TECHNISCHE UNIVERSITÄT MÜNCHEN
Computer-Aided Medical Procedures & Augmented Reality

# Ultrasound Mosaicing and Motion Modeling
## Applications in Medical Image Registration

Christian Wachinger

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

| | | |
|---|---|---|
| Vorsitzender: | | Univ.-Prof. Dr. D. Cremers |
| Prüfer der Dissertation: | 1. | Univ.-Prof. Dr. N. Navab |
| | 2. | Assoc. Prof. W. Wells, Ph.D., Harvard University, Boston, USA |

Die Dissertation wurde am 3. März 2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 31. August 2011 angenommen.

**Abstract**

The correct alignment of images is one of the key technologies in medical imaging. It allows for arranging images in a common reference frame and therefore to propagate complementary information between them. The process that calculates the transformation is referred to as registration. The automatic registration of images is important for several applications, with ultrasound mosaicing and motion modeling being described in more details. Image noise and artifacts, however, complicate the correct alignment. The registration of ultrasound images is considered to be especially challenging due to their inherent contamination with speckle noise and viewing angle dependency. In the scope of this thesis, we present contributions to ultrasound imaging and registration, together with advances in ultrasound mosaicing and motion modeling.

In *ultrasound imaging*, we present a new method for the envelope detection of radio-frequency data, which is part of the process for creating B-mode images. Further, we present an approach for the acoustic impedance estimation from multiple ultrasound views. For *registration*, we introduce a new probabilistic framework that incorporates context information. This enables us to review and mathematically deduce a large number of registration approaches within this framework. Moreover, we present a new technique for multi-modal image registration with structural images. Next to a theoretical analysis of properties for such structural representations, we introduce two specific examples, entropy and Laplacian images.

In *ultrasound mosaicing*, we propose a new approach for ultrasound alignment by applying simultaneous registration. To this end, we deduce a new class of multivariate similarity measures and derive efficient, gradient-based optimization techniques. We further adapt the registration to ultrasound by deducing ultrasound specific similarity measures. These are obtained by integrating intensity distributions that well characterize speckle statistics. Moreover, matching functions are introduced that separate reflectivity and scattering regions, which result from two different types of physical interactions of the ultrasound beam with the tissue. For *motion modeling*, we propose an image-based gating system with manifold learning. Additionally, we present a new registration technique for time-resolved data, considering the spatial and temporal components simultaneously to guarantee a smooth deformation field over time.

## Zusammenfassung

Die korrekte Anordnung von Bildern ist eine der Schlüsseltechnologien in der medizinischen Bildverarbeitung. Sie ermöglicht es die Bilder innerhalb eines gemeinsamen Koordinatensystems anzuzeigen und damit komplementäre Informationen zwischen den Bildern zu propagieren. Der Prozess, der die Transformation zwischen Bildern berechnet, wird als Registrierung bezeichnet. Die automatische Registrierung ist wichtig für etliche Anwendungen, wobei wir Ultraschall-Mosaicing und Bewegungsmodellierung genauer untersuchen. Bildrauschen und Artefakte erschweren jedoch die korrekte Überlagerung der Bilder. Die Registrierung von Ultraschallbildern ist in diesem Zusammenhang als besonders herausfordernd einzustufen, da Speckle-Rauschen und die Abhängigkeit vom Aufnahmewinkel zu einer inhärenten Beeinträchtigung der Bilder führt. Im Rahmen dieser Doktorarbeit stellen wir Beiträge zur Ultraschallbildgebung und Registrierung vor, zusammen mit Fortschritten in Ultraschall-Mosaicing und Bewegungsmodellierung.

Im Bereich der *Ultraschallbildgebung* stellen wir eine neue Methode zur Hüllenkurvendemodulation von Radiofrequenzdaten vor, welche zur Erstellung von B-Mode Ultraschallbildern benötigt wird. Des Weiteren präsentieren wir einen neuen Ansatz zur Berechnung des akustischen Widerstands aus mehreren Ultraschallbildern. Zur *Registrierung* stellen wir ein neues probabilistisches Rahmenwerk vor, dass auch Kontextinformationen berücksichtigt. Dies ermöglicht es uns eine Vielzahl von Registrieransätzen zu analysieren und mathematisch abzuleiten. Darüber hinaus schlagen wir einen neuen Ansatz zur multimodalen Registrierung unter Benutzung von Strukturbildern vor. Neben einer theoretischen Analyse der Merkmale solcher Repräsentationen, stellen wir zwei konkrete Beispiele, die Entopy- und Laplacebilder, vor.

Im Bereich *Ultraschall-Mosaicing* bringen wir einen neuen Ansatz zur Überlagerung von Ultraschallbildern durch simultane Registrierung ein. Dafür leiten wir eine neue Klasse von multivariaten Ähnlichkeitsmaßen ab und deduzieren effiziente, gradientenbasierte Optimierungsverfahren. Wir passen die Registrierung an die speziellen Erfordernisse von Ultraschall durch Ableitung ultraschallspezifischer Ähnlichkeitsmaße an. Die vorgeschlagene Kostenfunktion unterscheidet zwischen Reflexion und Streuung, welche die Ergebnisse zweier unterschiedlicher Arten von physikalischer Interaktion des Schalls mit dem Gewebe darstellen. Zur *Bewegungsmodellierung* stellen wir ein bildbasiertes Gating-system basierend auf manifold learning vor. Zusätzlich präsentieren wir eine neue Registriermethode zur Überlagerung zeitlich aufgelöster Daten. Die Methode berücksichtigt sowohl die räumliche als auch die zeitliche Komponente gleichzeitig und garantiert dadurch die Erzeugung eines glatten Deformationsfeldes.

**Acknowledgement**

I am most grateful to Professor Nassir Navab for having created his research group in Munich and that he started teaching and fascinating us for numerous subjects in medical imaging and computer vision. It was really a pleasure to have such a visionary and motivating personality as supervisor during my doctorate. Looking back, it is surprising how much trust Professor Navab put into my own research work, starting from day one. This was really helpful in order to pursue own research ideas, where the outcome is always unpredictable in the beginning. Moreover, I am very thankful that I had the opportunity to participate at a number of international conferences, encouraged by Professor Navab, which proved to be extremely helpful for my research work, and besides, gave me the opportunity to visit beautiful places all over the world. I would also like to thank Professor William Wells for accepting to be part of my thesis committee and for coming the long way to Munich for my defense. I am also grateful to Professor Daniel Cremers for heading the doctoral committee.

Spending most of the day in the office, it is best to share it with friends. I must admit that I really had a great time at the CAMP group with many friends around. Tassilo "Franz" Klein the master of the distributions, who knows lake Starnberg like his back pocket and, most importantly, a person who is able to drive a truck with 320 horsepower to get the delivery to Teheran done in time. Andreas Keil who gets as easily distracted from work as I do and who does not mind spending a day or two searching for the best investment fund to participate in the gold rush. His interpretation of work-life balance was always an inspiration. Also thanks for taking me to mountain bike tours with most parts of the mountain still covered in snow! Athanasios Karamalis with whom I had the pleasure to share the same bed and room during several trips and with whom I had such a great time in the office! Mehmet Yigitsoy for numerous discussions about manifold learning and registration. A persons who manages to make things work! Stefan and Darko for such nice conference trips. Our bright international visitors Mattias, Ramtin, and Michael. I am also very grateful for the delightful discussions with Diana, Selen, Ahmad, Tobias, Nicolas, Steffi, Martin, Loren, Olli, Jose, Pierre, and Selim. The tea group for keeping me away from work and giving me the power for long office hours. The soccer guys for so many nice moments in Englischer Garten.

Finally, all this work would not have been possible without the great support of my beloved parents and my entire family.

# Contents

# Chapter 1

# Introduction

This thesis is conducted in the interdisciplinary field of medical imaging. The purpose of medical imaging is to assist the medical treatment by creating images of the human body. Historically, the first device for medical imaging is the optical microscope. Followed by the introduction of X-ray imaging at the end of the 19th century and ultrasound imaging in the middle of the 20th century. Subsequently, three-dimensional imaging techniques such as computed tomography (CT) and magnetic resonance (MR) tomography developed. Moreover, nuclear imaging techniques emerged, which allow for measuring the extent of a disease-process, based on the cellular function and physiology. These imaging possibilities gave rise to a multitude of clinical applications. Examples are the examination of bone fractures with X-ray, the assessment of fetal growth and anatomy in obstetric ultrasound, full-motion heart scans with CT, and the staging of tumors with MR. Further, functional imaging provides a means to estimate neural activity by measuring hemodynamics. More technically involved are the guidance of surgical interventions, the support in dose delivery, and the study of dynamical process such as respiration and tumor growth with time-resolved imaging and longitudinal studies. A big advantage is that nowadays almost all the data is available in digital form. This facilitates the application of further processing steps such as filtering, segmentation, and registration, once the data is reconstructed. Additionally, more possibilities exist for the visualization, *e.g.* volume rendering.

In this dissertation, we consider ultrasound imaging in more details. It is an increasingly popular modality due to its low cost, high mobility, and non-ionizing radiation. Physicians also appreciate to use bedside ultrasound to answer specific questions in real-time. Moreover, 3D imaging capabilities are advancing and offer improved possibilities for visualization, tracking of surgical instruments, and modeling. The appearance of ultrasound images is, however, rather different to images of other modalities. While we discuss details later on, a simplistic view is to consider ultrasound images as gradient-like images, highlighting changes of tissue properties. In order to achieve good processing results, it is necessary to adapt methods to ultrasound and take its particularities into account. This requires, on the one hand, a precise understanding of the ultrasound acquisition process, and on the other hand, the familiarity with processing algorithms to adapt them accordingly.

## Registration

A drawback of ultrasound imaging is the limited field of view of the acquisitions. Large fetuses, organs, and bones cannot be captured within a single image. The combination of several images, as it is done for panorama imaging, is a possible remedy, but it requires the correct spatial alignment of the images. The process that calculates the alignment is referred to as registration. Due to the abundance of data currently available, the interest lies in automatic registration methods. The key components of a registration procedure are the transformation model, the similarity measure, and the optimization scheme. The alignment is not limited to images from the same modality and so called multi-modal registration helps to integrate information from multiple sensors. Since different imaging modalities base upon different physical principles and therefore probe different physical properties of tissue, the scene is displayed with different intensity values. Each imaging modality has, in this connection, its own characteristics. This allows for fusing complementary information, for instance, anatomical and functional data. Moreover, the distinction between structures and the identification of lesions may be straightforward in one modality, while it is challenging in another one. Essential for benefiting from the combination of the information of multiple images is, however, the correct registration.

## Probabilistic Modeling

As mentioned, one of the key components of registration is the similarity measure. For the derivation of and reasoning about similarity measures, a probabilistic approach proves advantageous. A famous quotation of Pierre-Simon Laplace about probability theory is [Laplace, 1812]:

> It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.

Obviously, we try to benefit from the most important object of human knowledge, but this is not just due to the advice of Monsieur Laplace. It essentially offers a convenient framework to theoretically reason about many problems in image processing. Probabilistic models allow for formulating the relationship between variables in terms of mathematical equations. Commonly it is used to express the relationship between the observed data and the unknown parameters. Statistics allow for quantifying the likelihood of parameter instantiations with respect to the observations, or to quote Pierre-Simon Laplace again [Scherzer, 2010]:

> Statistics is common sense expressed in terms of numbers.

In image registration, we want to deduce the unknown transformation from the observed images. Further, we are not only interested in quantifying the likelihood of the current transformation, but we want to achieve the most likely transformation. Therefore, tools form statistical inference such as maximum likelihood estimation can

be applied. Probabilistic models also enable to integrate noise into the formulation, considering that the image acquisition process is prone to errors and therefore a total similarity cannot be established. Finally, probabilistic models naturally extend to the consideration of prior knowledge, turning the maximum likelihood to a maximum a posteriori estimation.

The direct integration of noise assumptions into the probabilistic model is a major advantage for ultrasound registration. The reason is that the speckle noise in ultrasound causes distributions, which differ from the commonly assumed Gaußian distribution. In classification and segmentation of ultrasound images, a lot of work was performed on accurately characterizing speckle noise. One of the conclusions is that the scattering conditions change locally and therefore give rise to changing speckle patterns. Modeling intensity distribution in ultrasound is therefore dependent on the imaged scenery and best done locally. The probabilistic model allows for directly plugging in these noise assumptions and for deriving the corresponding likelihood functions. This leads to ultrasound specific similarity measures, which are better suited for estimating the similarity of ultrasound images.

## Manifolds

Manifolds are generalizations of curves and surfaces to arbitrary many dimensions. They locally resemble Euclidean spaces, on which one can do calculus. In medical imaging, working with manifolds gained a lot of interest because it enables to extend algorithms to non-linear spaces. One example is manifold learning, which is a non-linear dimensionality reduction technique. Considering that images are points in a high dimensional space with the dimensionality corresponding to the number of pixels. Images of the same organ, of the same patient, or of the same modality share a lot of information and therefore the corresponding points are not randomly distributed in high dimensional space. One can in fact assume that they lie on a manifold of lower dimensionality, which is embedded in the high dimensional space. We illustrate this for the cases of breathing gating, multi-modal registration, and patient positioning.

The theory of manifolds is equally important in the optimization of image registration. Commonly, we deal with elements, such as rigid, affine, projective, and diffeomorphic transformations, which do not form vector spaces. This can be easily seen because such transformations are not commutative, which is one of the axioms for vector spaces. One can rather think of such transformations as points lying on a manifold. More precisely, they belong to mathematical spaces called Lie groups. Lie groups are both, smooth manifolds and algebraic groups. For the optimization of entities living on Lie groups two general approaches exist. Either a constrained optimization is performed by embedding the Lie group in a Euclidean space, or the optimization is directly performed on the Lie group. Working on the Lie group intrinsically guarantees to remain on the geometric structure of the group and consequently enables the application of unconstrained optimization, as illustrated for simultaneous registration in this thesis.

## Ultrasound Mosaicing

The first medical application that we consider in more details is ultrasound mosaicing, which refers to the correct alignment and fusion of multiple ultrasound images. The usage of ultrasound mosaicing provides the sonographers not just with a compounded volume of higher quality; recent studies also state a couple of other clinical advantages that come along with the extended field of view. First, the spatial relationship among structures that are too large for a single volume is easier to understand [Kim et al., 2003]. Second, sonographers have the flexibility to visualize anatomical structures from a variety of different angles [Peetrons, 2002, Leung et al., 2005]. Third, size and distance measurements of large organs are possible [Ying and Sin, 2005, Kim et al., 2003]. Fourth, individual structures within a broader context can be identified by having an image of the whole examination area [Dietrich et al., 2002]. And last, because of the increased features in the compounded view, specialists that are used to other modalities than ultrasound can better understand the spatial relationships of anatomical structures [Henrich et al., 2003]; helping to bridge the gap between the modalities and making it easier to convey sonographic findings to other experts.

From a technical point of view, ultrasound mosaicing necessitates to devise algorithms that are able to cope with the low signal-to-noise ratio, with ultrasound specific artifacts, and with the viewing angle dependency. Moreover, we are not facing a pairwise but a groupwise registration problem, since we want to align a set of images. While one improvement to address these issues is the already mentioned design of ultrasound specific similarity measures, a further improvement is the application of simultaneous registration. Simultaneous registration is a specific method for groupwise registration, which aligns all the images at the same time. An example for illustrating the advantage of simultaneous registration is the acquisition of ultrasound volumes around the leg. The dominant structure of the leg is the bone, which appears as a semicircle in the US images. Aligning two images at a time easily leads to a total overlap of both images, because the registration algorithms overlays both semicircles. Considering, however, all images at the same time leads to a stabilization of the process. This equally holds for artifacts which may be present in few of the scans.

## Motion Modeling

The second clinical application we deal with is motion modeling. A typical cause for motion in humans is respiration, which we evaluate in more details. Respiration is a cyclic, irregular process that leads to deformations in the abdominal and thoracic region. The respiratory signal monitors the current breathing phase of the patient. For numerous applications, it is necessary to assign to each image the corresponding respiratory phase in which it was acquired. The application we consider is the creation of 4D images, for which we propose a novel, purely image-based gating technique, which does not require the laborious setup of external systems. Time-resolved images are extensively used to study cardiac [King et al., 2010, Klein and Huesman, 2002, Peyrat et al., 2010], lung [Bystrov et al., 2009, Castillo et al., 2010, Flampouri

et al., 2006] and liver [Rohlfing et al., 2004, von Siebenthal, 2008] motion. The motion can be calculated from time-resolved images with deformable registration. The created motion models lead to benefits for target localization in radiation therapy, since the exposition of healthy tissue to ionizing radiation can be minimized [Colgan et al., 2008]. Similarly, enhanced targeting enables to more accurately heat and destroy pathogenic tissue in high-intensity focused ultrasound [Tempany et al., 2003, Wu et al., 2005].

The creation of motion models necessitates the non-rigid alignment of time-resolved volumes. In contrast to mosaicing, where we are mainly interested in the effect of the transformation to correctly overlay images, the interest now shifts to the transformation itself. The standard approach is, first, to perform pairwise registrations, and second, to combine the deformation fields to create the motion model. This can result in inconsistent and non-smooth motion fields along the temporal direction [Sundar et al., 2009b]. Further, the pairwise registration approach has the disadvantage that either, all registrations are performed towards one target leading to large displacements fields, or, registrations between adjacent volumes are calculated causing the accumulation of errors in the image sequence [Castillo et al., 2010]. In our approach, we consider simultaneous deformable registration with a temporal deformation field. We parameterize the deformation with FFD B-splines, which guarantee a smooth deformation field. Directly estimating the temporal deformation field during the registration not only leads to a smooth deformation field, but also increases the robustness of the registration in regard to outliers. Moreover, the interest of a motion model lies not only in quantifying the motion between certain image time points, but to show the motion over the entire cycle. This is better reflected with the direct optimization of the temporal deformation field.

## 1.1 Thesis Overview

The manuscript is organized in four parts, as explained in the following.

**Part I: Ultrasound Imaging.** The first part provides an overview of ultrasound imaging and describes improvements; it is organized in three chapters. **Chapter two** describes the physical principles underlying the ultrasound acquisition and the processing chain to construct B-mode images. A special emphasis is placed on motivating various noise models to capture speckle statistics. These statistics change with varying speckle density and appearance of coherent structures. A possible improvement for the envelope detection of raw, radio-frequency ultrasound data is presented in **chapter three**. Instead of calculating the 1D analytic signal, we apply the 2D analytic signal to estimate the envelope. The improvement is evaluated through visual inspection, and further, by performing goodness-of-fit tests to a Nakagami distribution, which indicate improved statistical properties. The 2D analytic signal also leads to a more reliable estimation of local features on B-mode images, such as local orientation and local phase. In **chapter four**, we present a method to combine the information

from multiple ultrasound views. The fusion of information is not straightforward because of the viewing angle dependency and artifacts in ultrasound images. The proposed method does not try to combine intensity information from multiple views but estimates the physical quantity that largely determines the appearance of ultrasound images, the acoustic impedance. The acoustic impedance image can either directly be presented or ultrasound images from arbitrary views can be simulated.

