coverage increases the condition metric and thus indicates more robust displacement fields that are different from translation (\mathcal{K}_{SP}) and rigid body transformations (\mathcal{K}_{RE}), respectively. This relationship between the body coverage and the condition metric reflects the anthropomorphic consider-

		R_1	R_2	R_3	R_4	R_5
P	\mathcal{K}_{SP}	0.45 ± 0.02	0.51 ± 0.00	0.60 ± 0.01	0.63 ± 0.01	0.67 ± 0.01
	\mathcal{K}_{RE}	0.15 ± 0.01	0.29 ± 0.03	0.48 ± 0.06	0.57 ± 0.06	0.64 ± 0.05
S_1	\mathcal{K}_{SP}	0.29 ± 0.18	0.34 ± 0.15	0.39 ± 0.14	0.43 ± 0.13	0.49 ± 0.11
	\mathcal{K}_{RE}	0.28 ± 0.10	0.38 ± 0.06	0.48 ± 0.04	0.56 ± 0.04	0.67 ± 0.07
S_2	\mathcal{K}_{SP}	0.27 ± 0.07	0.38 ± 0.05	0.43 ± 0.06	0.45 ± 0.07	0.49 ± 0.07
	\mathcal{K}_{RE}	0.38 ± 0.16	0.48 ± 0.13	0.56 ± 0.07	0.61 ± 0.07	0.68 ± 0.08
S_3	\mathcal{K}_{SP}	0.43 ± 0.16	0.46 ± 0.16	0.51 ± 0.13	0.52 ± 0.11	0.54 ± 0.08
	\mathcal{K}_{RE}	0.44 ± 0.14	0.47 ± 0.12	0.56 ± 0.10	0.62 ± 0.09	0.68 ± 0.10
S_4	\mathcal{K}_{SP}	0.18 ± 0.09	0.29 ± 0.12	0.33 ± 0.10	0.37 ± 0.10	0.41 ± 0.10
	\mathcal{K}_{RE}	0.22 ± 0.15	0.35 ± 0.14	0.46 ± 0.11	0.56 ± 0.09	0.62 ± 0.09

Table 6.2: Condition metrics $\mathcal{K}_{\text{SP}}, \mathcal{K}_{\text{RE}}$ to quantify the fitness of body surface deformation fields \mathcal{U}^s computed with the DTR method. The metrics are evaluated for five different body surface regions $R_1 \subset R_2 \subset R_3 \subset R_4 \subset R_5$, cf. Fig. 6.5. The condition metrics are averaged across training deformation fields that include both thoracic and abdominal breathing. Reported are the metrics for four male subjects S_1 - S_4 and the NCAT respiration phantom (P).

ations in Section 6.3.2. By increasing the coverage, stationary regions and body parts that move in opposing directions are present in the displacement field and effectively increase the condition metric. Such regions are visible for the displacement fields depicted in Figs. 6.4 and 6.5. In particular, lateral body regions contribute to an increased condition metric.

More detailed quantitative results for the condition metrics for all subjects investigated in this section are provided in Table 6.2. Again, a higher body coverage corresponds to an increased condition metric. To quantify this relationship the Spearman’s rank correlation coefficient (RCC) was computed between the region-wise condition metrics and the number of displacement vectors in the corresponding regions as a body coverage surrogate assuming approximately equidistant surface sampling. The RCC was favored over the standard PCC to account for the potential non-linear relationship between the assessed variables. Across all subjects S_1 - S_4 and training samples, this resulted in statistical significant ($p \leq 0.05$) average RCCs of $\bar{\rho}_{\text{SP}} = 0.90 \pm 0.31$ and $\bar{\rho}_{\text{RE}} = 0.99 \pm 0.02$ for the \mathcal{K}_{SP} and \mathcal{K}_{RE} metric, respectively.

One reason for the inferior performance of the average \mathcal{K}_{SP} condition metric is its inherent inability to account for rotations. Thus, some training samples and regions are erroneously assigned a high condition number. This is reflected in the higher standard deviation compared to the RCC of the \mathcal{K}_{RE} metric. An example for such critical displacement fields is depicted in Fig. 6.4c. The surface deformation exhibits a strong outward directed abdominal component and an inward directed thoracic component. Such a deformation cannot be modeled by translations only, however, a rotation with pivot in the upper abdominal region can approximate this pattern to a certain degree. By design, the \mathcal{K}_{RE} condition metric accounts

		PCA			VR			WVR		
		$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
P	λ'_l	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00
	Λ_l	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
S_1	λ'_l	0.79	0.19	0.01	0.68	0.19	0.12	0.76	0.22	0.02
	Λ_l	0.79	0.98	0.99	0.68	0.88	0.99	0.76	0.98	0.99
S_2	λ'_l	0.76	0.23	0.00	0.38	0.28	0.33	0.72	0.28	0.01
	Λ_l	0.76	0.99	0.99	0.38	0.66	0.99	0.72	0.99	0.99
S_3	λ'_l	0.67	0.31	0.01	0.50	0.44	0.05	0.53	0.46	0.01
	Λ_l	0.67	0.98	0.99	0.50	0.94	0.99	0.53	0.98	0.99
S_4	λ'_l	0.85	0.11	0.03	0.56	0.13	0.30	0.84	0.12	0.03
	Λ_l	0.85	0.96	0.99	0.56	0.68	0.99	0.84	0.96	0.99

Table 6.3: Dimensionality reduction key figures for different motion model generation techniques. Shown are the relative variances $\lambda'_l = \lambda_l / \sum_{s=1}^S \lambda_s$ along with their cumulative values $\Lambda_l = \sum_{s=1}^l \lambda'_s$. Reported are the results for four male subjects S_1 - S_4 and the NCAT respiration phantom (P).

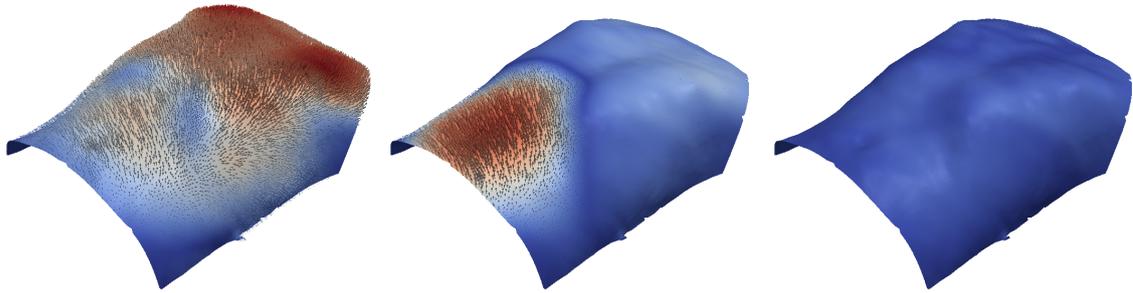
for this issue and yields the expected value. In fact, the high average RCC of 0.99 along with a low standard deviation of 0.02 for the \mathcal{K}_{RE} metric are strong indicators that the relationship between the condition metric and the body coverage can be described using a monotonic function, i. e. increasing the body coverage yields more robust deformations w. r. t. the fitness principle discussed in Section 6.3.2.

6.6.2 Sparse Linear Motion Models

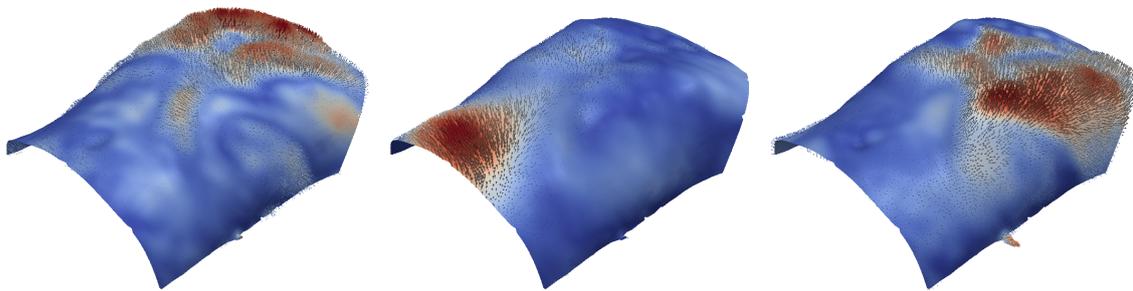
The second part of the experiments in this chapter is concerned with the evaluation of the proposed sparse linear motion model. Again, the evaluation is performed for surface data obtained from the NCAT phantom and real-life RI data from healthy volunteers.

Experimental Setup

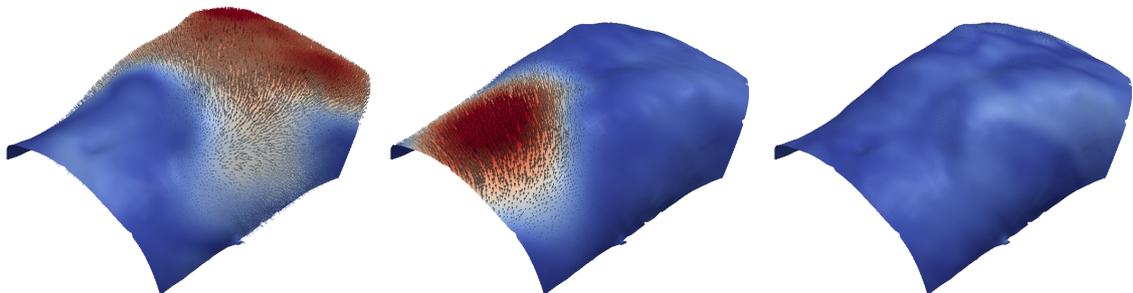
The training displacement fields that are required for building the motion model are computed with the DTR approach, cf. Section 6.6.1. As motivated in Section 6.3.2, Procrustes alignment is purposely not applied to the training data in order to retain affine components in the displacement fields that would otherwise be erroneously classified as rigid patient movement. The intrinsic dimensionality of the motion models was set to $L = 3$. This number was determined by analyzing PCA results of the investigated subjects and the NCAT phantom to yield a cumulative variance $\geq 99\%$. For computing the WVR model the weighting matrix was set to $\Lambda = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_L})$.



(a) S_2 PCA. $l = 1$ (left), $l = 2$ (middle), $l = 3$ (right).



(b) S_2 VR. $l = 1$ (left), $l = 2$ (middle), $l = 3$ (right).



(c) S_2 WVR. $l = 1$ (left), $l = 2$ (middle), $l = 3$ (right).

0 mm  ≥ 25 mm

Figure 6.6: Exemplary depiction of the three leading modes of variation e_l (left to right) for different models (top to bottom). The magnitude in [mm] of the individual displacement vectors is color coded. Shown are the models for standard PCA, conventional VR and the WVR technique. The PCA model exhibits global deformation modes whereas the VR models feature sparse modes of variation. The WVR model as proposed in this thesis allows for a differentiation of thoracic and abdominal breathing patterns.

Results

The key figures for the different motion model generation strategies with an intrinsic dimensionality of $L = 3$ are given in Table 6.3. For the NCAT phantom one single mode was found to be sufficient to describe external surface deformations induced by respiratory motion. Along with the registration accuracy found in the previous section, this confirms the fact that external surface deformations with the NCAT phantom only provide a simplified motion model unable to fully reflect the complex human respiration system. In contrast, for real-life data, three modes of variation are required to explain $\sim 99\%$ of the total variance. However, the variances are considerably different for the individual models. For the standard PCA model, the cumulative variances coincide with the interpretation of PCA to maximize the variance of training samples projected onto the principal axes e_l . In contrast, the conventional VR model inherently produces sparse principal axes that necessarily have less explained variances. In fact, the conventional VR model requires all three modes to yield a model that accounts for $> 95\%$ with the amount of explained variance in the two leading modes dropping below 70% for subjects S_2 and S_4 . Additionally, for these two subjects there is no monotonic increase in the individual variances, i. e. the second mode accounts for less variability compared to the third mode. This is a common issue with VR models and different sorting strategies have been proposed [Steg 06]. In contrast, the third mode with standard PCA and the proposed WVR model is considerably less important than the leading modes and most likely corresponds to artifacts in the training data.

A graphical representation of the motion modes that are obtained with the different model generation techniques is given in Fig. 6.6 for subject S_2 . These results coincide with the cumulative variances outlined in Table 6.3. The PCA model yields two important modes that exhibit global deformations corresponding to thoraco-abdominal breathing. In contrast, the conventional VR model produces three sparse motion modes that are hard to interpret from an anatomical point of view. Though exhibiting sparsity, the individual modes show no locality as there are distinct clusters with high displacement magnitude spread across the entire body surface. As outlined in Section 6.5.1 these observations can be directly linked to the nature of varimax rotations that rigorously enforce sparsity. Finally, the proposed WVR model yields two modes that correspond to abdominal and thoracic breathing. Thus this model exhibits sparsity as well as locality which is a desirable property for respiratory motion analysis. The leading two modes of variation for the PCA and the WVR model for the other subjects are shown in Fig. 6.7.

6.7 Discussion and Conclusion

This chapter was concerned with the foundations of 4-D shape priors for respiratory motion analysis. In particular, the basic strategies to set up sparse surface motion models, non-rigid surface registration schemes to match different respiration states, quantitative methods to assess elastic surface displacement fields for respiratory motion analysis and dimensionality reduction techniques to set up a surface motion model have been investigated.

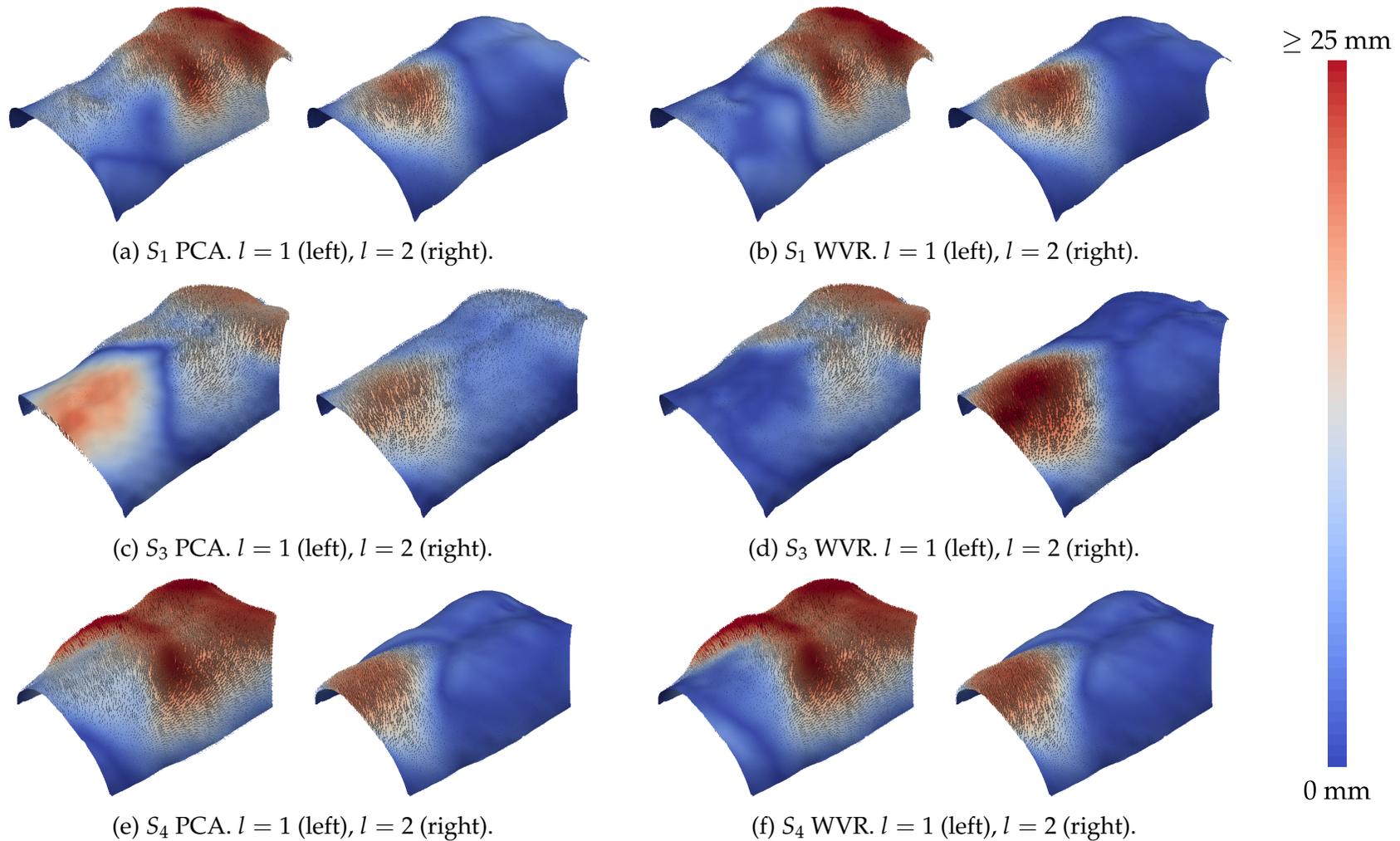


Figure 6.7: The two leading modes of variation e_l for conventional PCA and the proposed WVR model for subjects S_1, S_3 and S_4 . The third mode is omitted as it is of minor relevance, cf. Table 6.3. As with the modes for S_2 that are depicted in Fig. 6.6 the WVR model exhibits sparse motion modes that correspond to thoracic and abdominal movement.

One key finding in the experiments is that different non-rigid registration techniques can be used to compute dense displacement fields to describe surface deformations induced by respiratory motion. However, different registration techniques yield substantially different displacement fields. The problem with the investigated methods and most schemes published in literature is that they are based on point-set or surface alignment problems that do not account for matching salient anatomical landmarks. Such landmarks could be generated manually or via automatic key point selection strategies and provided as additional input for dedicated registration methods such as proposed by Daum [Daum 11]. Further, photometric information could be used to guide or assess the registration match [Baue 12c]. Thus, non-rigid body surface registration schemes for respiratory motion analysis is a topic for future research that must be especially concerned with the anatomical correctness of displacement fields. This also includes the incorporation of photometric information or salient landmarks into the motion models.

Regardless of the actual non-rigid registration techniques, this chapter introduced condition metrics to quantitatively assess the fitness of respiration induced body surface deformations for respiratory motion analysis. The results show that an increased body coverage allows to compute displacement fields that enable a robust differentiation between linear shifts and rotations caused by global patient movement on the one hand and respiratory motion on the other hand. In fact, the results demonstrate a correlation between the body coverage and the condition of the deformation fields. This motivates the usage of multi-view RI as proposed in Chapter 5. Further, future methods for model generation may incorporate the condition metric directly. For example, assigning each surface a weight proportional to the surface's condition metric would allow to use the weighted PCA framework proposed by Kriegel et al. [Krie 08] for a more robust model generation.

A major part of this chapter was devoted to motion model generation techniques that use non-rigidly registered body surfaces from different respiration states. Here, a key finding is that respiration induced external surface deformations are governed by only a few parameters. In fact, for linear models based on PCA only two components are required to describe $> 95\%$ of the variability seen in the training phase. This indicates that linear methods appropriately model external surface deformations and that non-linear techniques that commonly require a large number of training samples are not required per se. In this regard, no more than 15 training samples from two breathing cycles are required to build plausible linear models. If the training shapes are extracted from volumetric planning data with a limited number of samples, linear methods are thus suitable.

One problem with linear methods is that an intuitive and anatomical plausible interpretation of the modes of variation is not given per se. This issue was addressed by using methods known from factor analysis. The WVR proposed in this chapter allows for an unsupervised decomposition of the factors that govern respiration induced surface deformations into a thoracic and an abdominal component. This reflects fundamental anatomical considerations on the human respiratory motion system and allows for respiration analysis that differentiates between thoracic and abdominal breathing.

Real-time Respiratory Motion Analysis Using 4-D Shape Priors

7.1 Motivation	101
7.2 Related Work	102
7.3 4-D Shape Motion Model to RI Surface Alignment	105
7.4 Efficient Corresponding Point Search	112
7.5 Experiments and Results	115
7.6 Discussion and Conclusion	132

This chapter is concerned with *intra-procedural* respiratory motion analysis of RI body surface data using *pre-procedurally* obtained 4-D shape priors as investigated in Chapter 6. As the underlying methodological principle, a unifying framework that models respiratory motion analysis using 4-D shape priors as a point-set registration problem is derived. This framework was previously published for the clinical problems of motion-compensated patient alignment [Wasz 12b, Wasz 13] and continuous respiratory motion monitoring [Wasz 16].

7.1 Motivation

Accurate patient positioning and continuous respiratory motion monitoring are essential tasks for the success of computer assisted interventions. To support these tasks, RI technology was proposed over the last years, see Section 2.1.1 in the introductory part of this thesis for a detailed background.

Traditionally, RI based patient positioning systems establish the treatment table transform by registering an intra-procedural body surface obtained by RI to a static reference surface that was pre-procedurally acquired using RI or extracted from volumetric planning data such as CT or MRI. The problem with such methods is that they rely on rigid surface registration techniques and thus do not account for respiration induced free-form body surface deformations. For respiration states that do not coincide, error scales for rigid alignment up to 25 mm were reported [Plac 12]. Hence, gated positioning is required, i.e. the patient’s current respiration state must match the respiration phase of the static pre-procedural planning data. This implies the need for respiratory motion monitoring techniques to

generate respiration surrogates that uniquely encode the current respiration state w. r. t. complex body surface deformations. Further, such surrogates are an integral component for continuous beam gating or beam adjustment during radiation delivery after the patient was correctly aligned.

For respiratory motion monitoring, RI-based external respiration surrogates are commonly derived from the movement of local body surface regions. The problem with such conventional methods is that they are not based on true 3-D surface deformations but rather heuristically analyze plain 1-D depth measurements to derive a low dimensional surrogate. Further, these approaches usually require manual interaction which is potentially error prone and complicates clinical workflows. The related work in Section 7.2 covers these aspects in detail.

Alternative approaches to tackle the limitations of conventional techniques aim to generate multi-dimensional surrogates from dense displacement fields computed by non-rigid surface registration techniques, cf. Section 6.3.1 and the related work in Section 7.2. Such methods are fully automatic and provide true 3-D deformation analysis across the entire body surface region. However, directly generating respiration surrogates from dense displacement fields is complicated as they contain redundant information and the integral non-redundant components must be identified. As an important practical aspect, real-time capability is an open issue with non-rigid surface registration.

In this chapter, the limitations of existing approaches for both patient alignment as well as continuous respiratory motion monitoring are addressed by a joint framework using the principle of 4-D shape priors as introduced in Chapter 6. The core concept of this joint positioning and monitoring framework is to register the pre-procedurally trained dynamic 4-D surface motion model to the patient's current intra-procedural RI-based body surface. First, this enables a motion compensated patient alignment strategy that supersedes the need for a gated positioning approach. Second, the intrinsic parameters that govern the registered motion model define a non-redundant external respiration surrogate that is directly related to body surface deformations across the entire torso.

7.2 Related Work

Relevant literature for this chapter divides into several aspects. First, point-set registration schemes to align the motion models to RI data are outlined. Next, to address real-time constraints, related work on efficient corresponding-point search strategies as an integral part of the registration methods is reported. The section concludes with previously published methods for RI-based patient positioning and strategies to generate respiration surrogates from RI surface data are outlined.

Motion model registration From a methodological point of view the motion model registration task can be classified as a point-set or surface alignment problem. See the related work in Section 6.2 and the techniques in Section 6.4 for details on non-rigid surface registration. The most popular method for point-set registration is the ICP algorithm that was almost simultaneously proposed by Chen and

Medioni [Chen 92] and Besl and McKay [Besl 92]. The basic idea behind the ICP algorithm is to iteratively compute corresponding points between two data sets and to transform one data set to minimize a distance criterion that quantifies the match of corresponding points. Over the last years numerous extensions and improvements have been proposed, see the review by Rusinkiewicz and Levoy [Rusi 01].

Important variants are robust ICP schemes that aim to account for noise, outliers and partial matching issues. For example, Pulli proposed to reject those pairs of corresponding points that account for the worst $n\%$ of distances or that exhibit incompatible complementary information such as surface normals [Pull 99]. Further, rejecting pairs with a distance larger than the mean plus some multiple of the standard deviation of all distances have also been proposed [Zhan 94, Masu 96]. Another method is the robust ICP by Fitzgibbon that assigns a zero weight for points for which no suitable correspondence could be found [Fitz 03].

These robust methods have in common that they assign hard correspondences in a single-link sense, i. e. it is assumed that each point corresponds to exactly one point in the other data set. This assumption was relaxed by introducing soft assignments schemes that use correspondences between multiple points or even the combination of all points. For example, in the so-called expectation-maximization (EM)-ICP proposed by Granger and Pennec [Gran 02] each point in one data set is connected to all points in the other data set and the weights are determined by a robust version of the corresponding distances using for example a kernel function. This basic concept was further generalized by Tsin and Kanade who formulate the robust point-set registration problem as a so-called *kernel correlation* problem [Tsin 04]. Eventually, in this scheme the registration cost function is proportional to the correlation of two density estimates. This scheme was later extended by Bing and Vemuri [Jian 05] to a framework that interprets the point-set registration as aligning GMMs. A similar idea governs the CPD proposed by Myronenko and Song [Myro 10]. In the CPD framework only one point set is represented as GMM and the other point set is considered as data points, see Section 7.3.2 for a detailed description. A general taxonomy and unifying framework for these robust alignment strategies was recently developed by Jian and Vemuri [Jian 11].

A different robust ICP variant has been proposed by Maier-Hein et al. for aligning low SNR ToF range data [Maie 12]. This approach tackles the limitations of conventional ICP-like formulations that implicitly assume equal and isotropic covariances of the data points. A similar approach has been proposed by Balachandran and Fitzpatrick [Bala 09].

Efficient Corresponding Point Search Computing corresponding points across two data sets is a fundamental requirement for many motion model registration schemes. Due to iterative registration schemes and thousands of points, efficient strategies for computing correspondences are mandatory. For CPU architectures, space partitioning data structures like k-d trees [Aken 02] have shown to be beneficial. Besides space partitioning strategies, hardware acceleration using modern GPU architectures have become increasingly popular. For example, Garcia et al. showed that a GPU-based brute-force corresponding point search is superior to a CPU-based k-d tree look-up [Garc 08]. This is due to the parallel nature of

the corresponding point search that can be implemented efficiently using patterns known from the well understood problem of GPU-based matrix-matrix multiplication. In contrast, conventional nearest neighbor search acceleration strategies on the GPU are challenging due to the non-parallel and often recursive nature of construction and/or traversal of the underlying data structures. For instance, Qiu et al. [Qiu 09] achieved high frame-rates for GPU based k-d tree queries. However, the construction of the tree is performed on the CPU limiting performance when the tree must be constructed on a per-frame basis. Recently, space-partitioning strategies being specifically designed for GPU architectures have been proposed. One approach is the Random Ball Cover (RBC) proposed by Cayton [Cayt 10, Cayt 12]. The basic principle behind the RBC is a two-tier nearest neighbor search, each building on the brute-force primitive, to prune the search space.

Apart from space partitioning acceleration structures, so-called projective data association (PDA) schemes are commonly used for aligning point clouds or surface data [Blai 95, Benj 99, Dora 98, Rusi 02, Newc 11]. PDA schemes aim to accelerate the corresponding point search by projecting a given query point either to an auxiliary data structure [Benj 99] or directly into the target mesh or range image. This supersedes the need for a computational expensive 3-D search and allows for a direct corresponding point retrieval by indexing as pursued in the most publications [Blai 95, Dora 98, Rusi 02, Newc 11]. This direct approach that is also known as *reverse calibration* [Blai 95] gained increasing popularity with the Kinect Fusion framework [Newc 11]. Extensions to this method include verification steps that for example analyze mutual distances of corresponding points [Dora 98].

RI based patient positioning and respiratory motion analysis For aligning the patient to the reference planning data, the so-called ICP algorithm [Besl 92] is the predominant method [Scha 09, Baue 11, Plac 12]. To account for issues of the ICP with noisy and missing RI data Lindl et al. [Lind 13] instead use a robust alignment based on GMM matching proposed by Jian and Vemuri [Jian 11]. However, these methods only account for fine positioning and a coarse pre-alignment is usually required. For this purpose, feature-based approaches that are invariant to translations and rotations have been proposed. For example, Bauer et al. [Baue 11] investigate different surface descriptors for multi-modal patient pre-alignment and report a positioning error in the scale of 1.5° and 13 mm, respectively. Placht et al. [Plac 12] reported a superior performance of a feature-based pre-alignment compared to conventional centroid matching as initial guess for ICP-based fine tuning. They also investigated the stability of fine positioning w. r. t. respiratory motion and found an error of 25 mm for a respiration magnitude of 20 mm [Plac 12].

For RI-based continuous respiratory motion analysis or respiration surrogate generation a plurality of concepts have been proposed. The most basic approaches rely on plain 1-D depth data as for example proposed by Fayad et al. [Faya 09, Faya 11]. Such surrogates only account for distance changes towards the RI sensor and do not provide anatomical plausible tracking. More advanced strategies aim for generating a surrogate based on variations of 3-D points inside specific body surface regions. For example, Schaller et al. [Scha 08] or Xian and Siochi [Xia 12] derive respiration signals by fitting planes through the abdominal and thorax re-

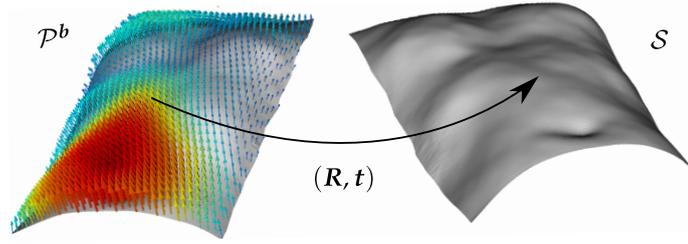


Figure 7.1: Motion model alignment similar to [Wasz 12b]. A model instance \mathcal{P}^b that reflects respiration induced deformations and a rigid-body transformation (\mathbf{R}, \mathbf{t}) that accounts for the table transform must be computed such that the transformed motion model is in congruence with the current RI surface data \mathcal{S} .

regions of the body surface. Similarly, Hughes et al. [Hugh 09] generate a surrogate by computing the volume of bounding boxes that enclose the thorax and the abdomen. These approaches have in common that they are heuristic body surface partitioning strategies based on points or patches and additionally require manual interaction to define the regions of interest. Further, though using 3-D surface data, no real 3-D tracking of landmarks is provided and these methods are not capable to describe complex surface deformations across the entire body.

To overcome the limitations of these methods, non-rigid surface registration techniques have been proposed for respiratory motion analysis. Schaerer et al. used a non-rigid variant of the ICP algorithm for surface matching and derived a multi-dimensional respiration signal from the principal direction of displacement vectors [Scha 12]. In contrast, Bauer et al. use a variational formulation for elastic surface registration and proposed to use the resulting displacement field directly as high-dimensional respiration surrogate [Baue 12b, Baue 12a]. Such high dimensional surrogates can for example be used for external-internal correlation models [Taub 14]. As an interesting approach, Bauer et al. proposed to use a photometric registration scheme for surface matching [Baue 12c]. With regard to salient body surface landmarks such as the mamillas, this photometric approach outperformed a conventional registration method that is solely based on geometric information.

7.3 4-D Shape Motion Model to RI Surface Alignment

The problem for motion compensated patient alignment as well as continuous respiratory motion monitoring is formulated in a unifying framework in this thesis. Mathematically, the task is to find a rotation and translation $(\mathbf{R} \in \text{SO}_3, \mathbf{t} \in \mathbb{R}^3)$ as well as a model parameter vector \mathbf{b} from Eq. (6.19) such that the transformed model instance \mathcal{P}^b is in congruence with the current RI surface \mathcal{S} , i. e. :

$$(\hat{\mathbf{R}}, \hat{\mathbf{t}}, \hat{\mathbf{b}}) = \underset{(\mathbf{R}, \mathbf{t}, \mathbf{b})}{\operatorname{argmin}} \sum_{n=1}^{|\mathcal{P}^b|} \Theta(\mathbf{R}\mathbf{x}_n^b + \mathbf{t}, \mathcal{S}), \quad \mathbf{x}_n^b \in \mathcal{P}^b, \quad (7.1)$$

where $\Theta(\cdot, \cdot)$ is a generic metric that quantifies the match of a transformed model point $\mathbf{R}\mathbf{x}_n^b + \mathbf{t}$ and the RI surface \mathcal{S} . A sketch of this problem is given in Fig. 7.1.

Distance metrics $\Theta(\cdot, \cdot)$ that are suitable for the real-time scenario at hand, including conventional single-link point-to-point (P2P) or advanced robust multi-link point-to-tangent (P2T) alignment, will be detailed subsequently in Sections 7.3.1 and 7.3.2.

Patient positioning and respiratory motion monitoring Interpreting Eq. (7.1) from a clinical point of view, the rigid body transformation (\mathbf{R}, \mathbf{t}) accounts for the table transform that aligns the patient to the reference state. In contrast, the model parameter \mathbf{b} accounts for the respiration induced deviations of the current RI surface data to the planning state. For a fixed \mathbf{b} , Eq. (7.1) thus corresponds to conventional rigid patient alignment that does not account for respiratory motion. On the other side, for fixed (\mathbf{R}, \mathbf{t}) , Eq. (7.1) corresponds to motion monitoring of an accurately positioned patient. Consequently, if all parameters are free, Eq. (7.1) translates to motion compensated patient positioning.

Model-based respiration surrogates For respiration monitoring and the PCA-based models as used in this thesis, the parameter vector \mathbf{b} constitutes a respiration surrogate that reflects the body surface extent. This property can be derived from the modes of variation \mathbf{e}_l contained in the model basis Φ that encode non-rigid body surface displacements and that synthesize a new model instance as:

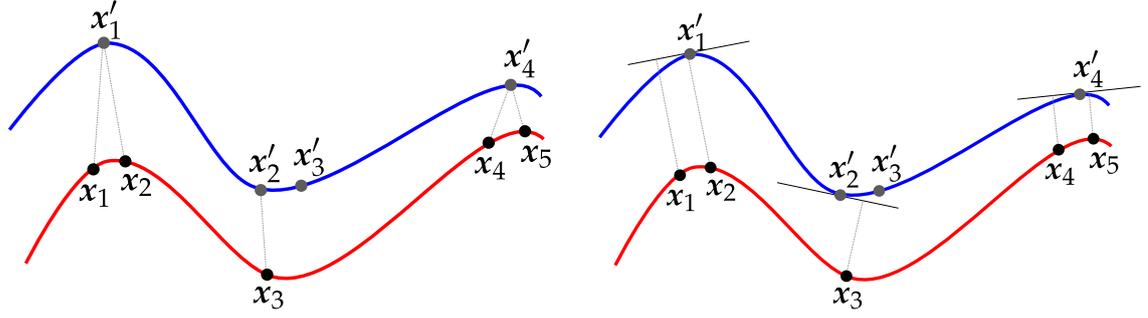
$$\mathbf{p} = \bar{\mathbf{p}} + \Phi \mathbf{b}.$$

See Eq. (6.19) for details and Figs. 6.7 and 7.1 for an illustration of such displacements. The individual parameters are now shifted according to the model plausibility criterion from Eq. (6.20) as $b'_l = b_l + 3\sqrt{\lambda_l}$. This enforces the parameters b'_l to be positive and to describe a surface extent relative to the most compact or fully exhale shape allowed by the model, i. e. $b_l = -3\sqrt{\lambda_l}$ [Wasz 16]. The l -th model-based surrogate $\zeta_l^M \in \mathbb{R}^+$ and the joint surrogate $\zeta_j^M \in \mathbb{R}^+$ are then given as:

$$\begin{aligned} \zeta_l^M &= \|b'_l \mathbf{e}_l\|_2 = b'_l, \\ \zeta_j^M &= \|\Phi \mathbf{b}'\|_2 = \|\mathbf{b}'\|_2. \end{aligned} \tag{7.2}$$

The last equalities follow directly from the unit length of \mathbf{e}_l and the orthonormality of Φ , respectively. Note that Eq. (7.2) would not hold for the raw parameters $b_l \in \mathbb{R}$ as a potential negative sign would be lost.

Optimization A well established method to solve Eq. (7.1) is an iterative scheme that alternately optimizes for (\mathbf{R}, \mathbf{t}) and \mathbf{b} [Coot 95]. The derivations of the solution to the alternating optimization steps and the investigated alignment metrics $\Theta(\cdot, \cdot)$ are detailed in Sections 7.3.1 and 7.3.2. Regardless of a particular alignment metric, this optimization scheme inherently requires that Φ does not represent translations or rotations as this would lead to ambiguities, i. e. there is no clear differentiation between global patient movement on the one side and local respiratory motion on the other side. This issue was investigated in Section 6.3.2 by means of condition metrics to quantify the fitness of surface displacement fields.



(a) P2P alignment. The source points x_n are connected to their closest target point x'_n . Note that different source points may be share the same target point.

(b) P2T alignment. The source points x_n are linked to their projection onto the corresponding tangent plane at target point x'_n .

Figure 7.2: Illustration of the P2P (left) and P2T (right) alignment strategies. The blue curve denotes the target surface and the red curve represents the source surface or point-set.

7.3.1 Conventional Single-Link Iterative Closest Point

In a single-link registration scheme each model point $x_n^b \in \mathcal{P}^b$ is connected to exactly one corresponding point $x'_n \in \mathcal{S}$ to evaluate the distance metric Θ . For a fixed model parameter \mathbf{b} the single-link iterative optimization of Eq. (7.1) corresponds to the ICP algorithm that has been proposed by Chen and Medioni [Chen 92] and Besl and McKay [Besl 92]. A variety of modifications and extensions for these algorithms have been published over the years, see the related work in Section 7.2.

One key difference for the various ICP variants is the alignment strategy that is either based on a P2P [Besl 92] or a point-to-plane also known as P2T distance metric [Chen 92]. These two approaches are illustrated in Fig. 7.2 and will be outlined in the following sections.

Point-to-Point Alignment

The popular P2P metric [Besl 92] quantifies the match of a model point to the RI surface as the squared Euclidean distance between corresponding points. A sketch of this principle is given in Fig. 7.2a. Mathematically, the P2P metric is expressed as:

$$\Theta^{\text{P2P}}(\mathbf{R}x_n^b + \mathbf{t}, \mathcal{S}) = \left\| \mathbf{R}x_n^b + \mathbf{t} - x'_n \right\|_2^2, \quad (7.3)$$

where the corresponding point x'_n is commonly computed by using a nearest neighbor relation according to:

$$x'_n = \operatorname{argmin}_{x \in \mathcal{S}} \left\| x_n^b - x \right\|_2^2. \quad (7.4)$$

However, approximative strategies for fast corresponding point computation as discussed in Section 7.4 have also been proposed.

For the P2P distance metric there exists a closed-form solution based on unit quaternions to estimate (\mathbf{R}, \mathbf{t}) [Horn 87]. As shown in prior work, this scheme

perfectly maps to modern GPU architectures for real-time computation [Wasz 12b, Baue 13b]. Computing the model parameter \mathbf{b} with (\mathbf{R}, \mathbf{t}) fixed corresponds to the conventional least-squares model approximation formulation from Eq. (6.19). By using the concatenation operator from Eq. (6.2) this is expressed as:

$$\mathbf{b} = \Phi^\top (\mathbf{x}' - \bar{\mathbf{p}}), \quad \mathbf{x}' = \mathcal{L} \left(\left\{ x'_1, \dots, x'_{|P^b|} \right\} \right). \quad (7.5)$$

A more generic formulation is derived in Eq. (7.15) for multiple weighted correspondences as a generalization of the single-link problem. Technically, Eq. (7.5) breaks down to the computation of scalar products that can be computed efficiently on GPUs using tree-based parallel reduction schemes [Harr 07].

Point-to-Tangent Alignment

A different way to measure the closeness of a point to a surface is based on the P2T metric that was proposed by Chen and Medioni for registration of range images [Chen 92]. This alignment strategy is depicted in Fig. 7.2b and can be formulated as:

$$\Theta^{\text{P2T}} (\mathbf{R}\mathbf{x}_n^b + \mathbf{t}, \mathcal{S}) = \left((\mathbf{R}\mathbf{x}_n^b + \mathbf{t} - \mathbf{x}'_n) \cdot \mathbf{n}_n \right)^2, \quad (7.6)$$

where $\mathbf{n}_n \in \mathcal{N}$ is the surface normal associated to $\mathbf{x}'_n \in \mathcal{S}$ from Eq. (7.4).

In contrast to the P2P metric, there exists no closed-form solution for estimation (\mathbf{R}, \mathbf{t}) and non-linear optimization techniques must be employed. However, for small rotation angles the optimization problem can be linearized and estimating (\mathbf{R}, \mathbf{t}) corresponds to a linear equation system that can be solved using singular value decomposition (SVD) [Low 04] or Cholesky decomposition (CD) [Newc 11]. As shown in Eq. (7.19) for multiple weighted correspondences as a generalization of single-link correspondences, the model parameter vector \mathbf{b} can be computed efficiently using CD of a positive definite matrix. The advantage of the P2T metric is that the convergence speed is improved without a significant increase in computation cost [Fitz 03].

7.3.2 Robust Multi-Link Coherent Point Drift

The problem with the conventional single-link ICP variants from the previous section is that they are not robust w. r. t. outliers, partial matching issues or false correspondences. Proposed techniques to cope with these problems are based on weighting and rejection of pairs of corresponding points based on heuristics, see the related work in Section 7.2. This led to the development of dedicated techniques that use a probabilistic assignment of correspondences between all combinations of points. One method that falls into this category is the Coherent Point Drift (CPD) algorithm [Myro 10] that was previously outlined in Section 6.4.3 for non-rigid body surface registration. However, the CPD method can also be used for a rigid-body point-set alignment and the motion model formulation can be seamlessly integrated as demonstrated in preceding work [Wasz 16].

The general idea behind the CPD framework is to model the point set registration as a probability density estimation problem with one point set representing

the centroids of a GMM and the second point set representing the data points. The registration is then performed by re-parameterizing the GMM centroids $\mathbf{x}_n^b \in \mathcal{P}^b$ to minimize the negative Log-Likelihood of the data points $\mathbf{x}'_m \in \mathcal{S}$:

$$\mathcal{J}_{\mathbf{R},\mathbf{t},\sigma}^{\text{CPD}} = - \sum_{m=1}^{|\mathcal{S}|} \log \left(\frac{(1-w)}{|\mathcal{P}^b|} \sum_{n=1}^{|\mathcal{P}^b|} p(\mathbf{x}'_m | \mathbf{R}\mathbf{x}_n^b + \mathbf{t}) + w \frac{1}{|\mathcal{S}|} \right), \quad (7.7)$$

where $w \in [0,1[$ is a weighting factor that accounts for outliers via the uniform distribution $|\mathcal{S}|^{-1}$. The likelihood $p(\mathbf{x}'_m | \mathbf{R}\mathbf{x}_n^b + \mathbf{t})$ of a d -dimensional data point \mathbf{x}'_m given a re-parameterized GMM centroid $\mathbf{R}\mathbf{x}_n^b + \mathbf{t}$ is computed as:

$$p(\mathbf{x}'_m | \mathbf{R}\mathbf{x}_n^b + \mathbf{t}) = \frac{1}{\sigma \sqrt{(2\pi)^d}} \exp \left(- \frac{\|\mathbf{R}\mathbf{x}_n^b + \mathbf{t} - \mathbf{x}'_m\|_2^2}{2\sigma^2} \right). \quad (7.8)$$

This GMM based alignment scheme is solved by the EM algorithm. As shown by Myronenko and Song [Myro 10], this eventually leads to an iterative optimization problem according to:

$$\mathcal{J}_{\mathbf{R},\mathbf{t},\sigma}^{\text{CPD}} = \sum_{n=1}^{|\mathcal{P}^b|} \sum_{m=1}^{|\mathcal{S}|} p(\tilde{\mathbf{R}}\mathbf{x}_n^{\tilde{b}} + \tilde{\mathbf{t}} | \mathbf{x}'_m) \left(\frac{\|\mathbf{R}\mathbf{x}_n^b + \mathbf{t} - \mathbf{x}'_m\|_2^2}{2\sigma^2} + \frac{d}{2} \log(\sigma^2) \right). \quad (7.9)$$

The posterior probabilities $p(\tilde{\mathbf{R}}\mathbf{x}_n^{\tilde{b}} + \tilde{\mathbf{t}} | \mathbf{x}'_m)$ of GMM centroids given the data points are computed using the previous parameter estimates marked with a tilde as:

$$p(\tilde{\mathbf{R}}\mathbf{x}_n^{\tilde{b}} + \tilde{\mathbf{t}} | \mathbf{x}'_m) = \frac{p(\mathbf{x}'_m | \tilde{\mathbf{R}}\mathbf{x}_n^{\tilde{b}} + \tilde{\mathbf{t}})}{p(\mathbf{x}'_m)} = \frac{\exp \left(- \frac{\|\tilde{\mathbf{R}}\mathbf{x}_n^{\tilde{b}} + \tilde{\mathbf{t}} - \mathbf{x}'_m\|_2^2}{2\tilde{\sigma}^2} \right)}{\sum_{k=1}^N \exp \left(- \frac{\|\tilde{\mathbf{R}}\mathbf{x}_k^{\tilde{b}} + \tilde{\mathbf{t}} - \mathbf{x}'_m\|_2^2}{2\tilde{\sigma}^2} \right) + c}, \quad (7.10)$$

where $c = (2\pi\tilde{\sigma}^2)^{\frac{d}{2}} \frac{w}{1-w} \frac{|\mathcal{P}^b|}{|\mathcal{S}|}$ follows from re-arranging the negative Log-Likelihood objective criterion outlined in Eq. (7.7).

Generalized CPD and Parameter Estimation

The CPD framework, or more specifically its inherent iterative optimization problem from Eq. (7.9), can be interpreted as a weighted multi-link generalization for the generic framework from Eq. (7.1). In fact, by replacing the probabilistic weights with a hard binary assignment the standard single-link formulation is retained. Further, using probabilistic weights but considering only single-links can be interpreted as an approximation of the *Kullback-Leibler* divergence between two point-sets represented as GMMs [Jian 05]. In fact, relaxing the number of correspondences is an efficient strategy to speed up the CPD scheme [Myro 10]. This generalization will be discussed in the subsequent section in the context of fast CPD approximation for RI surface data.

However, first, the solution for the transformation (\mathbf{R}, \mathbf{t}) and the model parameter \mathbf{b} will be outlined for the generic multi-link P2P and P2T alignment schemes.

In the following paragraphs, $\mathbf{x}'_{n,m} \in \mathcal{S}$ denotes the m -th correspondence for the n -th model point $\mathbf{x}_n^b \in \mathcal{P}^b$. Further, terms that are constant when optimizing w. r. t. $(\mathbf{R}, \mathbf{t}, \mathbf{b})$ are omitted for readability.

Point-to-Point The standard CPD scheme corresponds to a weighted multi-link P2P alignment, i. e. using Θ^{P2P} outlined in Eq. (7.3). Thus, the CPD distance metric for the generic registration framework from Eq. (7.1) can be derived according to:

$$\Theta^{\text{CPD-P2P}}(\mathbf{R}\mathbf{x}_n^b + \mathbf{t}, \mathcal{S}) = \sum_{m=1}^M \omega_{n,m} \left(\frac{\|\mathbf{R}\mathbf{x}_n^b + \mathbf{t} - \mathbf{x}'_{n,m}\|_2^2}{\sigma^2} + \frac{d}{2} \log(\sigma^2) \right), \quad (7.11)$$

with $\omega_{n,m}$ denoting weights that are given by the posterior probabilities from Eq. (7.10) and $M \ll |\mathcal{S}|$ denotes the number of correspondences. For finding the transformation (\mathbf{R}, \mathbf{t}) note that the multi-link formulation can be interpreted as a generalized weighted absolute orientation problem [Myro 10] which, for example, can be solved by using Horn's unit quaternion optimizer [Horn 87] as with the conventional single-link formulation.

Similarly, the solution to the model parameter \mathbf{b} can be deduced from a cost function using pair-wise weighted correspondences as:

$$\mathcal{J}_{\mathbf{b}}^{\text{CPD-P2P}} = \sum_{n=1}^{|\mathcal{P}^b|} \sum_{m=1}^M \omega_{n,m} \|\mathbf{x}_n^b - \mathbf{x}'_{n,m}\|_2^2 = \sum_{n=1}^N \sum_{m=1}^M \omega_{n,m} \left\| \sum_l^L b_l \mathbf{x}_{n,l} - \mathbf{x}'_{n,m} \right\|_2^2, \quad (7.12)$$

with $\mathbf{x}_{n,l} = (e_{3(n-1)+1,l}, e_{3(n-1)+2,l}, e_{3(n-1)+3,l})^\top \in \mathbb{R}^3$ and $\Phi = (e_{i,j})$ follows from the model instance synthesis in Eqs. (6.18) and (6.19). Eq. (7.12) can be written in a more compact form using matrix-vector notation according to:

$$\mathcal{J}_{\mathbf{b}}^{\text{CPD-P2P}} = \sum_{m=1}^M (\Phi \mathbf{b} - \mathbf{x}'_m)^\top \mathbf{W}_m (\Phi \mathbf{b} - \mathbf{x}'_m), \quad (7.13)$$

with weighting matrix:

$$\mathbf{W}_m = \text{diag}(\omega_{1,m}, \omega_{1,m}, \omega_{1,m}, \dots, \omega_{N,m}, \omega_{N,m}, \omega_{N,m}) \in \mathbb{R}^{3N \times 3N} \quad (7.14)$$

and $\mathbf{x}'_m = (\mathbf{x}'_{1,m}^\top, \dots, \mathbf{x}'_{N,m}^\top)^\top$. Computing the partial derivative w. r. t. \mathbf{b} and equating to zero yields [Pete 12]:

$$\sum_{m=1}^M (\Phi^\top \mathbf{W}_m \Phi) \mathbf{b} = \sum_{m=1}^M \Phi^\top \mathbf{W}_m \mathbf{x}'_m. \quad (7.15)$$

For uniform weights, i. e. $\mathbf{W}_m \propto Id$, the solution corresponds to the conventional least-squares estimator from Eqs. (6.19) and (7.5). Otherwise, \mathbf{b} can be computed via SVD or CD of the symmetric positive definite matrix $\sum_m (\Phi^\top \mathbf{W}_m \Phi)$.

Point-to-Tangent The CPD framework is also generic in a sense that it is not bound to a specific distance metric. This idea is based on the observation that the Log-Likelihood objective criterion from Eq. (7.8) can be seen as a robust version of the Euclidean distance as obtained from a Gaussian kernel function. Without loss of generality, as previously proposed in [Wasz 16], when using a robust version of the P2T alignment, a CPD-like metric can be derived according to:

$$\Theta^{\text{CPD-P2T}}(\mathbf{R}\mathbf{x}_n^b + \mathbf{t}, \mathcal{S}) = \sum_{m=1}^M \omega_{n,m} \left(\frac{((\mathbf{R}\mathbf{x}_n^b + \mathbf{t} - \mathbf{x}'_{n,m}) \cdot \mathbf{n}_{n,m})^2}{\sigma^2} + \frac{d}{2} \log(\sigma^2) \right), \quad (7.16)$$

and similar for the weights or posterior probabilities $\omega_{n,m}$ as in Eq. (7.10).

As with the P2P metric, finding an optimal rotation and translation (\mathbf{R}, \mathbf{t}) for the weighted multi-link P2T alignment involves the same scheme used for the conventional single-link formulation outlined in Eq. (7.6). This follows directly from the underlying cost function using point-wise distances. For estimating \mathbf{b} an optimization problem proposed in preceding work [Wasz 16] similar to the P2P formulation from Eq. (7.12) is derived according to:

$$\mathcal{J}_b^{\text{CPD-P2T}} = \sum_{n=1}^N \sum_{m=1}^M \omega_{n,m} \left(\left(\sum_l^L b_l \mathbf{x}_{n,l} - \mathbf{x}'_{n,m} \right) \cdot \mathbf{n}_{n,m} \right)^2, \quad (7.17)$$

where $\mathbf{x}_{n,l} = (e_{3(n-1)+1,l}, e_{3(n-1)+2,l}, e_{3(n-1)+3,l})^\top \in \mathbb{R}^3$ similar to the P2P distance from Eq. (7.12). Computing the partial derivative of Eq. (7.17) w. r. t. one specific motion model parameter b_k yields:

$$\frac{\partial \mathcal{J}_b^{\text{CPD-P2T}}}{\partial b_k} = \sum_{n=1}^N \sum_{m=1}^M \omega_{n,m} \left(\sum_{l=1}^L b_l \eta_{l,n,m} \eta_{k,n,m} - \zeta_{n,m} \eta_{k,n,m} \right) \stackrel{!}{=} 0, \quad (7.18)$$

where $\eta_{l,n,m} = \mathbf{x}_{n,l} \cdot \mathbf{n}_{n,m}$ and $\zeta_{n,m} = \mathbf{x}'_{n,m} \cdot \mathbf{n}_{n,m}$. For the complete model vector \mathbf{b} this can be expressed in matrix-vector notation as:

$$\sum_{n,m} \omega_{n,m} \begin{pmatrix} \eta_{1,n,m} \eta_{1,n,m} & \cdots & \eta_{1,n,m} \eta_{L,n,m} \\ \vdots & \ddots & \vdots \\ \eta_{L,n,m} \eta_{1,n,m} & \cdots & \eta_{L,n,m} \eta_{L,n,m} \end{pmatrix} \mathbf{b} = \sum_{n,m} \omega_{n,m} \begin{pmatrix} \zeta_{n,m} \eta_{1,n,m} \\ \vdots \\ \zeta_{n,m} \eta_{L,n,m} \end{pmatrix}. \quad (7.19)$$

This is a linear equation system in the form $\mathbf{A}\mathbf{b} = \mathbf{c}$ where the symmetric matrix $\mathbf{A} \in \mathbb{R}^{L \times L}$ is in practice positive definite which allows to solve for \mathbf{b} efficiently by using CD. Note that the matrix \mathbf{A} is positive definite for weights and diagonal elements > 0 . By definition, $\omega_{n,m} > 0$. Further, $\sum_n \sum_m \eta_{l,n,m}^2 \geq 0$. These scalars are zero iff *all* motion model directions $\mathbf{x}_{n,l}$ are perpendicular to the corresponding normals $\mathbf{n}_{n,m}$. Thus, *all* points would move parallel to the surface which is an invalid degenerated model that is of no relevance in practice.