**Part II: Registration.** The second part consists of two chapters and deals with image registration. In **chapter five**, we present a new probabilistic framework for image registration that incorporates context information. Theoretically, we consider neighborhood information by introducing a layer of latent random variables, which can be regarded as descriptors. This extended framework allows us to review and model a large number of image registration approaches, ranging from intensity- to feature-based ones. It further enables us to classify image registration techniques, based on the uniqueness of their descriptors. By considering larger parts of the registration procedure, not limited to the similarity measure, we can motivate certain pre-processing steps as being optimal in maximum likelihood sense. An extension of the framework to groupwise registration is also presented. A specific type of descriptors, capturing the structural information of images, is illustrated in **chapter six**. Calculating this descriptor densely on the entire image domain, we obtain a new image with each pixel intensity representing the structure in a local neighborhood. Under the assumption that images of the same scene, but from different modalities, have the same structural representations, we apply these descriptors for multi-modal registration. The presented structural representations are entropy and Laplacian images. The evaluation on multiple registration experiments shows the good performance of the approach.

**Part III: Ultrasound Moasicing.** The third part of this thesis deals with a specific application, ultrasound mosaicing, and is divided into two chapters. For ultrasound mosaicing, a group of images has to be correctly aligned. In **chapter seven**, we propose simultaneous registration for ultrasound mosaicing. In details, we deduce a new class of similarity measures from the previously illustrated probabilistic framework. Moreover, we derive efficient gradient-based optimization techniques for these similarity measures. The fastest convergence is achieved with a second-order optimization scheme. In **chapter eight**, we further focus on similarity measures and try to comprise the particularities of ultrasound. We propose to separate the calculation of similarity of regions with high reflectivity and texture. Moreover, we derive locally adaptive Nakagami-based likelihood functions for registration.

**Part IV: Motion Modeling.** The fourth part is split into two chapters and deals with the second application, motion modeling. Important for the analysis of motion is to identify the respiratory phase in which image frames were

acquired. In **chapter nine**, we present a purely image-based respiratory gating approach applying manifold learning. The gating is applicable in several scenarios, where our focus lies on the creation of 4D data. For this, frames are acquired at different positions over several respiratory cycles and retrospectively sorted to create consistent volumes over time. In **chapter ten**, these volumes are registered with a specifically designed deformable registration. It takes the smoothness along the temporal direction intrinsically into account by working with a 4D deformation field. Additionally, it is devised as simultaneous registration approach to achieve more robust results by taking all information into consideration. The calculated deformation field allows us to analyze the respiratory motion. Moreover, with the deformation field also in temporal direction, it is possible to interpolate images at arbitrary time points.

## 1.2 Contributions

In the course of this dissertation, several contributions to the field of medical image analysis have been made. These are shortly summarized in the following. Some of the contributions have already been presented to the scientific community, which is indicated accordingly.

- **2D Envelope Detection in Ultrasound:** We propose the application of the 2D analytic signal for the envelope detection of ultrasound radio-frequency data, instead of the commonly applied 1D analytic signal. Considering the image data in lateral, as well as, axial direction allows for a more robust estimation of the local amplitude. This is evaluated by visual inspection and an analysis of noise statistics. [Wachinger et al., 2011]

- **Acoustic Impedance Estimation:** We present an approach to estimate the acoustic impedance from multiple ultrasound images. We first identify large scale interfaces between tissues of different acoustic impedance. Second, we apply a model of ultrasound propagation to infer the change of acoustic impedance from the intensity of the reflection. A combined estimation over all images in a maximum likelihood framework results in the final estimate. [Wachinger et al., 2008b]

- **Contextual Probabilistic Framework:** We propose a contextual probabilistic framework to model image registration. It extends current probabilistic approaches by incorporating neighborhood information around a pixel location. A graphical model is utilized to organize the dependencies between random variables, where we integrate an additional layer of latent variables. These can be seen as descriptors, capturing the local image information and serving as input for the similarity calculation. Splitting the registration process in a description and a similarity stage allows us to put a multitude of registration approaches in a continuum, ranging from pure intensity- to feature-based reg-

istration. This is one of the first unified descriptions of iconic and geometric registration techniques in a probabilistic framework.

- **Structural Representations:** We propose a new approach for multi-modal image registration with structural representations. The structural information of patches is quantified and used to create new images, which can be applied in mono-modal registration. We present next to a theoretical analysis of structural representations two specific concepts: entropy and Laplacian images. Laplacian images are created with Laplacian eigenmaps on the space of image patches. [Wachinger and Navab, 2010b,Wachinger and Navab, 2010a,Wachinger and Navab, 2011]

- **Multivariate Similarity Measure:** We present accumulated pairwise estimates as new class of similarity measures for simultaneous registration. We deduce it from a maximum likelihood formulation of image registration. This class of similarity measures is especially interesting for a small number of overlapping images and large variations in the number of overlapping images. Relationships to existing multivariate similarity measures are established. [Wachinger and Navab, 2009b,Wachinger et al., 2008c]

- **Efficient Optimization for Simultaneous Registration:** We derive efficient, gradient-based optimization schemes for simultaneous registration on Lie groups. The fastest convergence is achieved with the efficient second-order optimization. [Wachinger and Navab, 2009b]

- **Ultrasound Specific Similarity Measures:** We present similarity measures specifically designed for the alignment of ultrasound images. We take the viewing angle dependent nature, as well as, the specific noise distributions of ultrasound images into account. [Wachinger and Navab, 2009a,Wachinger and Navab, 2008]

- **Image-Based Respiratory Gating:** We propose the application of manifold learning to construct a purely image-based gating system for respiratory motion. Image frames are considered as points in a high-dimensional space, which are mapped to a lower dimensional representation that serves as respiratory signal. This retrospective gating allows for reordering the frames to create consistent 4D data. Experiments are performed on ultrasound and magnetic resonance images. [Wachinger et al., 2010b,Yigitsoy et al., 2011a]

- **Spatio-Temporal Simultaneous Registration:** We devise a new registration algorithm, specifically designed to align images with a smooth deformation along the temporal direction, as it is required for respiratory motion sequences. Simultaneous deformable registration is performed with a probabilistic gradient-based optimization procedure and a FFD B-spline transformation model augmented with a temporal dimension. [Yigitsoy et al., 2011b]

Contributions that are not part of the main body of the thesis but attached in the appendix are:

- **Patient Position Detection in MRI:** For accurately modeling the specific absorption rate of the patient in magnetic resonance imaging, the current location of the patient in the scanner has to be determined. For this task, low resolution images from a move-during-scan protocol are available. We achieve very good classification results by embedding the slices in two dimensions with manifold learning. [Wachinger et al., 2010a]
- **Deformable Mosaicing in Whole-Body MRI:** The motivation of whole-body MRI is to produce images of the entire body in as few stations as possible to reduce scanning time. This can be achieved with large field of view images. Such images, however, give rise to system- and patient-specific distortions, mainly due to the inhomogeneity of the B0 field and the gradient non-linearity. We correct for these distortions in the overlapping region by performing nonrigid registration to a linearly weighted average image. [Wachinger et al., 2008a, Wachinger et al., 2009]

# Part I

# ULTRASOUND IMAGING

This part introduces the imaging modality we mainly focus on: ultrasound. We provide an overview about basic concepts for sound propagation, B-mode construction, and speckle statistics. These constitute the theoretical foundation for the adaptation of algorithms to ultrasound at later parts of the thesis. Additionally, we present two contributions with respect to ultrasound imaging: (i) the envelope detection with the 2D analytic signal and (ii) an approach towards acoustic impedance estimation.

# Chapter 2

# Ultrasound Overview

*In this chapter, we present a short introduction to ultrasound, highlighting developments on compounding and volumetric imaging. Further, physical models for ultrasound propagation are discussed, together with an illustration of the image formation on transducer arrays and the processing chain to create B-mode images. Finally, we motivate several distributions for modeling speckle statistics.*

## 2.1 Introduction

The work of the French physicist Paul Langevin during the First World War to detect enemy submarines is commonly considered as initiation of modern ultrasound. Working with a 150 kHz source, he noticed that small fishes were killed when entering the beam, and similarly, that intense pain is caused if a hand was put in the path [Cobbold, 2007]. The mathematical and physical foundations of acoustic wave propagation evolved early on in the 18th and 19th centuries with the two-volume treatise of [Rayleigh, 1877] being one of the most important milestones. Over the last fifty years, progress in this field has led to the development of systems that allow for depicting the acoustic structure of biological tissue, without causing damage when operated at moderate intensity. In 1942, Dussik published a paper in German that first suggested the usage of ultrasound for diagnostic purposes [Dussik, 1942]. The first 2D cross-sectional imaging system was proposed in 1952 in [Wild and Reid, 1952]. The same authors published in the same year the results of a pilot study for detecting breast tumors using ultrasound. Surprisingly, the first 3D imaging system for ultrasound was proposed as early as in 1954 in [Holmes et al., 1954]. The limited temporal and spatial resolution of the ultrasound systems led to volumes of lower quality, the basic idea of mechanically moving the probe to create volumetric images, however, was out soon after 2D imaging. Even earlier than 3D imaging, the combination of several ultrasound images was proposed to enhance the brightness of structures that the beam does not hit perpendicularly and to image structures behind strong reflectors. This technique was referred to as compound scanning and Howry and Gordon stated in 1964 that is was originally proposed in 1952 [Howry and Gordon, 1964]. Figure 2.1 illustrates a schematic overview of compound scanning

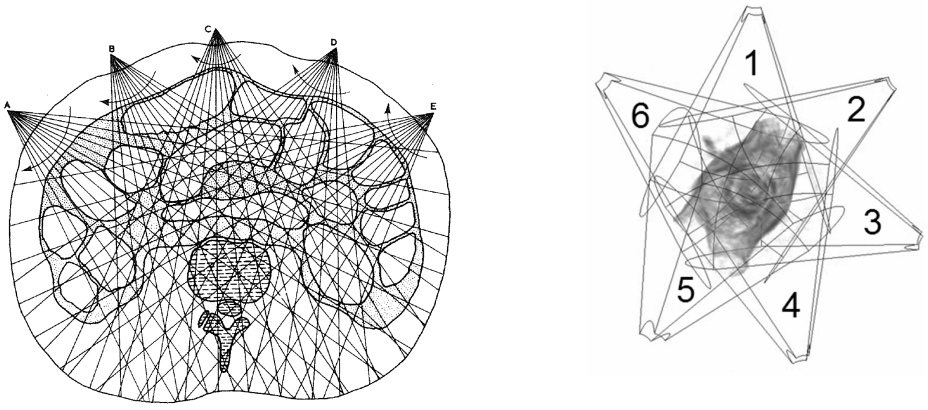



Figure 2.1: Left: Illustration from 1959 showing the idea of compound scanning [Brown, 1959]. The probe is placed at various locations around the abdominal region, A-E. At each position, ultrasound sweeps are acquired. Right: Current ultrasound compound created from six volumetric scans acquired with a wobbler probe.

from 1959 and a recent volumetric compound. While the limitations in acquisition speed and resolution impeded the development of clinical solutions in the 1950s, subsequent developments enabled the design of high performance ultrasound systems, which are nowadays common in clinics. Such systems offer more possibilities for 3D ultrasound imaging and compound scanning, making these areas still active fields of research, even more than half a century after the initial proposition.

Generally, ultrasound is defined as sound with a frequency above the audible range of humans, which is at frequencies above around 20 kHz. It has many advantages in comparison to other imaging modalities which have led to its widespread use in clinical practice; it is (i) harmless at low power, (ii) portable, (iii) a real-time modality, and (iv) cost effective. The recent introduction of 2D array US transducers in the market makes further applications possible, due to the instantaneous acquisition of ultrasound volumes. However, ultrasound imaging has a number of disadvantages including: (i) a limited field of view, (ii) occlusions behind structures with high acoustic impedance, (iii) a low signal-to-noise ratio (SNR), and (iv) position/viewing angle dependency. The already mentioned compounding of several views helps reducing some of these shortcomings.

In the focus of this dissertation is volumetric ultrasound imaging, with a variety of acquisition techniques being available. The 3D systems that are mainly used throughout this thesis are freehand systems and wobbler probes. These use a regular 1D array probe that is moved during the acquisition, requiring high imaging frequency. For 3D freehand ultrasound, the position of the transducer is determined by an optical or a magnetic tracking system to place the ultrasound frames at the right spatial location. Freehand systems offer a lot of flexibility because the probe can be moved to arbitrary positions; however, they require an experienced user to sample the

region of interest without gaps. The wobbler probe consists of a 1D array transducer that continuously oscillates to acquire fan like volumes. An alternative that allows for instantaneous volumetric imaging is the usage of 2D arrays. In contrast to commonly utilized piezoelectric transducers, capacitive transducers become popular for 3D imaging. These capacitive micromachined ultrasound transducers (CMUT) offer superior and efficient volumetric imaging. This is possible due to a wide bandwidth, the ease of manufacture (wafer fabrication process), and the high electromechanical conversion efficiency. The instantaneous acquisition with 2D arrays is especially interesting for 4D imaging, because breathing motion can, for instance, alter the underlying scene during the acquisition with 1D arrays. Serious drawbacks of 2D array systems are their limited availability, high price, and very limited access to data, especially raw data, impeding their application in most of our experiments. Since we are interested in 4D data for motion modeling, we propose a new technique for the acquisition of breathing-affected 4D ultrasound with wobbler probes in chapter 9.

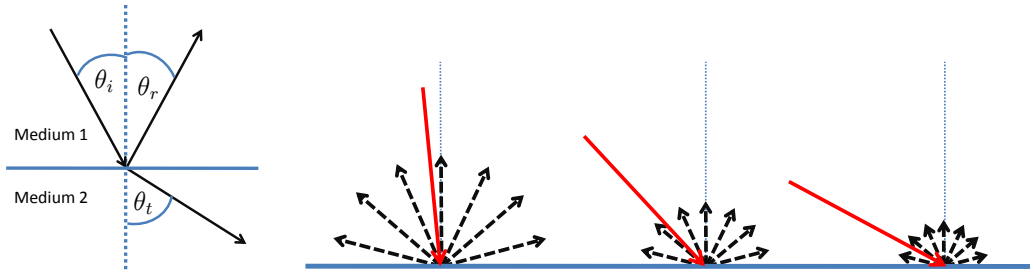
From a research point of view, ultrasound is quite different to other imaging modalities, because the system can directly be operated by oneself. Such practical experiments help to develop a better understanding of ultrasound imaging. This is necessary because ultrasound, even more than other imaging modalities, requires the specific adaptation of processing techniques to achieve good results. The challenge lies in devising new algorithms that take the particularities of ultrasound into account. This, on the one hand, involves taking a detailed look on the mathematical derivation of approaches in order to be aware of implicit assumptions, and on the other hand, requires a good knowledge of the process of ultrasound imaging, to check whether the assumptions are consistent with the imaging scenario. We will present contributions in this regard for ultrasound registration in parts III and IV. In the next sections, we will provide a concise description of ultrasound propagation, image formation, and speckle statistics.

## 2.2 Modeling Ultrasound Propagation

In order to describe the wave propagation in fluids, the particle velocity function $v$ has to be known at all spatial and temporal points, together with two thermodynamic functions [Cobbold, 2007]. The flow field can be fully determined by solving these equations with respect to initial and boundary conditions. Combining the Navier-Stokes equation with the equation of state and the continuity equation, it is possible to derive a wave equation for inviscid fluids [Cobbold, 2007]

$$\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} - \nabla^2 P = f, \tag{2.1}$$

with the speed of sound $c$ and the source distribution $f$. The pressure field $P$ is related to the velocity function $\nabla P = -\rho\frac{\partial v}{\partial t}$ through the density $\rho$. Current short, high-amplitude transmit pulses in diagnostic applications give rise to nonlinear effects in ultrasound propagation, which are not modeled by equation (2.1). Various approaches exist to incorporate nonlinear effects in the propagation process, with




(a) Snell (b) Lambertian

Figure 2.2: Left: Reflection and refraction as expressed in Snell's law. $\theta_i$, $\theta_r$, $\theta_t$ angles of incoming, reflected, transmitted waves. Right: Reflection due to Lambert's emission law. Red: incoming intensity, black: emitted intensity.

the KZK and Westervelt equations being two popular examples [Huijssen et al., 2003, Pinton et al., 2009, Karamalis et al., 2010]. Such equations allow for an accurate simulation, however, they include tissue dependent parameters (*e.g.* diffusivity, density, nonlinearity), which in practice are difficult to determine. Due to this drawback and the high computational complexity, we decided to work with a simpler model, which we derive in the following.

Consider the case where the acoustic wave hits a boundary between two media, producing reflection and refraction, see figure 2.2(a). These two phenomena can be considered as special cases of scattering, with the object boundary being large in comparison to the wavelength, yielding a redirection of incident radiation in an organized manner [Cobbold, 2007]. The formulation of boundary conditions that constrain the particle velocity component normal to the interface to be continuous allows the derivation of the following relationships from the wave equation

$$\frac{\sin\theta_i}{c_1} = \frac{\sin\theta_r}{c_1} = \frac{\sin\theta_t}{c_2} \tag{2.2}$$

with $\theta_i$, $\theta_r$, $\theta_t$ the angles of incoming, reflected, transmitted waves and $c_1$, $c_2$ the speed of sound in the first and second medium, respectively. It follows that $\theta_i = \theta_r$, and further, according to Snell's law for acoustics

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{c_1}{c_2}. \tag{2.3}$$

It is also possible to derive the relationship between the reflected $P_r$ and incident $P_i$ pressure amplitudes, expressed in terms of the pressure amplitude reflection coefficient $\zeta_P$

$$\zeta_P = \frac{P_r}{P_i} = \frac{Z_2\cos\theta_i - Z_1\cos\theta_t}{Z_2\cos\theta_i + Z_1\cos\theta_t} \tag{2.4}$$

where the acoustic impedance $Z_1$ is the product of density and speed of sound $Z_1 = \rho_1 \cdot c_1$, analogously for $Z_2$. Similarly, the intensity reflection coefficient $\zeta_I$ is

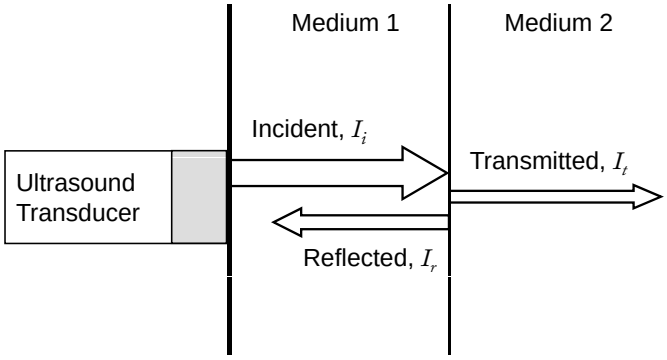



Figure 2.3: Reflection and transmission of the incident sound wave after hitting an interface. The relative intensities are determined by the acoustic impedances of the media.

defined as

$$\zeta_I = \frac{I_r}{I_i} = \frac{P_r^2}{P_i^2} = \left( \frac{Z_2 \cos\theta_i - Z_1 \cos\theta_t}{Z_2 \cos\theta_i + Z_1 \cos\theta_t} \right)^2 , \tag{2.5}$$

with $I_i$ the incoming and $I_r$ the reflected intensity.