Fast CPD Approximation for RI Surface Data

A crucial issue with the CPD framework is run-time performance. In particular, linking each model point \mathbf{x}_n^b with all surface points $\mathbf{x}'_n \in \mathcal{S}$ is the major bottleneck

in terms of computational complexity. Solutions proposed in the original work [Myro 10] to cope with this performance problem are to use the *Fast Gauss Transform* [Gree 91] or an approximated scheme by switching to truncated Gaussians: instead of using all surface points $x'_n \in \mathcal{S}$ for one individual model point x_n^b only a small subset of closest points $x'_m \in \mathcal{S}_{x_n^b} \subset \mathcal{S}$ is considered. Similar to the original work by Myronenko and Song [Myro 10] the rationale behind this scheme is that the posteriors as weighting factors rapidly decay with increasing distances. Thus, the closest points exhibit the highest weights and are the most important points for the optimization scheme. The key step towards real-time computation is that for a point x_n^b its closest neighbors $\mathcal{S}_{x_n^b}$ can be computed efficiently using a projective data association scheme that is inherent to the high coverage body surface models from the multi-view fusion framework outlined in Chapter 5 of this thesis. This approach will be detailed in Section 7.4.2 in the context of efficient corresponding point search strategies.

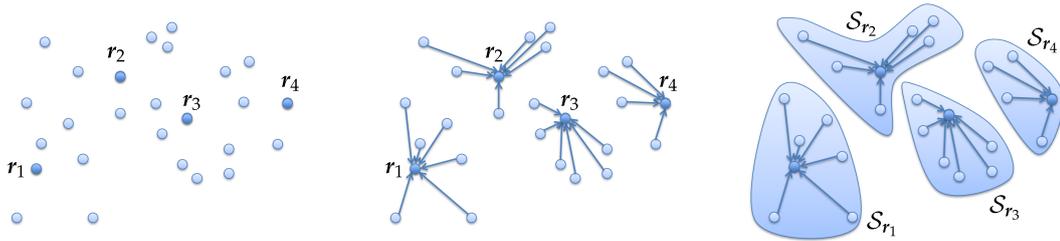
7.4 Efficient Corresponding Point Search

The key aspect towards real-time capability of the investigated model registration schemes is the corresponding point (CP) search. This applies for both the conventional single-link ICP as well as the approximated multi-link CPD. In particular for the latter strategy there is also the need to compute a set of nearest neighbors (NNs) instead of just one single corresponding point as with the conventional ICP. In this thesis, the CP search problem is formulated as a set of *query* points \mathcal{P}^Q given by the transformed model instance from Eq. (7.1) for which the corresponding *database* points defined by the current RI surface \mathcal{S} have to be found.

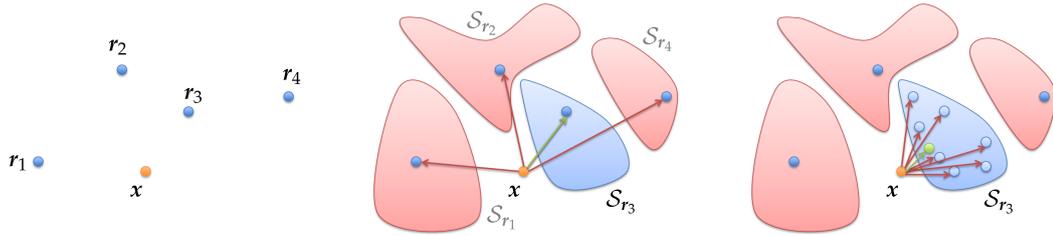
The most trivial approach to tackle this problem is the brute force (BF) primitive that tests all query points against all database points. This BF scheme has a computational complexity of $O(|\mathcal{P}^Q| \cdot |\mathcal{S}|)$ which is prohibitive in the real-time respiratory motion analysis scenario covered in this thesis. For a typical number of model points $|\mathcal{P}^Q| \sim 1.0 \times 10^4$ and number of surface points $|\mathcal{S}| \sim 3.6 \times 10^5$ cf. Section 6.6 this would imply approximately three billion point comparisons. Thus, efficient CP search strategies are inevitable and the following two sections outline CP frameworks that have been identified to comply with real-time constraints imposed by the respiratory motion analysis scenario investigated in this thesis.

7.4.1 Random Ball Cover

The RBC is a framework for efficient NN search that tackles the inherent performance problem by a reduction of the search complexity as well as a computation scheme that is suitable for the parallel architecture of modern GPUs. The RBC framework was originally proposed by L. Cayton for high-dimensional database queries and supports both the search for one closest point as well as the retrieval of a set of nearest neighbors [Cayt 10, Cayt 12]. The RBC framework provides the option for an exact closest point search or, if small errors are tolerable, a so-called



(a) RBC construction. A set of representatives $r_i \in \mathcal{S}$ is randomly selected from the database points \mathcal{S} (left). For each database point its closest representative r_i is computed (middle) and local clusters $\mathcal{S}_{r_i} \subset \mathcal{S}$ identified by the representatives are formed (right).



(b) RBC query. For a query point $x \in \mathcal{P}^Q$ (left) the local cluster \mathcal{S}_{r_i} defined by the closest representative r_i is determined (middle). The nearest neighbor for the query point is then computed by a local search in the cluster identified in the previous step (right).

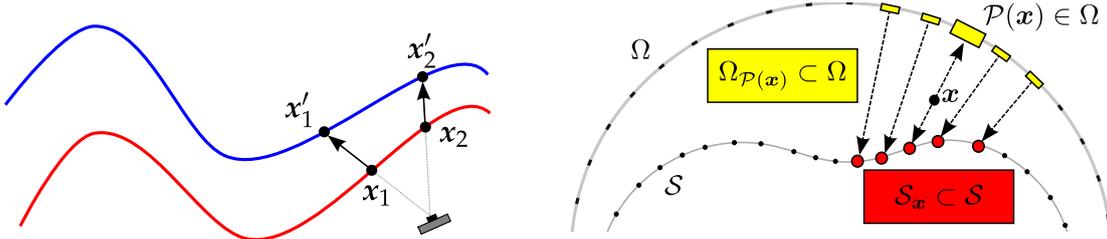
Figure 7.3: Simplified 2-D illustration of the approximative RBC framework similar to [Baue 13b]. The construction scheme is depicted in the top and the query scheme in the bottom figure. Note the two-tier approach to prune the search space.

one-shot approximative search that is typically faster [Cayt 12]. The original core concept of the *one-shot* RBC framework is a space partitioning data structure that subdivides the set of database points into potentially intersecting local clusters with equal cardinality. The individual clusters $\mathcal{S}_{r_i} \subset \mathcal{S}$ are uniquely identified by so-called *representatives* $r_i \in \mathcal{S}$ that are randomly selected from the database points \mathcal{S} . However, this implies a sorting of database entries for each representative or the need for multiple BF runs [Cayt 10]. Both are potential performance bottle-necks in a real-time scenario.

In preceding work, an approximate but performance optimized version of the *one-shot* approach was proposed [Neum 11, Baue 13b]. This scheme trades accuracy against run-time by assigning each database point $x \in \mathcal{S}$ to *exactly* one representative r_i during RBC construction, i. e. the local clusters are disjoint. This scheme is illustrated in Fig. 7.3a. The modified RBC NN query is basically consistent with the original approach proposed by Cayton [Cayt 10, Cayt 12] and consists of a two-stage method that for a given query point $x \in \mathcal{P}^Q$ prunes the search space by first identifying the nearest representative r_i followed by a closest point search in the representative's corresponding cluster \mathcal{S}_{r_i} , see Fig. 7.3b.

7.4.2 Projective Data Association

One inherent problem with corresponding point search strategies that rely on special data structures is that these data structures must be constructed as a very first



(a) Single-shot with a conventional pinhole camera model following Fig. 7.2.

(b) Projection and local search with a camera model similar to manifold ray-casting.

Figure 7.4: Projective data association schemes. Simple projection for pinhole camera models (left) and projection with subsequent local search for manifold ray casting (right).

step. As outlined in the related work from Section 7.2 and as will be shown later in the experiments for the RBC this is an overhead that cannot be neglected in a real-time scenario. Further, such strategies are often optimized for corresponding point retrieval from a discrete set of points and obtaining the corresponding point on a continuous surface is a non-trivial task.

By design, PDA schemes do not suffer from these problems. The basic idea with PDA methods is to project a query point $x \in \mathcal{P}^Q$ into the RI sensor domain Ω by using the projective geometry inherent to the camera, cf. Eq. (3.7). Let $\Omega_{P(x)}^r \subset \Omega$ similar to Eq. (3.8) denote the local index neighborhood with radius r around the query point's projection $P(x) \in \Omega$ into the RI sensor domain Ω . The nearest neighbors of x are then given by the set of surface points $\mathcal{S}_x \subset \mathcal{S}$ that correspond to the indices contained in $\Omega_{P(x)}^r \subset \Omega$, i. e.

$$\mathcal{S}_x = \left\{ x_i \mid i \in \Omega_{P(x)}^r \right\}. \quad (7.20)$$

If the radius r of the index neighborhood is set to zero, i. e. only one single point is considered, then this corresponds to the original reverse calibration scheme [Blai 95, Rusi 01]. This approach is schematically depicted in Fig. 7.4a for a conventional pinhole camera model. Selecting the radius $r > 0$ results in a PDA with subsequent local search. Similar to the high coverage surface reconstruction in Section 5.4.2, this local search can be interpreted in terms of manifold learning techniques that aim for a mapping that preserves mutual distances in the high- and low-dimensional representation of data points. Thus, the idea is that, on a local scale, mutual distances in $\Omega_{P(x)}^r$ are a strong clue for mutual distances in \mathcal{S}_x . This approach is depicted in Section 7.4.2 for the manifold ray casting surface reconstruction technique from Section 5.4.2.

By design, PDA schemes can operate on continuous surface representations by using interpolation techniques w. r. t. the sensor domain Ω .

7.5 Experiments and Results

Contrary to the structure in the preceding sections, the experiments first investigate the accuracy and run-time performance of corresponding point search strategies in order to establish a baseline strategy for the application of 4-D shape priors for respiratory motion analysis. The experiments for the latter are subdivided into a study on real-time continuous respiratory motion monitoring and motion compensated patient positioning.

7.5.1 Corresponding Point Search Strategies

The first part of the evaluation is concerned with evaluating the RBC and PDA corresponding point search strategies outlined in Section 7.4 with regard to their performance in a respiratory motion analysis scenario.

Experimental Setup

For the experiments, the pre-procedurally trained motion model of subject S_1 and its intra-procedural test sequences with 117 frames covering regular breathing was chosen exemplarily. The motion model parameter was fixed to $\mathbf{b} = \mathbf{0}$ which is a reasonable initial estimate for the model registration problem from Eq. (7.1).

The experiments are conducted on the RI surface representations \mathcal{S} that is directly obtained from the manifold ray casting method from Section 5.4.2 with $\Omega \in \mathbb{R}^{640 \times 480}$ as well as a surface data representation \mathcal{S}' with $\Omega \in \mathbb{R}^{320 \times 240}$ and \mathcal{S}'' with $\Omega \in \mathbb{R}^{160 \times 120}$ that are computed from smoothing and resampling the depth measurements in $\mathcal{R} \equiv \mathcal{S}$. For \mathcal{S}' the smoothing Gaussian was parameterized with $\sigma = 2$ and for \mathcal{S}'' with $\sigma = 4$. The rationale behind this experiment is to investigate the performance of the different NN search strategies in a scale-space and multi-resolution scenario which is a common approach in numerical optimization for improved convergence and robustness concerning local extrema. This results in the number of query points as $|\mathcal{P}^b| \sim 10^4$ and the number of data base points as $|\mathcal{S}| = 640 \cdot 480 \sim 3.0 \times 10^5$, $|\mathcal{S}'| = 320 \cdot 240 \sim 7.7 \times 10^4$ and $|\mathcal{S}''| = 160 \cdot 120 \sim 2.0 \times 10^4$.

As baseline, the brute-force primitive is chosen to yield *exact* closest points in an Euclidean distances sense. The RBC and PDA frameworks are then assessed against the BF baseline w. r. t. their inherently free parameter, i. e. the number of representatives $|\{r_i\}|$ with RBC and the local neighborhood radius r as with PDA. The experiments are concerned with assessing the competing strategies w. r. t. their capability to yield *exact* closest points as well as their run-time characteristics.

Results

The very basic accuracy and run-time trends for the RBC and PDA strategies are illustrated in Fig. 7.5 for the original surface data \mathcal{S} . The figure shows the ratio of correspondences that coincide with the NN results of the BF baseline method and the NN query times. Both aspects are assessed for varying the inherent free

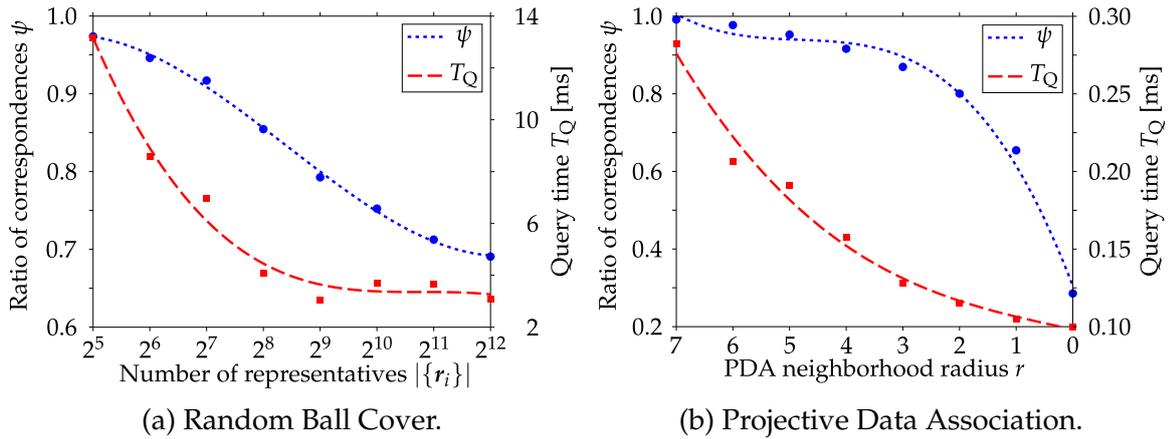


Figure 7.5: Trends for RBC (left) and PDA (right) corresponding point search for varying the free parameter, i. e. the number of representatives with RBC and the neighborhood radius with PDA. Plotted is the ratio ψ of correspondences that are in congruence with the BF baseline against the NN query time T_Q . The dashed and dotted curves denote a polynomial fitted to the data points. Note the substantially different axes scales of the RBC and PDA plots regarding both ratio of correspondences and query time.

parameter of the investigated strategies, i. e. the number of representatives with RBC and the radius of the local index neighborhood with PDA.

The trend graphs show that the number of correctly identified corresponding points directly depends on the number of representatives and neighborhood radius, respectively. For the RBC a low number of representatives yields a high accordance to the BF results and the PDA benefits from a high neighborhood radius. In this context, note that if the number of representatives is chosen to one then the RBC coincides with the BF primitive. Similarly, the PDA strategy resembles the BF method if the radius is chosen to include all available surface points. As with the number of correctly identified closest points, the run-time of one single NN query depends on the number of representatives and the radius of the local index neighborhood, respectively. This is self-explaining for the PDA and for the RBC this is due to the fact that a low number of representatives do not provide a proper search space partitioning. However, the most important finding from the run-time graph is the far superior performance of the PDA approach compared to the RBC framework. Selecting an exemplary threshold of $\psi = 0.9$, i. e. 90% correctly identified closest points w. r. t. the BF primitive, the PDA exhibits a query time of ~ 0.15 ms and outperforms the RBC by a factor of approximately 46 given an RBC query time of ~ 6.9 ms. Yet, this does not include the RBC construction time that adds an additional 6.6 ms overhead.

The metrics reported above are detailed in Table 7.1 for the full and down-sampled RI surface data \mathcal{S} , \mathcal{S}' and \mathcal{S}'' . Additionally, the table provides the 95-th percentile of the Euclidean distances between BF results and RBC and PDA NN query results, respectively. This metric is used to quantify the severity of mismatches. This is an important issue as incorrectly identified *closest* points may for example denote the second or third closest points and thus still provide a suit-

		Brute Force	Random Ball Cover								Projective Data Association							
			Number of representatives $ \{r_i\} =$								PDA radius $r =$							
			2^5	2^6	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}	7	6	5	4	3	2	1	0
\mathcal{S}''	ψ	1.00	0.97	0.96	0.95	0.93	0.90	0.87	0.82	0.78	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.59
	P_{95} [mm]	0.00	0.00	0.00	4.03	4.11	4.88	8.59	9.10	9.36	0.00	0.00	0.00	0.00	0.00	0.00	1.18	10.8
	T_C [ms]	–	1.41	1.32	0.94	0.97	1.07	1.23	2.00	2.94	–	–	–	–	–	–	–	–
	T_Q [ms]	5.21	0.47	0.21	0.12	0.12	0.10	0.15	0.25	0.46	0.19	0.11	0.08	0.07	0.05	0.04	0.04	0.03
\mathcal{S}'	ψ	1.00	0.98	0.98	0.98	0.97	0.95	0.94	0.92	0.89	1.00	1.00	1.00	1.00	0.98	0.93	0.83	0.43
	P_{95} [mm]	0.00	0.00	0.00	0.00	0.00	1.28	2.02	2.08	2.65	0.00	0.00	0.00	0.00	0.00	1.68	4.62	8.58
	T_C [ms]	–	1.97	1.95	1.97	1.97	2.32	2.80	3.81	6.23	–	–	–	–	–	–	–	–
	T_Q [ms]	20.2	1.66	1.62	1.30	0.94	0.80	0.68	0.88	0.81	0.24	0.16	0.13	0.11	0.10	0.09	0.08	0.08
\mathcal{S}	ψ	1.00	0.97	0.95	0.92	0.85	0.79	0.75	0.71	0.69	0.99	0.98	0.95	0.92	0.87	0.80	0.65	0.29
	P_{95} [mm]	0.00	0.00	0.67	1.47	2.97	3.84	4.51	5.24	5.85	0.00	0.07	0.67	1.85	3.14	4.49	6.15	8.21
	T_C [ms]	–	6.58	6.59	6.61	6.67	7.71	9.50	13.0	20.8	–	–	–	–	–	–	–	–
	T_Q [ms]	80.6	13.1	8.57	6.93	4.05	3.05	3.72	3.66	3.08	0.28	0.21	0.19	0.16	0.13	0.12	0.10	0.10

Table 7.1: Accuracy and run-time performance of different corresponding point search strategies. The brute force primitive yields exact closest points and constitutes the baseline for assessing the RBC and PDA search strategies. Reported accuracy metrics are the ratio of correspondences that coincide with the BF results (ψ) and the 95-th percentile of the Euclidean distances between BF results and the corresponding approximative results across all query points (P_{95}). For assessing the run-time performance, the construction times T_C (RBC only) and query times T_Q are reported. For the RBC the influence of the number of representatives $|\{r_i\}|$ is evaluated whereas the PDA approach is evaluated w. r. t. the radius r in pixels of the local neighborhood. All metrics are averaged over an exemplarily selected respiration sequence consisting of 117 frames.

able *corresponding* point. The outlined data basically reflect the relationship between the accuracy and run-time performance as depicted in the trend graph from Fig. 7.5. For full resolution data, the RBC must be parameterized with $|\{r_i\}| \leq 2^7$ representatives to yield $> 90\%$ correct matches and the PDA approach requires a neighborhood radius of $r \geq 4$. Interestingly, the rate of decrease and the range of correctly identified correspondences is more distinct with the PDA method. For 95% of query points, the distance to the BF baseline results are in average less than 1.47 mm with RBC and 1.85 mm with PDA. This error is in the scale of the noise level of typical RI sensors, cf. Section 4.7.

Regarding run-time performance, the experiments confirm the complexity of $O(|\mathcal{P}^Q| \cdot |\mathcal{S}|)$ for the brute force primitive and different resolution levels. For the RBC, the results demonstrate the overhead of the acceleration structure construction that is directly related to the number of representatives. In practical applications, this usually requires a trade-off between low construction times on the one hand and low query times on the other hand. For ICP-like algorithms where the RBC data structure is built once and queried often, one favors low query times and accepts the additional overhead during construction. An interesting outcome of the experiments is the comparison of run-times between the different resolution levels for the RBC and the PDA scheme. While both construction and query run-times can be substantially decreased for the RBC this does not hold true for the PDA approach. This is due to the fact that the PDA method by design does not depend on the number of data base points. Run-time differences between the resolution levels are most likely due to cache hit issues with texture memory that was used to represent the surface data \mathcal{S} .

7.5.2 Continuous Respiratory Motion Monitoring

The next part of the evaluation in this chapter is concerned with the applicability of 4-D shape priors for continuous respiratory motion monitoring. The experiments and results have been previously published in [Wasz 16] and are twofold. First, the general principles and methodology of the proposed respiratory motion framework are assessed. In particular, this includes the ability of 4-D motion models to adapt to unseen RI body surface data, the robustness of the approximated CPD registration scheme in the presence of noise and outliers, the run-time characteristics and the performance of the model-based surrogates from Eq. (7.2) compared to conventional RI surrogates. Additionally, the respiration surrogates obtained by the model-based formulation are compared to a respiration signal obtained from an impedance pneumography (IP) sensor that does not rely on body surface motion but instead directly measures the change of lung volume. For the experiments that are concerned with respiration surrogates, the focus is on the assessment of the proposed WVR models and their ability to automatically differentiate between distinct breathing patterns.

Experimental Setup

For the experiments the approximated CPD model registration scheme from Section 7.3.2 was used for registering the shape motion model to the unseen RI body

surface. The CPD method was parameterized with $w = 0.99$ and a 5×5 PDA neighborhood $\mathcal{S}_{x_n^b}$, i. e. each model point x_n^b is connected to its 25 nearest neighbors $x'_{n,m} \in \mathcal{S}_{x_n^b}$. The initial guess for the model parameter was set to $\mathbf{b} = \mathbf{0}$. Convergence of the iterative optimization scheme is assumed when there are no considerable changes in the cost function which is expressed in this experiment as:

$$\left| 1 - \left| \mathcal{J}_{\mathbf{b},\sigma}^{\text{CPD}} / \mathcal{J}_{\mathbf{b}',\sigma'}^{\text{CPD}} \right| \right| < \epsilon, \quad (7.21)$$

where a prime denotes the estimates from the previous iteration. The convergence parameter was heuristically determined and set to $\epsilon = 10^{-2}$.

Setup and data for methodological assessment For the methodological experiment, high resolution RI data from subjects S_1, \dots, S_4 were captured using two Microsoft Kinect RI sensors and reconstructed using the multi-view fusion framework introduced in Chapter 5 of this thesis. The sensor specifications and fusion parameters can be found in Section 5.5.2. The surface post-processing pipeline including cropping and Laplacian smoothing as well as details on the body surface deformations are detailed in the foundation chapter on 4-D shape priors in Section 6.6.1. The resulting key figures of the trained 4-D motion models are summarized in Table D1.

For testing, the subjects were instructed to perform one cycle of thoracic and abdominal and two cycles of regular breathing. These sequences were then analyzed at the full frame-rate of the RI sensor, i. e. 30 Hz. The detailed listing of the number of frames for each cycle and each subject is provided in Table D1. In contrast to the surfaces used in the training phase, neither cropping nor mesh decimation nor Laplacian smoothing was applied. The rationale behind this scheme is multi-modality, i. e. the employed algorithms must be robust w. r. t. different spatial sampling, mesh topology or partial matching.

Setup and data for the IP experiment For the second experiment, three additional subjects S_5 – S_7 were captured using two Asus Xtion PRO RI sensors with a frame-rate of ~ 30 Hz and a resolution of $\Omega = \mathbb{R}^{320 \times 240}$. The same multi-view fusion, pre-processing, model generation and registration methods as for subjects S_1 – S_4 were used, cf. Section 6.6.2. Simultaneously, a respiration signal based on IP was acquired using a PMM2 sensor (Siemens Healthcare, Erlangen, Germany). The IP sensor sends high frequency current through biopotential electrodes that are attached to the left and right side of the thorax and the side of the belly. Based on the change of impedance caused by the change of lung volume, a 1-D respiration signal was obtained at 250 Hz. To match the RI frame-rate, the IP signal was uniformly downsampled to 30 Hz.

Similar to the first experiment with subjects S_1 – S_4 , the subjects S_5 – S_7 were instructed to perform free abdominal and thoracic breathing to train the WVR model. This resulted in 6–7 training samples, see Table D2 for a detailed listing on the number of frames and the key figures of the corresponding WVR models.

For testing, a series of different breathing instructions were to be performed over a period of about 7 min. The breathing patterns consisted of (i) abdominal

and thoracic breathing with a duration of about 60 s, each, (ii) breathing at fixed frequencies of 0.5 Hz and 0.17 Hz for both shallow (60 s) and strong respiration (30 s) and (iii) breath holds of 15 s over a total time of 120 s. Using this protocol, a total number of more than 10 000 testing frames were collected per subject, see Table D2 for a detailed listing.

Though being acquired simultaneously, the RI surrogates and the IP signal exhibited a temporal shift caused by latencies in the employed computer and recording system. To bring both signals into congruence the phase-wise shift that maximizes the PCC between the shifted IP signal and the model surrogate $\xi_{\{J,A,T\}}^M$ yielding the highest correlation was estimated. The rationale behind this signal alignment scheme is that the temporal shift is not uniform across the entire evaluation period but at least one model surrogate is supposed to explain the IP signal within one evaluation phase.

Evaluation concepts The ability of pre-procedurally trained 4-D motion models to adapt to unseen intra-procedural respiration states is quantified using the model-to-surface (M2S) registration error. This error is calculated as the distance between the points $\mathbf{x}_n^b \in \mathcal{P}^b$ of the registered motion model and intra-procedural RI surface data \mathcal{S} :

$$\mathcal{E}_n^{\text{M2S}} = \min \left\{ \|\mathbf{x}_n^b - \mathbf{x}'\|_2, \mathbf{x}' \in \mathcal{S} \right\}. \quad (7.22)$$

For parameters \mathbf{b} that were estimated using the P2T distance or the CPD framework, the M2S error \mathcal{E}^{M2S} is an unbiased metric as it does not reflect the corresponding cost functions that are optimized, cf. Eqs. (7.6), (7.11) and (7.16).

To assess the robustness of the proposed framework in the presence of noise and outliers, the RI data was corrupted according to $\mathcal{R}^\Delta(\mathbf{i}) = \mathcal{R}(\mathbf{i}) + \Delta r_i$ prior to computing the corresponding surface representation $\mathcal{S}^\Delta(\mathbf{i})$. The pixel-wise offsets Δr_i were drawn from a standard normal distribution to simulate noise. For 25% of the offsets which correspond to $|\Delta r_i| > 1.15$ mm according to the standard normal distribution, an arbitrarily selected multiplication by a factor of 5 was applied to simulate outliers. This is a simplified noise model compared to the pre-processing experiments in Section 4.7.1 that incorporated systematic errors and localized defect regions. However, as shown in the corresponding evaluation, pre-processing techniques are effective means to remove these artifacts and only temporal noise and a limited number of outliers remain. The robustness of an alignment strategy is then quantified by the point-wise model-to-model (M2M) difference as:

$$\mathcal{E}_n^{\text{M2M}} = \|\mathbf{x}_n^b - \mathbf{x}_n^{b^\Delta}\|_2, \quad (7.23)$$

where $\mathbf{x}_n^b \in \mathcal{P}^b$ are the model points registered to the original RI surface data \mathcal{S} and $\mathbf{x}_n^{b^\Delta} \in \mathcal{P}^{b^\Delta}$ denote the model points \mathcal{P}^{b^Δ} aligned to the corrupted data \mathcal{S}^Δ .