For the simulation of ultrasound and also for the estimation of acoustic impedance, the dependency on the transmission angle $\theta_t$ in equation (2.5) is problematic. Supposing that we know the location and geometry of the transducer together with the normal of the boundary, we are able to calculate the incidence angle $\theta_i$. Directly calculating the transmission angle from the image information is, however, not possible. Snell's law allows for relating the transmission to the incidence angle

$$\cos\theta_t = \sqrt{1 - \left( \frac{c_2 \sin\theta_i}{c_1} \right)^2} , \tag{2.6}$$

providing a possible solution to determine $\theta_t$. Unfortunately, this requires the knowledge of the speed of sound in both media, which is commonly not available. These issues led to the introduction of an alternative model for ultrasound reflection, which combines parts of Snell's law with Lambert's emission law in optics [Wein et al., 2008]. Supposing the beam hits the boundary perpendicularly and no refraction occurs, the reflected intensity is

$$I_r = I_i \cdot \left( \frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 , \tag{2.7}$$

see figure 2.3. Lambert's emission law states that the radiant intensity from a diffusely reflecting surface is directly proportional to the cosine of the incidence angle, see figure 2.2(b), yielding the following relationship

$$I_r = I_i \cdot \cos^m \theta_i \cdot \left( \frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 , \tag{2.8}$$

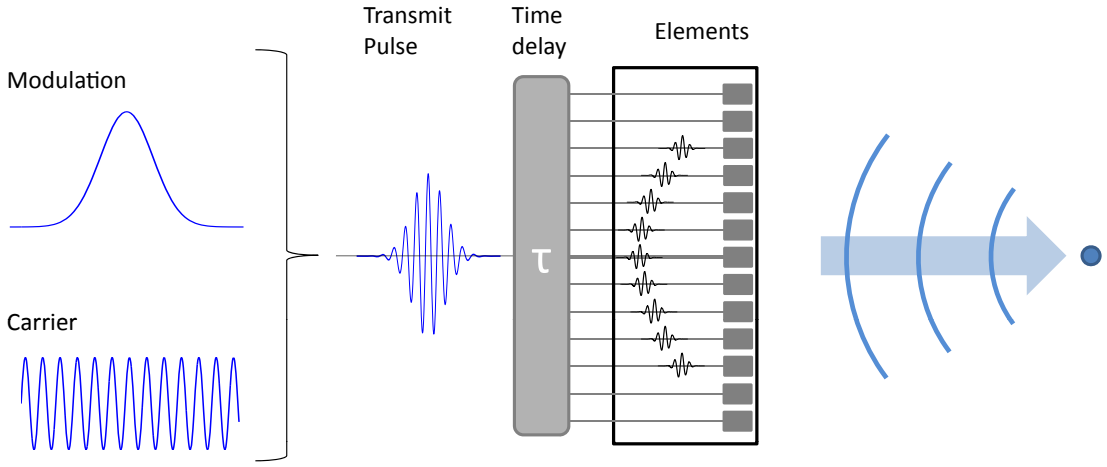



Figure 2.4: The combination of the modulation function and the carrier wave results in the transmit pulse. For focusing and beam steering, various time delays are added. The activation of the elements with the pulse pattern leads to the creation of a sound wave in the medium. The current center location is indicated as bold line.

with $m$ modeling the diffusivity of the surface [Wein et al., 2008]. It is this equation that is used for modeling ultrasound reflection throughout the thesis.

Apart from reflection, there are also other types of interactions such as scattering and absorption. Absorption leads to a loss of energy of the ultrasound wave by conversion to heat or chemical energy. Scattering is caused by microscopic inhomogeneities in the tissue and leads to a redirection of energy along paths that differ from the one followed by the incoming wave. An active scatterer acts as a separate sound source that emits waves in all directions, causing speckle patterns typical for ultrasound. The loss due to absorption and scattering is summarized as attenuation. Attenuation is frequency dependent and follows an exponential decay [Hedrick et al., 2004, Zagzebski, 1996]

$$P(x) = P_{\max} \cdot e^{-\alpha x} \tag{2.9}$$

with $P_{\max}$ the initial sound pressure, $P(x)$ the pressure after traversed distance $x$, and $\alpha$ the attenuation coefficient.

For structures being visible in ultrasound images, the interaction of the ultrasound beam with tissue must have led to the redirection of some of the energy back to the transducer. This energy either comes from scattering or reflection. Since they are both caused by inhomogeneities in the tissue, they are the result of the same physical interaction, once on a large scale, and once on a small scale, with respect to the wavelength $\lambda$. In medical ultrasound, typically $0.1 \leq \lambda \leq 1.0$ mm holds. As mentioned above, reflection can be seen as a special case of scattering on large object boundaries, yielding a redirection of energy in an organized manner. Nevertheless, it is important to distinguish between these two types of interaction, because reflection is viewing angle dependent, while scattering is only to very limited extent dependent on the direction of insonification. This motivates us to design a new ultrasound

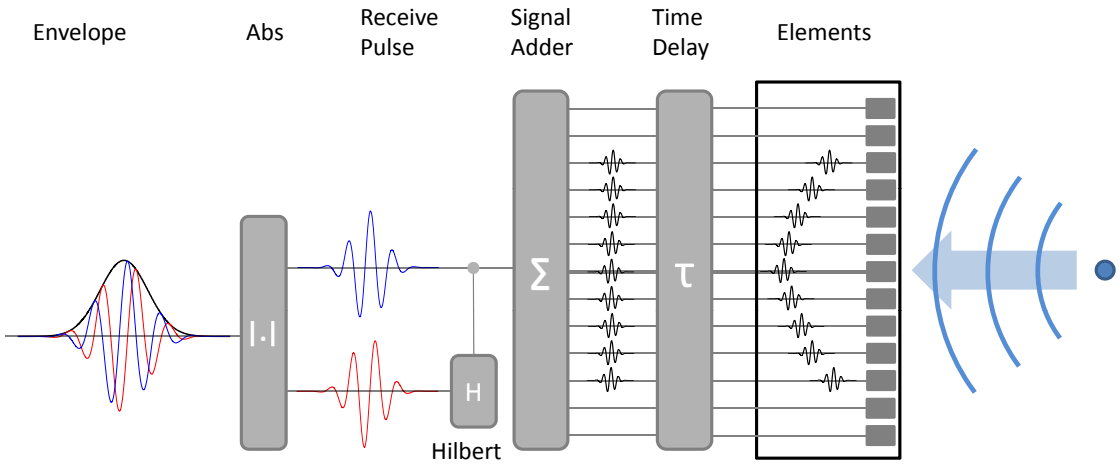



Figure 2.5: From right to left. The reflected sound waves in the medium are detected by the transducer elements. Each measured pulse is time delayed to compensate for differences in traveled distance. All received pulses are weighted and summed up to create the receive pulse for one location. The envelope of the pulse is determined by calculating the absolute value of the analytic signal.

similarity measure that considers reflection and scattering parts separately, see chapter 8. The statistics of the scattering process are further described in section 2.4.

## 2.3 Image Formation and B-mode Conversion

In the last section, we described models for the propagation of sound waves, constituting the physical basis of ultrasound imaging. In this section, we provide further details about the image formation and conversion to B-mode images, which is widely covered in textbooks [Zagzebski, 1996, Hedrick et al., 2004, Cobbold, 2007]. Today's ultrasound probes are mainly transducer arrays, consisting of a group of closely arranged piezoelectric elements, where each element can be excited separately. This allows for electronic beam steering and focusing by delaying the firing of crystals. Further, a dynamic aperture can be created by flexibly activating a number of elements for transmission and reception. A schematic overview of ultrasound transmission and reception is shown in figures 2.4 and 2.5, respectively. The transmit pulse is created by convolving the modulation function with a carrier wave. In the example shown, we convolve a sinusoidal wave with a Gaußian modulator. Subsequently, a specific time delay is added to each element to account for focusing and steering. The created wavefront propagates in the medium and is reflected at inhomogeneities.

After the transmission, the elements are in receive mode and detect arriving waves. Since the traveled distance of the returning pulses is different for the various elements, delays have to be added to compensate for this. The delays change dynamically while echoes from deeper reflectors arrive. In the beam former, the echoes are amplified and accumulated with an additional weighting. This leads to the creation of a single

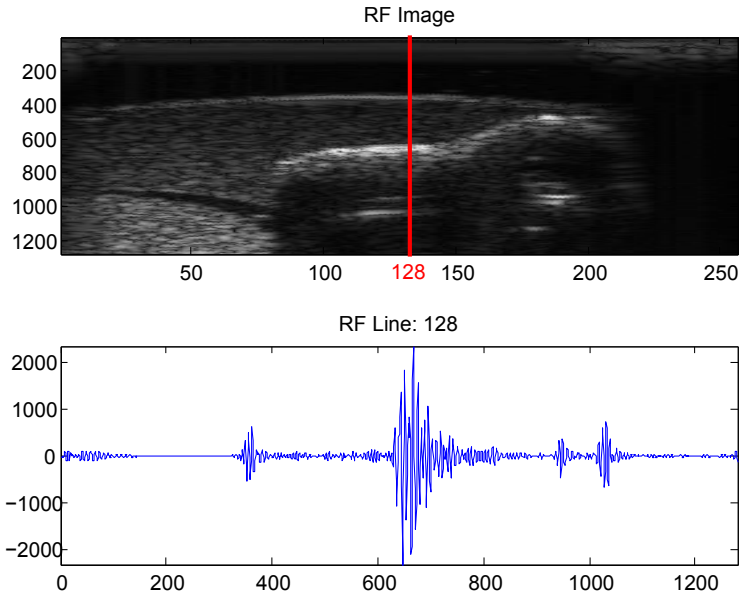



Figure 2.6: Top: RF image after envelope detection and dynamic range compression. Bottom: Plot of RF scan line 128, highlighted in red.

receive signal for each position. Current transducers commonly operate with 128 elements, which leads in combination with slight beam steering to 256 different receive signals per image. The signal is subsequently passed through a pre-amplifier and a time gain compensation (TGC), which emphasizes signals from deeper regions to compensate for attenuation. The resulting signal is commonly referred to as radio-frequency (RF) signal and can be accessed in ultrasound systems. Figure 2.6 illustrates one RF scan line of an ultrasound image.

The images presented on the screen of the ultrasound system are B-mode images. A sequence of processing steps is necessary for transforming RF signals to B-mode images, see figure 2.7. We illustrate the processing chain for RF to B-mode conversion on actual images in figure 2.8. Starting from the gain compensated signal, a first step is the frequency compounding. This is an optional step to increase the quality of images by multi-frequency decomposition and compounding [Cincotti et al., 2001]. The next step is the demodulation, which extracts the information-bearing signal from the modulated carrier wave. In ultrasound processing, the demodulation is commonly performed with an envelope detection. Therefore, the absolute value of the analytic signal, consisting of the original signal and the Hilbert transformed signal, is calculated, see figure 2.5. 1D envelope detection is performed for each scan line separately. In chapter 3, we present the 2D envelope detection with the 2D analytic signal. Next, the dynamic range of the signal is compressed by applying a non-linear intensity mapping. Typically a logarithmic scaling is applied, referred to as log-compression. In [Kaplan and Ma, 1994], the following formula is proposed

$$Y = \alpha \log(X) + \beta \tag{2.10}$$

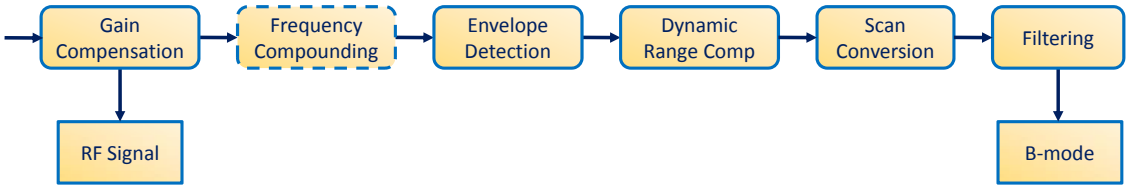
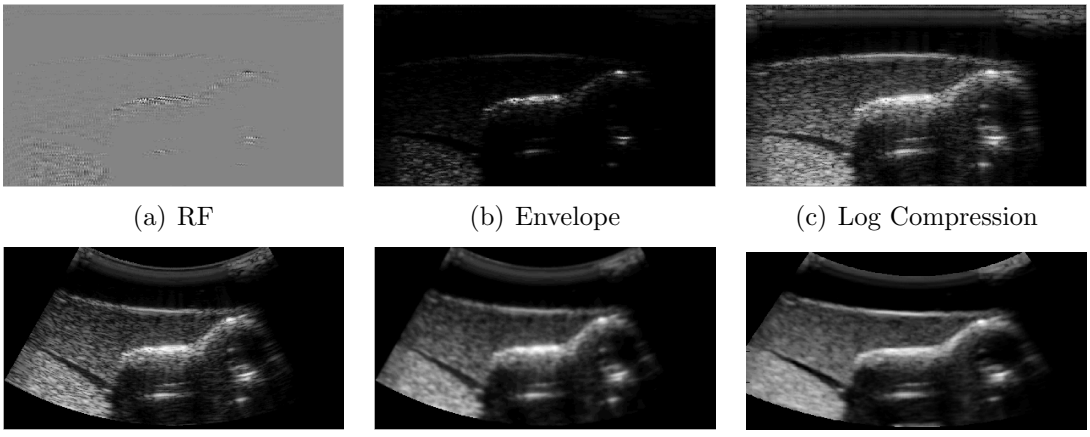



Figure 2.7: Exemplar ultrasound processing pipeline. Gain compensated signals are available as RF data. Further steps are as indicated. The B-mode image is the final result, presented to the operator.



(a) RF (b) Envelope (c) Log Compression

(d) Scan Conversion (e) Diffusion Filtering (f) B-mode US System

Figure 2.8: Images for illustrating the RF to B-mode processing chain. For comparison, we also show the B-mode image of the US system.

with the received signal $X$, the compressed signal $Y$, and the compression parameters $\alpha, \beta$. In [Wein et al., 2008], a log-compression with

$$Y = \frac{\log(1 + \alpha X)}{\log(1 + \alpha)} \tag{2.11}$$

is proposed. The next step is to place scan lines at their correct geometric location, especially important for curved and phased array transducers. This conversion from polar to Cartesian coordinate system is done by the scan conversion. Unavoidable is the interpolation of intensity values at this stage. Finally, noise reduction filters are applied. The final image of the presented processing chain is shown in figure 2.8(e), with anisotropic diffusion used for filtering [Abd-Elmoniem et al., 2002]. The B-mode image, as presented on the ultrasound system, is shown in figure 2.8(f). We observe that the result of our replicated processing chain is very similar to the B-mode image of the system. Unfortunately, ultrasound manufacturers provide very limited information about proprietary processing steps that are included in the processing chain to enhance the final result, but the main steps are as presented.

## 2.4 Ultrasound Statistics

Modeling the statistics of an imaging modality is related to modeling its noise statistics. For ultrasound, the dominant noise is speckle, which occurs at structures that are rough with respect to the scale of the wavelength. Typical imaging modalities affected by speckle patterns are optics, radar imagery, and ultrasound. In optics, it is the rough surface of an object that produces speckle. In ultrasound, microscopic inhomogeneities in the tissue are responsible for speckle [Cobbold, 2007]. These inhomogeneities are due to the cellular nature of tissue. The granular texture due to speckle can, for instance, be appreciated in figure 2.8. Previously described compounding techniques, which combine images that contain uncorrelated speckle, can be used to increase the signal to noise ratio (SNR), which is degraded by speckle [Burckhardt, 1978]. There was, however, also work conducted on accurately characterizing the statistics of speckle patterns. While the initial intention was to design specific filters to increase the SNR, it soon became clear that one can benefit from working with speckle to assess the tissue microstructure, to measure tissue strain distributions in elastography, or to calculate flow velocity fields [Trahey et al., 1987]. In chapter 8, we will apply the speckle statistics described in the following to design appropriate likelihood functions for ultrasound registration.

The cause of speckle is a signal that consists of a number of independently phased additive complex components [Goodman, 2006, Dutt, 1995], with each component having an amplitude $a_n$ and a phase $\phi_n$. The single components are complex values because they are phasors $\mathbf{a}_n$ that represent sinusoidal signals. The addition of these components creates a random walk and results in an analytic signal $\mathbf{A}$

$$\mathbf{A} = A \cdot e^{i\phi} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathbf{a}_n = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} a_n e^{i\phi_n} \tag{2.12}$$

with $N$ the number of scatterers and $\frac{1}{\sqrt{N}}$ the normalization factor. The result of this addition is dependent on the phases and amplitudes of the single components, with either destructive or constructive interference occurring, see Figure 2.9. The resulting length $A$ of the component is the amplitude of the observed wave, corresponding to the output of the envelope detector described in section 2.3. Let the real and imaginary parts of $\mathbf{A}$ be

$$\mathcal{R} = \mathrm{Re}(\mathbf{A}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} a_n \cos(\phi_n) \tag{2.13}$$

$$\mathcal{I} = \mathrm{Im}(\mathbf{A}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} a_n \sin(\phi_n). \tag{2.14}$$

### 2.4.1 Large Number of Scatterers

For the deduction of noise statistics, we assume amplitudes and phases to be statistically independent from each other [Dutt, 1995]. Further, let the phases be uniformly

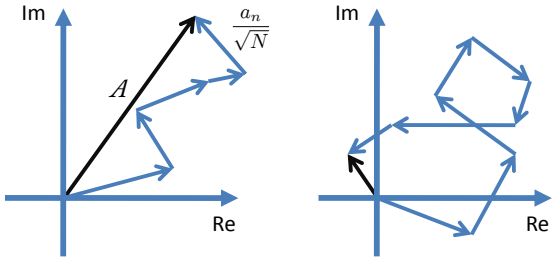



Figure 2.9: Summation of random phasors, illustrating constructive (left) and destructive (right) interference.

distributed in the interval $[-\pi, \pi]$ and let the number of scatters $N$ per resolution cell be very large, see first image in figure 2.10. Since the real and imaginary parts are calculated by the summation of a large number of independent random variables, the central limit theorem indicates that the sum is asymptotically Gaußian distributed. The assumptions of independence and uniform distribution yield the expected values of $\mathcal{R}$ and $\mathcal{I}$ to be zero [Goodman, 2006]. Moreover, the cross correlation is zero and the variances of real and imaginary part are identical, $\sigma_{\mathcal{R}}^2 = \sigma_{\mathcal{I}}^2 = \sigma^2$ , leading to

$$p_{\mathcal{R},\mathcal{I}}(\mathcal{R},\mathcal{I}) = \frac{1}{2\pi\sigma^2} e^{-\frac{\mathcal{R}^2+\mathcal{I}^2}{2\sigma^2}}. \tag{2.15}$$

Our interest lies in finding the distribution of the amplitude $A$, which is related to $\mathcal{R}$ and $\mathcal{I}$ in polar coordinates by the expressions

$$A = \sqrt{\mathcal{R}^2 + \mathcal{I}^2} \tag{2.16}$$

$$\phi = \arctan\left(\mathcal{I}/\mathcal{R}\right) \tag{2.17}$$

and

$$\mathcal{R} = A\cos\phi \tag{2.18}$$

$$\mathcal{I} = A\sin\phi. \tag{2.19}$$

The joint distribution of $A$ and $\phi$ is calculated with the fundamental theorem for joint densities [Papoulis and Pillai, 1991, p.143]

$$p_{A,\phi}(A,\phi) = p_{\mathcal{R},\mathcal{I}}(A\cos\phi, A\sin\phi) \cdot |J|^{-1} \tag{2.20}$$

with the determinant of the Jacobian

$$|J|^{-1} = \begin{vmatrix} \partial\mathcal{R}/\partial A & \partial\mathcal{R}/\partial\phi \\ \partial\mathcal{I}/\partial A & \partial\mathcal{I}/\partial\phi \end{vmatrix} = \begin{vmatrix} \cos\phi & -A\sin\phi \\ \sin\phi & A\cos\phi \end{vmatrix} = A. \tag{2.21}$$

We achieve the joint distribution for the amplitude and phase

$$p_{A,\phi}(A,\phi) = \frac{A}{2\pi\sigma^2} e^{-\frac{A^2}{2\sigma^2}}, \tag{2.22}$$

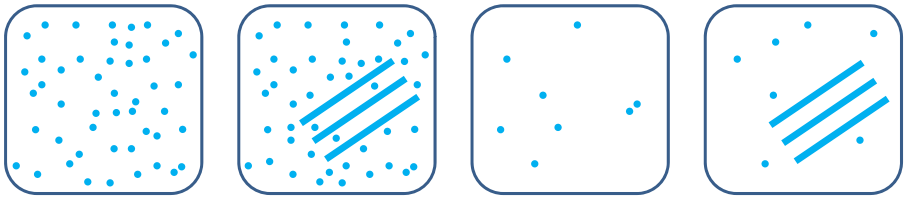


Figure 2.10: Illustration of scattering scenarios in a resolution cell. From left to right: (i) large number of scatterers, (ii) with coherent component, (iii) limited number of scatterers, (iv) with coherent component.

and further we find the marginal statistics of $A$ with

$$p(A) = \int_{-\pi}^{\pi} p(A, \phi)\, \mathrm{d}\phi = \frac{A}{\sigma^2} e^{-\frac{A^2}{2\sigma^2}} \tag{2.23}$$

for $A \geq 0$, which is known as the *Rayleigh* density function. We simplify the notation by writing $p(A)$ instead of $p_A(A)$.

### 2.4.2 Large Number of Scatterers with Constant Phasor

A different scenario is to consider once again a random phasor sum, but this time in combination with a known constant phasor. The constant phasor is due to a coherent component in the resolution cell, which may arise due to unresolved periodically located scatterers or strong specular scattering [Dutt, 1995]. This is illustrated in the second image in figure 2.10. The calculation of the real part changes in this case to

$$\mathcal{R} = A_0 + \frac{1}{\sqrt{N}} \sum_{n=1}^{N} a_n \cos(\phi_n) \tag{2.24}$$

with $A_0$ the amplitude of the known phasor. Since we assume again a large number of random phasors, the real and imaginary parts are asymptotically Gaußian. This time, however, with the mean of the real part being the known phasor. A similar deduction as above, presented in [Goodman, 2006], yields the following distribution for the amplitude

$$p(A) = \frac{A}{\sigma^2} e^{-\frac{A^2 + A_0^2}{2\sigma^2}} \cdot I_0\left(\frac{A A_0}{\sigma^2}\right) \tag{2.25}$$

for $A \geq 0$ and with $I_0$ the modified Bessel function of the first kind of order zero. This is the *Rician* density function derived in [Rice, 1945].

### 2.4.3 Limited Number of Scatterers

The Rayleigh density is a good model for the echo envelope if the number of scatters is large ($> 10$ per resolution cell). The Rice distribution models the scenario where there is an additional coherent backscatter. The drawback of both approaches is the

presumption of a high scatter density, which limits their applicability. In order to be able to model situations with lower scatterer density, once again a random walk is considered, but with a limited number of phasors, see third image in figure 2.10. In [Jakeman and Pusey, 1976], the negative binomial distribution is proposed to model the number of scatterers. This leads to the amplitude following a *K distribution*

$$p(A) = 2\left(\frac{A}{2}\right)^{\alpha} \frac{b^{\alpha+1}}{\Gamma(\alpha)} \cdot K_{\alpha-1}(bA) \tag{2.26}$$

with $b = \sqrt{\frac{4\alpha}{E[A^2]}}$ and $K_\alpha$ the modified Bessel function of the second kind of order $\alpha$. The parameter $\alpha$ can be used to model the effective number of scatterers per resolution cell. In the limit for $\alpha \to \infty$ this turns into the Rayleigh distribution. Assuming a biased random walk, the generalized K distribution is shown to be an appropriate model [Jakeman and Pusey, 1976].

### 2.4.4 Limited Number of Scatterers with Constant Phasor

While the K distribution accounts for low scatter densities, and the Rician distribution incorporates the presence of a coherent component, we will now describe a combination of both. The extension for obtaining a model for a small number of scatterers and a coherent signal is similar to the extension of the Rayleigh to the Rice distribution. The scenario is illustrated in the last image of figure 2.10. This extension leads to the *homodyned K distribution* [Dutt and Greenleaf, 1994]. The explicit formula is complicated and a more convenient way to represent the distribution is through the following integral

$$p(A) = \int_0^\infty p_r(A/x)p_\gamma(x)\,\mathrm{d}x, \tag{2.27}$$

with $p_r$ being the Rice distribution with variance $\frac{\sigma^2 x}{\alpha}$ and $p_\gamma$ the Gamma distribution

$$p_\gamma(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x}. \tag{2.28}$$

The K distribution is a special case of the homodyned K distribution. Further, the homodyned K distribution converges to the Rice distribution for $\alpha \to \infty$.

### 2.4.5 Nakagami Distribution

In the last sections, we have presented various scattering scenarios and we have seen that the homodyned K distribution is the most versatile model. A problem with this distribution is, however, the lack of a tractable density function. An alternative is the Nakagami distribution [Nakagami, 1960], proposed for modeling ultrasound envelope statistics in [Shankar, 2000]. The Nakagami distribution belongs to the exponential family and is defined as

$$p(A) = \frac{2m^m A^{2m-1}}{\Gamma(m)\omega^m} \exp\left(-\frac{m}{\omega}A^2\right), \forall A \in \mathbb{R}_+, \tag{2.29}$$

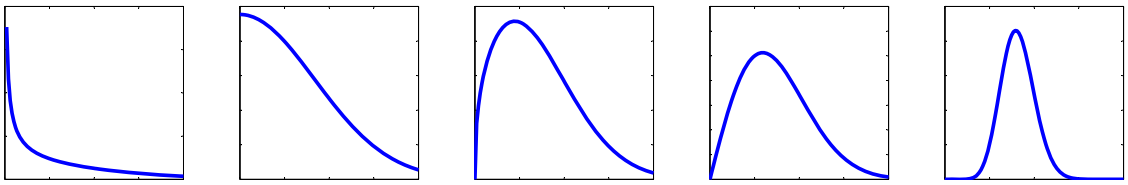


Figure 2.11: Plot of Nakagami distribution for various shape parameters $m$ and fixed scaling parameter $\omega = 0.7$. From left to right: $0 < m < 0.5$ (pre-Rician), $m = 0.5$ (generalized Rician), $0.5 < m < 1$ (generalized Rician), $m = 1$ (Rayleigh), $m > 1$ (Rician).

with shape parameter $m$ and scale parameter $\omega$. The main motivation for using the Nakagami distribution is to have an easy formula to calculate an approximation of the previously presented distributions. For $m = 1$, this corresponds to the Rayleigh distribution. For varying $m$, the Nakagami distribution can model pre- and post-Rayleigh as well as generalized Rician distributions [Shankar et al., 2001]. Further, it is a good approximation to the homodyned K distribution [Destrempes and Cloutier, 2010]. A drawback is that the parameters of Nakagami do not seem to have a physical meaning, in contrast to the K distribution, which may be of value for tissue characterization. Figure 2.11 illustrates the Nakagami distribution for various shape parameters and fixed scale.

The Nakagami distribution is applied in several studies. [Shankar et al., 2002] show that the parameters of the Nakagami distribution are suitable for tissue classification. [Larrue and Noble, 2011] suggest a variant of the previous classification method, using a small window kernel regression to guarantee locality. [Bouhlel and Sevestre-Ghalila, 2009] propose an ultrasound specific auto-model by embedding the Nakagami distribution into a Markov random field (MRF), facilitating the classification of cancerous breast tissue. Similarly, [Klein et al., 2011] developed a MRF-based feature descriptor for tissue classification and image registration. We apply the Nakgami distribution to assess the statistical properties of envelope data in chapter 3 and we present a similarity measure based on the Nakagami distribution in chapter 8.

Important to note is that if the amplitude is assumed to be Nakagami distributed, then the intensity $I = A^2$ is Gamma distributed [Papoulis and Pillai, 1991, p.95]. The Gamma density is

$$p(I) = \frac{m^m I^{m-1}}{\Gamma(m)\omega^m} \exp\left(-\frac{m}{\omega} I\right), \forall I \in \mathbb{R}_+, \tag{2.30}$$

following the notation in [Shankar, 2000]. Working with the Gamma distribution is sometimes easier than working with the Nakagami distribution because software environments, such as Matlab, have certain functionalities for Gamma functions integrated.

### 2.4.6 Statistics of B-mode Images

Most commonly, ultrasound data is not available in form of RF but B-mode data. The entire analysis provided in this chapter is focusing on the amplitude $A$, the result of the envelope detection. We have seen in section 2.3 that a sequence of processing steps is applied to the envelope to create B-mode images. It would be necessary to model each of the processing steps to make a clear prediction about the statistics on B-mode images. We have already mentioned that ultrasound manufacturers use proprietary processing steps to increase the image quality. Since speckle is perceived as noise, these processing steps focus on removing speckle. This questions the validity of the previously presented noise models, which characterize speckle in the images. Consequently, it is better to either directly work on envelope images or to create B-mode images oneself, to control the processing steps.

This is, however, not always possible because the access is limited to B-mode data. In the literature, several works were published that try to at least model part of the processing chain. While the interpolation of intensity values is difficult to incorporate and there is uncertainty about the applied smoothing filters, the interest lies in replicating the dynamic range compression. Therefore, a log-compression, as shown in equations (2.10) and (2.11), is assumed. Generally, one can distinguish between approaches that try to decompress the B-mode images and approaches that try to model the distribution of the B-mode. [Prager et al., 2003] decompress images to detect speckle with the homodyned K distribution. An iterative procedure is proposed to determine the multiplicative compression factor $\alpha$. [Kaplan and Ma, 1994] derive the Fisher-Tippett distribution for B-mode images, assuming Rayleigh envelope statistics. [Dutt and Greenleaf, 1996] derive an elaborate density function for B-mode, assuming envelope statistics following the K distribution. For ultrasound registration, the log-compression was integrated in the imaging model in [Cohen and Dinstein, 2002, Myronenko et al., 2009], further discussed in chapter 8.

# Chapter 3

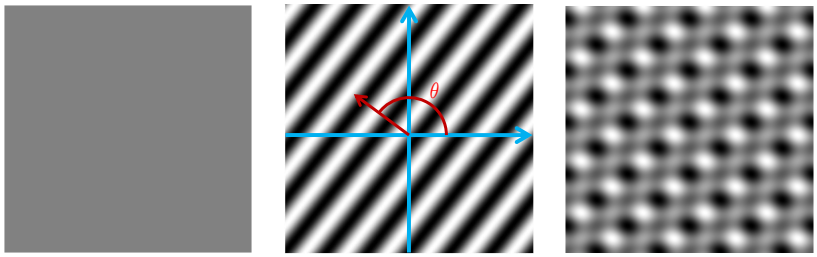
# 2D Analytic Signal on Ultrasound Images

*In ultrasound imaging, the local phase is applied for identifying structures in B-mode images and the local amplitude is used for envelope detection of RF data. Recently, the extension of the analytic signal from 1D to 2D, covering also intrinsic 2D structures, was proposed. We illustrate the advantages of this improved concept for RF demodulation by performing goodness-of-fit tests to a Nakagami distribution, indicating a clear improvement of statistical properties. Moreover, we demonstrate benefits for the estimation of local features on B-mode images.*

## 3.1 Introduction

The analytic signal (AS) enables to extract local, low-level features from images. It has the fundamental property of split of identity, meaning that it separates qualitative and quantitative information of a signal in form of the *local phase* and the *local amplitude*, respectively. These quantities further fulfill invariance and equivariance properties [Felsberg and Sommer, 2001], allowing for an extraction of structural information that is invariant to brightness or contrast changes in the image. Exactly these properties lead to a multitude of applications in computer vision and medical imaging, such as registration [Carneiro and Jepson, 2002, Grau et al., 2007, Mellor and Brady, 2005, Zang et al., 2007, Zhang et al., 2007], detection [Estepar et al., 2006, Mulet-Parada and Noble, 2000, Szilágyi and Brady, 2009, Xiaoxun and Yunde, 2006], segmentation [Ali et al., 2008, Hacihaliloglu et al., 2008, Wang et al., 2009a], and stereo [Fleet et al., 1991]. Phase-based processing is particularly interesting for ultrasound images because they are affected by significant brightness variations, as shown in [Grau et al., 2007, Hacihaliloglu et al., 2008, Mellor and Brady, 2005, Mulet-Parada and Noble, 2000], as well as, chapters 4 and 8.

For 1D, the local phase is calculated with the 1D analytic signal. For 2D, several extensions of the analytic signal are proposed, with the *monogenic signal* [Felsberg and Sommer, 2001] presenting an isotropic extension. The description of the signal's structural information (phase and amplitude) is extended by a geometric component,




(a) i0D (b) i1D (c) i2D

Figure 3.1: Illustration of 2D signals with different intrinsic dimensionality. For i1D, we show the local orientation $\theta$.

the *local orientation*. The local orientation indicates the orientation of intrinsic 1D (i1D) structures in 2D images, see figure 3.1. This already points to the limitation of the monogenic signal; it is restricted to the subclass of i1D signals. Recently, an extension to the monogenic signal, referred to as *2D analytic signal* [Wietzke et al., 2009], was proposed that permits the analysis of i2D signals. Therefore, the 2D signal analysis is embedded into 3D projective space, and a new geometric quantity, the *apex angle*, is introduced. The 2D analytic signal also has the advantage of more accurately estimating local features from i1D signals [Wietzke et al., 2009].

In this chapter, we illustrate the advantages of the calculation of the 2D analytic signal for RF and B-mode ultrasound images. Instead of performing the demodulation of RF signals for each scan line separately, as described in section 2.3, we perform the demodulation in its 2D context with 2D Hilbert filters of first- and second-order. This leads to advantages in the envelope detection. Since all further processing steps of the creation of the B-mode image are based on the envelope, the improvement of the 2D envelope detection propagates to the quality of the B-mode image. Moreover, the result from the 2D envelope detection bears better statistical properties, as we illustrate with goodness-of-fit tests towards a Nakagami distribution, with its implications to classification, segmentation, and registration. Finally, we evaluate the advantages of the 2D analytic signal for estimating local features on B-mode images. All experiments are performed on clinical ultrasound images.

## 3.2 2D Analytic Signal

There are various concepts to analyze the phase of signals, such as Fourier phase, instantaneous phase, and local phase [Granlund and Knutsson, 1995]. We are primarily interested in the last two. For 1D signals, $g \in L^2(\mathbb{R})$, the instantaneous phase is defined as the argument of the analytic signal, $\arg(g + i \cdot \mathcal{H}\{g\})$, with $\mathcal{H}$ being the Hilbert transform. Since real signals consist of a superposition of multiple signals of different frequencies, the instantaneous phase, although local, can lead to wrong estimates. The signal has to be split up into multiple frequency bands, by means of bandpass filters, to achieve meaningful results, as further described in

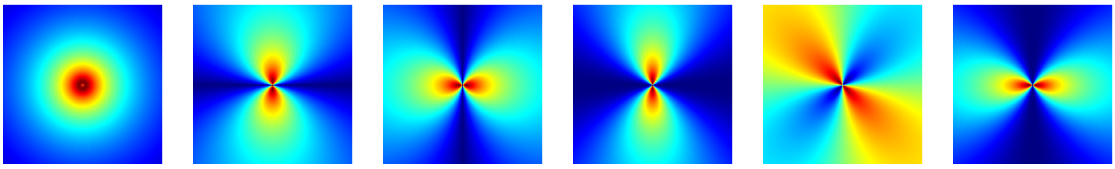


Figure 3.2: Magnitude of 2D Hilbert transforms with log-Gabor kernels in frequency domain. From left to right: $B$, $B \odot H^1_x$, $B \odot H^1_y$, $B \odot H^2_{xx}$, $B \odot H^2_{xy}$, $B \odot H^2_{yy}$.

section 3.2.2.

Considering 2D signals, $f \in L^2(\mathbb{R}^2)$, the intrinsic dimension expresses the number of degrees of freedom to describe local structures [Zetsche and Barth, 1990]. Intrinsic zero dimensional (i0D) signals are constant signals, i1D signals are lines and edges, and i2D are all other patterns in 2D, see figure 3.1. The monogenic signal is restricted to i1D signals and calculated with the two-dimensional Hilbert transform, also referred to as the Riesz transform. In the frequency domain, the first-order 2D Hilbert transform is obtained with the multiplication of

$$H^1_x(\mathbf{u}) = i \cdot \frac{x}{||\mathbf{u}||}, \quad H^1_y(\mathbf{u}) = i \cdot \frac{y}{||\mathbf{u}||}, \quad \mathbf{u} = (x, y) \in \mathbb{C}\backslash\{(0,0)\} \tag{3.1}$$

with $i = \sqrt{-1}$. For the calculation of the 2D analytic signal, higher order Hilbert transforms are used [Wietzke et al., 2009]. The Fourier multipliers of the second-order Hilbert transform [1] are

$$H^2_{xx}(\mathbf{u}) = -\frac{x \cdot x}{||\mathbf{u}||^2}, \quad H^2_{xy}(\mathbf{u}) = -\frac{x \cdot y}{||\mathbf{u}||^2}, \quad H^2_{yy}(\mathbf{u}) = -\frac{y \cdot y}{||\mathbf{u}||^2}, \tag{3.2}$$

again with $\mathbf{u} = (x, y) \in \mathbb{C}\backslash\{(0,0)\}$. In contrast to [Wietzke et al., 2009], we do not present the formulas of the Hilbert transforms in spatial but in frequency domain, which is more versatile for filtering, see section 3.2.2. Throughout the chapter we use upper case letters for filters and signals in frequency domain, and lower case ones for their representation in spatial domain.