For comparing the model-based surrogates to conventional surface monitoring techniques, two baseline surrogates that encode abdominal (ξ_A^{RI}) and thoracic (ξ_T^{RI}) surface motion, respectively, are used. Here, two circular regions $\Omega_j \in \Omega$ with $j \in \{A,T\}$ and a diameter of approximately 50 mm were manually selected for each of the subjects. The regions were placed at positions where the maximum

respiration magnitude w. r. t. thoracic and abdominal motion is expected. Similar to [Faya 11, Xia 12], the respiration surrogates $\bar{\zeta}_j^{\text{RI}}$ are calculated as the mean distance of 3-D points to the camera center according to:

$$\bar{\zeta}_j^{\text{RI}} = |\Omega_j|^{-1} \sum_{i \in \Omega_j} \|\mathcal{S}(i)\|_2. \quad (7.24)$$

Results

Results for the adaptability and robustness tests as well as the comparison with base-line RI surrogates are conducted for subjects S_1 – S_4 . The comparison of the model-based surrogates and the IP signal is performed using subjects S_5 – S_7 .

Adaptability and robustness The M2S error metric from Eq. (7.22) that quantifies the ability of 4-D shape priors to adapt to unseen RI body surface data of the corresponding subject are depicted in Fig. 7.6a. Here, the first (Q_1), second (Q_2) and third (Q_3) quartiles of $\mathcal{E}_n^{\text{M2S}}$ across all model points x_n^b were calculated. The quartiles were finally averaged across all breathing sequences for the purpose of a compact representation. For all subjects and respiration sequences the average median (Q_2) distance is always below 1.0 mm and across all subjects and respiration sequences a mean Q_2 distance of 0.5 mm was obtained. Note that an exact surface reconstruction is not to be expected as the dimensionality reduction step in the model generation phase is parameterized in a way to account for shape variations caused by respiratory motion and not by noise. Further, Laplacian smoothing applied to the training samples acts as an additional low-pass component that filters out small structures and surface variations. The reconstruction errors of approximately 1.0 mm are in the scale of the noise level of the employed RI sensor for an acquisition distance of ~ 1 m, cf. Section 4.7.

The M2M metric \mathcal{E}^{M2M} from Eq. (7.23) to assess the robustness of the proposed registration methods in the presence of noise and outliers is depicted in Fig. 7.6b for the conventional P2T least-squares estimator from Eq. (7.6) and the proposed robust CPD approximation from Eq. (7.16). Again, the quartile differences $\mathcal{E}_n^{\text{M2M}}$ across the entire surface were averaged over the individual breathing sequences. The differences w. r. t. the ground-truth estimation using non-corrupted data are substantially smaller with the robust CPD alignment strategy compared to the conventional least-squares estimator and, except for subject S_4 with thoracic breathing as a prominent error increase, the median difference never exceeds 0.2 mm. As an overall measure across all subjects and sequences, the robust CPD estimator yields an average Q_2 M2M error of 0.06 mm and outperforms the conventional approach that has a mean Q_2 M2M distance of 0.36 mm by a factor of approximately 6.

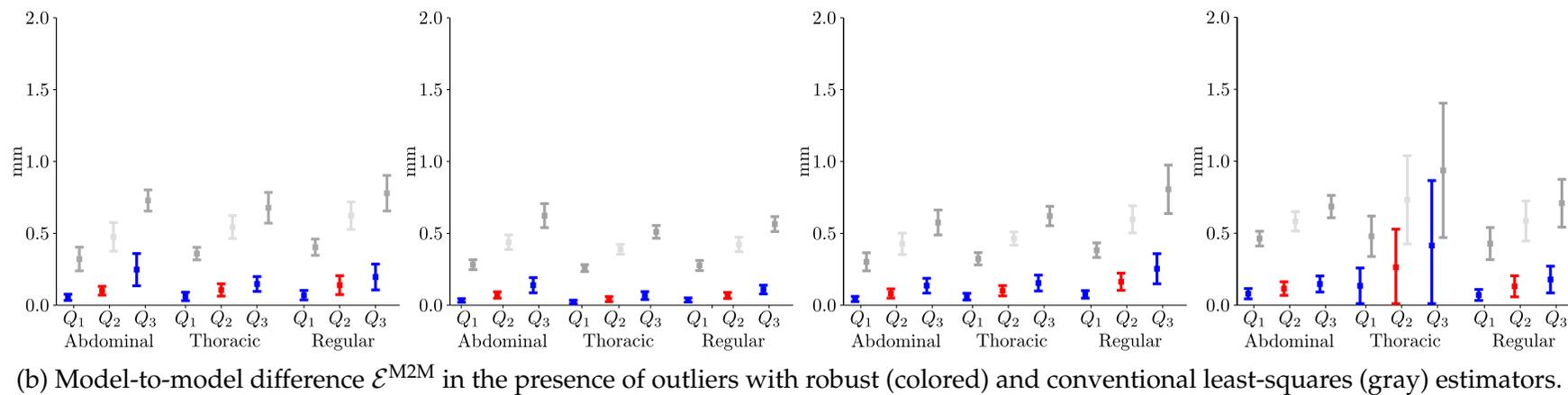
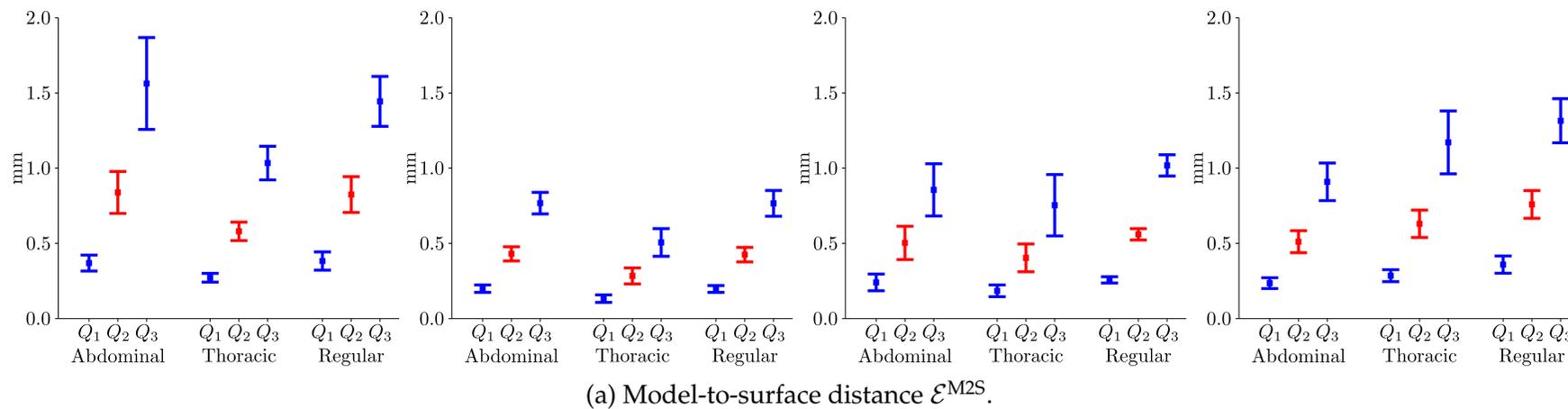


Figure 7.6: Model-to-surface distance and model-to-model differences for the P2T metric. Shown are the first to third quartiles Q_1 – Q_3 averaged across all frames for abdominal, thoracic and regular breathing patterns and subjects S_1 (left) to S_4 (right).

Table 7.2: Performance evaluation of the proposed framework for model-based continuous motion monitoring. Reported are the average number of iterations and the resulting computation times for conventional P2P and the proposed P2T approximated CPD alignment schemes (cf. Eqs. (7.11) and (7.16)). The metrics are averaged over all testing frames.

	CPD P2T		CPD P2P	
	Iterations	Run-time [ms]	Iterations	Run-time [ms]
S_1	15.8 ± 5.5	18.5 ± 5.5	29.1 ± 3.6	28.9 ± 3.5
S_2	9.67 ± 1.4	12.2 ± 1.8	34.0 ± 5.3	33.9 ± 4.9
S_3	12.1 ± 3.1	15.6 ± 3.6	32.2 ± 8.1	33.0 ± 7.6
S_4	14.6 ± 6.1	16.2 ± 5.8	37.8 ± 6.2	37.2 ± 5.8

Run-time evaluation The number of required iterations and the resulting run-times for the approximated CPD registration schemes using conventional P2P versus P2T alignment are given in Table 7.2. In average, the P2T scheme requires 13 iterations which corresponds to ~ 16 ms. For comparison, the conventional P2P metric requires about 33 iterations with an average run-time of ~ 33 ms. Though being computationally more expensive for one single iteration, the P2T scheme has a superior convergence rate compared to P2P alignment. Ultimately, this results in a substantially lower run-time for the P2T method, outperforming the conventional P2P scheme by a factor of 2.0 in average. However, regardless of the distance metric, the multi-link formulations from Eqs. (7.15) and (7.19) can be computed very efficiently on GPU architectures as the underlying summation of matrices can be calculated based on tree-based reduction schemes [Harr 07]. Including the run-times for multi-view RI data fusion of approximately 5 ms as listed in Table 5.2, continuous respiratory motion monitoring using 4-D shape priors with the approximated CPD P2T registration scheme achieves run-times of about 20 ms or 50 Hz on the off-the-shelf hardware used in this thesis, see Appendix B.

Model-based vs. conventional RI-based respiration surrogates Table 7.3 lists the PCCs between the baseline surrogate $\tilde{\zeta}_j^{\text{RI}}$ from Eq. (7.24) and the model-based surrogates $\tilde{\zeta}_j^{\text{M}}$ from Eq. (7.2) for the conventional PCA-based model and the proposed WVR model. For the WVR model a PCC of 0.91 with subject S_2 and S_3 for abdominal breathing is obtained at the minimum. In contrast, the surrogates obtained by the conventional PCA model partially fail to differentiate between distinct respiration patterns with PCCs as low as 0.10 as with S_3 and abdominal breathing. This can be traced back to the global nature of the conventional PCA modes, see Fig. 6.6 for an illustration of the deformation modes. The global deformations induced by a certain mode often must be corrected by a complementary mode such that the resulting model instance matches the RI surface data. This results in a score on the model parameter b_j defining the surrogate $\tilde{\zeta}_j^{\text{M}}$ though there is no anatomical equivalent. In contrast, the WVR model does not suffer from this problem due to the inherent sparsity constraints and ability to describe local sur-

Table 7.3: PCCs $\rho_{\{A,T\}}$ between conventional RI-based surrogates $\zeta_{\{A,T\}}^{\text{RI}}$ and model-based surrogates $\zeta_{\{A,T\}}^{\text{M}}$ for the abdominal (A) and thoracic (T) mode. Reported are the values for subjects S_1 - S_4 and sequences of abdominal, thoracic and regular breathing patterns. The evaluation further opposes the standard PCA vs. the proposed WVR approach. All reported PCCs are statistical significant ($p \leq 0.01$).

		Thoracic		Abdominal		Regular	
		ρ_T	ρ_A	ρ_T	ρ_A	ρ_T	ρ_A
S_1	PCA	1.00	0.99	1.00	1.00	0.97	0.99
	WVR	1.00	1.00	1.00	1.00	0.98	1.00
S_2	PCA	1.00	0.31	0.86	1.00	0.96	0.97
	WVR	1.00	0.97	0.91	1.00	0.99	0.99
S_3	PCA	0.98	0.13	0.10	0.98	0.96	0.79
	WVR	0.98	0.95	0.91	1.00	0.97	0.98
S_4	PCA	0.98	1.00	1.00	0.98	0.99	0.94
	WVR	0.98	1.00	1.00	0.99	0.99	0.99

face deformations. Across all breathing sequences, subjects and surrogates a PCC of 0.91 for the PCA model and 0.98 for the WVR model was obtained.

Comparison to IP respiration signals Fig. 7.7 exemplarily depicts the shift corrected IP signal and the proposed model-based RI surrogates for subject S_5 over the entire evaluation period of approximately 7 min. Noticeable are the distinct breathing instructions and varying degrees of correlation between the IP signal and the different model-based surrogates. This is most prominent for phase P_2 where the abdominal model surrogate ζ_A^{M} contradicts the IP sensor regarding both signal scale and shape whereas the thoracic surrogate ζ_T^{M} follows the IP signal. Overall, the thoracic and joint model surrogates $\zeta_{T,J}^{\text{M}}$ best resemble the IP signal which can best be seen for the distinct spikes in the breath hold phase P_7 .

For quantification, the PCCs between the shifted IP signal ζ^{IP} and the model-based respiration surrogates $\zeta_{J,T,A}^{\text{M}}$ were computed for the individual sequences P_t . The results are listed in Table 7.4. For the joint signal ζ_J^{M} , an average PCC of 0.96 ± 0.04 with $\rho_J \geq 0.97$ for 19 out of 21 evaluations and a minimum of 0.92 was found across the individual phases and all subjects. In contrast, the thoracic and abdominal surrogates show substantial variations regarding the correlation with the IP signal with an average PCC of 0.79 ± 0.35 and 0.84 ± 0.28 for the abdominal and thoracic surrogate, respectively. Here, the most prominent results are for subject S_6 and phase P_1 where there is no correlation in the thoracic mode ($\rho_T = -0.04$) but a strong correlation in the abdominal part ($\rho_A = 0.97$).

As an overall performance metric, the PCCs across the entirety of available breathing samples were computed. This avoids any bias that might occur in the phase-wise correlation due to the signal shift or the mean and variance normalization inherent to the PCC. Here, PCCs of 0.93 to 0.96 for the joint, 0.93 to 0.99

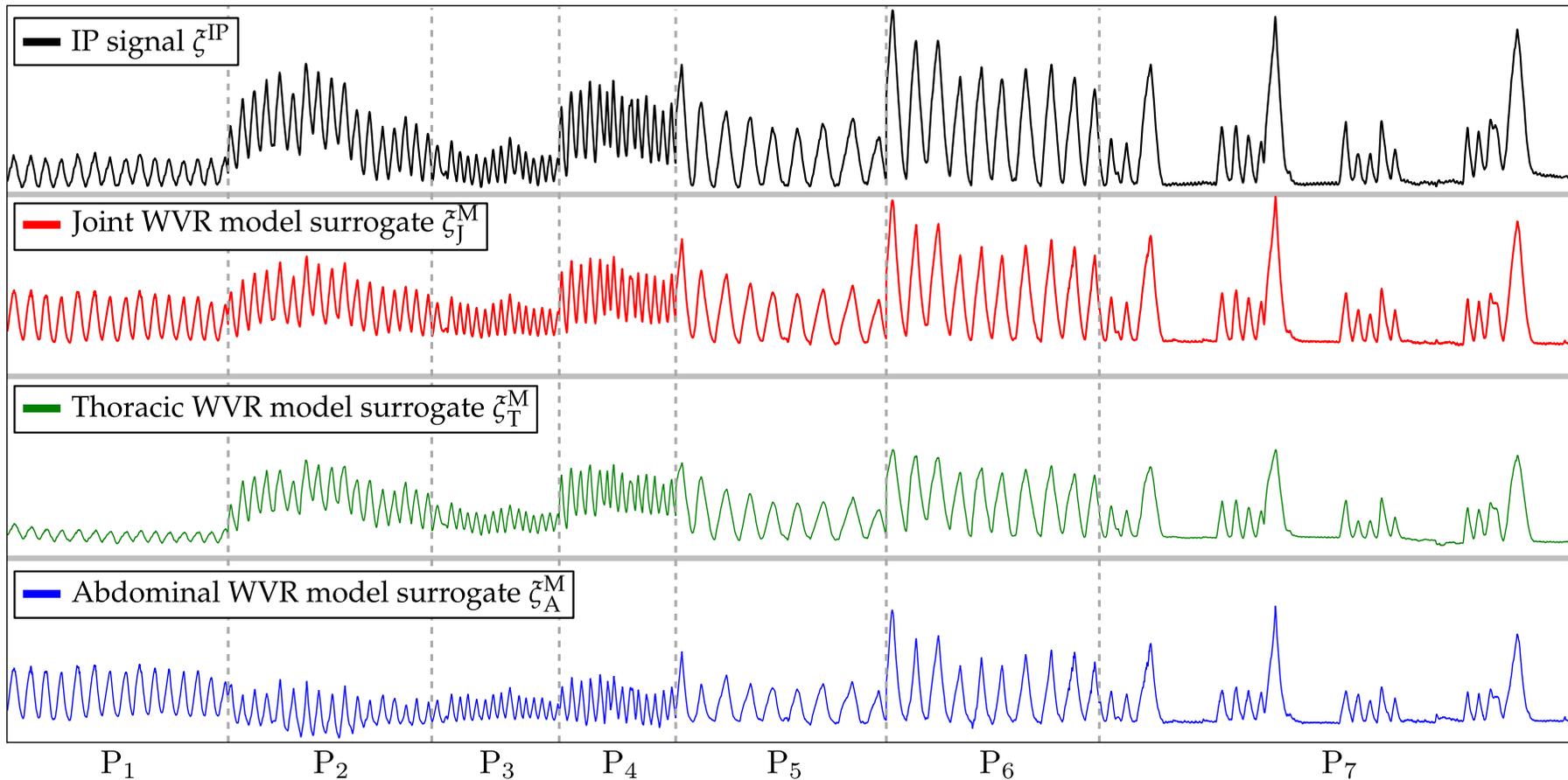


Figure 7.7: Qualitative visualization of the IP signal ζ^{IP} compared to the joint (ζ_J^{M}), thoracic (ζ_T^{M}) and abdominal (ζ_A^{M}) WVR model surrogates for subject S_5 and different respiration patterns. The signals divide into abdominal (P_1), thoracic (P_2), fast shallow (P_3), fast strong (P_4), slow shallow (P_5), slow strong (P_6) and breath hold (P_7) breathing instructions. For visualization, the IP signal is scaled to $[\min(\zeta^{\text{IP}}), \max(\zeta^{\text{IP}})]$ and the model surrogates to $[\min(\zeta_J^{\text{M}}, \zeta_T^{\text{M}}, \zeta_A^{\text{M}}), \max(\zeta_J^{\text{M}}, \zeta_T^{\text{M}}, \zeta_A^{\text{M}})]$.

	All phases			P ₁			P ₂			P ₃			P ₄			P ₅			P ₆			P ₇		
	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A	ρ_J	ρ_T	ρ_A
S_5	0.96	0.97	0.73	0.98	0.85	0.98	0.94	0.96	0.63	0.98	0.98	0.92	0.97	0.99	0.80	0.99	0.98	0.94	0.99	0.99	0.96	0.99	0.98	0.98
S_6	0.93	0.93	0.90	0.97	-0.04	0.97	0.98	0.96	0.97	0.98	0.55	0.96	0.99	0.93	0.97	0.99	0.93	0.98	1.00	0.99	0.99	0.98	0.88	0.96
S_7	0.96	0.99	0.16	0.92	0.99	0.86	0.97	0.99	-0.47	0.97	0.97	0.84	0.98	0.98	0.09	0.98	0.97	0.96	0.99	1.00	0.63	0.98	0.98	0.74

Table 7.4: PCCs (ρ_i) between the shifted IP sensor signal and WVR model-based joint (ρ_J), thoracic (ρ_T) and abdominal (ρ_A) surrogates. Correlation coefficients are reported for phases of different breathing patterns as well as the entire evaluation period. The individual phases consist of abdominal (P₁), thoracic (P₂), fast shallow (P₃), fast strong (P₄), slow shallow (P₅), slow strong (P₆) and breath hold (P₇) breathing instructions, cf. Fig. 7.7 for an exemplary illustration. The surrogate yielding the best PCCs to estimate the signal shift is highlighted. All relevant PCCs with $|\rho_i| > 0.1$ are statistically significant ($p \leq 0.01$).

for the thoracic, and 0.16 to 0.90 for the abdominal surrogates were found. Compared to the phase-wise evaluation, the performance of the joint signal slightly decreased whereas the thoracic surrogate shows a higher correlation w. r. t. the IP signal. This effect is solely due to the mathematical foundations of the PCC and can exemplarily be described for subject S_5 . As depicted in Fig. 7.7 for the abdominal P_1 test sequence, there is a high correlation with the joint surrogate whereas the thoracic surrogate exhibits a slight downward shift resulting in a lower correlation. This observation is also quantified in Table 7.4. However, when considering this particular sequence in the context of the entirety of available samples, the thoracic surrogate better reflects the IP signal compared to the joint surrogate that, for example, does not feature the sharp increase in the surrogate response between sequences P_1 and P_2 .

7.5.3 Motion Compensated Patient Alignment

The concluding part of the evaluation addresses the problem of motion compensated patient positioning using 4-D shape priors. The evaluation is particularly concerned with (i) the benefits of using a motion compensated alignment scheme compared to conventional positioning, (ii) the effect of varying the body surface coverage as motivated in Section 6.3.2, and (iii) the degree of initial misalignment. Throughout the experiments, an additional focus is on the performance of the P2P and P2T metrics that drive the model to RI surface registration.

Experimental Setup

The experimental setup and the data follow the respiratory motion analysis experiment for subjects S_1 – S_4 in Section 7.5.2, see also Table D1 for a detailed listing of the key parameters. As the subjects were instructed not to move between the training and the testing phase, the ground-truth transformation for the evaluation can be derived from the identity transform. This scheme further allows to use the surrogates, metrics and general findings from the preceding experiment to assess the alignment accuracy and precision. For the conventional positioning, the mean body surface, i. e. $\mathbf{b} = \mathbf{0}$ was chosen as static reference.

Evaluation concepts For evaluation, the trained motion models were initially misaligned by a set of transformations $T^{\text{MA}} \equiv (\mathbf{R}^{\text{MA}}, \mathbf{t}^{\text{MA}})$ with rotation $\mathbf{R}^{\text{MA}} \in \text{SO}_3$ and translation $\mathbf{t}^{\text{MA}} \in \mathbb{R}^3$. The parameterizations of these transformations are motivated by initial misalignments and residual positioning errors with automatic RI based coarse alignment strategies as for example proposed by Bauer et al. [Baue 11] or Placht et al. [Plac 12]. The detailed transformation specifications are listed in Table 7.5.

Conventionally, the accuracy and precision of a patient alignment system is assessed by the difference between the estimated and the ground truth transform w. r. t. the translational and rotational components [Baue 11, Plac 12, Wasz 12b]. Though this is an intuitive practical assessment regarding the treatment table transformation input parameters, this evaluation scheme does not account for the actual alignment error in its entirety. This is due to the fact that, in general, rotations and

	\mathbf{t}_i^{MA} [mm]			\mathbf{R}_i^{MA} [°]			Initial \mathcal{E}^{TRE} [mm]			
	t_x	t_y	t_z	r_x	r_y	r_z	S_1	S_2	S_3	S_4
T_0^{MA}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T_1^{MA}	2.0	-1.0	3.0	1.0	-1.0	2.0	30.0	29.4	30.9	28.3
T_2^{MA}	7.0	-5.0	8.0	3.0	-2.0	2.0	64.0	62.7	67.5	60.3
T_3^{MA}	27.0	-37.0	15.0	10.0	-9.0	7.0	223.6	218.8	234.8	210.1
T_4^{MA}	57.0	-35.0	33.0	-17.0	20.0	15.0	432.8	413.8	466.3	407.4

Table 7.5: Parameterization of the transformations $T_i^{\text{MA}} \equiv (\mathbf{R}^{\text{MA}}, \mathbf{t}^{\text{MA}})$ defining the initial misalignment for motion compensated patient positioning. Translations and rotations are defined along and around the x , y and z axis, respectively. The table further lists the corresponding initial TRE \mathcal{E}^{TRE} for subjects S_1 - S_4 .

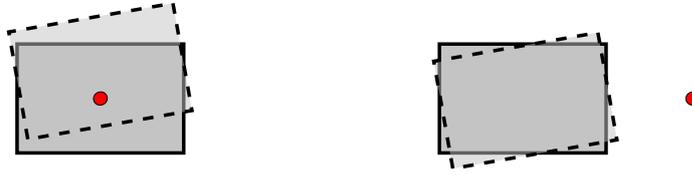


Figure 7.8: Alignment assessment based on rotation and translation components only. Two rectangles are rotated around different pivots (red dots) and vertically shifted. The left figure corresponds to a *patient*-based whereas the right figure is a *LINAC*-based frame of reference. Though the rotation angle and the shift are the same, fundamentally different alignment matches are obtained.

translations exhibit non-comparable physical units. An exception are small rotations that can be approximated by translations as pursued in the linearization of the P2T alignment metric in Section 7.3.1. Further, rotation and translation errors depend on the frame of reference for evaluation, i. e. *patient*-based or *LINAC*-based, and may further cancel each other out to a certain degree. A sketch of this problem is depicted in Fig. 7.8.

In this thesis, the estimated alignment is instead assessed by the target registration error (TRE). For a known ground truth transformation $T^{\text{GT}} \equiv (\mathbf{R}^{\text{GT}}, \mathbf{t}^{\text{GT}})$, the estimated transformation $\hat{T} \equiv (\hat{\mathbf{R}}, \hat{\mathbf{t}})$ from Eq. (7.1) and the n -th model point $x_n^{b'} \in \mathcal{P}^{b'}$, the TRE is calculated according to:

$$\mathcal{E}_n^{\text{TRE}} = \left\| \left(\mathbf{R}^{\text{GT}} x_n^{b'} + \mathbf{t}^{\text{GT}} \right) - \left(\hat{\mathbf{R}} x_n^{b'} + \hat{\mathbf{t}} \right) \right\|_2. \quad (7.25)$$

Here, $\mathcal{P}^{b'}$ denotes the model with the largest surface extent analogous to the model-based surrogates from Eq. (7.2). The model with the largest extent is favored over the estimated model with parameter $\hat{\mathbf{b}}$ as the ground truth model parameter is unknown and b' provides an upper bound for the surface extent and thus the TRE. This evaluation scheme further implicitly accounts for the internal registration accuracy as all landmarks inside the body must necessarily be contained within the

surrounding body surface. The final error metric for this experiment is calculated as the maximum of all TREs:

$$\mathcal{E}^{\text{TRE}} = \max \left\{ \mathcal{E}_n^{\text{TRE}} \mid n = 1 \dots |\mathcal{P}^{b'}| \right\}. \quad (7.26)$$

This maximum scheme is preferred over the mean TRE across all model points in order to establish an upper bound for the registration error.

Results

Fig. 7.9 depicts the results for conventional and motion compensated patient alignment as well as the effect of varying degrees of body surface coverage and initial misalignments. The individual aspects of the experiments and key findings depicted in this figure are detailed in the following dedicated paragraphs.

General aspects of conventional vs. motion compensated alignment Results for conventional alignment using the P2P metric as a baseline for the evaluation are shown in Fig. 7.9a. Besides a high target registration error that frequently exceeds 10.0 mm the plots show that for large initial misalignment configurations $T_{\{3,4\}}^{\text{MA}}$ the registration often fails. This is a general problem of ICP-like registration schemes that motivates the employment of coarse pre-alignment techniques, see the related work in Section 7.2.

The results for motion compensated positioning are reported in Fig. 7.9b for the P2P metric and in Fig. 7.9c for the P2T metric. For all investigated scenarios, the motion compensated positioning approaches outperform the conventional baseline alignment scheme. As an overall performance indicator using the largest body surface coverage R_1 and the medium initial misalignment T_2^{MA} , an average \mathcal{E}^{TRE} of about 19.5 mm for conventional and approximately 4.7 mm for motion compensated alignment using the P2P distance metric was obtained across all evaluation frames and subjects. For the P2T metric, a TRE of 11.0 mm for conventional and 3.6 mm for motion-compensated alignment was obtained. In addition to this average error reduction by a factor of more than 3 the plots also depict a substantially smaller error spread for the motion compensated alignment approaches. A general finding from these experiments is that the registration schemes using the P2T metric outperform the P2P-based variants regarding both the stability w. r. t. initial misalignments and the final TRE.

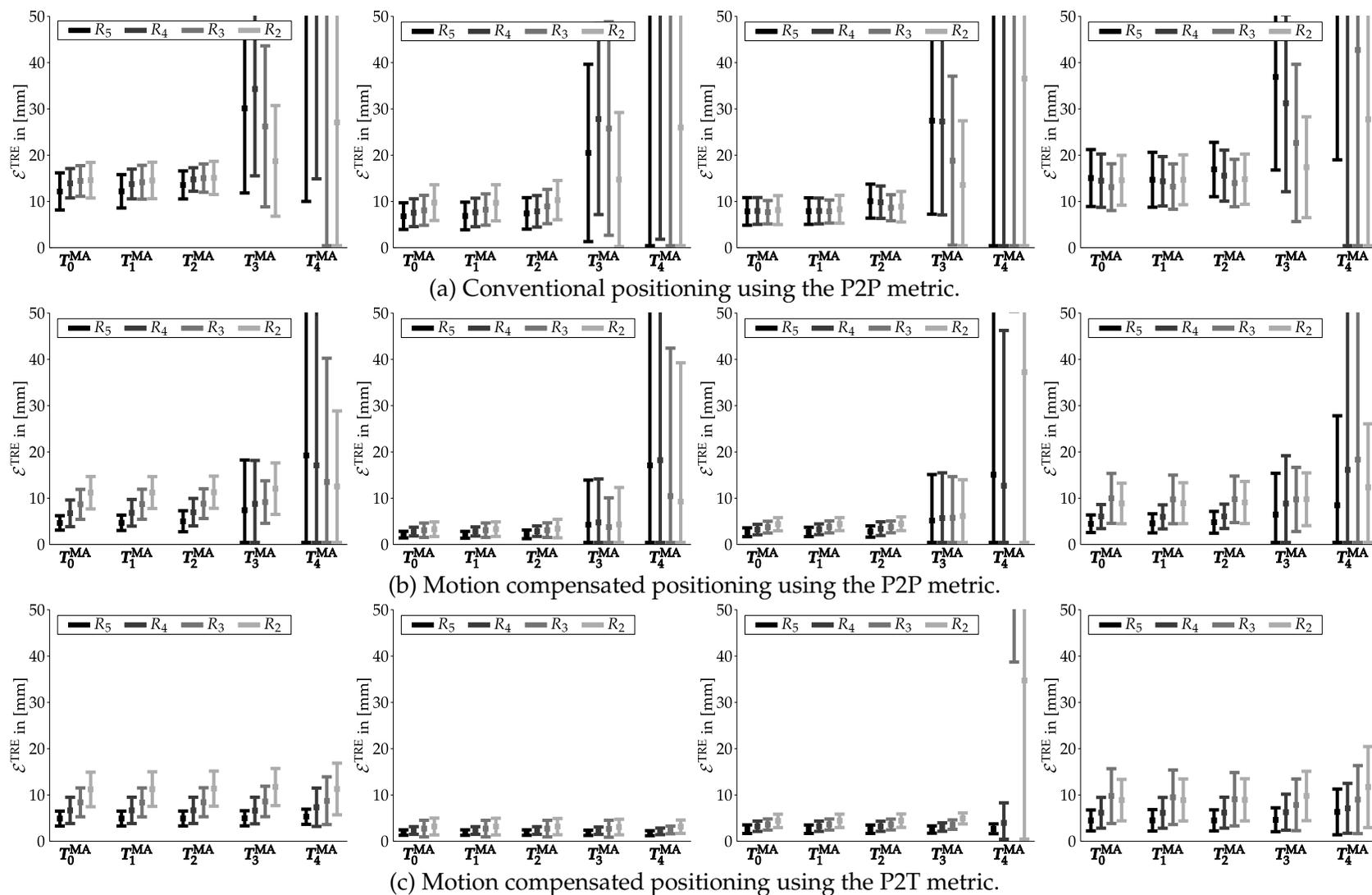


Figure 7.9: Conventional and motion compensated positioning for subjects S_1 - S_4 (left to right). Shown is the TRE \mathcal{E}^{TRE} for different initial misalignments T_i^{MA} , cf. Table 7.5 and varying degrees of body surface coverage $R_2 \subset R_3 \subset R_4 \subset R_5$, cf. Fig. 6.5.

	Thoracic				Abdominal				Regular			
	$\overline{\mathcal{E}^{\text{TRE}}}$	$\rho_{\mathcal{E}^{\text{M2S}}}$	$\rho_{\zeta_J^{\text{M}}}$	$\rho_{\mathcal{K}_{\text{RE}}}$	$\overline{\mathcal{E}^{\text{TRE}}}$	$\rho_{\mathcal{E}^{\text{M2S}}}$	$\rho_{\zeta_J^{\text{M}}}$	$\rho_{\mathcal{K}_{\text{RE}}}$	$\overline{\mathcal{E}^{\text{TRE}}}$	$\rho_{\mathcal{E}^{\text{M2S}}}$	$\rho_{\zeta_J^{\text{M}}}$	$\rho_{\mathcal{K}_{\text{RE}}}$
S_1	6.33	0.90	-0.89	-0.74	3.18	0.63	0.45	0.45	4.74	0.77	0.45	0.48
S_2	2.36	0.45	-0.15	0.12	1.09	0.67	-0.49	-0.51	1.99	0.43	0.18	0.11
S_3	2.18	0.51	0.04	-0.19	2.07	0.84	-0.71	-0.89	3.12	0.09	-0.33	0.27
S_4	7.22	-0.50	-0.50	-0.51	3.34	0.38	-0.19	-0.54	4.35	0.32	-0.56	0.11

Table 7.6: Sequence-wise mean target registration errors $\overline{\mathcal{E}^{\text{TRE}}}$ in [mm] and correlation coefficients of the TREs to selected metrics for the largest body surface region R_5 and the identity initial misalignment T_0^{MA} .

Body coverage and initial misalignment An important finding from the results in Fig. 7.9 is that both the P2P and P2T based motion compensated alignment techniques are in general more robust w. r. t. large initial misalignments. Additionally, the degree of body surface coverage now constitutes an important factor with the most prominent result being subject S_3 and the P2T metric where only the large misalignment configuration T_4^{MA} could not be recovered for small body coverage $R_{\{3,2\}}$. For small to medium initial misalignments $T_0^{\text{MA}}-T_2^{\text{MA}}$ the TREs for motion compensated positioning are mostly governed by the degree of body surface coverage and not the initial alignment error. In fact, for any degree of body surface coverage and across all subjects, the mutual correlation coefficients between the frame-wise TREs for these misalignments is always ≥ 0.89 for the P2T metric and ≥ 0.71 for the P2P metric. These high correlation coefficients with the P2T metric indicate that the registration procedure always converges to the global optimum that, however, does not necessarily coincide with the reference alignment position. This can best be seen for the identity transformation T_0^{MA} where the registration is performed from the reference position and converges to a systematic offset. For the individual subjects this effect varies and due to the different error spread that captures samples covering the entire breathing spectrum, the error is at least partially caused by the respiration state. A detailed analysis of this problem is provided by Table 7.6 that lists the sequence-wise P2T-based mean TREs and the correlation coefficients of the TREs to a set of selected metrics that were calculated in the preceding experiment concerned with continuous respiratory motion analysis, i. e. these metrics are not biased regarding the transformation estimation but are solely based on respiration. As can be seen, there is no metric that consistently correlates with the TRE. Whereas for subject S_1 and the thoracic sequence the TREs highly correlate with the M2S distance \mathcal{E}^{M2S} , i. e. the model is not fully capable to describe the unseen respiration states, there is no correlation for subject S_4 and the thoracic sequence that exhibits a comparable TRE level. Similarly, the TRE does not consistently correlate with the respiration magnitude expressed as the joint signal ζ_J^{M} or the condition \mathcal{K}_{RE} of the body surface displacement field computed from the estimated model parameter \hat{b} .

For larger displacements T_3^{MA} and T_4^{MA} , the registration procedure frequently converges to an invalid position which results in both an increased median error as

well as an increased error spread. In particular, the P2P alignment schemes suffer from this problem whereas the P2T scheme is consistently more robust w. r. t. such large initial displacements.

Regarding the body surface coverage, Fig. 7.9 shows that a higher coverage in general yields more accurate results. In particular, this holds true for the motion compensated approaches. However, this effect is not equally distinct for all subjects. The most extreme effect of a low body coverage can be observed for subject S_3 where the largest initial misalignment transformation T_4^{MA} could not be recovered for the robust P2T alignment scheme. For conventional rigid alignment the body coverage does not constitute the most dominant error source as the effect of non-matching shapes influences the target registration error.

7.6 Discussion and Conclusion

This chapter was concerned with the practical application of 4-D shape priors for respiratory motion analysis. In particular, a unifying point-set registration framework that enables the clinical applications of continuous respiration monitoring and motion compensated patient positioning was proposed.

As a fundamental component of the investigated registration methods, the RBC framework and PDA schemes for efficient corresponding point search were assessed. Though the RBC yields acceptable run-times in the motion analysis scenario investigated in this thesis, it is outperformed by the PDA scheme that exploits the projective geometry inherent to body surfaces that were reconstructed using the manifold ray-casting technique proposed in Chapter 5 of this thesis. This underlines the benefits of domain specific knowledge and data structures that are designed for both algorithmic as well as hardware acceleration. From these experiments the RBC must be considered as a generic database query framework for arbitrary dimensions and metrics rather than the optimal acceleration structure for high-performance corresponding point search for RI based point-set or surface registration problems. However, future research must be concerned with the performance of the RBC framework and PDA schemes when incorporating supplementary information such as photometric data or salient landmarks into 4-D shape motion models. Further, the RBC framework is a promising candidate for the registration of surfaces with arbitrary topology and that are not governed by projective properties.

For continuous respiratory motion monitoring, 4-D shape priors built upon the WVR principle in conjunction with the approximated CPD scheme using the P2T metric are very promising. The main features here are the fully automatic and non-invasive nature, high temporal resolution of 50 Hz that easily satisfies clinical run-time requirements of few hundred milliseconds, surface reconstruction errors in the scale of the RI sensor's noise level and the congruence with existing respiration surrogates or signals. For the latter, the experiments demonstrate a high correlation of ≥ 0.97 w. r. t. conventional RI-based external surrogates. The conclusion from this result is twofold. On the one hand, 4-D shape priors may supersede conventional methods that require manual interaction or that rely on heuristics. On the other hand, the investigated conventional methods are simple

yet able to generate respiration surrogates that are as informative as the complex model-based formulation. This is a direct result of the non-rigid surface registration methods that govern the shape motion models and that are designed to use geometric information exclusively. When using complementary information such as photometric data or anatomical landmarks to guide the registration process it is expected that there is a divergence between the surrogates obtained by 4-D shape priors and the conventional formulations. Regarding respiration signals that are not based on surface deformations, the experiments show that the proposed model-based surrogates show a high correlation with a signal obtained from an IP sensor. The results show that an RI-based respiratory motion analysis regarding change of lung volume requires surrogates based on distinct spatially localized surface displacements as well as their superposition. This is enabled by the proposed WVR model. However, which surrogate best correlates to the IP signal depends on the breathing pattern and the subject. This is no flaw in the model-based respiration analysis framework but instead indicates that the 1-D IP signal as a measure of lung volume change is not capable to describe the entire spectrum of surface deformations that is inherent to the model-based formulation. Ultimately, this leads to the fundamental question on the actual usage of the obtained respiration surrogates or signals. In radiation therapy, for simple beam gating, a 1-D surrogate that describes the lung volume may be sufficient whereas for more advance beam adjustment methods complex external deformations for predicting internal motion may be required. For both scenarios, continuous respiratory motion analysis based on 4-D shape priors is a promising direction. However, the fitness for clinical purposes must be evaluated in future research.

For patient positioning the results show that incorporating prior knowledge on respiration-induced body surface deformations into the registration process allows to substantially decrease the alignment error. Here, the experiments demonstrated an error reduction by a factor of about 3 compared to conventional alignment techniques that neglect respiratory motion. Though outperforming conventional systems, it must be investigated if a maximum TRE of about 4 mm obtained with motion compensated alignment is sufficient for clinical demands. This TRE, that is to some extent due to a systematic offset in the registration process, could not be traced back to a single cause in this thesis. Potential candidates including a shape model that is not fully able to adapt to unseen respiration states or ill-conditioned body surface deformations only partially account for this issue and are not consistent for all subjects and evaluation sequences. Future research must thus be concerned with a detailed error analysis including the investigation of partial surface matching issues, missing salient geometric features on the body surface, as well as noise and the general acquisition characteristics and suitability of the used RI sensors. In this regard, patient alignment using RI technology, both conventional and motion compensated, should be considered as a non-intrusive initial step for subsequent verification procedures that may involve ionizing radiation. In such a multi-step alignment, the radiation exposure of the patients may be reduced substantially.

Part IV

Outlook and Summary

Outlook

This chapter is concerned with future directions regarding both hardware and methodological aspects as well as challenges toward clinical translation.

Real-time Range Imaging

RI Devices In this thesis all experiments were conducted using an off-the-shelf Microsoft Kinect RI sensor that was originally developed for entertainment purposes and thus is not suitable for clinical applications. In fact, one crucial issue toward clinical translation of RI-based respiratory motion analysis is the availability of certified RI devices that feature real-time frame-rates, sub-millimeter accuracy and a high failure safety. Further, potential issues due to temperature drifts and systematic offsets as for example observed for ToF devices [Baue 13a], sub-surface scattering for various skin types and robustness w. r. t. ionizing radiation must be investigated. For a multi-camera setup as used in this thesis, a crucial issue is the availability of easy-to-use calibration systems.

Pre-processing and Multi-view Range Data Fusion One question that must be investigated in future research is whether a dedicated RI pre-processing pipeline is necessary at all. This is to be expected for simple respiration surrogates that analyze plain depth measurements or the most basic ICP surface registration methods as summarized in Section 7.2. However, the robust CPD-based point registration scheme investigated in Section 7.3 is explicitly designed to account for noise and outliers. Thus, pre-processing methods may be dispensable and run-times can be decreased substantially. An interesting question regarding both run-time and accuracy is whether a dedicated pre-processing pipeline followed by some basic analysis method, or one single advanced analysis technique that inherently accounts for noise and outliers is more efficient.

In a similar fashion, it must be investigated if a dedicated explicit surface reconstruction from the fused volumetric distance transform must be performed at all. For example, Placht et al. [Plac 12] and similarly Bauer et al. [Baue 12a] used a distance transform to accelerate the closest point search for rigid and non-rigid surface registration, respectively. However, this scheme cannot be transferred directly to the distance transforms as used in this thesis as they are only *approximative* and data is only available inside a small support region around the surface. Future research should therefore investigate the effect of increasing the support region or different fusion strategies that are of less approximative nature. Further, it must be

investigated if robust multi-link registration methods are possible with implicitly given surface data or if a different class of registration methods must be used.

Respiratory Motion Analysis using 4-D Shape Priors

The general question and main future research direction regarding the principle of 4-D shape priors for respiratory motion analysis is how they fit into clinical routine and how they can be integrated into established workflows. This applies for both the *pre-procedural* training phase as well as the *intra-procedural* analysis stage. In particular, the targeted level of accuracy and robustness must be investigated. In addition to these general issues, the following paragraphs outline future research directions and challenges that are particularly concerned with the methodology inherent to the proposed framework of 4-D shape priors and its applications.

Anatomical Plausible Deformation Fields One fundamental question for 4-D shape priors is how to compute the body surface deformations that govern the patient specific motion models. The non-rigid registration methods employed in this thesis, and similar with the majority of related publications, establish the surface match by optimizing an objective function that is rather abstract and simplifying from an anatomical point of view. Indeed, for the techniques investigated in this thesis, different displacement fields were obtained for the same torso deformation. Future research must be explicitly concerned with the question if anatomically correct deformation fields are required and how such constraints can be enforced. A promising approach would be to incorporate additional information based on photometric data as proposed by Bauer et al. [Bau12c] or salient geometric landmarks. For example, the mamillas as captured by conventional color cameras may constitute a strong clue for a correct match in the thoracic region whereas a surface feature descriptor that encodes the topography of the navel region may serve as an anchor point for the abdominal body part. Such landmarks could be used as hard constraints in a non-rigid registration formulation as for example proposed by Daum [Daum11].

In addition to the problem regarding the anatomical correctness of surface deformations, future research must also be concerned with the question whether surface deformations that are solely due to respiratory motion are sufficient or if additional deformation components caused by different patient postures must be incorporated in the motion models.

Model Generation Using Dimensionality Reduction A different aspect concerns the dimensionality reduction step and the derivation of the variation modes encoded by the proposed surface motion models. For sparse modes that are required for an analysis of distinct breathing patterns, one interesting approach would be to compute the factor rotations not by using the PCA modes derived from the displacement fields but instead by using PCA modes computed from auxiliary data. Such data may for example encode the local respiration magnitude or manually labeled information to enforce sparsity w. r. t. user defined regions of interest. This may also be interpreted as a semi-supervised extension to

the conventional factor rotation principle. However, dedicated formulations and optimization schemes to generate a sparse respiratory motion model may also be investigated. Here, a natural candidate is the sparse PCA proposed by Zou et al. [Zou06].

In this regard, future research may also investigate non-linear dimensionality reduction techniques to generate 4-D shape priors. Such methods only assumed a minor role in this thesis as they typically require a large number of training samples which is not given if they are extracted from tomographic planning data. Such a limited number of training samples is a hard constraint for applications where the respiration state has to be analyzed instantaneously in a single-shot fashion as for example pursued with motion compensated patient positioning proposed in this thesis. However, as a promising research direction for continuous motion monitoring, incremental learning techniques to extend the model on-the-fly should be investigated. Additionally, non-linear methods hold great potential if the surface displacement fields that form the training data contain deformations both due to respiratory motion as well as different patient postures.

Motion Compensated Positioning and Continuous Respiration Analysis The key component for the utilization of 4-D shape priors is the registration of the *pre-procedurally* trained motion models to *intra-procedurally* acquired RI data. As with the problem of anatomical plausible deformation fields in the training phase, future research must be concerned with the extension of the model registration scheme with complementary information such as photometric data or salient geometric landmarks. In particular, the corresponding point search might benefit from such an extension and the integration of landmark or feature matching terms holds great potential for an improved convergence rate, increased attraction range of the optimizer and more robust results in the absence of distinct geometric information as for example demonstrated in [Haas13a]. However, care must be taken to not sacrifice the run-time performance which is particularly important for complex geometric features.

Besides complementary information, an interesting future research direction is the integration of *problem-specific* prior knowledge in the optimization process. An example for this principle is continuous motion monitoring where the initial guess for the model parameter was always set to the zero vector, i.e. the mean shape. However, given the high frame-rates of contemporary RI sensors and the low run-times of the proposed framework, a smooth transition between successive respiration states can be safely assumed and the result from the previous analysis could serve as a more suitable initial guess. Further, formulating the problem of continuous motion analysis as a Markov chain would account for a series of preceding respiration states with the potential of increased robustness and smoothness of the obtained respiration surrogates. In a similar fashion, samples from continuous respiration states could be used to replace the single-shot motion compensated positioning problem as proposed in this thesis. The rationale here is that for a reclined patient the sought treatment table transform must be the same for all samples despite the respiration state. This could be implemented using majority voting, a weighted average or formulated in a probabilistic framework. Be-

sides such temporal prior knowledge the cost function for motion compensated positioning could give high priority to stationary model points. For example, this could be achieved by assigning each model point a weight that is inversely proportional to the corresponding deformation magnitude.

GPU Computing

As a general remark, GPU technology is evolving very quickly and several new possibilities have been recently enabled by new hardware architectures. In particular, the ever-increasing number of available streaming multiprocessors and compute cores allow for more parallel processing. This is promising for the algorithms and methods investigated in this work as they inherently feature a high degree of data parallelism which in turn allows for faster computation on future hardware generations. Further, the availability of vector-style half-precision floating point operations implemented by the latest GPU architectures is a very promising technique for iterative methods such as the investigated surface registration approaches. Here, the first iterations might be computed very efficiently in low-precision accuracy to provide an initial estimate for subsequent iterations that are performed in computationally more demanding single- or double-precision.

Summary

The analysis and management of respiratory motion is a crucial factor for a plurality of medical applications. This thesis investigated RI-based methods for real-time respiratory motion analysis. The scientific contributions of this thesis and an overview of the proposed methods were introduced in Chapter 1. In particular, this thesis investigated efficient methods for RI data pre-processing and multi-view sensor fusion as well as machine learning techniques and dedicated surface registration schemes for model-based real-time respiratory motion analysis. A particular focus of this thesis was on GPU architectures to accelerate computation steps and ultimately enable real-time capability.

Introduction and Background Part I of this thesis was concerned with the general motivation of this thesis. In particular, the medical background for respiratory motion in therapeutic and diagnostic applications, as well as a literature survey of existing respiration management systems that already found their way in clinical routine, were given in Chapter 2. This chapter also introduced the paradigm of using RI technologies for respiratory motion analysis and provided a background on the working principles of contemporary RI sensors that are used both in daily routine and for the experiments in this thesis. Based upon the general requirements of using RI systems for respiratory motion analysis, this chapter covered the paradigm of using GPU architectures to accelerate computing steps. In particular, it was shown that the high degree of data level parallelism that is inherent to many computation steps in RI-based respiratory motion analysis perfectly fits the GPU programming models and hardware features.

Real-time Range Imaging The focus of Part II of this work was on efficient concepts for real-time RI processing as the prerequisites for respiratory motion management. The acquisition geometry in RI, the basic mathematical notation and nomenclature as well as the software framework that was used throughout this thesis were introduced in Chapter 3.

Chapter 4 was concerned with pre-processing techniques to address the low SNR of range data and typical artifacts that are often observed with RI sensors such as ToF cameras or the Microsoft Kinect device. In particular, missing data in defect regions that often occur at shiny surfaces, as well as temporal and spatial noise that are due to the physical working principles of RI sensors were identified as the most important error sources. Based on these observations, a pre-processing

pipeline dedicated to RI data enhancement and high-performance image processing on GPUs was developed. Specifically, the pipeline comprised the concepts of normalized convolution for restoring missing data and bilateral filtering principles for both temporal as well as edge-preserving noise reduction. The results in a highly dynamic scenario using synthetic range data computed from the well-known NCAT phantom showed that this pipeline can reconstruct the ideal range data with an error scale of <1.0 mm and a run-time of approximately 2.0 ms. An important result from the experiments is that temporal denoising methods must be used with care in a respiratory motion monitoring scenario. This is due to the fact that such methods essentially correspond to an averaging or superposition of respiration phases which is an error source especially during inhale and exhale, respectively. An interesting result of this chapter was concerned with the run-time performance of edge-preserving filtering. The results demonstrated that conventional bilateral filtering, depending on the kernel size, often outperforms modern constant time methods such as the concept of guided filtering. This was shown to be due to the complex infrastructure required for the guided filtering that hinders and efficient usage of the GPU computing resources.

Besides pre-processing methods for data enhancement, the fusion of range data obtained by multiple sensors is an important prerequisite for further surface data analysis. For this purpose, a novel framework that enables the reconstruction of a high coverage body surface model was proposed in Chapter 5 of this work. The core idea of this approach was to represent the individual RI surface data implicitly using SDFs and to reconstruct an explicit representation using ray-casting techniques. It was shown that conventional ray-casting methods based on a pinhole camera model are not capable to reconstruct a high coverage body surface model that includes the frontal and lateral body surface. This is due to the divergent viewing rays that are inherent to a pinhole camera model. To cope with this issue this thesis proposed an approach where the individual viewing rays do not originate from one single optical center but instead from multiple points on a 2-D manifold that surrounds the patient. This enabled the reconstruction of body surfaces with 180° coverage and a 2-D topology as known from real-life RI sensors. It was shown that this inherent 2-D topology allows to apply dedicated RI data enhancement techniques to increase the SNR as well as the employment of high performance methods for surface neighborhood analysis. The experiments on synthetic data using the NCAT phantom showed that the proposed approach is capable to reconstruct the ideal surface with an error in the scale of approximately 0.2 mm given run-times of less than 5.0 ms. For real-life RI data from male subjects, a qualitative analysis showed that the reconstructed surfaces exhibit distinct anatomical landmarks.

Respiratory Motion Analysis Using 4-D Shape Priors Part III of this thesis was concerned with the principle of using 4-D shape priors for respiratory motion analysis. Building upon this concept, a unifying framework for the clinical problems of motion compensated patient positioning and continuous respiratory motion analysis was developed.

The foundations and theoretical background of 4-D surface motion models were covered in Chapter 6. This thesis proposed to non-rigidly register patient specific body surfaces acquired at discrete respiration states in conjunction with dimensionality reduction techniques to train continuous surface motion models. An important aspect of non-rigidly registered body surfaces is the fitness of the corresponding displacement fields for respiratory motion analysis. This thesis investigated different non-rigid surface registration techniques and developed condition metrics that allow to quantitatively assess displacement fields regarding their stability to differentiate between shifts and rotations caused by global patient movement on the one hand and respiration induced deformations on the other hand. The experimental results for this issue show that displacement fields covering the entire body surface exhibit a better condition compared to displacement fields of small body surface regions. This is a strong argument for using high coverage body surface data as for example proposed in Chapter 5 of this thesis.

Regarding dimensionality techniques for training motion models, one key finding of this thesis is that linear dimensionality reduction techniques such as PCA are sufficient to capture the entire range of breathing patterns with only two modes of variation. In contrast to typical non-linear dimensionality reduction techniques, a small number of training surfaces are thus sufficient for model generation. This is of particular importance if the training samples are limited as for example with surfaces extracted from tomographic planning data acquired by CT or MRI. However, it was shown that conventional PCA-based models exhibit abstract and global deformation modes that hinder an intuitive manual interpretation as well as the automatic derivation of anatomical plausible respiration surrogates. To cope with this issue, this thesis proposed a framework based on factor rotations that allows to transform the conventional PCA models to so-called WVR models that enable an unsupervised decomposition of body surface displacement fields into a thoracic and an abdominal component.