### 3.2.1 Structural and Geometrical Features

The proposed extension of the 2D analytic signal is obtained by an embedding in 3D projective space. This allows for a differentiation of geometrical features (local orientation, local apex) and structural features (local phase, local amplitude). The filtered signal $F_p$, the first-order Hilbert transformed signals $F_x, F_y$, and the second-order Hilbert transformed signals $F_{xx}, F_{xy}, F_{yy}$ are calculated with the bandpass filter $B$ and the pointwise multiplication $\odot$ in frequency domain as

$$\begin{bmatrix} F_p \\ F_x \\ F_y \end{bmatrix} = \begin{bmatrix} B \odot F \\ H^1_x \odot B \odot F \\ H^1_y \odot B \odot F \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} F_{xx} \\ F_{xy} \\ F_{yy} \end{bmatrix} = \begin{bmatrix} H^2_{xx} \odot B \odot F \\ H^2_{xy} \odot B \odot F \\ H^2_{yy} \odot B \odot F \end{bmatrix}. \tag{3.3}$$

[1] We want to thank the authors of [Wietzke et al., 2009] for discussions.

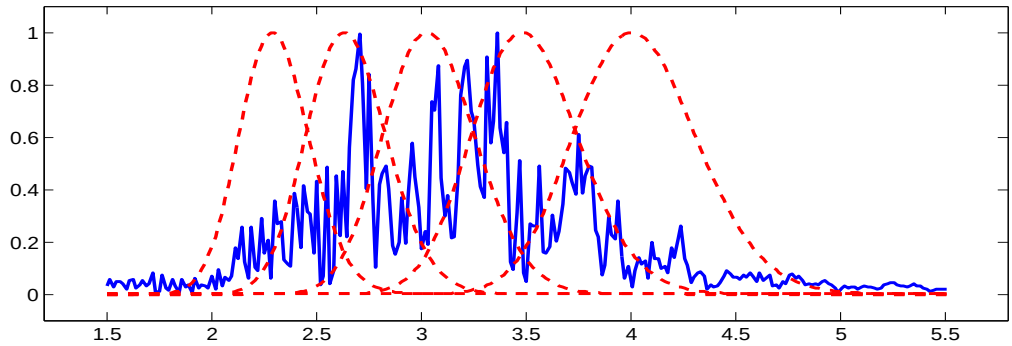



Figure 3.3: Log-Gabor filter bank consisting of 5 filters (red) and ultrasound signal spectrum (x-axis: frequency in MHz). Ultrasound acquisition frequency: 3.3 MHz.

We illustrate the Hilbert transforms in frequency domain multiplied with log-Gabor bandpass filters in figure 3.2. In order to enable an interpretation of second-order Hilbert transformed signals in projective space, an isomorphism between the Hesse matrix and a vector valued representation is used [Wietzke et al., 2009], leading to $f_s = \frac{1}{2}[f_{xx} + f_{yy}]$, $f_+ = f_{xy}$, and $f_{+-} = \frac{1}{2}[f_{xx} - f_{yy}]$.

Finally, the local features are calculated as follows. The apex angle $\alpha$, which differentiates between features of different intrinsic dimensionality, is

$$\alpha = \arccos \frac{\sqrt{f_+^2 + f_{+-}^2}}{||f_x||}. \tag{3.4}$$

With the apex angle, the homogeneous signal component $f_h$ of the signal $f_p$ in projective space is defined as

$$f_h = \sqrt{\frac{1 + \cos\alpha}{2}}. \tag{3.5}$$

The local orientation $\theta$, local phase $\phi$, and local amplitude $A$ are calculated with

$$\theta = \frac{1}{2} \arctan \frac{f_+}{f_{+-}}, \tag{3.6}$$

$$\phi = \operatorname{atan2}\left(\sqrt{[f_h^{-1} f_x]^2 + [f_h^{-1} f_y]^2}, f_p\right), \tag{3.7}$$

$$A = \frac{1}{2}\sqrt{f_p^2 + [f_h^{-1} f_x]^2 + [f_h^{-1} f_y]^2}. \tag{3.8}$$

For i1D signals, the homogeneous component is $f_h = 1$, and the formulas above reduce to the ones known from the monogenic signal.

### 3.2.2 Frequency Selection

Each signal $f$ can be described with the Fourier series, decomposing the signal into components of different frequencies, each one having its own phase and amplitude. The direct application of the Hilbert transform on the original signal, which

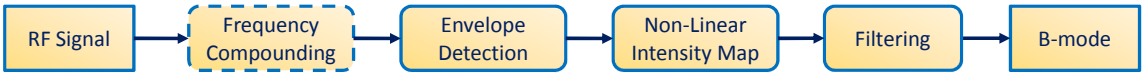



Figure 3.4: Exemplar ultrasound processing pipeline for RF to B-mode conversion.

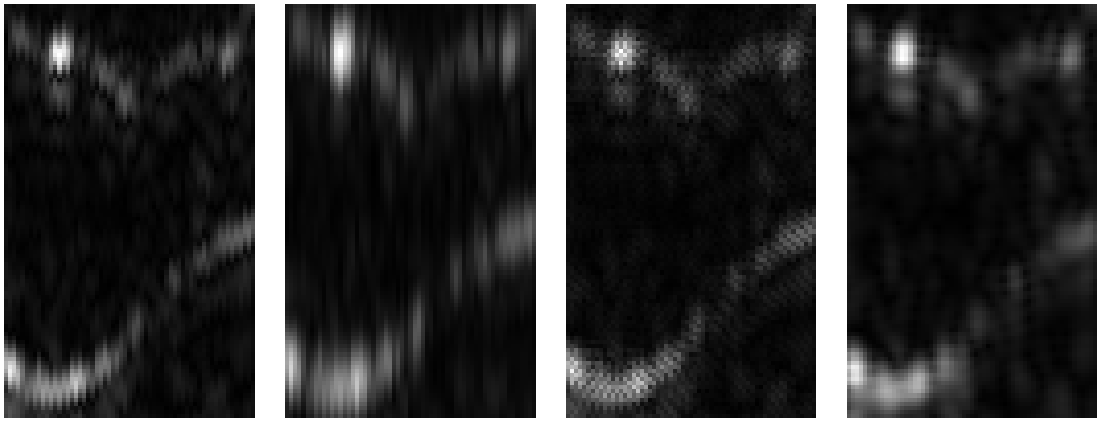
presents an accumulation of local signals from different frequencies, does therefore not adequately extract local features. Theoretically, we would need to calculate the analytic signal for infinitely narrow bandwidths, *i.e.*, Dirac deltas in the frequency domain. Following the uncertainty principle this results in filters with global support. Bandpass filters present an appropriate approximation for localization in spatial and frequency domain.

[Felsberg and Sommer, 2001] apply the difference of Poisson kernels for frequency selection. An interesting property of the Poisson filter is that it creates a linear scale-space [Felsberg and Sommer, 2004]. Another filter that is commonly applied, especially in ultrasound, is the log-Gabor filter [Boukerroui et al., 2004, Grau et al., 2007, Hacihaliloglu et al., 2008, Mulet-Parada and Noble, 2000]. Also in our analysis on ultrasound images, we achieve better results with the log-Gabor filter so that we concentrate on it in the following. A drawback of the log-Gabor filter is, however, that it has no analytic expression in the spatial domain. This is also the reason why we present the Hilbert transforms in equations (3.1) and (3.2) in frequency and not in spatial domain, as it is done in [Wietzke et al., 2009].

Important for the design of the filter bank is to create filters, so that the transfer function of each filter overlaps sufficiently with its neighbors, in order to have a uniform coverage of the spectrum. A filter bank with five log-Gabor filters is illustrated in figure 3.3. A study of alternative bandpass filters is presented in [Boukerroui et al., 2004]. For the further analysis, it is either possible to focus on the signal at one specific scale, or accumulate all responses from various scales, as it is *e.g.* done for the phase congruency [Kovesi, 1999].

## 3.3 2D Analytic Signal on RF Data

We saw in chapter 2 that the pipeline of the RF to B-mode conversion consists of multiple steps, including amongst others demodulation, non-linear intensity mapping, and filtering. Figure 3.4 illustrates a reduced processing chain. The demodulation is one of the central parts and extracts the information-bearing signal from a modulated carrier wave. In ultrasound processing, the demodulation is commonly performed by an envelope detection. Hereby, the amplitude of the analytic signal is calculated for each of the 1D scan lines separately. Interestingly, calculating the amplitude of the 1D analytic signal is equivalent to the *instantaneous amplitude*. In the literature of ultrasound imaging, it is noted that the quality of ultrasound images can be increased by multi-frequency decomposition and compounding of the received signal, simply referred to as *frequency compounding* [Cincotti et al., 2001]. This is equivalent to the *local amplitude* estimation. This constitutes an interesting analogy, between the advantage of the frequency compounded signal to the normal one, on the one



(a) 1D AS (b) 1D ASF (c) 2D AS (d) 2D ASF

Figure 3.5: Magnified region of envelope detected 2D image for various envelopes.

hand, and the advantage of the local amplitude in comparison to the instantaneous amplitude, on the other hand. We have neither seen this analogy noted in the literature before, nor the application of local amplitude and local phase techniques to RF data.

In contrast to the usually separate processing of each scan line, we consider all scan lines at once and construct the 2D analytic signal to estimate the local amplitude. This enables improved envelope detection because the signal is analyzed in its 2D context by also considering information in lateral direction. The balance between influence from lateral and axial direction can be adjusted by the bandwidth in each direction of the bandpass filter, where the smaller spacing in axial direction should be considered accordingly.

### 3.3.1 Envelope and B-mode Results

We perform experiments on multiple RF images acquired from the neck with a linear transducer at 3.3 MHz. The sampling frequency of the RF data is 40 MHz. We compare the envelope detection for: (i) 1D analytic signal (1D AS), (ii) 1D analytic signal with filter bank (1D ASF), (iii) monogenic signal (MS), and (iv) monogenic signal with filter bank (MSF), (v) 2D AS, and (vi) 2D ASF. Exemplarily, we show the frequency spectrum of one scan line together with the log-Gabor filter bank in figure 3.3. We present magnified regions of the various envelope images in figure 3.5. Note that we do not show the results of the MS, because the more interesting improvement is for 2D AS. However, we include them into the analysis of noise statistics in section 3.3.2. We can clearly observe that the 2D analytic signal leads to a more accurate and consistent extraction of structures. This becomes particularly clear on the circular structure on the top left, which appears rather ellipsoidal on the estimates from the 1D analytic signals. We also note the positive influence of the filter bank for the estimation of the 2D analytic signal.

We perform an RF to B-mode conversion of local amplitude images $A$ with a

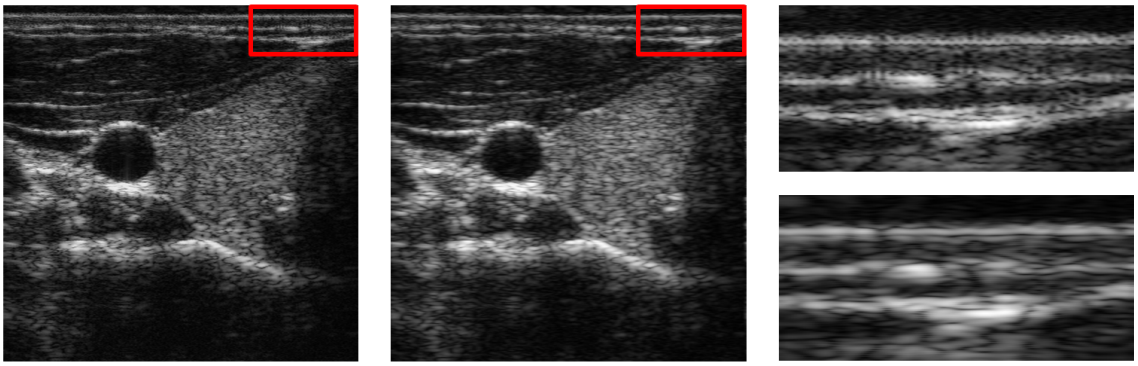
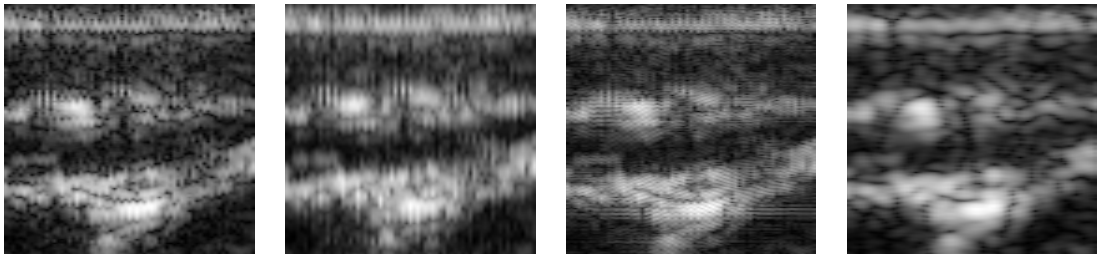


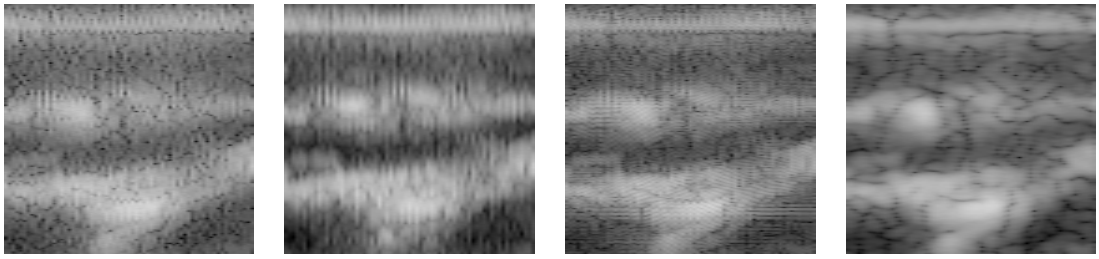
Figure 3.6: Log-compressed envelopes from 1D ASF (left) and 2D ASF (middle). Magnification of red region (top right: 1D AS, bottom right: 2D AS).



(a) 1D AS (b) 1D ASF (c) 2D AS (d) 2D ASF

Figure 3.7: Magnified regions of images after log-compression.

log-compression including a translation of 25, $\log(A + 25)$. The results for 1D ASF and 2D ASF are shown in figure 3.6, together with a magnification of the red region. Further results for the various envelopes are presented in figure 3.7. The B-mode image resulting from the 2D analytic signal clearly shows more consistent structures and less noise. Typically, further filtering steps are applied to the log-compressed image to improve its visual appearance. These processing steps are proprietary to the manufacturer and generally not publicly accessible. Ultrasonix (Redmond, Canada), however, distributes a particular research system with a specific SDK including their post-processing filter, called MUCRO. We apply MUCRO to the log-compressed images, with the results shown in figure 3.8. Even after the application of MUCRO, the advantages of the images from the 2D analytic signal are clearly visible. This is not self-evident because the post-processing methods are designed to be applied to 1D envelope detected images, still leaving room for improvement by adapting the post-processing to 2D envelope estimation. Finally, one of the reasons for applying the post-processing filtering is to establish consistency between scan lines, which we already achieve by the 2D envelope detection.



(a) 1D AS (b) 1D ASF (c) 2D AS (d) 2D ASF

Figure 3.8: Magnified regions of images after log-compression and MUCRO.

### 3.3.2 Analysis of Envelope Statistics

Next to the visual assessment of the 2D envelope detection, we also analyze the statistical properties of the data. In section 2.4, we discussed different statistical distributions to model ultrasound data. Among them there is Rayleigh, Rician [Shankar et al., 1993, Wagner et al., 1983], K distribution, generalized K distribution, homodyned K distribution, pre-Rician K [Jakeman and Pusey, 1976, Lord, 1954], as well as, Rician Inverse of the Gaußian [Eltoft, 2003]. As already mentioned, the homodyned K distribution allows modeling various scattering scenarios, but its inherent complexity limits the practical applicability. In order to address this issue, the Nakagami distribution [Nakagami, 1960] was proposed, because it admits an explicit analytical expression. It is used in various applications to model backscatter characteristics of US envelope data for segmentation and classification, see [Destrempes et al., 2009, Shankar et al., 2002] and references therein. In the following, we analyze the effects of the 2D envelope detection on the speckle statistics with the Nakagami model. In particular, we quantify the impact of the 2D analytic signal with goodness-of-fit (GOF) tests, and show the potential for the aforementioned applications based on example images.

The Nakagami distribution with shape $m$ and scale $\omega$ parameters is

$$p(x \mid m, \omega) = \frac{2m^m x^{2m-1}}{\Gamma(m)\omega^m} \exp\left(-\frac{m}{\omega}x^2\right), \forall x \in \mathbb{R}_+. \tag{3.9}$$

A goodness-of-fit test evaluates if the data $d_1, \ldots, d_n$, under the assumption of i.i.d. samples, comes from the given theoretical probability distribution $p$ [D'Agostino and Stephens, 1986]. Note that conventional GOF tests are restricted to the case of single distributions. For inhomogeneous regions in the image, however, a mixture of Nakagami is more appropriate, see figure 3.10 for an illustration of a misfit of a single Nakagami to mixture Nakagami data as well as a perfect mixture fit. Consequently, we can only achieve good results with the GOF test on homogeneous image regions. The mixture case has to be further evaluated, with similar results to be expected.

For the GOF test, the range of the data is partitioned into $M$ bins $\beta_i$, $i = 1, \ldots, M$, with $N_i$ and the number of samples per bin. Moore suggests to divide the data into $M = 2n^{\frac{2}{5}}$ bins [D'Agostino and Stephens, 1986]. Furthermore, we assume the bins

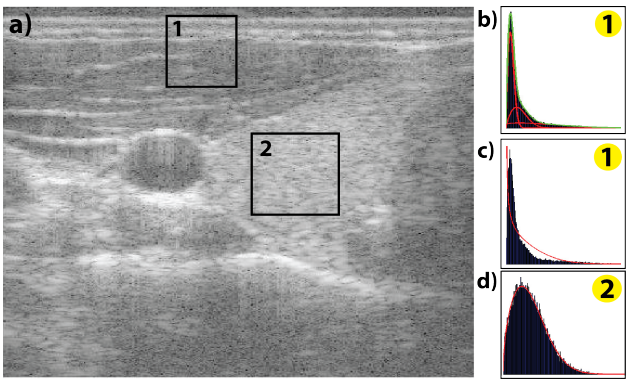



Figure 3.9: RF image with sample distributions estimated for two areas. Region 1 contains a mixture of Nakagami, region 2 a single Nakagami. Whereas MLE can fit nicely in region 2 (d) it expectedly performs poorly in region 1 (c), that can only be represented properly by mixture (b).

to be equiprobable as suggested in [Bock and Krischer, 1998]. In this regard, we let $p_i$ be the integral of the distribution in the range $\beta_i$ given the parameters of the distribution $\theta = \{m, \omega\}$

$$p_i = \int_{\beta_i} p(x \mid \theta)\, \mathrm{d}x. \tag{3.10}$$

Hence, $p_i$ expresses the likelihood of a sample to be in the bin $\beta_i$ (identical for all bins). The test statistics underlying the GOF test is the sum of differences between observed and expected outcome frequencies

$$X^2 = \sum_{i=1}^{M} \frac{(N_i - np_i)^2}{np_i}. \tag{3.11}$$

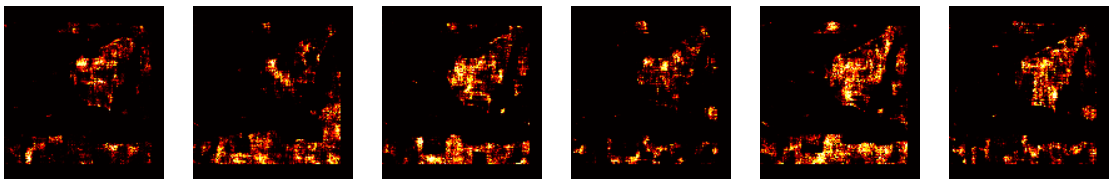
This yields a quadratic form in $N_i$ that has approximately a $\chi$ distribution with $M - N - 1$ degrees of freedom and $N = 2$ the number parameters of the distribution. In order to assess the GOF quantitatively, we employ the P-value based hypothesis test. The P-value indicates the significance level of accepting the null hypothesis $H_0$. In our case, $H_0$ is the hypothesis that the observations are Nakagami distributed, leading to the following calculation of the P-value

$$P = \int_{X^2}^{\infty} \chi^2(M - N - 1)\, \mathrm{d}x, \tag{3.12}$$

employing equation (3.11) as the lower bound of integration.

### 3.3.3 Rao-Robson Statistic

Given the data, we first have to estimate the parameters $m, \omega$ of the Nakagami distribution before the GOF test is performed. This is, however, opposing the general



(a) 1D AS (b) 1D ASF (c) MS (d) MSF (e) 2D AS (f) 2D ASF

Figure 3.10: The P-values are calculated for all patches of an envelope image. Pixel brightness indicates P-value. We perform the calculation for various envelope detection techniques. Comparing the P-value images to the B-mode image in figure 3.9, we see that the bright regions correspond to homogeneous regions in the US image.

assumption that the parameters of the distribution are a-priori given before the test is performed. Therefore, another quadratic form in $N_i$ has to be used, with the Rao-Robson statistic being one possibility [D'Agostino and Stephens, 1986]. Considering the parametric form of the distribution $p(x|\theta)$ and the maximum likelihood estimate $\hat{\theta}$, the Rao-Robson statistic is

$$RR = V^\top(\hat{\theta})Q(\hat{\theta})V(\hat{\theta}) \tag{3.13}$$

with

$$V(\theta) = \frac{N_i - np_i}{(np_i)^{1/2}} \tag{3.14}$$

$$Q(\theta) = I + D(\theta)[J(\theta) - D^\top(\theta)D(\theta)]^{-1}D^\top(\theta) \tag{3.15}$$

$$D_{ij}(\theta) = p_i(\theta)^{\frac{1}{2}}\frac{\partial p_i(\theta)}{\partial \theta_j} \tag{3.16}$$

$J(\theta)$ is the $N \times N$ Fisher information matrix and $I$ is an $M \times M$ identity matrix. The Rao-Robson statistic is $\chi^2$ distributed with $M - N - 1$ degrees of freedom, leading to P-values computed by

$$P = \int_{RR}^{\infty} \chi^2(M - N - 1)\,\mathrm{d}x \tag{3.17}$$

with the Rao-Robson statistic $RR$ as lower bound of integration.

### 3.3.4 Statistical Results

We perform the Rao-Robson GOF test on local patches, densely throughout the image. Plotting the results for all patches therefore creates a new image with the intensity values being the P-values. We show these images in figure 3.10 for the various envelope detection schemes. The brighter the images, the higher the P-values, and consequently the better for statistical applications because we achieve better

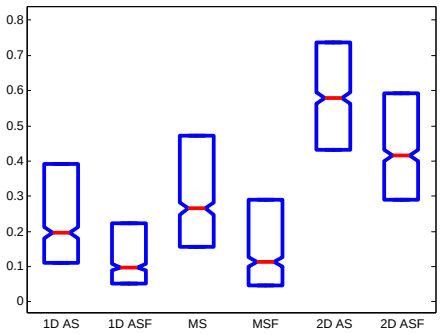



Figure 3.11: Box plot of P-values for different envelope detections.

fits. We note that the bright regions are corresponding to the homogeneous areas in the ultrasound image because only these areas are appropriately modeled with a single distribution, as discussed previously. Additionally, we calculate the statistics of the P-values, visualized in the box plot of figure 3.11. The red line indicates the median and the box is constructed from the interquartile range. Our results therefore show that the envelope detection without the filter bank produces better fits, which makes sense, because log-Gabor filters influence the distribution. More importantly, however, we note the improvement from 1D AS to MS, and further from MS to 2D AS. This shows on the one hand, the advantage of applying 2D Hilbert transforms in contrast to 1D ones, and on the other hand, the advantage of the 2D analytic signal in contrast to the monogenic signal. This confirms the visually improved results for 2D envelope detection from the previous section.

## 3.4 2D Analytic Signal on B-mode Images

Next to the benefits of the 2D analytic signal for the demodulation of RF data, it also allows for a more accurate estimation of local features on B-mode images [Wietzke et al., 2009]. This has the potential to increase the quality of follow-up applications such as registration [Grau et al., 2007, Mellor and Brady, 2005, Zhang et al., 2007], segmentation [Hacihaliloglu et al., 2008], and detection [Mulet-Parada and Noble, 2000], which use the local features as input. To demonstrate this, we calculate the local orientation on B-mode images showing a biopsy needle. In figure 3.12, we illustrate the local orientation that is estimated from the monogenic signal and the 2D analytic signal, both with filtering. The estimation from the monogenic signal shows no consistent orientation information in the region of the needle. In contrast, the improved concept of the 2D analytic signal indicates a consistent result.

## 3.5 Conclusion

We demonstrated that the application of the 2D analytic signal has multiple advantageous for RF and B-mode data. The demodulation of RF signals with the 2D

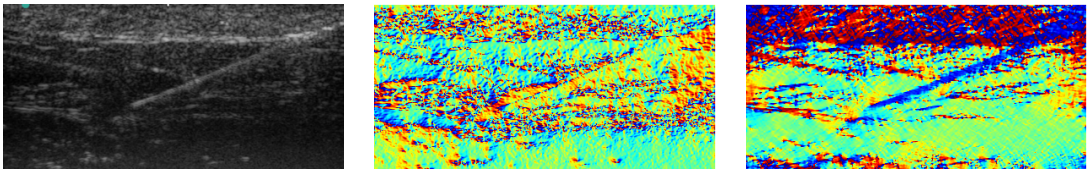


Figure 3.12: Ultrasound image with biopsy needle (left). Calculated local orientation for monogenic signal (middle) and 2D analytic signal (right).

analytic signal enables a more consistent extraction of structures, because the signal is analyzed in its natural 2D context. We further showed that the improved envelope detection enables the creation of B-mode images of enhanced quality. To validate this, we applied a proprietary post-processing filtering for ultrasound on the log-compressed images and compared the result of 1D and 2D analytic signal. Moreover, we validated the improved statistical properties of envelope data resulting from the 2D analytic signal by performing goodness-of-fit tests to a Nakagami distribution. Finally, the advanced signal model of the 2D analytic signal leads to benefits in the estimation of local features in B-mode images, as we have illustrated for the case of needle detection.

For the demodulation, we focused on scans from a linear transducer. For curved linear transducers, the application of 2D Hilbert transforms without a previous scan conversion can be achieved with the polar Fourier transform [Averbuch et al., 2006]. This, together with the incorporation of mixture models in the statistical analysis, remains as future work.

# Chapter 4

# Towards Acoustic Impedance Estimation

*Reflection of sound waves, due to acoustic impedance mismatch at interfaces of media, is a principal physical property which allows for imaging with ultrasound. In this chapter, we investigate the reconstruction of the acoustic impedance from ultrasound images. This reconstruction is an alternative to common spatial compounding methods that also combine the information from multiple images. We use local phase information to determine regions of strong reflection from an ultrasound image. A model of ultrasound imaging is used for computing the acoustic impedance (up to scale) from areas of high reflectivity. The acoustic impedance image can either be directly visualized or be used in the simulation of ultrasound images from an arbitrary point of view.*

## 4.1 Introduction

Spatial compounding of several views, acquired from different positions, helps to reduce shortcomings of ultrasound imaging, as detailed in chapter 2. When working with 2D US images, compounding from different positions poses a problem because all the scans have to be in one plane. Therefore, multi-angle compounding with beam steering is typically performed, where the probe remains fixed [Wilhjelm et al., 2004]. Moving to 3D imaging, compounding from different positions offers much more flexibility [Grau et al., 2007, Wachinger et al., 2007]. The prerequisite for spatial compounding is to know the relative positions of the acquired images. This can either be obtained by tracking the ultrasound transducer or by image registration. In part III about ultrasound mosaicing, we focus on registration strategies for the correct alignment of US images. In this chapter, we assume the alignment to be known and concentrate on the last stage of mosaicing, the combination of the information.

The major problem in compounding ultrasound images is the view-dependent nature of ultrasound. If we would deal with several CT images of the same object, the compounding could be done by calculating the mean value. The dominant features of ultrasound due to reflection and attenuation are, however, dependent on

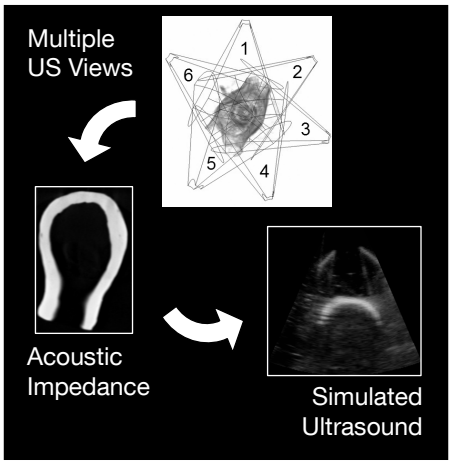



Figure 4.1: Schematic illustration of spatial compounding with acoustic impedance estimation. First, the impedance is estimated from multiple ultrasound images. Second, an ultrasound image is simulated from an arbitrary point of view.

the direction of insonification. Averaging intensity values is not optimal because strong echoes from small incident angles (transducer perpendicular to the surface) would be degraded by weak echoes from large incidence angles. Therefore, in the literature, several methods for spatial compounding have been proposed, which we are going to discuss in section 4.1.2. Instead of following one of the established methods, we introduce a novel approach, which is based on the estimation of the acoustic impedance of the imaged scene. From each image, we will reconstruct an acoustic impedance image, which we subsequently average to get an estimation for the whole imaged area. The reconstructed acoustic impedance images can either directly be presented to the physician or can be used for the simulation of ultrasound images from an arbitrary point of view, see Figure 4.1.

In mathematical terms, we deal with an inverse problem, because we infer cause from effect. We probe the living tissue with sound waves and measure the effect in form of reflections. The cause of the reflections is changing physical properties - the change in acoustic impedance - which we try to estimate. The reconstruction of tissue properties is common in ultrasound tomography [Greenleaf and Bahn, 1981], where dedicated devices are used. In [Gemmeke and Ruiter, 2007], 384 sending and 1600 receiving transducers are arranged around rings, with the object being placed in the middle. Such setups allow, next to measuring reflection, also assessing transmission and scattering. Although such systems are very interesting, they have not yet passed the stage of research prototypes. Consequently, we investigate for one of the first times the possibilities of using standard ultrasound probes for the reconstruction of the acoustic impedance. This is challenging because fewer measurements are available.

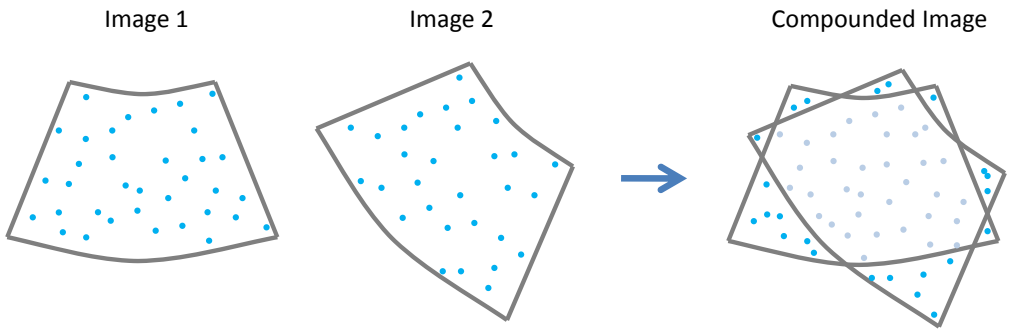



Figure 4.2: Illustrated is the spatial compounding of two ultrasound images. The SNR can be improved by averaging the intensity information of correctly aligned images, because speckle is reduced. [Jespersen et al., 1998] indicate that images are virtually fully decorrelated for relatively small rotation angles of around 10 degrees.

### 4.1.1 Clinical Value of Compounding

The clinical value of US compounding is mainly a result of increased quality and extended field of view (FOV) of the images presented to the physician. We have already discussed advantages that come along with the extended FOV in chapter 1. The quality can be improved by compounding images from different positions, because the direction dependent speckle noise is reduced and therefore the SNR is increased [Wilhjelm et al., 2004], see figure 4.2. Moreover, occlusion artifacts below structures with high acoustic impedance can be removed and the boundary continuity is enhanced. The positive effects of spatial compounding for diagnosis of atherosclerotic plaques [Huber et al., 2002, Jespersen et al., 2000] and breast cancer [Anderson et al., 1997] have already been reported. It also helps for administering epidural anesthesia by especially improving the depiction of key structures such as ligamentum flavum and epidural space [Tran et al., 2008]. [Grau and Noble, 2005] work on the combination of several acquisitions from different positions of the heart. It is not possible to depict the whole heart in a single acquisition; however, scans from particular acoustic windows can be acquired to show specific cardiac structures. The combination of these acquisitions into a single volume can be of great benefit in clinical practice.

### 4.1.2 Related Work

There are several articles that address ultrasound compounding with a few of them taking the view dependency of ultrasound into account. [Wilhjelm et al., 2004] use multi-angle spatial compounding with beam steering, for which the transducer stays at the same spatial location. They are able to reduce the angle dependency and speckle noise by combining multiple images. They compare a number of methods for compounding including mean, median, root-mean-squared value, and geometric mean. The highest SNR is achieved when using the mean. In a recent work also

based on beam steering, [Tran et al., 2008] improve the spatial compounding by using a combination of median- and gradient-based approaches. The median is used if at a certain location more than half of the images have a high feature-content, otherwise the gradient-weighted average is calculated. The problem we see with this approach is the use of thresholding to decide whether a pixel has a high feature-content or not. [Behar et al., 2003] propose a new method for spatial compounding by using three ultrasound transducers simultaneously. The transducer in the middle acts as sender and receiver, the remaining two only act as receiver. With their method they were able to improve visibility, detectability, and lateral resolution. During their experiments, various averaging methods were investigated, with the best results for the averaging of intensities.

[Leotta and Martin, 1999] propose a weighting scheme based on the incidence angle of the ultrasound beam on a reflecting surface. This technique leads to significantly improved results in comparison to using the mean value, but is based on an initial fitting of a surface to the data, which is a challenge for complex images. [Grau and Noble, 2005] use multiscale information about local structure definition and orientation to weight the contributions of different images. They obtain these image characteristics by calculating the local phase, as described in chapter 3, which is to a certain extent invariant to image contrast, making it particularly interesting for US images. While this approach is interesting for image registration [Grau et al., 2006], the compounding is rather cumbersome [Grau and Noble, 2005]. As can be seen, ultrasound compounding is a non-trivial exercise and still an active field of research.

## 4.2 Impedance Estimation Framework

Core to our method is the estimation of the acoustic impedance of the region depicted in the ultrasound image. As we will see, acoustic impedance images are related to CT attenuation values expressed in Hounsfield units. Further, they do no longer exhibit view-dependent artifacts or emphasized boundaries. We first formulate the problem as maximum likelihood estimation, second we present the ultrasound simulation function, third we identify regions of high reflectivity, and finally, we describe visualization approaches.

### 4.2.1 Maximum Likelihood Estimation

We formulate the acoustic impedance estimation as maximum likelihood (ML) estimation. For this, we take advantage of recently proposed methods for ultrasound simulation [Wein et al., 2008, Shams et al., 2008]. They allow us to define a US simulation function $s$, producing one of the $n$ simulated US images $\hat{\mathcal{U}} = \{\hat{u}_1, \ldots, \hat{u}_n\}$, by taking the corresponding transformation in $\mathcal{T} = \{T_1, \ldots, T_n\}$ and the acoustic impedance image $z$ as input

$$s : (z, T_j) \longmapsto \hat{u}_j. \tag{4.1}$$

(a) Original US image (b) Phase image (c) Threshold of phase image (d) Impedance estimation (e) Smooth imp. estimation

Figure 4.3: Processing steps for acoustic impedance estimation of a clay model.

The likelihood function, which indicates how well the simulated US images $\hat{\mathcal{U}}$ match the real ones $\mathcal{U} = \{u_1, \ldots, u_n\}$, is

$$\begin{aligned} \mathcal{L}(z) &= P(\mathcal{U}|z, \mathcal{T}, \varepsilon) && (4.2)\\ &= \prod_j P(u_j|z, T_j, \varepsilon) && (4.3)\\ &= \prod_j P(u_j - s(z, T_j) = \varepsilon), && (4.4) \end{aligned}$$

with the random variable $\varepsilon$ modeling the noise and incorporating the assumption of independent US images. In order to proceed with the ML estimation, $\arg\max_z \mathcal{L}(z)$, we have to choose a distribution for the noise.

In section 2.4, we described several distributions for modeling ultrasound speckle. For the acoustic impedance estimation, we focus on large scale interfaces, which are not well characterized by the previously presented noise models. Moreover, we do not work on RF but on B-mode images. Due to these problems in correctly modeling the distribution, we decided to remove speckle in a preprocessing step and assume a Gaußian distribution, which leads to the following least-squares formulation

$$\log \mathcal{L}(z) \propto -\frac{1}{n} \sum_{i=1}^{n} \left(u_j - s(z, T_j)\right)^2 . \tag{4.5}$$

In the next section, we describe details of the ultrasound simulation function $s$.