The actual application of 4-D motion models for respiratory motion analysis was covered in Chapter 7 of this work. In particular, this chapter introduced a framework that unifies the problems of motion compensated patient positioning and continuous respiratory motion analysis. The governing idea here is that these tasks can be formulated mathematically as the registration of pre-procedurally trained 4-D motion models to intra-procedurally acquired RI surface data. For this purpose, a dedicated registration scheme based on the CPD principle was developed. This method formulates the model registration task as a density estimation problem and can essentially be interpreted as a multi-link ICP variant that is robust to outliers that are common with RI.

This thesis performed a thorough investigation of the methodology that governs the CPD method. First, for the point correspondence problem being an integral component of ICP-like methods, this work investigated the RBC and PDA schemes for fast nearest neighbor queries. In contrast to the RBC being a generic space partitioning strategy designed for GPU architectures, the PDA scheme exploits the 2-D topology and projective properties that are inherent to the high coverage RI surface data as proposed in Chapter 5 of this thesis. For a typical corresponding points query using a 25 point neighborhood, the PDA scheme exhibits a

run-time of about 0.15 ms and outperforms the RBC by a factor of approximately 46. As a further essential component of ICP-like methods, this work investigated the P2P and the P2T criterion to quantify the alignment match of the motion model and the RI surface. It was shown that the iterative CPD scheme using the P2T metric consistently requires less than half of the number of iterations compared to using the P2P metric. Moreover, the results demonstrated that a P2T metric is superior to the P2P metric in terms of misalignment and initial guess for the sought registration parameters. Regarding the registration accuracy for the P2T metric, the experiments demonstrated that the second quartile of the M2S to quantify the closeness of the model to the surface is consistently less than 0.5 mm. Further, using range data that was corrupted by more than 5.0 mm for 25% of the measurements, it was shown that the CPD-like registration scheme is robust to outliers.

For continuous respiratory motion analysis, this thesis proposed anatomical plausible respiration surrogates that are derived from the model's inherent parameters. The governing idea here is that 4-D shape priors essentially encode body surface displacement fields that may be interpreted as a surrogate that reflects the body surface extent. For the proposed WVR models, in contrast to standard PCA models, it was shown the proposed surrogates yield a statistically significant PCC of more than 0.97 compared to conventional RI-based surrogates that measure depth variations in manually selected thoracic or abdominal regions. Further, the experiments conducted in this thesis show that the model-based surrogates highly correlate with a respiration signal obtained from an IP sensor that measures the change of lung volume. Besides the application of 4-D shape priors for continuous respiratory motion monitoring, this thesis also showed that incorporating prior knowledge on respiration-induced body surface deformations substantially decreases the alignment error for patient positioning. This motion compensated positioning strategy reduces the error by a factor of about 3 compared to conventional alignment schemes that do not account for respiratory motion.

Outlook and Summary The thesis is concluded in Part IV with an outlook and the summary at hand. In particular, future research directions regarding both hardware and methodological aspects were discussed and challenges towards clinical translation of the proposed methods investigated.

Part V
Addendum

A GPU Programming Paradigms: Case Study Confidence Maps

The general principles and issues of GPU programming can best be understood by considering a specific problem. For this purpose, the problem of associating a reliability or confidence indicator to range measurements is investigated. This problem exhibits the most important aspects that are of relevance for GPU programming in the RI based respiratory motion analysis problem investigated in this thesis.

One specific confidence indicator \mathcal{C} for range data is based on the angle between the surface normal \mathcal{N} at a certain 3-D point and the direction from which this point was observed, see Section 3.2.2 for details and other reliability indicators. Neglecting the data representation details and RI sampling principles that are covered in Section 3.2, all data in range imaging is organized on a 2-D regular grid with $N_1 \times N_2$ indices. 3-D surface points \mathcal{S} can be computed from the 1-D range measurements \mathcal{R} as obtained from the RI sensor using the corresponding observation direction. This observation or viewing direction is data independent and uniquely defined on each grid element by the intrinsic parameters of the RI sensor. In contrast, the surface normals \mathcal{N} must be computed on a per-frame basis from the instantaneous 3-D surface data. Surface normals are calculated by using the cross product of partial derivatives that, by taking advantage of the regular grid layout, can be calculated using central, backward or forward differences of neighboring surface points in \mathcal{S} . The problem of computing confidence values exhibits a high degree of data parallelism as the same instructions are issued for all data points. Further, the computation has typically to be performed for several hundreds thousands of elements and thus perfectly fits the massive parallel computation paradigm of GPGPU computing. However, care has to be taken w. r. t. the hierarchical memory model in order to achieve maximum performance.

A naïve approach would pre-compute the data-independent acquisition directions and re-use them (1) for computing the 3-D surface points from the 1-D range measurements and (2) to compute the final confidence measure. This approach is outlined in Algorithm 1. Though re-using the pre-computed viewing directions reduces the number of arithmetic operations this approach lacks an efficient memory management as the data must be stored in off-chip global memory. To make things worse, surface normals computation using finite differences as outlined in Line 11 requires the neighboring surface points at positions $i_1 \pm 1$ and $i_2 \pm 1$ where ± 1 depends on whether forward, central or backward differences are used. The naïve approach would load this data from global memory for each index pair (i_1, i_2) independently, despite the fact that this data can be re-used across neighboring in-

Algorithm 1 Naive parallel confidence computation. Data $(\mathcal{V}, \mathcal{S})$ is re-used for the purpose of less arithmetic operations at the cost of excessive access to global memory (G) for both scalar valued as well as three-tuple data.

```

1: Input: Range measurements  $\mathcal{R} \in \mathbb{R}^{N_1 \times N_2}$  ▷ G
2: Input: Pre-computed 3-D viewing directions  $\mathcal{V} \in \mathbb{R}^{N_1 \times N_2 \times 3}$  ▷ G
3: Output: Confidence map  $\mathcal{C} \in \mathbb{R}^{N_1 \times N_2}$  ▷ G
4: for  $i_1 = 0 \rightarrow N_1 - 1$  do in parallel
5:   for  $i_2 = 0 \rightarrow N_2 - 1$  do in parallel
6:      $\mathcal{S}(i_1, i_2) = \text{CalcCoord}(\mathcal{R}(i_1, i_2), \mathcal{V}(i_1, i_2))$  ▷ G → G
7:   end for
8: end for
9: for  $i_1 = 0 \rightarrow N_1 - 1$  do in parallel
10:  for  $i_2 = 0 \rightarrow N_2 - 1$  do in parallel
11:     $\mathcal{N}(i_1, i_2) = \text{CalcNormal}(\mathcal{S}(i_1, i_2), \mathcal{S}(i_1 \pm 1, i_2 \pm 1))$  ▷ G → G
12:     $\mathcal{C}(i_1, i_2) = \text{CalcConfidence}(\mathcal{N}(i_1, i_2), \mathcal{V}(i_1, i_2))$  ▷ G → G
13:  end for
14: end for

```

Algorithm 2 Block-wise parallel confidence computation with a fixed block size of $B_1 \times B_2$ elements. The method uses a dedicated sub-routine `ComputeConfidence` listed in Algorithm 3 that performs on-the-fly calculation of intermediate results for solely scalar valued global memory (G) access. Static sensor parameters are kept in constant memory (C).

```

1: Input: Range measurements  $\mathcal{R} \in \mathbb{R}^{N_1 \times N_2}$  ▷ G
2: Input: Static camera parameters  $\mathbf{p}$  ▷ C
3: Output: Confidence map  $\mathcal{C} \in \mathbb{R}^{N_1 \times N_2}$  ▷ G
4: for  $i'_1 = 0 \rightarrow \lceil N_1/B_1 \rceil$  do in parallel
5:   for  $i'_2 = 0 \rightarrow \lceil N_2/B_2 \rceil$  do in parallel
6:      $\mathbf{B} = [i'_1 B_1, \dots, (i'_1 + 1)B_1] \times [i'_2 B_2, \dots, (i'_2 + 1)B_2] \in \mathbb{N}^{B_1 \times B_2}$ 
7:      $\mathcal{C}(\mathbf{B}) = \text{ComputeConfidence}(\mathcal{R}, \mathbf{p}, \mathbf{B})$  ▷ G → G
8:   end for
9: end for

```

dices. This follows directly from the definition of finite differences. In fact, though the problem's inherent degree of parallelism is optimally exploited, such an approach is not efficient in practice as most of the time is spent in loading data from and writing data to global memory.

An approach that takes advantage of the GPU architecture now comprises the following two concepts: first, computations are performed multi-threaded but block-wise and the data is shared across a block and re-used by the individual threads. Second, intermediate results, i. e. the pixel-wise 3-D observation directions, surface points and normals, are not stored persistently in global memory but kept in volatile memory, i. e. registers or shared memory. In fact, the input as well as the output of the confidence computation routine is a scalar valued func-

Algorithm 3 Block-wise parallel confidence computation with a fixed block size of $B_1 \times B_2$ elements. The block is processed concurrently in i_1 direction but sequentially along the i_2 direction. The algorithm exhibits few access to global memory (G), re-uses data across neighboring threads using shared memory (S) and keeps intermediate results in local registers (R). Note that each *global* index $i_{(1,2)}$ has a unique *local* identifier $t_{(1,2)} \in 0, \dots, B_{(1,2)} - 1$ inside the block, cf. Fig. 2.5.

```

1: Input: Range measurements  $\mathcal{R} \in \mathbb{R}^{N_1 \times N_2}$                                 ▷ G
2: Input: Static camera parameters  $\mathbf{p}$                                        ▷ C
3: Input: Block configuration  $\mathbf{B} \in \mathbb{N}^{B_1 \times B_2}$ 
4: Output: Confidence map  $\mathcal{C} \in \mathbb{R}^{N_1 \times N_2}$                                 ▷ G
5: for all  $i_1 \in \mathbf{B}$  do in parallel
6:    $S[0][t_1] = \mathcal{R}(i_1, \min(i_2 \in \mathbf{B}))$                                 ▷ G → S
7:    $S[1][t_1] = \mathcal{R}(i_1, \min(i_2 \in \mathbf{B}) + 1)$                             ▷ G → S
8:   for all  $i_2 \in \mathbf{B}$  do
9:      $[v_1, \dots, v_4] = \text{CalcViewDirs}(\mathbf{p}, i_1 \pm 1, i_2 \pm 1)$             ▷ C → R
10:     $[c_1, \dots, c_4] = \text{CalcCoords}(S[0, 1][t_1 \pm 1], [v_1, \dots, v_4])$     ▷ S → R
11:     $\mathbf{n} = \text{CalcNormal}([c_1, \dots, c_4])$                                     ▷ R → R
12:     $\mathbf{v} = \text{CalcViewDir}(i_1, i_2)$                                           ▷ C → R
13:     $\mathcal{C}(i_1, i_2) = \text{CalcConfidence}(\mathbf{v}, \mathbf{n})$                             ▷ R → G
14:     $S[0][t_1] = S[1][i_1]$                                                   ▷ S → S
15:     $S[1][t_1] = \mathcal{R}(i_1, i_2 + 2)$                                        ▷ G → S
16:   end for all
17: end for

```

tion and redundant computations are potentially cheaper than access to global memory [NVIDIA13].

This GPU friendly approach is outlined in Algorithm 2 and for block-wise computation uses a dedicated sub-routine `ComputeConfidence` that is listed in Algorithm 3. Compared to the naive approach, this version exhibits a substantially reduced global memory access which is due to the facts that intermediate results are calculated on-the-fly and data is shared across neighboring threads as listed in Line 10. However, it is worth noticing that not the 3-D coordinates itself are shared but only the corresponding 1-D range measurements. This entails an overhead w. r. t. arithmetic operations that in fact are redundant but comes with only one third of shared memory requirements for both size and bandwidth. As outlined in Section 2.3.1 resource requirements can influence the number of tasks that are executed in parallel and, though featuring a high bandwidth, shared memory is rather slow compared to registers. Resource saving is also achieved by using a mixed parallel and sequential processing approach as listed in Lines 5 and 8. This results in memory requirements that solely depend on the size of the block in the i_1 direction with full data sharing along the i_2 dimension, see also the data fetching from global memory in Lines 14 and 15. Such a mixed parallel and sequential approach that is often referred to as *cascading* or *persistent blocks* is commonly used in GPU computing and has proven to increase performance [Mart12].

The importance of a proper design w. r. t. the GPU memory model is underlined by a run-time analysis of the naive and the optimized approach. For an image resolution of 640×480 pixels and using a manually determined optimal block configuration on a GTX 680 GPU, the naive approach has a run-time of 0.6 ms whereas the optimized version features a superior run-time of only 0.06 ms, for details on the test system and profiling principle see Appendix B. Though the absolute values are low for both approaches and the difference is of no practical relevance when considering the confidence computation as a standalone module, a speedup of a factor of 10 is of high relevance for filter pipelines in a multi-sensor setup or complex iterative algorithms that require several hundreds of steps. For respiratory motion analysis this also means that latency is reduced, thus allowing for an instantaneous system response with minimal latencies.

B GPU Performance Assessment Setup

Performance studies in this thesis are conducted on an off-the-shelf Windows 7 desktop PC equipped with an Intel Core i7 3770K CPU and an NVIDIA GTX 680 GPU (CUDA version 5.0) operating in *Windows Display Driver Model* (WDDM) mode, i. e. the GPU is not dedicated to general purpose computing only and thus is subject to scheduling on the driver side. Further, both the employed CPU as well as the GPU have a so-called *boost* mode that, permitting among others thermal conditions, automatically overclocks the device which potentially results in fluctuations of measured run-times. To account for these issues, the reported run-times are averaged over several runs.

C Range Measurement Representations

This section provides a detailed description of the two different types of range measurement representation as introduced in Section 3.2.

Radial Depth Measurements From the definition of camera coordinate systems in Eq. (3.4) and the projection operator from Eq. (3.7) it follows that the relationship between a world space coordinate \mathbf{x}_W and the corresponding *radial* depth measurement $\mathcal{R}_k^<(i)$ at position $i = P_k(\mathbf{x}_W)$ for the k -th camera can be expressed as:

$$\mathcal{R}_k^<(P_k(\mathbf{x}_W)) = \left\| \mathbf{T}_k \begin{pmatrix} \mathbf{x}_W \\ 1 \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \right\|_2 = \|\mathbf{x}_k\|_2. \quad (9.1)$$

Here, the zero vector $\mathbf{0} \in \mathbb{R}^3$ encodes the *local* camera origin \mathbf{o}_k that is independent from the actual pixel index i . Further, $\|\cdot\|_2$ denotes the L^2 -norm. Of course, this model neither accounts for occlusions nor reflects the actual range sampling principle of modern RI sensors. Yet, it provides the basis for the reconstruction of a 3-D points given a set of depth measurements in accordance with the generic 3-D reconstruction model from Eq. (3.13) as:

$$\mathcal{S}_k(i) = \mathbf{x}_{k,i} = \mathcal{R}_k^<(i) \cdot \mathbf{v}_{k,i}^<. \quad (9.2)$$

The viewing rays $\mathbf{v}_{k,i}^<$ describe the *inverse* projection, i. e. all points $\mathbf{x}_{k,i}$ that are perspectively projected onto the same pixel index \mathbf{i} . Given a point \mathbf{x}_k defined in the k -th camera space, it follows from the projection in Eq. (3.5) that:

$$\mathbf{x}_k = \mathbf{C}_k^{-1} \tilde{\mathbf{i}} = \lambda \mathbf{C}_k^{-1} \begin{pmatrix} \mathbf{i} \\ 1 \end{pmatrix}, \lambda \in \mathbb{R}. \quad (9.3)$$

Thus, all points lying on the line defined by $\lambda \mathbf{C}_k^{-1} (\mathbf{i}^\top, 1)^\top$ are projected to the same pixel index and the viewing ray $\mathbf{v}_{k,i}^<$ at pixel index \mathbf{i} is consequently given as:

$$\mathbf{v}_{k,i}^< = \frac{\mathbf{v}_{k,i}^<}{\|\mathbf{v}_{k,i}^<\|_2}, \mathbf{v}_{k,i}^< = \mathbf{C}_k^{-1} \begin{pmatrix} \mathbf{i} \\ 1 \end{pmatrix}. \quad (9.4)$$

From the definition of the *camera matrix* in Eq. (3.6) it follows directly that its inverse \mathbf{C}_k^{-1} is given by a closed form solution:

$$\mathbf{C}_k^{-1} = \begin{pmatrix} \frac{1}{f_{k,1}} & 0 & -\frac{1}{f_{k,1}} c_{k,1} \\ 0 & \frac{1}{f_{k,2}} & -\frac{1}{f_{k,2}} c_{k,2} \\ 0 & 0 & 1 \end{pmatrix}. \quad (9.5)$$

Thus, the *radial* viewing rays $\mathbf{v}_{k,i}^<$ as described in Eq. (9.4) can be simplified in a straight forward manner to:

$$\mathbf{v}_{k,i}^< = \frac{\mathbf{v}_{k,i}^<}{\|\mathbf{v}_{k,i}^<\|_2}, \mathbf{v}_{k,i}^< = \left(\frac{i_1 - c_{k,1}}{f_{k,1}}, \frac{i_2 - c_{k,2}}{f_{k,2}}, 1 \right)^\top. \quad (9.6)$$

Orthogonal Depth Measurements Compared to *radial* depth measurements *orthogonal* depth measures are an alternative yet equivalent depth representation defined as:

$$\mathcal{R}_k^\perp (P_k(\mathbf{x}_W)) = (0, 0, 1)^\top \left(\mathbf{T}_k \begin{pmatrix} \mathbf{x}_W \\ 1 \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \right) = (0, 0, 1)^\top \mathbf{x}_k = x_{k,3}. \quad (9.7)$$

Given the fact that both representations are equivalent and thus must reconstruct the same 3-D point, i. e. :

$$\mathcal{S}_k(\mathbf{i}) = \mathcal{R}_k^<(\mathbf{i}) \cdot \mathbf{v}_{k,i}^< = \mathcal{R}_k^\perp(\mathbf{i}) \cdot \mathbf{v}_{k,i}^\perp, \quad (9.8)$$

it follows from the definition of *radial* range values in Eq. (9.1) and *orthogonal* depth measurements in Eq. (9.7) that the following relationship holds true:

$$\mathbf{v}_{k,i}^\perp = \left(\frac{x_{k,1}^2}{x_{k,3}^2} + \frac{x_{k,2}^2}{x_{k,3}^2} + 1 \right)^{\frac{1}{2}} \cdot \mathbf{v}_{k,i}^<. \quad (9.9)$$

Using the perspective projection model from Eq. (3.5) that describes the relationship between \mathbf{x}_k and the pixel index \mathbf{i} , this translates to:

$$\mathbf{v}_{k,i}^\perp = \left(\left(\frac{i_1 - c_{k,1}}{f_{k,1}} \right)^2 + \left(\frac{i_2 - c_{k,2}}{f_{k,2}} \right)^2 + 1 \right)^{\frac{1}{2}} \cdot \mathbf{v}_{k,i}^<. \quad (9.10)$$

This is the non-normalized viewing ray derived in Eq. (9.6).

D Data Specification for Respiratory Motion Analysis

This section provides the detailed specification of the test data used for the respiratory motion analysis experiments in Chapter 6 and Chapter 7.

Table D1: Training and evaluation data specification for subjects S_1 – S_4 . Listed are the number of frames in thoracic (T), abdominal (A) and regular (R) breathing sequences. For the training stage, the normalized cumulative variances $\lambda'_i = \sum_{s=1}^i \lambda_s / \sum_{s=1}^3 \lambda_s$ are additionally reported for PCA and WVR models.

	Training								Testing		
	Frames		λ'_1		λ'_2		λ'_3		Frames		
	T	A	WVR	PCA	WVR	PCA	WVR	PCA	T	A	R
S_1	6	5	0.76	0.79	0.98	0.98	0.99	0.99	167	112	272
S_2	6	6	0.72	0.76	0.99	0.99	0.99	0.99	118	122	324
S_3	5	5	0.53	0.67	0.98	0.98	0.99	0.99	235	156	354
S_4	7	6	0.84	0.85	0.96	0.96	0.99	0.99	89	168	328

Table D2: Data for the IP sensor experiment. Shown are the number of frames in the thoracic (T) and abdominal (A) training sequences and the resulting WVR model variances λ'_i (cf. Table 6.3). For the testing stage, the number of frames in phases of abdominal (P_1), thoracic (P_2), fast shallow (P_3), fast strong (P_4), slow shallow (P_5), slow strong (P_6) and breath hold (P_7) breathing instructions are listed.

	Training					Testing						
	Frames		WVR			Frames ($ P_i \cdot 10^3$)						
	T	A	λ'_1	λ'_2	λ'_3	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇
S_5	7	6	0.64	0.93	0.97	1.6	1.5	0.9	0.9	1.5	1.6	3.4
S_6	7	6	0.87	0.97	0.99	1.6	1.6	0.8	0.8	1.5	1.7	3.4
S_7	7	6	0.50	0.94	0.97	1.5	1.6	0.9	0.8	1.6	1.5	3.3

List of Symbols

General Symbols

$T \equiv (R, t)$	3-D Transformation matrix . . . 19, 20, 21, 22, 56, 57, 59, 127, 128, 129, 130, 131, 132, 150, 151
$R \in \text{SO}(N)$	N-D Rotation Matrix . . . 19, 20, 57, 82, 86, 87, 105, 106, 107, 108, 109, 110, 111, 127, 128, 153
$t \in \mathbb{R}^N$	N-D Translation Vector . . . 19, 20, 57, 82, 105, 106, 107, 108, 109, 110, 111, 127, 128, 153
$x \in \mathbb{R}^3$	3-D Point . . . 20, 21, 22, 56, 57, 58, 59, 60, 61, 79, 82, 83, 84, 105, 107, 108, 109, 110, 111, 112, 113, 114, 119, 120, 121, 128, 150, 151, 154
$I \in \mathbb{R}^{N \times N}$	N-D Identity Matrix 20
$n \in \mathbb{R}^3, \ n\ _2 = 1$	3-D Normal 23, 108, 111
\mathcal{J}	Generic cost function 36, 109, 110, 111, 119

Range Imaging and Surface Representations

$\Omega \subset \mathbb{R}^2$	RI domain . . . 20, 21, 22, 23, 24, 30, 31, 32, 34, 35, 36, 37, 38, 41, 42, 43, 50, 56, 57, 61, 62, 63, 72, 82, 114, 115, 119, 120, 121, 153
$i \in \Omega$	Pixel index . . . 20, 21, 22, 23, 24, 30, 31, 32, 34, 35, 36, 37, 38, 39, 41, 42, 46, 56, 58, 61, 62, 82, 83, 114, 120, 121, 147, 148, 149, 150, 151
$P : \mathbb{R}^3 \rightarrow \Omega$	Projection operator . . . 21, 22, 57, 59, 83, 114, 150, 151
$\mathcal{R} : \Omega \rightarrow \mathbb{R}^+$	RI range/depth measurements . . . 22, 23, 24, 30, 31, 32, 33, 35, 36, 38, 39, 41, 42, 57, 61, 62, 63, 67, 68, 69, 115, 120, 147, 148, 149, 150, 151
$\mathcal{S} : \Omega \rightarrow \Psi$	RI surface . . . 22, 23, 30, 42, 46, 56, 57, 58, 59, 60, 61, 62, 63, 82, 83, 105, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 147, 148, 150, 151
$\Psi \subset \mathbb{R}^3$	RI surface codomain 22, 58, 61, 153
$v \in \mathbb{R}^3$	Generic viewing ray for RIs . . . 22, 24, 61, 62, 150, 151
$\mathcal{N} : \Omega \rightarrow \mathbb{R}^3$	RI surface normals 23, 24, 108, 147, 148
$\mathcal{C} : \Omega \rightarrow [0, 1]$	RI confidence map . . . 23, 24, 25, 30, 41, 58, 59, 147, 148, 149
$\mathcal{M} : \Omega \rightarrow \{0, 1\}$	RI mask 24, 30, 31, 32, 33, 34

$\Gamma \subset \mathbb{R}^3$	Domain of an implicit surface representation . . . 56, 57, 58, 59, 60, 63, 64, 65, 68, 69, 71, 154
$\mathcal{D} : \Gamma \rightarrow [-1, +1]$	Signed distance function . . . 56, 58, 59, 60, 61, 68, 71, 83, 84
$\mathcal{W} : \Gamma \rightarrow [0, 1]$	Weights associated to a signed distance function . . . 56, 58, 59, 60, 69, 71, 72

Motion Models and Model Registration

$\mathcal{P} = \{x_n\}$	Point set . . . 79, 82, 84, 85, 105, 107, 108, 109, 110, 112, 113, 114, 115, 118, 120, 128, 129, 154
$\mathcal{U} = \{u_n\}$	Displacement field . . . 79, 81, 82, 83, 85, 91, 92, 94, 95, 154
$u \in \mathbb{R}^3$	Displacement vector 79, 82, 154
\mathcal{L}	Linearization operator 79, 81, 85, 108, 154
$v = \mathcal{L}(\mathcal{U})$	High dimensional representation of \mathcal{U} 81, 85
$p = \mathcal{L}(\mathcal{P})$	High dimensional representation of \mathcal{P} . . . 85, 86, 88, 89, 106, 108
$\Phi \in \mathbb{R}^{3N \times L}$	PCA basis 85, 86, 87, 106, 108, 110
$b \in \mathbb{R}^L$	Model parameter . . . 86, 89, 105, 106, 107, 108, 109, 110, 111, 112, 115, 119, 120, 121, 127, 128, 129, 131
$\Theta : (x_1, x_2) \rightarrow \mathbb{R}^+$	Generic distance between corresponding points . . . 105, 106, 107, 108, 110, 111
$\xi \in \mathbb{R}$	Respiration surrogate/signal . . . 106, 120, 121, 123, 124, 125, 131

List of Abbreviations

AAPM	American Association Of Physicists In Medicine	9, 10, 16
AP	Anterior-posterior	63, 65
BF	Brute Force	112, 113, 115, 116, 117, 118
CC	CUDA Core	15, 16
CCD	Charge-coupled Device	21
CD	Cholesky Decomposition	108, 110, 111
CP	Corresponding Point	112
CPD	Coherent Point Drift	84, 91, 92, 93, 103, 108, 109, 110, 111, 112, 118, 119, 120, 121, 123, 132, 137, 143, 144
CPU	Central Processing Unit	14, 15, 26, 49, 103, 104, 150, 157
CT	Computed Tomography	3, 7, 9, 14, 21, 53, 78, 90, 101, 143
CTV	Clinical Target Volume	8
CUDA	Compute Unified Device Architecture	15, 16, 150, 157
DFT	Discrete Fourier Transform	32
DTR	Distance Transform Registration	91, 92, 93, 94, 95, 96
EM	Expectation-maximization	103, 109
FFT	Fast Fourier Transform	32
FLOPS	Floating Point Operations Per Second	14, 157
FPGA	Field-programmable Gate Array	26
GMM	Gaussian Mixture Model	84, 103, 104, 109
GPGPU	General Purpose Computing On GPUs	6, 147
GPU	Graphics Processing Unit	3, 4, 5, 7, 14, 15, 16, 26, 27, 28, 29, 34, 37, 49, 50, 52, 54, 55, 56, 72, 103, 104, 108, 112, 123, 140, 141, 142, 143, 147, 148, 149, 150, 157, 159
ICP	Iterative Closest Point	77, 102, 103, 104, 105, 107, 108, 112, 118, 129, 137, 143, 144
IGRT	Image Guided External Beam Radiation Therapy	8, 9, 10, 157
IP	Impedance Pneumography	118, 119, 120, 121, 124, 125, 126, 127, 133, 144, 152, 158, 159
KPCA	Kernel PCA	78, 88, 89, 157
LDA	Linear Discriminant Analysis	85
LEM	Laplacian Eigenmaps	78, 87, 88
LINAC	Linear Particle Accelerator	8, 21, 53, 128
LLE	Locally Linear Embedding	78, 87, 88
M2M	Model-to-model	120, 121, 122, 158
M2S	Model-to-surface	120, 121, 122, 131, 144, 158

ML	Medio-lateral	63
MRI	Magnetic Resonance Imaging	9, 78, 87, 90, 101, 143
NCAT	Nurbs-based Cardiac-torso	40, 41, 42, 63, 89, 90, 91, 93, 95, 96, 98, 142, 157
NN	Nearest Neighbor	112, 113, 115, 116
P2P	Point-to-point	106, 107, 108, 109, 110, 111, 123, 127, 129, 130, 131, 132, 144, 158
P2T	Point-to-tangent	106, 107, 108, 109, 111, 120, 121, 122, 123, 127, 128, 129, 130, 131, 132, 144, 158
PA	Procrustes Alignment	80, 81, 82
PCA	Principal Component Analysis	77, 78, 85, 86, 87, 88, 89, 96, 97, 98, 99, 100, 106, 123, 124, 138, 139, 143, 144, 152, 154, 157
PCC	Pearson Correlation Coefficient	44, 65, 95, 120, 123, 124, 126, 127, 144, 159
PDA	Projective Data Association	104, 114, 115, 116, 117, 118, 119, 132, 143, 158
RBC	Random Ball Cover	104, 112, 113, 114, 115, 116, 117, 118, 132, 143, 144, 158
RBF	Radial Basis Function	89
RCC	Spearman's Rank Correlation Coefficient	95, 96
RI	Range Imaging	3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 39, 40, 41, 42, 43, 46, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 65, 68, 69, 70, 71, 72, 77, 78, 80, 81, 82, 84, 86, 89, 90, 91, 93, 100, 101, 102, 104, 105, 106, 107, 109, 112, 114, 115, 116, 118, 119, 120, 121, 123, 124, 127, 132, 133, 137, 139, 141, 142, 143, 144, 147, 150, 153, 157, 158, 159
RITK	Range Imaging Toolkit	26
SDF	Signed Distance Function	55, 56, 57, 58, 59, 60, 64, 68, 71, 142, 157, 159
SI	Superior-inferior	9, 63
SM	Streaming Multiprocessor	15, 16, 140
SNR	Signal-to-noise Ratio	27, 28, 103, 141, 142
SSM	Statistical Shape Model	76, 77, 78
SVD	Singular Value Decomposition	108, 110
ToF	Time-of-Flight	13, 14, 23, 24, 27, 28, 29, 40, 72, 82, 103, 137, 141
TRE	Target Registration Error	128, 129, 130, 131, 133, 159
VGA	Video Graphics Array	25
VR	Varimax Rotation	86, 87, 96, 97, 98
WVR	Weighted Varimax Rotation	87, 89, 96, 97, 98, 99, 100, 118, 119, 123, 124, 125, 126, 132, 133, 143, 144, 152, 157

List of Figures

1.1	Framework overview	5
2.1	IGRT system	8
2.2	Working principles of RI sensors	11
2.3	Comparison of CPU and GPU FLOPS and memory bandwidth	14
2.4	<i>Kepler</i> GPU architecture	15
2.5	CUDA execution model	16
3.1	Multi-view RI acquisition geometry	20
3.2	Interpolation strategies	21
3.3	Confidence illustration	25
3.4	The Range Imaging Toolkit	26
4.1	Noise and artifacts in range imaging	28
4.2	Embedding regions for invalid measurements	34
4.3	Numerical issues with guided filtering	37
4.4	Virtual test environment using the NCAT phantom	40
4.5	Corrupted synthetic RI data	42
4.6	Comparison of normalized convolution and spectral deconvolution	44
4.7	Pre-processing results for edge-preserving denoising	45
4.8	Denoising comparison of bilateral and guided filtering	46
4.9	Temporal denoising	47
4.10	Quantization and spatial filtering	48
4.11	Run-time comparison of the bilateral and the guided filter	51
5.1	SDF embedding with multi-view RI	56
5.2	Surface reconstruction by ray casting	60
5.3	Respiration phantom reconstruction accuracy	64
5.4	Reconstruction error for different temporal integration parameters	66
5.5	Reconstruction error with noise and artifacts	67
5.6	Raw multi-view reconstruction results for real-life RI data	69
5.7	Post-processed multi-view reconstruction results for real-life RI data	70
6.1	Non-rigid vs. rigid body surface displacement fields	80
6.2	Kernel matrices with KPCA	88
6.3	The 4-D NCAT phantom	90
6.4	Body surface deformation fields	92
6.5	Condition of body surface deformation fields	94
6.6	Modes of variation for different linear techniques	97
6.7	Modes of variation for PCA and WVR	99

7.1	Motion model to RI data alignment	105
7.2	P2P and P2T alignment	107
7.3	RBC nearest neighbor search	113
7.4	PDA nearest neighbor search	114
7.5	RBC and PDA trends	116
7.6	M2M and M2S for continuous motion analysis	122
7.7	Qualitative visualization of an IP signal and model-based surrogates	125
7.8	Alignment assessment issues	128
7.9	Conventional and motion compensated patient positioning	130

List of Tables

4.1	GPU run-times for RI pre-processing	50
5.1	Reconstruction errors for different SDF discretizations	68
5.2	Run-times for the multi-view RI data fusion	71
6.1	Non-rigid surface registration statistics	93
6.2	Condition metrics to quantify surface deformation fields	95
6.3	Key figures for dimensionality reduction	96
7.1	Accuracy and run-times of corresponding point search strategies . . .	117
7.2	Run-time performance of model-based motion monitoring	123
7.3	PCCs for conventional and model-based respiration surrogates . . .	124
7.4	PCCs for different model-based respiration surrogates	126
7.5	Initial misalignments for motion compensated positioning	128
7.6	TREs for motion compensated alignment	131
D1	Data specification for subjects S_1 – S_4	152
D2	Data for the IP sensor experiment	152

Bibliography

- [Aach 01] T. Aach and V. Metzler. “Defect Interpolation in Digital Radiography - How Object-Oriented Transform Coding Helps”. In: *SPIE Medical Imaging*, pp. 4322:824–835, Jul 2001.
- [Aken 02] T. Akenine-Möller and E. Haines. *Real-Time Rendering*. A. K. Peters, Natrick, MA, USA, 2 Ed., 2002.
- [Ambe 07] B. Amberg, S. Romdhani, and T. Vetter. “Optimal Step Nonrigid ICP Algorithms for Surface Registration”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Jun 2007.
- [Aria 07] P. Arias, G. Randall, and G. Sapiro. “Connecting the Out-of-Sample and Pre-Image Problems in Kernel Methods”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Jun 2007.
- [Auri] V. Aurich and J. Weule. “Non-Linear Gaussian Filters Performing Edge Preserving Diffusion”. In: G. Sagerer, S. Posch, and F. Kummert, Eds., *German Association for Pattern Recognition Symposium (DAGM '95)*.
- [Bala 09] R. Balachandran and J. M. Fitzpatrick. “Iterative Solution for Rigid-Body Point-Based Registration with Anisotropic Weighting”. In: *SPIE Medical Imaging*, pp. 72613D–72613D–10, Mar 2009.
- [Baue 11] S. Bauer, J. Wasza, S. Haase, N. Marosi, and J. Hornegger. “Multi-modal Surface Registration for Markerless Initial Patient Setup in Radiation Therapy using Microsoft’s Kinect Sensor”. In: *IEEE International Conference on Computer Vision Workshops (ICCV Workshops '11)*, pp. 1175–1181, Nov 2011.
- [Baue 12a] S. Bauer, B. Berkels, S. Ettl, O. Arold, J. Hornegger, and M. Rumpf. “Marker-less Reconstruction of Dense 4-D Surface Motion Fields Using Active Laser Triangulation for Respiratory Motion Management”. In: N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '12)*, pp. 414–421, LNCS 7510, Part I, Oct 2012.
- [Baue 12b] S. Bauer, B. Berkels, J. Hornegger, and M. Rumpf. “Joint ToF Image Denoising and Registration with a CT Surface in Radiation Therapy”. In: A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein, Eds., *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM '12)*, pp. 98–109, LNCS 6667, Part I, May 2012.
- [Baue 12c] S. Bauer, J. Wasza, and J. Hornegger. “Photometric Estimation of 3-D Surface Motion Fields for Respiration Management”. In: T. Tolxdorff, T. M. Deserno, H. Handels, and H. P. Meinzer, Eds., *Bildverarbeitung für die Medizin (BVM '12)*, pp. 105–110, Mar 2012.

- [Baue 13a] S. Bauer, A. Seitel, H. Hofmann, T. Blum, J. Wasza, H. P. Balda, M. Meinzer, N. Navab, J. Hornegger, and L. Maier-Hein. *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, Chap. Real-Time Range Imaging in Health Care: A Survey, pp. 228–254. *LNCS 8200*, 2013.
- [Baue 13b] S. Bauer, J. Wasza, F. Lugauer, D. Neumann, and J. Hornegger. *Consumer Depth Cameras for Computer Vision - Research Topics and Applications*, Chap. Real-Time RGB-D Mapping and 3-D Modeling on the GPU Using the Random Ball Cover, pp. 27–48. *Advances in Computer Vision and Pattern Recognition*, 2013.
- [Bay 06] H. Bay, T. Tuytelaars, and L. Gool. “SURF: Speeded Up Robust Features”. In: A. Leonardis, H. Bischof, and P. A., Eds., *European Conference on Computer Vision (ECCV '06)*, pp. 404–417, *LNCS 3951*, Part I, May 2006.
- [Belk 03] M. Belkin and P. Niyogi. “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”. *Neural Computation*, Vol. 15, No. 6, pp. 1373–1396, Jun 2003.
- [Beng 03] Y. Bengio, J. F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. “Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering”. In: S. Thrun, L. K. Saul, and B. Schölkopf, Eds., *International Conference on Neural Information Processing Systems (NIPS '03)*, pp. 177–184, Dec 2003.
- [Benj 99] R. Benjemaa and F. Schmitt. “Fast Global Registration of 3D Sampled Surfaces using a Multi-z-buffer Technique”. *Image and Vision Computing*, Vol. 17, No. 2, pp. 113–123, Feb 1999.
- [Bern 15] K. Bernatowicz, P. Keall, P. Mishra, A. Knopf, A. Lomax, and J. Kipritidis. “Quantifying the impact of respiratory-gated 4D CT acquisition on thoracic image quality: A digital phantom study”. *Medical Physics*, Vol. 42, No. 1, pp. 324–334, Jan 2015.
- [Bert 05] C. Bert, K. G. Metheany, K. Doppke, and G. T. Chen. “A Phantom Evaluation of a Stereo-vision Surface Imaging System for Radiotherapy Patient Setup”. *Medical Physics*, Vol. 32, No. 9, pp. 2753–2762, Sep 2005.
- [Bert 11] C. Bert and M. Durante. “Motion in Radiotherapy: Particle Therapy”. *Physics in Medicine and Biology*, Vol. 56, No. 16, pp. R113–R144, Aug 2011.
- [Besl 92] P. J. Besl and N. D. McKay. “A Method for Registration of 3-D Shapes”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239–256, Feb 1992.
- [Blai 04] F. Blais. “Review of 20 Years of Range Sensor Development”. *Journal of Electronic Imaging*, Vol. 13, No. 1, pp. 231–243, Jan 2004.
- [Blai 95] G. Blais and M. D. Levine. “Registering Multiview Range Data to Create 3D Computer Objects”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 8, pp. 820–824, Aug 1995.

- [Blan 99] V. Blanz and T. Vetter. "A Morphable Model for the Synthesis of 3D Faces". In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*, pp. 187–194, Aug 1999.
- [Boyk 01] Y. Boykov, O. Veksler, and R. Zabih. "Fast Approximate Energy Minimization Via Graph Cuts". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 11, pp. 1222–1239, Nov 2001.
- [Brad 08] G. R. Bradski and A. Kaehler. *Learning OpenCV*. O'Reilly Media, Sebastopol, CA, USA, 1 Ed., 2008.
- [C 03] V. B. C., J. Ye, Y. Chen, and C. M. Leonard. "Image Registration via Level-set Motion: Applications to Atlas-based Segmentation". *Medical Image Analysis*, Vol. 7, No. 1, pp. 1–20, Mar 2003.
- [Camp 13] M. Camplani, T. Mantecon, and L. Salgado. "Depth-Color Fusion Strategy for 3-D Scene Modeling With Kinect". *IEEE Transactions on Cybernetics*, Vol. 43, No. 6, pp. 1560–1571, Dec 2013.
- [Cayt 10] L. Cayton. "A Nearest Neighbor Data Structure for Graphics Hardware". In: *International Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS '10)*, pp. 1–10, Sep 2010.
- [Cayt 12] L. Cayton. "Accelerating Nearest Neighbor Search on Manycore Systems". In: *IEEE International Parallel & Distributed Processing Symposium (IPDPS '12)*, pp. 402–413, May 2012.
- [Chap 06] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. The MIT Press, Cambridge, MA, USA, 1 Ed., 2006.
- [Chen 07] J. Chen, S. Paris, and F. Durand. "Real-time Edge-aware Image Processing with the Bilateral Grid". *ACM Transactions on Graphics*, Vol. 26, No. 3, Article No. 103, Jul 2007.
- [Chen 92] Y. Chen and G. Medioni. "Object Modeling by Registration of Multiple Range Images". *Image and Vision Computing*, Vol. 10, No. 3, pp. 145–155, Apr 1992.
- [Chun 16] H. Chung, P. R. Poulsen, P. J. Keall, S. Cho, and B. Cho. "Reconstruction of implanted marker trajectories from cone-beam CT projection images using interdimensional correlation modeling". *Medical Physics*, Vol. 43, No. 8Part1, pp. 4643–4654, Aug 2016.
- [Cizn 12] M. Ciznicki, M. Kierzynka, K. Kurowski, B. Ludwiczak, K. Napierala, and J. Palczynski. "Efficient Isosurface Extraction Using Marching Tetrahedra and Histogram Pyramids on Multiple GPUs". In: R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Wasniewski, Eds., *International Conference on Parallel Processing and Applied Mathematics (PPAM '12)*, pp. 343–352, LNCS 7204, Part II, Sep 2012.
- [Coot 95] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Active Shape Models - Their Training and Application". *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp. 38–59, Jan 1995.

- [Coot 99] T. F. Cootes and C. J. Taylor. "A Mixture Model for Representing Shape Variation". *Image and Vision Computing*, Vol. 17, No. 8, pp. 567–573, Jun 1999.
- [Corr 11] A. Corrigan, F. F. Camelli, R. Löhner, and J. Wallin. "Running Unstructured Grid-based CFD Solvers on Modern Graphics Hardware". *International Journal for Numerical Methods in Fluids*, Vol. 66, No. 2, pp. 221–229, May 2011.
- [Crow 84] F. C. Crow. "Summed-area Tables for Texture Mapping". In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '84)*, pp. 207–212, Jul 1984.
- [Curl 96] B. Curless and M. Levoy. "A Volumetric Method for Building Complex Models from Range Images". In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pp. 303–312, Aug 1996.
- [Daum 11] V. Daum. *Model-Constrained Non-Rigid Registration in Medicine*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2011.
- [Davi 02] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. "A Minimum Description Length Approach to Statistical Shape Modeling". *IEEE Transactions on Medical Imaging*, Vol. 21, No. 5, pp. 525–537, May 2002.
- [Deri 90] R. Deriche. "Fast Algorithms for Low-level Vision". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, pp. 78–87, Jan 1990.
- [Dora 98] C. Dorai, G. W., A. K. Jain, and C. Mercer. "Registration and Integration of Multiple Object Views for 3D Model Construction". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 83–89, Jan 1998.
- [Duda 00] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, New York, NY, USA, 2 Ed., 2000.
- [Edel 10] T. Edeler, K. Ohliger, S. Hussmann, and A. Mertins. "Time-of-Flight Depth Image Denoising Using Prior Noise Information". In: *IEEE International Conference on Signal Processing (ICSP '10)*, pp. 119–122, Oct 2010.
- [Ehrh 11] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels. "Statistical Modeling of 4D Respiratory Lung Motion Using Diffeomorphic Image Registration". *IEEE Transactions on Medical Imaging*, Vol. 30, No. 2, pp. 251–265, Feb 2011.
- [Evan 08] P. M. Evans. "Anatomical Imaging for Radiotherapy". *Physics in Medicine and Biology*, Vol. 53, No. 12, pp. R151–R191, May 2008.
- [Faya 09] H. Fayad, T. Pan, C. Roux, C. C. Le Rest, O. Pradier, J. F. Clement, and D. Visvikis. "A Patient Specific Respiratory Model based on 4D CT Data and a Time of Flight Camera (TOF)". In: *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS MIC '09)*, pp. 2594–2598, Oct 2009.

- [Faya 11] H. Fayad, T. Pan, J. F. Clement, and D. Visvikis. "Technical Note: Correlation of Respiratory Motion between External Patient Surface and Internal Anatomical Landmarks". *Medical Physics*, Vol. 38, No. 6, pp. 3157–3164, Jun 2011.
- [Fisc 14] P. Fischer, T. Pohl, and J. Hornegger. "Real-Time Respiratory Signal Extraction from X-Ray Sequences using Incremental Manifold Learning". In: *IEEE International Symposium on Biomedical Imaging (ISBI '14)*, pp. 915–918, Apr 2014.
- [Fitz 03] A. W. Fitzgibbon. "Robust Registration of 2D and 3D Point Sets". *Image and Vision Computing*, Vol. 21, No. 13, pp. 1145–1153, Dec 2003.
- [Fole 90] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, Boston, MA, USA, 2 Ed., 1990.
- [Fosk 05] M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi. "Large Deformation Three-dimensional Image Registration in Image-guided Radiation Therapy". *Physics in Medicine and Biology*, Vol. 50, No. 24, pp. 5869–5892, Dec 2005.
- [Fran 09] M. Frank, M. Plaue, and F. A. Hamprecht. "Denoising of Continuous-Wave Time-Of-Flight Depth Images using Confidence Measures". *Optical Engineering*, Vol. 48, No. 7, pp. 077003–077003–13, Jul 2009.
- [Fris 02] S. F. Frisken and R. N. Perry. "Efficient Estimation of 3D Euclidean Distance Fields from 2D Range Images". In: *IEEE Symposium on Volume Visualization and Graphics (VVS '02)*, pp. 81–88, Oct 2002.
- [Garc 08] V. Garcia, E. Debreuve, and M. Barlaud. "Fast k Nearest Neighbor Search Using GPU". In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR '08 Workshops)*, pp. 1–6, Jun 2008.
- [Getr 12] P. Getreuer. "Total Variation Inpainting using Split Bregman". *Image Processing On Line*, Vol. 2, pp. 147–157, Jul 2012.
- [Gran 02] S. Granger and X. Pennec. "Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration". In: A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., *European Conference on Computer Vision (ECCV '02)*, pp. 418–432, LNCS 2353, Part IV, Apr 2002.
- [Gree 91] L. Greengard and J. Strain. "The Fast Gauss Transform". *SIAM Journal on Scientific and Statistical Computing*, Vol. 12, No. 1, pp. 79–94, 1991.
- [Grim 15] R. Grimm, S. Fürst, M. Souvatzoglou, C. Forman, J. Hutter, I. Dregely, S. I. Ziegler, B. Kiefer, J. Hornegger, K. T. Block, and S. G. Nekolla. "Self-gated MRI Motion Modeling for Respiratory Motion Compensation in Integrated PET/MRI". *Medical Image Analysis*, Vol. 19, No. 1, pp. 110–120, Jan 2015.
- [Haas 13a] S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, A. Schneider, M. Kranzfelder, H. Feußner, and J. Hornegger. "3-D Operation Situs Reconstruction with Time-of-Flight Satellite Cameras Using Photogeometric Data Fusion". In: K. Mori, I. Sakuma, Y. Sato, C. Barillot, and

- N. Navab, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '13)*, pp. 356–363, LNCS 8149, Part I, Oct 2013.
- [Haas 13b] S. Haase, J. Wasza, T. Kilgus, and J. Hornegger. “Laparoscopic Instrument Localization using a 3-D Time-of-Flight/RGB Endoscope”. In: *IEEE Workshop on Applications of Computer Vision (WACV '13)*, pp. 449–454, Jan 2013.
- [Harr 07] M. Harris, S. Sengupta, and J. D. Owens. “Parallel Prefix Sum (Scan) with CUDA”. In: *GPU Gems 3*, pp. 851–876, Addison-Wesley, Boston, MA, USA, 2007.
- [Hart 04] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2 Ed., 2004.
- [Haus 11] G. Häusler and S. Ettl. “Limitations of Optical 3D Sensors”. In: *Optical Measurement of Surface Topography*, pp. 23–48, Springer, 2011.
- [Haus 88] G. Häusler and W. Heckel. “Light Sectioning with Large Depth and High Resolution”. *Applied Optics*, Vol. 27, No. 24, pp. 5165–5169, 1988.
- [He 10] K. He, J. Sun, and X. Tang. “Guided Image Filtering”. In: K. Daniilidis, P. Maragos, and N. Paragios, Eds., *European Conference on Computer Vision (ECCV '10)*, pp. 1–14, LNCS 6311, Part I, Sep 2010.
- [He 13] K. He, J. Sun, and X. Tang. “Guided Image Filtering”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 6, pp. 1397–1409, Jun 2013.
- [Heim 09] T. Heimann and H. P. Meinzer. “Statistical Shape Models for 3D Medical Image Segmentation: A Review”. *Medical Image Analysis*, Vol. 13, No. 4, pp. 543–563, Aug 2009.
- [Hilg 03] K. B. Hilger, R. Larsen, and M. C. Wrobel. “Growth Modeling of Human Mandibles Using Non-Euclidean Metrics”. *Medical Image Analysis*, Vol. 7, No. 4, pp. 425–433, Dec 2003.
- [Hof 03] H. Hof, K. K. Herfarth, M. Münter, A. Hoess, J. Motsch, M. Wannemacher, and J. Debus. “Stereotactic Single-dose Radiotherapy of Stage I non small cell Lung Cancer (NSCLC)”. *International Journal of Radiation Oncology, Biology, Physics*, Vol. 56, No. 2, pp. 335–341, Jun 2003.
- [Hone 11] P. Honeine and C. Richard. “Preimage Problem in Kernel-Based Machine Learning”. *IEEE Signal Processing Magazine*, Vol. 28, No. 2, pp. 77–88, Mar 2011.
- [Hopp 92] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. “Surface Reconstruction from Unorganized Points”. In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '92)*, pp. 71–78, Jul 1992.
- [Horn 87] B. K. P. Horn. “Closed-form Solution of Absolute Orientation Using Unit Quaternions”. *Journal of the Optical Society of America A*, Vol. 4, No. 4, pp. 629–642, 1987.

- [Huan 06] X. Huang, N. Paragios, and D. N. Metaxas. "Shape Registration in Implicit Spaces Using Information Theory and Free Form Deformations". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 8, pp. 1303–1318, Aug 2006.
- [Hugh 09] S. Hughes, J. McClelland, S. Tarte, D. Lawrence, S. Ahmad, D. Hawkes, and D. Landau. "Assessment of two Novel Ventilatory Surrogates for Use in the Delivery of Gated/tracked Radiotherapy for Non-small Cell Lung Cancer". *Radiotherapy and Oncology*, Vol. 91, No. 3, pp. 336–341, Jun 2009.
- [Huhl 10] B. Huhle, T. Schairer, P. Jenke, and W. Straßer. "Fusion of Range and Color Images for Denoising and Resolution Enhancement with a Non-local Filter". *Computer Vision and Image Understanding*, Vol. 114, No. 12, pp. 1336–1345, Dec 2010.
- [Izad 11] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera". In: *ACM Symposium on User Interface Software and Technology (UIST '11)*, pp. 559–568, Oct 2011.
- [Jian 05] B. Jian and B. C. Vemuri. "A Robust Algorithm for Point Set Registration Using Mixture of Gaussians". In: *IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1246–1251 Vol. 2, Oct 2005.
- [Jian 11] B. Jian and B. C. Vemuri. "Robust Point Set Registration Using Gaussian Mixture Models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 8, pp. 1633–1645, Aug 2011.
- [Kais 58] H. F. Kaiser. "The Varimax Criterion for Analytic Rotation in Factor Analysis". *Psychometrika*, Vol. 23, No. 3, pp. 187–200, 1958.
- [Kazh 06] M. Kazhdan, M. Bolitho, and H. Hoppe. "Poisson Surface Reconstruction". In: A. Sheffer, K. Polthier, D. W. Fellner, W. Hansmann, W. Purgathofer, and F. Sillion, Eds., *Eurographics Symposium on Geometry Processing (SGP '06)*, pp. 61–70, Jun 2006.
- [Keal 06] P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, R. C. R., V. H. M. B., V. S. S., W. J. W., and Y. E. "The Management of Respiratory Motion in Radiation Oncology Report of AAPM Task Group 76a". *Medical Physics*, Vol. 33, No. 10, pp. 3874–3900, Oct 2006.
- [Kell 07] M. Keller, J. Orthmann, A. Kolb, and V. Peters. "A Simulation Framework for Time-Of-Flight Sensors". In: *IEEE International Symposium on Signals, Circuits and Systems (ISSCS '07)*, pp. 1–4, Vol. 1, Jul 2007.
- [Kell 09] M. Keller and A. Kolb. "Real-time Simulation of Time-of-Flight Sensors". *Simulation Modelling Practice and Theory*, Vol. 17, No. 5, pp. 967–978, May 2009.
- [Kell 13] M. Keller, D. Lefloch, M. Lambers, S. Izadi, T. Weyrich, and A. Kolb. "Real-time 3D Reconstruction in Dynamic Scenes using Point-based Fusion". In: *International Conference on 3D Vision (3DV '13)*, pp. 1–8, Jun 2013.