### 4.2.2 Ultrasound Simulation

As we have discussed in section 2.2, ultrasound imaging can be described by the reflection at tissue interfaces and the exponential loss of intensity within the tissue. The reflected intensity at a location $\mathbf{x}$ is calculated by running along the scan line with direction $\mathbf{d}$ and evaluating

$$I_r(\mathbf{x}) = I_i(\mathbf{x}) \cdot \cos^m \theta_i(\mathbf{x}) \cdot \left( \frac{z(\mathbf{x}) - z(\mathbf{x} - \Delta\mathbf{d})}{z(\mathbf{x}) + z(\mathbf{x} - \Delta\mathbf{d})} \right)^2 , \tag{4.6}$$

similar to equation (2.8). The distance between scan line points is indicated by $\Delta\mathbf{d}$. The incidence angle $\theta_i$, which is the angle between the US beam and the normal of

the surface, is calculated with the scalar product

$$\cos\theta_i(\mathbf{x}) = \left|\mathbf{d}.\frac{\nabla z(\mathbf{x})}{|\nabla z(\mathbf{x})|}\right| \tag{4.7}$$

where $\nabla$ is the spatial derivative operator. In order to perform the optimization, we make the assumption that the incidence angle for the impedance and ultrasound image are roughly the same

$$\cos\theta_i(x) = \left|\mathbf{d}.\frac{\nabla z(\mathbf{x})}{|\nabla z(\mathbf{x})|}\right| \approx \left|\mathbf{d}.\frac{\nabla u(\mathbf{x})}{|\nabla u(\mathbf{x})|}\right|. \tag{4.8}$$

Since the orientation of the interfaces in impedance and ultrasound image should be the same, this approximation is reasonable. Considering, however, that US images are very noisy, this can lead to problems. In our estimation framework, we directly access the local orientation information delivered by the analytic signal for this purpose, see section 4.2.3. As final step in the simulation, a log-compression is applied to the images, see equation (2.11), so that the reflectivity regions of the simulated US images $\hat{u}(\mathbf{x})$ are

$$\hat{u}(\mathbf{x}) = \frac{\log(1+\alpha\cdot I_r(\mathbf{x}))}{\log(1+\alpha)} \tag{4.9}$$

with $\alpha$ parameterizing the log-compression.

### 4.2.3 Filtering and Identifying Interfaces

In order to be able to calculate the acoustic impedances for various tissues, we have to identify the tissue boundaries. On these boundaries we perform the acoustic impedance estimation. In a first step, we deal with the speckle in ultrasound images. As discussed in section 4.2.1, we consider speckle as noise for the acoustic impedance estimation, which we want to remove. A multitude of approaches for speckle reduction can be found in the literature such as Gaußian filtering, coherence-enhancing diffusion filtering, and despeckling filters based on the envelope of the US image [Grau et al., 2007]. We achieved good results with median filtering, which has superior speckle reduction properties compared to Gaußian smoothing [Yang and Fox, 2004]. We denote the filtered images as $\tilde{u}_j$.

For the identification of boundaries, we use the analytic signal as presented in chapter 3. The local phase provides us with structural information independent of the brightness and contrast. Further, we use the local orientation to calculate the incidence angle. For the creation of the filter bank we use log-Gabor filters. A local phase image is illustrated in figure 4.3(b). We threshold the phase image to extract the reflectivity part $\Omega$ from the ultrasound image, see figure 4.3(c).

### 4.2.4 Acoustic Impedance Estimation

For the estimation of the acoustic impedance, we combine the results from the last sections. Inserting the simulation in the maximum likelihood estimation, together

with the filtered image and restricting the comparison to interfaces, yields the following optimization

$$\arg\min_{z_j} \sum_{\mathbf{x}\in\Omega} \left( \tilde{u}_j(\mathbf{x}) - \frac{\log\left(1+\alpha\cdot\cos^m\theta_i(\mathbf{x})\cdot\left(\frac{z_j(\mathbf{x})-z_j(\mathbf{x}-\Delta\mathbf{d})}{z_j(\mathbf{x})+z_j(\mathbf{x}-\Delta\mathbf{d})}\right)^2\right)}{\log(1+\alpha)} \right)^2 . \tag{4.10}$$

The acoustic impedance in the reflection coefficient $\zeta_I(\mathbf{x}) = \left(\frac{z(\mathbf{x})-z(\mathbf{x}-\Delta\mathbf{d})}{z(\mathbf{x})+z(\mathbf{x}-\Delta\mathbf{d})}\right)^2$, is recursively defined, depending on the previous estimation $z(\mathbf{x}-\Delta\mathbf{d})$. Consequently, we need an initial value to make the calculation along the scan line. When taking *e.g.* acquisitions in a water bath, we can directly use the acoustic impedance of water. But it is not always possible to determine a proper initialization and as such, in general, a reconstruction up to scale is possible. This is sufficient for visualization and US simulation. In the future, further information from the US image such as tissue estimation from RF data [Moradi et al., 2007] or speckle [Aja-Fernandez et al., 2007] could be integrated to make the estimation more precise. Since we estimate the acoustic impedance per scan line, an averaging with neighboring scan lines while propagating the values between the interfaces leads to smoother estimations, see figure 4.3(e). We start the estimation at the beginning of the scan line with the initial values. Locations between interfaces are assigned impedance values, as calculated at the last interface.

In equation 4.10, we ignore the incident intensity $I_i$. Although attenuation and scattering lead to a decrease in intensity as the ultrasound beam penetrates the tissue farther, the time gain compensation simulates that everywhere in the image the same incident intensity is present, causing a constant $I_i$.

Moreover, we separate the direct estimation of the global impedance $z$ in equation (4.5), and instead, estimate for each US image $u_j$ an acoustic impedance $z_j$. In section 4.1.2, we argued that compounding of ultrasound images is not a trivial task. In contrast, compounding of estimated acoustic impedance images is straightforward because these images hold a correspondence between intensity value and tissue type. The global acoustic impedance image $z$ is consequently the mean of the estimates $z_j$, obtained from equation (4.10)

$$z = \frac{1}{n}\sum_{j=1}^{n} z_j. \tag{4.11}$$

Problems can still occur when structures with high acoustic impedance such as bones cause occlusion in the underlying region. For the detection of occlusions, the intensity term in equation (4.6) can be used to make a reliable compounding possible

### 4.2.5 Visualization

Once the global acoustic impedance image $z$ is estimated, we have to find ways to visualize it for the physician. One possibility would be to directly present the acoustic

(a) Original US image (b) Phase image (c) Threshold of phase image (d) Impedance estimation (e) Smooth imp. estimation

Figure 4.4: Processing steps for acoustic impedance estimation of the first forearm image.

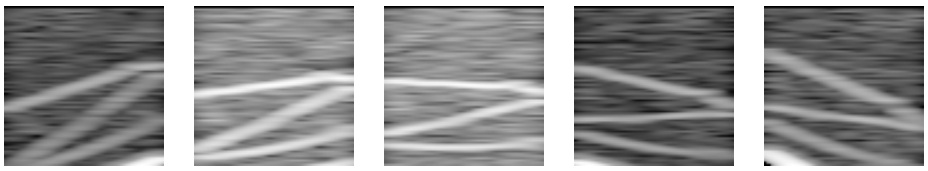
(a) Original US image (b) Phase image (c) Threshold of phase image (d) Impedance estimation (e) Smooth imp. estimation

Figure 4.5: Processing steps for acoustic impedance estimation of the second forearm image.

impedance image, but this may be of limited clinical value, because physicians are not used to these images and may have problems interpreting them. A better way may be to create artificial ultrasound views. The simulation of US from the acoustic impedance image is feasible because it is the acoustic impedance that determines the structure of the US images. It has the advantage, that US views can be simulated that were initially not recorded, and from positions that are physically not possible, *e.g.* below the skin. We use a recently introduced method [Shams et al., 2008], designed for simulating ultrasound images from CT data, to simulate US images from acoustic impedance, see Figure 4.8(h) for an example. A final possibility would be to convert the acoustic impedance image to a CT image. Therefore, the mapping from Hounsfield units to acoustic impedance values, as used in simulation [Wein et al., 2008], would have to be inverted.

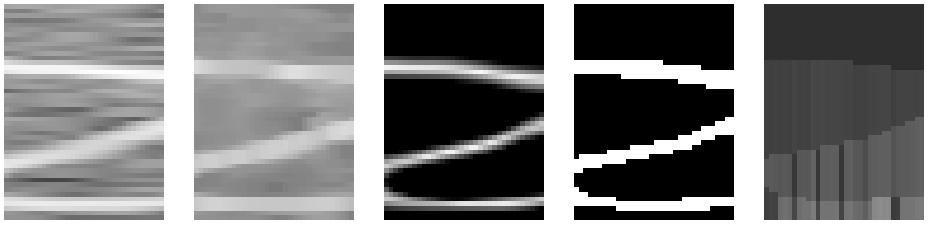
## 4.3 Estimation Results

We present results for the acoustic impedance estimation for three data sets. The first one is an image of a clay model, see figure 4.3(a), the second one consists of two scans of a human forearm, see figures 4.4(a) and 4.5(a), and the third one consists of five simulated US images from a human abdomen, see figure 4.6. The acquisitions for the first and second data sets were done in a water bath, because we wanted to avoid tissue deformation due to probe pressure, so that we could focus on the acoustic impedance estimation. We used a linear array ultrasound transducer for



(a) US −20° (b) US −10° (c) US 0° (d) US 10° (e) US 20°

Figure 4.6: Five ultrasound images simulated from CT from −20° to 20°.



(a) Original US image (b) Filtered image (c) Phase image (d) Threshold of phase image (e) Impedance estimation

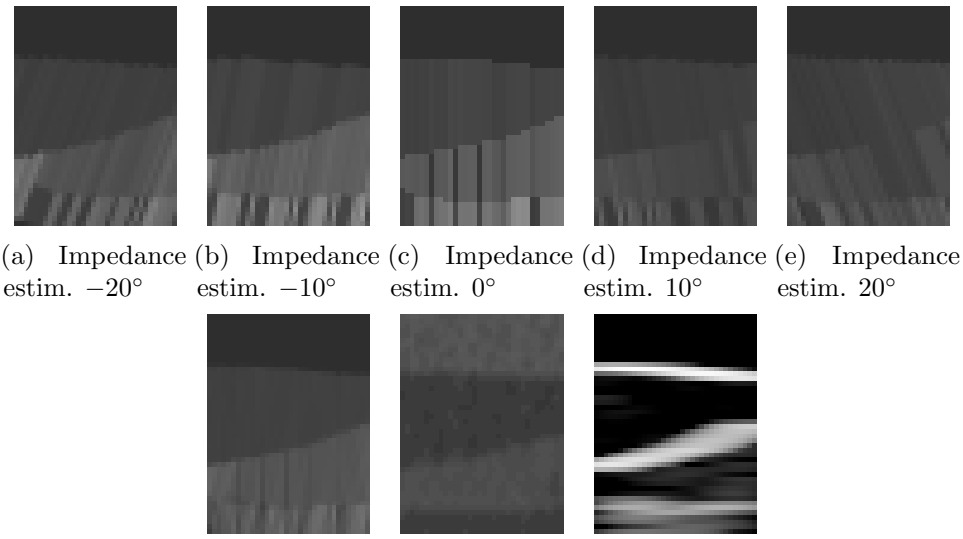
Figure 4.7: Processing steps for acoustic impedance estimation of simulated ultrasound at 0°.

the acquisitions.

The steps for acoustic impedance estimation, as described in section 4.2, are shown in figure 4.3 for the clay model. The image is filtered with a median filter with a window size $10 \times 10$. Then, the phase is calculated on the filtered image, where we use a wavelength of 250mm for the log-Gabor filter, see figure 4.3(b). We apply thresholding on the phase image to obtain a mask, showing us the regions with high reflectivity, see figure 4.3(c). Determining the threshold is not critical and we performed all our experiments with a value of 0.7. For the calculation of the acoustic impedance, we set the log-compression parameter $\alpha = 700$ and the exponent $m = 0.1$. The original and smoothed estimations are shown in figures 4.3(d) and 4.3(e).

In figures 4.4 and 4.5, the estimation steps for both forearm images are shown. We use a wavelength of 180mm for the log-Gabor filter and $\alpha = 1000$ for the log-compression.

For the third data set, we simulate US from a CT image shown in figure 4.8(g), using a recently introduced method [Shams et al., 2008], which produces realistic ultrasound images from CT data. The images are acquired from different viewing angles varying from −20° to 20° to create a realistic spatial compounding scenario. We show as an example the processing steps for the estimation of the simulated US images acquired at 0° in figure 4.7, with a wavelength of 60mm, $\alpha = 83$, and $m = 0.1$. After analog processing with the four remaining images and mapping them back to the reference frame at 0°, see figures 4.8(a) - 4.8(e), we can calculate the global acoustic impedance image, see figure 4.8(f). It is the average of the separate estimations and one observes the improved quality. When comparing it to the original CT image, which can to some extent be seen as ground truth, the good quality of



(a) Impedance estim. −20° (b) Impedance estim. −10° (c) Impedance estim. 0° (d) Impedance estim. 10° (e) Impedance estim. 20°

(f) Imp. mean (g) CT (h) US from imp. mean

Figure 4.8: Acoustic impedance estimation in the overlap area from the 5 simulated US images, transformed in reference coordinate system. Mean of estimations, and in comparison the original CT. Last, simulation of US image from the global acoustic impedance image.

the reconstruction becomes apparent. Finally, we use the global impedance image to simulate an ultrasound image, see figure 4.8(h).

## 4.4 Conclusion

This work is a first approach to estimate the acoustic impedance from standard B-mode ultrasound images. The proposed method deduces the change of impedance from reflections at tissue interfaces. Crucial is the reliable detection of these interfaces, with the local phase being one possibility. An extension leading to more robust results may be the segmentation of the entire image domain. With this it could be possible to impose further constraints, guaranteeing the same impedance estimates for one tissue. A drawback of the presented analysis is the utilization of B-mode images. RF data would give more precise measurements, which are necessary for an accurate calculation. Finally, the applied model for ultrasound reflection is only an approximation, as already indicated in section 2.2 and further evaluated in section 8.4. Consequently, we consider the presented analysis as a first step into the direction of acoustic impedance estimation with the essential building blocks being identified. Better results could be achieved by improving each of these blocks. Also interesting would be the integration of additional data from elastography and speckle analysis to better characterize tissue.

# Part II

# REGISTRATION

This part describes the key methodology of this thesis: image registration. We introduce a new contextual probabilistic framework to model image registration. It is more versatile than existing frameworks and allows for a unified description of geometric and iconic registration. Further, it enables to review a large number of registration techniques and to arrange them in a continuum, limited by pure intensity- and feature-based registration. Subsequently, we propose a new technique for multi-modal image registration with structural representations.

# Chapter 5

# A Contextual Maximum Likelihood Framework

*In this chapter, we review a large number of registration techniques with the help of a novel probabilistic framework. This framework considers, in contrast to previous ones, local neighborhood information. We integrate the neighborhood information into the framework by adding layers of latent random variables, characterizing the descriptive information of each image. This extension has multiple advantages. It allows for a unified description of geometric and iconic registration, with the consequential analysis of similarities. It enables to arrange registration techniques in a continuum, limited by pure intensity- and feature-based registration. With this wide spectrum of techniques combined, we can model hybrid registration approaches. The probabilistic coupling allows further to deduce optimal descriptors and to model the adaptation of description layers during the process, as it is done for deformable registration and joint registration/segmentation.*

## 5.1 Introduction

Image registration is a key technology in medical image analysis and computer vision. It is not only applied for fusing the information from multiple images, but it is also a prerequisite for many processing tasks. The goal of registration is to establish a spatial relationship between a pair or a group of images. This is done by optimizing a similarity criterion between a fixed and a transformed image with respect to a chosen transformation model. Since registration is the main methodological component of this thesis, we review related registration approaches in this chapter. This is done with a new probabilistic model. We will frequently refer to this model in the following chapters in order to illustrate how our contributions about registration fit into this model.

A common classification in registration is to distinguish between geometry- and intensity-based approaches. Geometric approaches establish the spatial relationship between images based on extracted features, landmarks, surfaces, or point clouds. Intensity-based or iconic approaches directly operate on the images by comparing

their pixel intensities or photometric properties. In order to get an overview of the plethora of registration techniques, several survey articles are available [Brown, 1992, Maintz and Viergever, 1998, Lester and Arridge, 1999, Hill et al., 2001, Hajnal et al., 2001, Zitova and Flusser, 2003, Pluim et al., 2003]. Next to a listing of existing techniques, these articles as well present a rough categorization. For intensity-based registration unifying, probabilistic frameworks [Viola, 1995, Roche et al., 2000, Zöllei, 2006] were proposed. These frameworks are essential in better understanding and categorizing different types of registration approaches. With a strict deduction from a mathematical framework, it is possible to detect implicitly incorporated assumptions. Discovering such assumptions allows for a better adaptation of registration to specific applications and to justify the adequacy of an approach in a specific scenario. Concepts from probability theory, such as maximum likelihood or a posteriori estimation, were in this context shown to be very useful to reason about image registration. The limitation of currently existing probabilistic frameworks is, however, that they focus on modeling the similarity measure.

Looking at registration in practice, we observe that processing steps such as gradient calculation, multi-scale analysis, and noise reduction are applied to the images, before performing the alignment. Further, the comparison of single pixel information is prone to noise, leading to the introduction of context and spatial information in registration. With the presented contextual, probabilistic framework we are able to model these approaches. Moreover, we can model geometric approaches through the introduction of layers of latent random variables. Dealing with these representations allows for differentiating between pure image processing steps, such as smoothing and gradient calculation, and the estimation of the similarity between images. This helps to classify registration techniques and identify commonalities.

In section 5.2, we describe existing unifying approaches in detail and provide further insights and interpretations, together with the corresponding probabilistic graphical models. In section 5.3, we present the contextual framework, in which the graphical model allows a descriptive integration of neighborhood information. In section 5.4, we review several geometric and iconic techniques in the context of the proposed framework. This permits to establish a continuum of registration approaches ranging from geometry- to intensity-based approaches. Moreover, we describe the dynamic adaptation of description layers. In section 5.6, we formulate the extension of the framework to groupwise registration.

## 5.2 Probabilistic Modeling of Image Registration

In this section, we provide further insights into existing probabilistic frameworks for image registration proposed in [Viola and Wells, 1997, Roche et al., 2000, Zöllei et al., 2003], with a detailed discussion in the corresponding theses [Viola, 1995] and [Zöllei, 2006]. Most of the approaches model the registration as maximum likelihood (ML) estimation or, more generally, as maximum a posteriori (MAP) estimation. In image analysis, we usually consider probabilities from a Bayesian perspective, because it allows a more general notion of probability. In the center is Bayes' theorem that we

repeat shortly due to its frequent application in image registration.