- [Kipr 15] J. Kipritidis, G. Hugo, E. Weiss, J. Williamson, and P. J. Keall. "Measuring interfraction and intrafraction lung function changes during radiation therapy using four-dimensional cone beam CT ventilation imaging". *Medical Physics*, Vol. 42, No. 3, pp. 1255–1267, Mar 2015.
- [Kirs 11] M. Kirschner, M. Becker, and S. Wesarg. "3D Active Shape Model Segmentation with Nonlinear Shape Priors". In: G. Fichtinger, A. Martel, and T. Peters, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '11)*, pp. 492–499, LNCS 6892, Part II, Sep 2011.
- [Klin 10] T. Klinder, C. Lorenz, and J. Ostermann. "Prediction Framework for Statistical Respiratory Motion Modeling". In: T. Jiang, N. N., J. P. W. Pluin, and M. A. Viergever, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '10)*, pp. 327–334, LNCS 6363, Part III, Sep 2010.
- [Knut 93] H. Knutsson and C. F. Westin. "Normalized and Differential Convolution: Methods for Interpolation and Filtering of Incomplete and Uncertain Data". In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '93)*, pp. 515–523, Jun 1993.
- [Koch 04] N. Koch, H. H. Liu, G. Starkschall, M. Jacobson, K. Forster, Z. Liao, R. Komaki, and C. W. Stevens. "Evaluation of Internal Lung Motion for Respiratory-gated Radiotherapy Using MRI: Part I - Correlating Internal Lung Motion With Skin Fiducial Motion". *International Journal of Radiation Oncology, Biology, Physics*, Vol. 60, No. 5, pp. 1459–1472, Dec 2004.
- [Kohl 13] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. "ToF Meets RGB: Novel Multi-Sensor Super-Resolution for Hybrid 3-D Endoscopy". In: K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '13)*, pp. 139–146, LNCS 8149, Part I, Oct 2013.
- [Kolb 10] A. Kolb, E. Barth, R. Koch, and R. Larsen. "Time-of-Flight Cameras in Computer Graphics". *Computer Graphics Forum*, Vol. 29, No. 1, pp. 141–159, Mar 2010.
- [Korr 12] S. S. Korreman. "Motion in Radiotherapy: Photon Therapy". *Physics in Medicine and Biology*, Vol. 57, No. 23, pp. R161–R191, Dec 2012.
- [Kren 09] M. Krenkli, S. Gaiano, E. Mones, A. Ballarè, D. Beldi, C. Bolchini, and G. Loi. "Reproducibility of Patient Setup by Surface Image Registration System in Conformal Radiotherapy of Prostate Cancer". *Radiation Oncology*, Vol. 4, No. 9, pp. 1–10, Feb 2009.
- [Krie 08] H. P. Kriegel, P. Kröger, E. Schubert, and A. Zimek. "A General Framework for Increasing the Robustness of PCA-Based Correlation Clustering Algorithms". In: B. Ludäscher and N. Mamoulis, Eds., *International Conference on Scientific and Statistical Database Management (SS-DBM '08)*, pp. 418–435, LNCS 5069, Jul 2008.

- [Kwok 04] J. T. Y. Kwok and I. W. Tsang. "The Pre-image Problem in Kernel Methods". *IEEE Transactions on Neural Networks*, Vol. 15, No. 6, pp. 1517–1525, Nov 2004.
- [Ladi 10] A. Ladikos, C. Cagniard, R. Ghotbi, M. Reiser, and N. Navab. "Estimating Radiation Exposure in Interventional Environments". In: T. Jiang, N. N., J. P. W. Pluin, and M. A. Viergever, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '10)*, pp. 237–244, LNCS 6363, Part III, Sep 2010.
- [Law 06] M. H. C. Law and A. K. Jain. "Incremental Nonlinear Dimensionality Reduction by Manifold Learning". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 3, pp. 377–391, Mar 2006.
- [Lind 10] M. Lindner, I. Schiller, A. Kolb, and R. Koch. "Time-of-Flight Sensor Calibration for Accurate Range Sensing". *Computer Vision and Image Understanding*, Vol. 114, No. 12, pp. 1318–1328, Dec 2010.
- [Lind 13] B. L. Lindl, R. G. Müller, S. Lang, M. D. Herraiz Lablanca, and S. Klock. "TOPOS: A New Topometric Patient Positioning and Tracking System for Radiation Therapy Based on Structured White Light". *Medical Physics*, Vol. 40, No. 4, p. 042701, Apr 2013.
- [Liu 04] H. H. Liu, N. Koch, G. Starkschall, M. Jacobson, K. Forster, Z. Liao, R. Komaki, and C. W. Stevens. "Evaluation of Internal Lung Motion for Respiratory-gated Radiotherapy Using MRI: Part II - Margin Reduction of Internal Target Volume". *International Journal of Radiation Oncology, Biology, Physics*, Vol. 60, No. 5, pp. 1473–1483, Dec 2004.
- [Liu 07] H. H. Liu, P. Balter, T. Tutt, B. Choi, J. Zhang, C. Wang, M. Chi, D. Luo, T. Pan, S. Hunjan, G. Starkschall, I. Rosen, K. Prado, Z. Liao, J. Chang, R. Komaki, J. D. Cox, R. Mohan, and L. Dong. "Assessing Respiration-Induced Tumor Motion and Internal Target Volume Using Four-Dimensional Computed Tomography for Radiotherapy of Lung Cancer". *International Journal of Radiation Oncology, Biology, Physics*, Vol. 68, No. 2, pp. 531–540, Jun 2007.
- [Liu 10] X. Liu, I. Oguz, S. M. Pizer, and G. S. Mageras. "Shape-correlated Deformation Statistics for Respiratory motion prediction in 4D Lung". In: *SPIE Medical Imaging*, pp. 76252D–76252D, Feb 2010.
- [Liu 12] J. Liu, X. Gong, and J. Liu. "Guided Inpainting and Filtering for Kinect Depth Maps". In: *International Conference on Pattern Recognition (ICPR '12)*, pp. 2055–2058, Nov 2012.
- [Lore 87] W. E. Lorensen and H. E. Cline. "Marching Cubes: A High Resolution 3D Surface Construction Algorithm". In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '87)*, pp. 163–169, Jul 1987.
- [Low 04] K. L. Low. "Linear Least-Squares Optimization for Point-to-Plane ICP Surface Registration". Tech. Rep. TR04-004, University of North Carolina at Chapel Hill, Feb 2004.

- [Maie 12] L. Maier-Hein, A. M. Franz, T. R. dos Santos, M. Schmidt, M. Fangerau, H. Meinzer, and J. M. Fitzpatrick. "Convergent Iterative Closest-Point Algorithm to Accomodate Anisotropic and Inhomogenous Localization Error". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 8, pp. 1520–1532, Aug 2012.
- [Maie 13] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov. "Optical Techniques For 3D Surface Reconstruction in Computer-assisted Laparoscopic Surgery". *Medical Image Analysis*, Vol. 17, No. 8, pp. 974–996, Dec 2013.
- [Mart 12] P. J. Martin, L. F. Ayuso, R. Torres, and A. Gavilanes. "Algorithmic Strategies for Optimizing the Parallel Reduction Primitive in CUDA". In: *International Conference on High Performance Computing and Simulation (HPCS '12)*, pp. 511–519, Jul 2012.
- [Masu 96] T. Masuda, K. Sakaue, and N. Yokoya. "Registration and Integration of Multiple Range Images for 3-D Model Construction". In: *International Conference on Pattern Recognition (ICPR '96)*, pp. 879–883, Vol. 1, Aug 1996.
- [McCl 13] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King. "Respiratory Motion Models: A Review". *Medical Image Analysis*, Vol. 17, No. 1, pp. 19–42, Jan 2013.
- [Mika 98] S. Mika, B. Schölkopf, A. Smola, K. R. Müller, M. Scholz, and G. Rätsch. "Kernel PCA and De-noising in Feature Spaces". In: M. J. Kearns, S. A. Solla, and D. A. Cohn, Eds., *Conference on Advances in Neural Information Processing Systems 11*, pp. 536–542, 1998.
- [Mull 11] K. Müller, S. Bauer, J. Wasza, and J. Hornegger. "Automatic Multi-modal ToF/CT Organ Surface Registration". In: H. Handels, J. Ehrhardt, T. M. Deserno, H. P. Meinzer, and T. Tolxdorff, Eds., *Bildverarbeitung für die Medizin (BVM '11)*, pp. 154–158, Mar 2011.
- [Myro 10] A. Myronenko and X. Song. "Point Set Registration: Coherent Point Drift". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 12, pp. 2262–2275, Dec 2010.
- [Neum 11] D. Neumann, F. Lugauer, S. Bauer, J. Wasza, and J. Hornegger. "Real-time RGB-D Mapping and 3-D Modeling on the GPU using the Random Ball Cover Data Structure". In: *IEEE International Conference on Computer Vision Workshops (ICCV Workshops '11)*, pp. 1161–1167, Nov 2011.
- [Newc 11] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. "KinectFusion: Real-time Dense Surface Mapping and Tracking". In: *IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11)*, pp. 127–136, Oct 2011.
- [NVID 13] NVIDIA. *CUDA C Programming Guide 5.5*. NVIDIA, 2013.

- [Nyla 07] L. Nyland, M. Harris, and J. Prins. "Fast n-Body Simulation with CUDA". In: *GPU Gems 3*, pp. 677–696, Addison-Wesley, Boston, MA, USA, 2007.
- [Papi 13] B. W. Papiez, M. P. Heinrich, L. Risser, and J. A. Schnabel. "Complex Lung Motion Estimation via Adaptive Bilateral Filtering of the Deformation Field". In: K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '13)*, pp. 25–32, LNCS 8151, Part III, Oct 2013.
- [Pari 06] S. Paris and F. Durand. "A Fast Approximation of the Bilateral Filter Using a Signal Processing Approach". In: A. Leonardis, H. Bischof, and P. A., Eds., *European Conference on Computer Vision (ECCV '06)*, pp. 568–580, LNCS 3954, Part IV, May 2006.
- [Pari 09] S. Paris and F. Durand. "A Fast Approximation of the Bilateral Filter Using a Signal Processing Approach". *International Journal of Computer Vision*, Vol. 81, No. 1, pp. 24–52, Jan 2009.
- [Park 98] S. Parker, P. Shirley, Y. Livnat, C. Hansen, and P. P. Sloan. "Interactive Ray Tracing for Isosurface Rendering". In: *Conference on Visualization (VIS '98)*, pp. 233–238, Oct 1998.
- [Paul 02] M. Pauly, M. Gross, and L. P. Kobbelt. "Efficient Simplification of Point-sampled Surfaces". In: *Conference on Visualization (VIS '02)*, pp. 163–170, Oct 2002.
- [Pav107] G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, and C. Chamzas. "Methods for 3D Digitization of Cultural Heritage". *Journal of Cultural Heritage*, Vol. 8, No. 1, pp. 93–98, Jan 2007.
- [Peng 10] J. L. Peng, D. Kahler, J. G. Li, S. Samant, G. Yan, R. Amdur, and C. Liu. "Characterization of a Real-time Surface Image-guided Stereotactic Positioning System". *Medical Physics*, Vol. 37, No. 10, pp. 5421–5433, Oct 2010.
- [Penn 08] J. Penne, C. Schaller, J. Hornegger, and T. Kuwert. "Robust Real-time 3D Respiratory Motion Detection Using Time-of-Flight Cameras". *International Journal of Computer Assisted Radiology and Surgery*, Vol. 3, No. 5, pp. 427–431, Nov 2008.
- [Perl 85] K. Perlin. "An Image Synthesizer". In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '85)*, pp. 287–296, Jul 1985.
- [Pete 12] K. B. Petersen and P. M. S. "The Matrix Cookbook". Nov 2012. Version 20121115.
- [Pirk 11] K. Pirker, M. Ruther, H. Bischof, and G. Schweighofer. "Fast and Accurate Environment Modeling Using Three-dimensional Occupancy Grids". In: *IEEE International Conference on Computer Vision Workshops (ICCV Workshops '11)*, pp. 1134–1140, Nov 2011.
- [Plac 12] S. Placht, J. Stancanella, C. Schaller, M. Balda, and E. Angelopoulou. "Fast Time-of-Flight Camera Based Surface Registration for Radiotherapy Patient Positioning". *Medical Physics*, Vol. 39, No. 1, pp. 4–17, Jan 2012.

- [Poll 15] S. Pollock, R. Keall, and P. Keall. "Breathing guidance in radiation oncology and radiology: A systematic review of patient and healthy volunteer studies". *Medical Physics*, Vol. 42, No. 9, pp. 5490–5509, Sep 2015.
- [Pori 08] F. Porikli. "Constant Time O(1) Bilateral Filtering". In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Jun 2008.
- [Pric 12] G. J. Price, J. M. Parkhurst, P. J. Sharrock, and C. J. Moore. "Real-time Optical Measurement of the Dynamic Body Surface for Use in Guided Radiotherapy". *Physics in Medicine and Biology*, Vol. 57, No. 2, pp. 415–436, Jan 2012.
- [Pull 99] K. Pulli. "Multiview Registration for Large Data Sets". In: *International Conference on 3-D Digital Imaging and Modeling (3DIM '99)*, pp. 160–168, Oct 1999.
- [Qiu 09] D. Qiu, S. May, and A. Nüchter. "GPU-Accelerated Nearest Neighbor Search for 3D Registration". In: M. Fritz, B. Schiele, and J. H. Piater, Eds., *IEEE International Conference on Computer Vision Systems (ICVS '09)*, pp. 194–203, LNCS 5815, Oct 2009.
- [Rams 93] P. H. Ramsey, J. L. Hodges, and J. P. Shaffer. "Significance Probabilities of the Wilcoxon Signed-rank Test". *Journal of Nonparametric Statistics*, Vol. 2, No. 2, pp. 133–153, 1993.
- [Rowe 00] S. T. Roweis and L. K. Saul. "Nonlinear Dimensionality Reduction by Locally Linear Embedding". *Science*, Vol. 290, No. 5500, pp. 2323–2326, Dec 2000.
- [Rusi 01] S. Rusinkiewicz and M. Levoy. "Efficient Variants of the ICP Algorithm". In: *International Conference on 3-D Digital Imaging and Modeling (3DIM '01)*, pp. 145–152, May 2001.
- [Rusi 02] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. "Real-time 3D Model Acquisition". In: *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '02)*, pp. 438–446, Jul 2002.
- [Salv 10] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. "A State of the Art in Structured Light Patterns for Surface Profilometry". *Pattern recognition*, Vol. 43, No. 8, pp. 2666–2680, Aug 2010.
- [Sans 09] G. Sansoni, M. Trebeschi, and F. Docchio. "State-of-The-Art and Applications of 3D Imaging Sensors in Industry, Cultural Heritage, Medicine, and Criminal Investigation". *Sensors*, Vol. 9, No. 1, pp. 568–601, Jan 2009.
- [Scha 08] C. Schaller, J. Penne, and J. Hornegger. "Time-of-Flight Sensor for Respiratory Motion Gating". *Medical Physics*, Vol. 35, No. 7, pp. 3090–3093, Jul 2008.
- [Scha 09] C. Schaller, A. Adelt, J. Penne, and J. Hornegger. "Time-of-Flight Sensor for Patient Positioning". In: *SPIE Medical Imaging*, p. Vol. 7261, Mar 2009.

- [Scha 12] J. Schaerer, A. Fassi, M. Riboldi, P. Cerveri, G. Baroni, and D. Sar-
rut. "Multi-dimensional Respiratory Motion Tracking from Markerless
Optical Surface Imaging Based on Deformable Mesh Registration".
Physics in Medicine and Biology, Vol. 57, No. 2, pp. 357–373, Jan 2012.
- [Sche 07a] K. Scherbaum, M. Sunkel, H. P. Seidel, and V. Blanz. "Prediction of
Individual Non-Linear Aging Trajectories of Faces". *Computer Graphics
Forum*, Vol. 26, No. 3, pp. 285–294, Sep 2007.
- [Sche 07b] H. Scherl, B. Keck, M. Kowarschik, and J. Hornegger. "Fast GPU-
Based CT Reconstruction using the Common Unified Device Archite-
cture (CUDA)". In: *IEEE Nuclear Science Symposium Conference Record
(NSS '07)*, pp. 4464–4466, Oct 2007.
- [Scho 02] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, Cam-
bridge, MA, USA, 2002.
- [Scho 07] P. J. Schöffel, W. Harms, G. Sroka-Perez, W. Schlegel, and C. P. Karger.
"Accuracy of a Commercial Optical 3D Surface Imaging System for
Realignment of Patients for Radiotherapy of the Thorax". *Physics in
Medicine and Biology*, Vol. 52, No. 13, p. 3949, Jul 2007.
- [Scho 98] B. Schölkopf, A. Smola, and K. R. Müller. "Nonlinear Component
Analysis as a Kernel Eigenvalue Problem". *Neural Computation*, Vol. 10,
No. 5, pp. 1299–1319, Jul 1998.
- [Sega 01] W. P. Segars. *Development of a New Dynamic NURBS-based Cardiacorso
(NCAT) Phantom*. PhD thesis, University of North Carolina at Chapel
Hill, 2001.
- [Shot 13] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake,
M. Cook, and R. Moore. "Real-time Human Pose Recognition in Parts
from Single Depth Images". *Communications of the ACM*, Vol. 56, No. 1,
pp. 116–124, Jan 2013.
- [Silb 11] N. Silberman and R. Fergus. "Indoor Scene Segmentation Using a
Structured Light Sensor". In: *IEEE International Conference on Computer
Vision Workshops (ICCV Workshops '11)*, pp. 601–608, Nov 2011.
- [Sjos 07] K. Sjostrand, E. Rostrup, C. Ryberg, R. Larsen, C. Studholme,
H. Baezner, J. Ferro, F. Fazekas, L. Pantoni, D. Inzitari, and G. Walde-
mar. "Sparse decomposition and modeling of anatomical shape varia-
tion". *IEEE Transactions on Medical Imaging*, Vol. 26, No. 12, pp. 1625–
1635, Dec 2007.
- [Smis 11] J. Smisek, M. Jancosek, and T. Pajdla. "3D with Kinect". In: *IEEE
International Conference on Computer Vision Workshops (ICCV Workshops
'11)*, pp. 1154–1160, Nov 2011.
- [Soti 13] A. Sotiras, C. Davatzikos, and N. Paragios. "Deformable Medical Im-
age Registration: A Survey". *IEEE Transactions on Medical Imaging*,
Vol. 32, No. 7, pp. 1153–1190, Jul 2013.

- [Sozo 94] P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. C. Di Mauro. "A Non-linear Generalisation of PDMs using Polynomial Regression". In: E. Hancock, Ed., *British Machine Vision Conference (BMVC '94)*, pp. 397–406, Sep 1994.
- [Spie 09] M. Spiegel, D. Hahn, V. Daum, J. Wasza, and J. Hornegger. "Segmentation of Kidneys Using a New Active Shape Model Generation Technique Based on Non-rigid Image Registration". *Computerized Medical Imaging and Graphics*, Vol. 33, No. 1, pp. 29–39, Jan 2009.
- [Steg 06] M. B. Stegmann, K. Sjöstrand, and R. Larsen. "Sparse Modeling of Landmark and Texture Variability using the Orthomax Criterion". In: *SPIE Medical Imaging*, pp. 61441G1–61441G.12, Mar 2006.
- [Suzu 03] K. Suzuki, I. Horiba, and N. Sugie. "Linear-time Connected-component Labeling based on Sequential Local Operations". *Computer Vision and Image Understanding*, Vol. 89, No. 1, pp. 1–23, Jan 2003.
- [Taub 14] O. Taubmann, J. Wasza, C. Forman, P. Fischer, J. Wetzl, A. Maier, and J. Hornegger. "Prediction of Respiration-Induced Internal 3-D Deformation Fields From Dense External 3-D Surface Motion". In: *International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS '14)*, pp. 33–34, Jun 2014.
- [Tene 00] J. B. Tenenbaum, V. De Silva, and J. C. Langford. "A Global Geometric Gramework for Nonlinear Dimensionality Reduction". *Science*, Vol. 290, No. 5500, pp. 2319–2323, Dec 2000.
- [Thir 98] J. P. Thirion. "Image Matching as a Diffusion Process: an Analogy with Maxwell's Demons". *Medical Image Analysis*, Vol. 2, No. 3, pp. 243–260, Sep 1998.
- [Thod 03] H. H. Thodberg. "Minimum Description Length Shape and Appearance Models". In: C. Taylor and J. A. Noble, Eds., *International Conference on Information Processing in Medical Imaging (IPMI '03)*, pp. 51–62, LNCS 2732, Jul 2003.
- [Toma 98] C. Tomasi and R. Manduchi. "Bilateral Filtering for Gray and Color Images". In: *IEEE International Conference on Computer Vision (ICCV '98)*, pp. 839–846, Jan 1998.
- [Tree 99] G. M. Treece, R. W. Prager, and A. H. Gee. "Regularised Marching Tetrahedra: Improved Iso-Surface Esxtraction". *Computers & Graphics*, Vol. 23, No. 4, pp. 583–598, Aug 1999.
- [Tsin 04] Y. Tsin and T. Kanade. "A Correlation-Based Approach to Robust Point Set Registration". In: T. Pajdla and J. Matas, Eds., *European Conference on Computer Vision (ECCV '04)*, pp. 558–569, LNCS 3023, Part III, May 2004.
- [Twin 01] C. J. Twining and C. J. Taylor. "Kernel Principal Component Analysis and the Construction of Non-linear Active Shape Models". In: *British Machine Vision Conference (BMVC '01)*, pp. 23–32, Sep 2001.

- [Uzum 03] M. Üzümcü, A. F. Frangi, J. H. C. Reiber, and B. P. F. Lelieveldt. "Independent Component Analysis in Statistical Shape Models". In: *SPIE Medical Imaging*, pp. 375–383, May 2003.
- [Viol 01] P. Viola and M. Jones. "Rapid Object Detection Using a Boosted Cascade of Simple Features". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '01)*, pp. 511–518, Vol. 1, Dec 2001.
- [Wach 12] C. Wachinger, M. Yigitsoy, E. J. Rijkhorst, and N. Navab. "Manifold Learning for Image-based Breathing Gating in Ultrasound and MRI". *Medical Image Analysis*, Vol. 16, No. 4, pp. 806–818, May 2012.
- [Wang 04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity". *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612, Apr 2004.
- [Wasz 11a] J. Wasza, S. Bauer, S. Haase, M. Schmid, S. Reichert, and J. Hornegger. "RITK: The Range Imaging Toolkit - A Framework for 3-D Range Image Stream Processing". In: P. Eisert, J. Hornegger, and K. Polthier, Eds., *Workshop on Vision, Modeling & Visualization (VMV '11)*, pp. 57–64, Oct 2011.
- [Wasz 11b] J. Wasza, S. Bauer, and J. Hornegger. "High Performance GPU-based Preprocessing for Time-of-Flight Imaging in Medical Applications". In: H. Handels, J. Ehrhardt, T. M. Deserno, H. P. Meinzer, and T. Tolxdorff, Eds., *Bildverarbeitung für die Medizin (BVM '11)*, pp. 324–328, Mar 2011.
- [Wasz 11c] J. Wasza, S. Bauer, and J. Hornegger. "Real-time Preprocessing for Dense 3-D Range Imaging on the GPU: Defect Interpolation, Bilateral Temporal Averaging and Guided Filtering". In: *International Conference on Computer Vision Workshops (ICCV Workshops '11)*, pp. 1221–1227, Nov 2011.
- [Wasz 12a] J. Wasza, S. Bauer, S. Haase, and J. Hornegger. "Sparse Principal Axes Statistical Surface Deformation Models for Respiration Analysis and Classification". In: T. Tolxdorff, T. M. Deserno, H. Handels, and H. P. Meinzer, Eds., *Bildverarbeitung für die Medizin (BVM '12)*, pp. 316–321, Mar 2012.
- [Wasz 12b] J. Wasza, S. Bauer, and J. Hornegger. "Real-time Motion Compensated Patient Positioning and Non-rigid Deformation Estimation Using 4-D Shape Priors". In: N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '12)*, pp. 576–583, LNCS 7511, Part II, Oct 2012.
- [Wasz 13] J. Wasza, S. Bauer, and J. Hornegger. "Real-Time Respiratory Motion Analysis Using Manifold Ray Casting of Volumetrically Fused Multi-view Range Imaging". In: K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '13)*, pp. 116–123, LNCS 8150, Part II, Sep 2013.

- [Wasz 16] J. Wasza, P. Fischer, H. Leutheuser, T. Oefner, C. Bert, A. Maier, and J. Hornegger. "Real-time Respiratory Motion Analysis Using 4-D Shape Priors". *IEEE Transactions on Biomedical Engineering*, Vol. 63, No. 3, pp. 485–495, Mar 2016.
- [Welc 67] P. D. Welch. "The use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms". *IEEE Transactions on Audio and Electroacoustics*, Vol. 15, No. 2, pp. 70–73, Jun 1967.
- [Will 00] C. K. I. Williams and M. Seeger. "The Effect of the Input Density Distribution on Kernel-based Classifiers". In: *International Conference on Machine Learning (ICML '00)*, pp. 1159–1166, Jul 2000.
- [Will 12] T. Willoughby, J. Lehmann, J. A. Bencomo, S. K. Jani, L. Santanam, A. Sethi, T. D. Solberg, W. A. Tomé, and T. J. Waldron. "Quality Assurance for Nonradiographic Radiotherapy Localization and Positioning Systems: Report of Task Group 147". *Medical Physics*, Vol. 39, No. 4, pp. 1728–1747, Apr 2012.
- [Wilt 13] N. Wilt. *The Cuda Handbook: A Comprehensive Guide to GPU Programming*. Addison-Wesley, Boston, MA, USA, 2013.
- [Xia 12] J. Xia and R. A. Siochi. "A Real-time Respiratory Motion Monitoring System Using KINECT: Proof of Concept". *Medical Physics*, Vol. 39, No. 5, pp. 2682–2685, May 2012.
- [Yan 06] H. Yan, F. F. Yin, G. P. Zhu, M. Ajlouni, and J. H. Kim. "The Correlation Evaluation of a Tumor Tracking System Using Multiple External Markers". *Medical Physics*, Vol. 33, No. 11, pp. 4073–4084, Nov 2006.
- [Yang 09] Q. Yang, K. H. Tan, and N. Ahuja. "Real-time O(1) Bilateral Filtering". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 557–564, Jun 2009.
- [Yosh 10] S. Yoshizawa, A. Belyaev, and H. Yokota. "Fast Gauss Bilateral Filtering". *Computer Graphics Forum*, Vol. 29, No. 1, pp. 60–74, Mar 2010.
- [Yu 98] C. X. Yu, D. A. Jaffray, and J. W. Wong. "The Effects of Intra-fraction Organ Motion on the Delivery of Dynamic Intensity Modulation". *Physics in Medicine and Biology*, Vol. 43, No. 1, pp. 91–104, Jan 1998.
- [Zeng 12] M. Zeng, F. Zhao, J. Zheng, and X. Liu. "A Memory-Efficient KinectFusion Using Octree". In: S. M. Hu and R. R. Martin, Eds., *Conference on Computational Visual Media (CVM '12)*, pp. 234–241, LNCS 7633, Nov 2012.
- [Zhan 12] Z. Zhang. "Microsoft Kinect Sensor and Its Effect". *IEEE MultiMedia*, Vol. 19, No. 2, pp. 4–10, Feb 2012.
- [Zhan 94] Z. Zhang. "Iterative Point Matching for Registration of Free-form Curves and Surfaces". *International Journal of Computer Vision*, Vol. 13, No. 2, pp. 119–152, Oct 1994.

-
- [Zhen 10] W. S. Zheng, J. Lai, and P. C. Yuen. "Penalized Preimage Learning in Kernel Principal Component Analysis". *IEEE Transactions on Neural Networks*, Vol. 21, No. 4, pp. 551–570, Apr 2010.
- [Zhu 08] X. Zhu. "Semi-supervised Learning Literature Survey". Tech. Rep. TR-1530, University of Wisconsin-Madison, Jul 2008.
- [Zito 03] B. Zitova and J. Flusser. "Image Registration Methods: A Survey". *Image and Vision Computing*, Vol. 21, No. 11, pp. 977–1000, Oct 2003.
- [Zou 06] H. Zou, T. Hastie, and R. Tibshirani. "Sparse Principal Component Analysis". *Journal of computational and graphical statistics*, Vol. 15, No. 2, pp. 265–286, 2006.