### 5.2.1 Bayes' Theorem

Bayes' theorem describes how one conditional probability depends on its inverse. Given observed data, a common task is to express the likelihood of a set of parameters describing the underlying model. With Bayes' theorem it is possible to express this posterior probability in the form of a likelihood and a prior term

$$p(\text{Parameter} \mid \text{Data}) = \frac{p(\text{Data} \mid \text{Parameter}) \cdot p(\text{Parameter})}{p(\text{Data})} \tag{5.1}$$

with the normalization factor

$$p(\text{Data}) = \int p(\text{Data} \mid \text{Parameter}) \cdot p(\text{Parameter}) \;\; \mathbf{d}\text{Parameter} \tag{5.2}$$

which ensures the integral of $p(\text{Parameter} \mid \text{Data})$ to be one. Short and concise, Bayes' theorem can be formulated as

$$\text{Posterior} \propto \text{Likelihood} \cdot \text{Prior}. \tag{5.3}$$

Often, one is not just interested in quantifying the probability of the parameters given the data, or the inverse, but one tries to find the most probable parameters that make the model fit best to the observed data. Depending on whether the posterior or the likelihood term is maximized, this results in an MAP or ML estimation, respectively.

### 5.2.2 Image Registration Frameworks

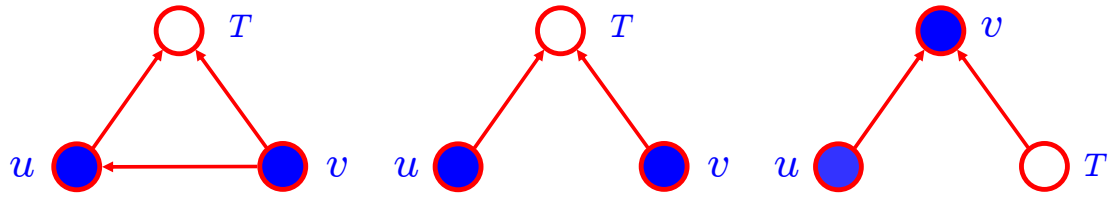
In order to describe image registration from a probabilistic point of view, we consider each image to be a random variable $U$. The probability of the appearance of a concrete sample image $u$ is $p(U = u)$, with the simplified notation $p(u)$. Considering further that an image is defined on a grid $\Omega$, each spatial location $U(x)$ with $x \in \Omega$ is a random variable. Taking the set of intensity values $\mathcal{I}$, *e.g.* $\mathcal{I} = \{0, 1, \ldots, 255\}$, the probability of a location having a certain intensity is $p(U(x) = i)$ with $i \in \mathcal{I}$. The goal of registration is to find the transformation $T$ that expresses the spatial relationship between two images $u$ and $v$

$$u(x) = v(T(x)). \tag{5.4}$$

This is the underlying model of the generative [Bishop and Lasserre, 2007], joint probability that we maximize

$$\hat{T} = \arg\max_{T \in \mathcal{T}} p(u, v, T), \tag{5.5}$$

with $\mathcal{T}$ being the space of transformations and $\hat{T}$ the optimal transformation with respect to the model. This is the full probability model, which is a joint probability distribution for all observable and unobservable quantities.




(a) $p(u, v, T)$ (b) $p(T|u, v)$ (c) $p(v|u, T)$

Figure 5.1: Probabilistic graphical models of the listed probabilities.

We apply the product rule and Bayes' theorem to the joint probability density function (PDF) in equation (5.5) to relate to the commonly utilized conditional probabilities

$$p(u, v, T) = p(T \mid u, v) \cdot p(u, v) \tag{5.6}$$

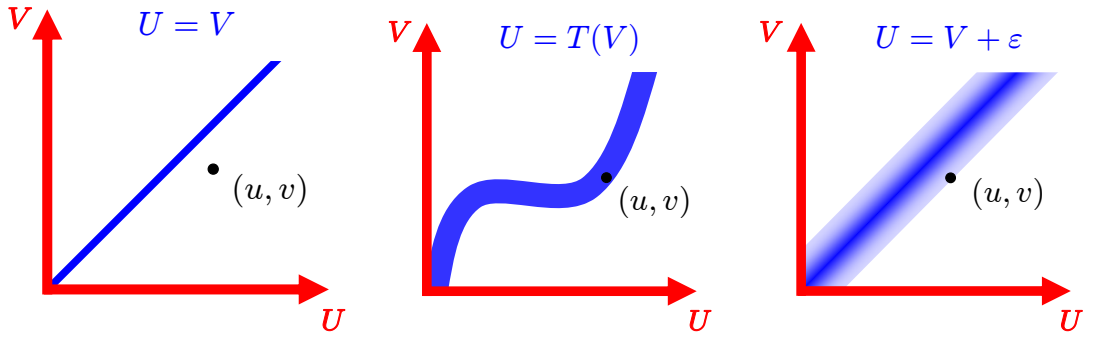
$$= \frac{p(u, v \mid T) \cdot p(T)}{p(u, v)} \cdot p(u, v) \tag{5.7}$$

$$= p(v \mid u, T) \cdot p(u \mid T) \cdot p(T). \tag{5.8}$$

Maximizing the term $p(T \mid u, v)$ corresponds to an MAP estimation used *e.g.* in [Zöllei et al., 2007] and [Toews et al., 2005]. Applying Bayes' theorem, one identifies the likelihood term $p(u, v \mid T)$ and the prior on the transformation $p(T)$. [Viola, 1995, Roche et al., 2000] further apply the product rule to the likelihood term to get $p(v \mid u, T)$. In the literature, the probabilities $p(u, v \mid T)$ and $p(v \mid u, T)$ are not distinguished and plainly considered as likelihood terms in an ML estimation. In figure 5.1, we show the corresponding probabilistic graphical models [Bishop, 2008, Bishop, 2006] for the aforementioned probabilities, where filled nodes correspond to *observed* random variables and empty ones to *latent* variables. Graphical models have the advantage that dependencies between random variable are easily observable, like in this case between $u$ and $v$.

### 5.2.3 Interpretation of Probability Maximization

The relationship between the maximization of the probabilities in equations (5.6 - 5.8) and the correct registration of images is not trivial, but may become more clear thanks to further interpretation. We start with the term $p(v \mid u, T)$, which is the most intuitive one, to demonstrate that the maximum likelihood estimation makes sense. Consider the probability density function $p(V = v)$, defined over the set of images on the grid $\Omega$. Without further knowledge, it is meaningful to assume $p$ to be a uniform distribution, because each observation $v$ is equally likely. Adding, however, knowledge in the form of the observation of the image $u$, whose relation to $v$ is known from our model in equation (5.4), the conditional distribution $p(V = v \mid u)$ has peaks for images $v$ being equal to $u$. When further considering that the images




(a) Identity transformation (b) Translation (c) Identity with noise model

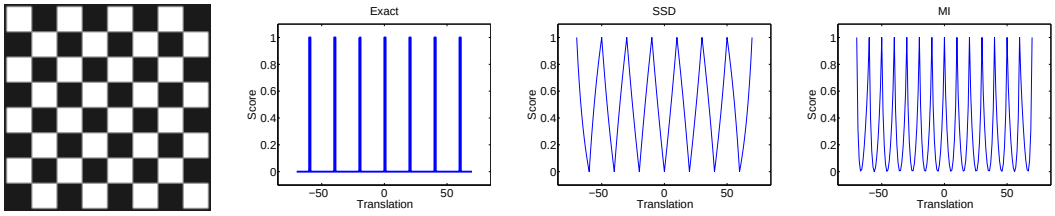
Figure 5.2: Illustration of the joint density $p(u, v|T)$ conditioned on the transformation.

are not correctly aligned, the correct transformation $T$ leads to the maximization of the likelihood $p(V = v \mid u, T)$ of the appearance of image $v$.

For the interpretation of $p(u, v \mid T)$ consider the joint density over the images $p(U = u, V = v)$. Initially it may seem strange that the joint density is maximized in case of correctly aligned images. The situation becomes clearer when considering again that the joint density is linked to the model in equation (5.4). Assuming the transformation to be the identity, $T = \mathrm{Id}$, the only images that fulfill the model equation are identical images, so $u = v$. This is illustrated in figure 5.2(a), where the histogram of the joint distribution consists only of the bisecting line. Considering actual transformations, such as translations, the histogram changes accordingly. Interesting in this regard is that the number of non-zero entries of the joint probability increases since the comparison of the images can only be done in the overlapping region $\Omega_{uv} = \Omega_u \cap \Omega_v$, which is the intersection of both image grids. The image region which is not part of the overlap $\Omega_u - \Omega_{uv}$ can be arbitrary. This is schematically illustrated in figure 5.2(b) with an increased blue region. Also note that all entries in the joint histogram have the same value, since the images are either equal or not. The registration process can then be considered as finding for a given image pair $(u, v)$ a transformation so that the images are equal in the overlapping area or more descriptive, that the point $(u, v)$ in the diagram is part of the blue region.

For the term $p(T \mid u, v)$ consider the marginal probability of the transformation $p(T)$, which is once again uniformly distributed without prior knowledge. Conditioning the distribution on the known images $p(T \mid u, v)$, where we know from our model in equation (5.4) that $T$ relates them, the maximum should be achieved for the correct alignment. To illustrate this, we translate the chessboard pattern shown in figure 5.3(a) along one of the axes. The distribution $p(T \mid u, v)$ is shown in figure 5.3(b), where we have peaks at the positions of exact match. Note that the likelihood is identical for each of the peaks, even if the overlapping region is smaller.

Next, we look at the joint distribution over all random variables in the problem,




(a) Chessboard (b) Exact Matching (c) SSD (d) MI

Figure 5.3: Illustration of $p(T \mid u, v)$ for the translation of the chessboard image along one of its axes. For each of the diagrams we consider a different underlying model, expressed by the probability.

$p(u, v, T)$. An illustration is, unfortunately, more challenging because we have to consider a three-dimensional probability. However, we can relate it back to the two-dimensional distribution $p(u, v \mid T)$. The cutting plane of $p(u, v, T)$ for a specific $T$ is exactly $p(u, v \mid T)$ weighted by $p(T)$. So stacking all joint distributions $p(u, v \mid T)$ with the appropriate weighting leads to $p(u, v, T)$, see figure 5.4. Analogously is the relation to $p(T \mid u, v)$, which corresponds to a line in 3D space for a specific pair of images $(u, v)$, also illustrated in figure 5.4.

The last interpretation we want to present is of a more theoretical nature and is provided in [Zöllei, 2006, Zöllei et al., 2003]. An information-theoretic framework is introduced and its relationship to the maximum likelihood approach is shown. It bases upon the utilization of two information-theoretic quantities, the Kullback-Leibler (KL) divergence $D$ and the Shannon entropy $H$, between a source $p_S$ and model $p_M$ distribution

$$\hat{T} = \arg\min_{T \in \mathcal{T}} D(p_S || p_M) + H(p_S). \tag{5.9}$$

In [Zöllei et al., 2003], it is shown that the ML approach is equivalent to the following minimization

$$\hat{T} \approx \arg\min_{T \in \mathcal{T}} D(p(u, v, T^*) || p(u, v, T)) \tag{5.10}$$

with $p(u, v, T^*)$ the source distribution, the ground truth transformation $T^*$, and $p(u, v, T)$ the distribution of the current alignment. So, in order to show that the maximization of the likelihood is meaningful, we can equivalently show that the minimization in equation (5.10) is meaningful. This is straightforward, because the Kullback-Leibler divergence is the lower, the more similar the distributions are. And the distributions are the more similar, the better the images are aligned, so the better $T$ approximates $T^*$.

The information-theoretic framework is interesting not just as another interpretation of the ML framework; it also presents a different probabilistic view on the problem of image registration. More precisely, it enables us to analyze the work of [Leventon et al., 1998] and [Chung et al., 2002], who learn the joint intensity distribution from previous registrations and use it to evaluate the current alignment.

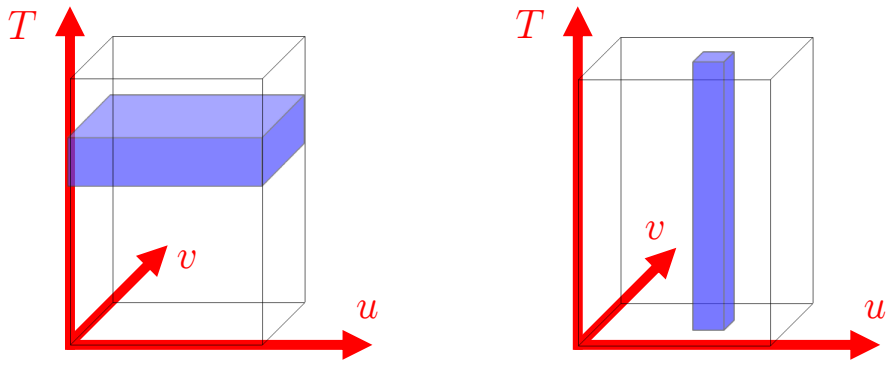



Figure 5.4: Illustration of the joint probability $p(u, v, T)$. The cutting plane in the left figure corresponds to $p(u, v \mid T)$, for a specific $T$. The line on the right figure corresponds to $p(T \mid u, v)$ for a specific pair $(u, v)$ (since the visualization of an infinitely thin line is challenging, we represent it as a cuboid).

Further, the derivation of mutual information (MI) [Wells et al., 1996, Collignon et al., 1995] becomes easier than the rather challenging deduction from an ML approach with an unspecified channel in [Roche et al., 2000].

We shortly present the integration of MI since it contains a nice explanation of the underlying assumptions [Zöllei et al., 2003]. Maximizing mutual information is equivalent to the following maximization

$$\hat{T} \approx \arg\max_{T} D(p(u, v, T^*)||p(u)p(v, T)). \tag{5.11}$$

Interesting is the discrepancy in comparison to equation (5.9), where the KL divergence is to be minimized. The maximization of mutual information corresponds to moving away from the situation of $u$ and $v$ being independent, so $p(u, v, T) = p(u)p(v, T)$. This is justified because we know that the independence of the images is false, and therefore, a correct registration impedes such an alignment. In contrast, the minimization in equation (5.9) tries to equal the current and ground truth distributions. Hence, instead of moving away from the wrong alignment, it forces the registration to approximate the alignment that we know is right.

### 5.2.4 Integration of Noise and Intensity Mapping

The exact equality of the images is practically not possible due to noise, distortion, and interpolation artifacts. This is taken into account by introducing an additive and stationary Gaußian white noise $\varepsilon$ to the imaging model [Viola, 1995, Roche et al., 2000]

$$u(x) = v(T(x)) + \varepsilon. \tag{5.12}$$

Considering once again the joint density $p(u, v \mid T)$ in figure 5.2(c), there is no longer just the bisecting line for identical images, but a gradation starting from this line. This gives similar images still a noticeable likelihood. The optimal similarity criterion

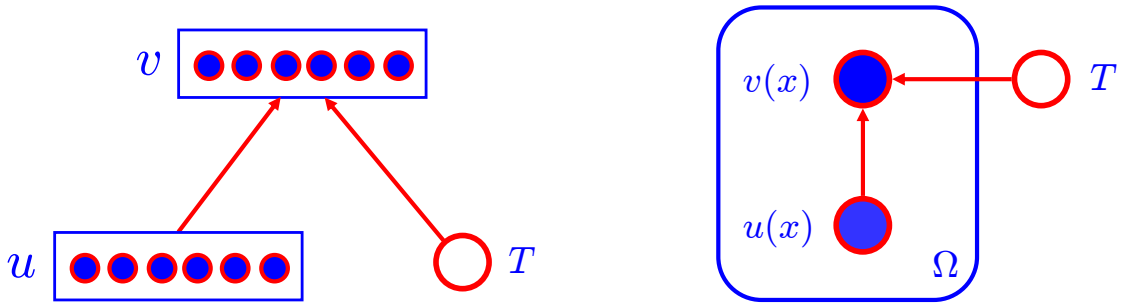



Figure 5.5: Left: Graph showing the probability $p(v \mid u, T)$, where each of the images consists of a random variable for each location $x \in \Omega$, in this case 5. Right: Assumption of i.i.d. coordinate samples, $\prod p(v(x) \mid u(x), T)$, illustrated as plate.

considering this imaging model is sum of squared differences (SSD) [Viola, 1995]. In figure 5.3(c), we show the similarity plot for the chess pattern using SSD, which is corresponding to the histogram of $p(T \mid u, v)$ with respect to this new imaging model. The chosen type of noise enables the adaptation to specific applications, which is further analyzed in chapter 8 for ultrasound images.

A further extension of the model presents the consideration of an intensity mapping $f$, allowing for a functional relationship [Viola, 1995, Roche et al., 2000] between pixel intensities

$$u(x) = f(v(T(x))) + \varepsilon. \tag{5.13}$$

Standard similarity measures can then be classified according to the type of similarity function they are able to model: SSD (identity), correlation coefficient (affine), and correlation ratio (functional) [Roche et al., 1998].

Similarity measures like mutual information or joint entropy do not fit into this category, because they assume a statistical, instead of a functional, relationship. The ability of MI to deal with more complex similarity relationships is illustrated in figure 5.3(d). MI does not only show maxima, where white squares are correlating with white squares, but also where white ones correlates with black ones. For certain situations this could be desired. However, also notice that the increased modeling capability of mutual information may create ambiguous situations, introducing further local minima in the cost function [Roche et al., 2000]. The similarity measures have therefore to be chosen dependent on the concrete application.

### 5.2.5 I.I.D. Coordinate Samples

A general assumption in the discussed unifying approaches [Viola, 1995, Roche et al., 2000, Zöllei et al., 2003] are independent and identically distributed (i.i.d.) coordinate samples. With this assumption, equation (5.6) is simplified to

$$p(u, v, T) = \prod_{x \in \Omega} p(u(x), v(x), T). \tag{5.14}$$

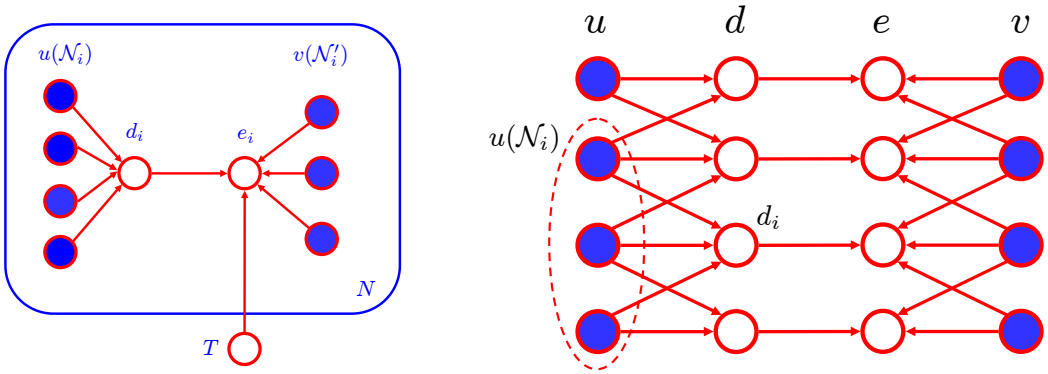
This is illustrated in figure 5.5 for the probability $p(v \mid u, T)$. Since each of the spatial locations in the images corresponds to a random variable, we use the plate visualization, as proposed by [Bishop, 2006], because it permits a more compact representation of the graph.

Relating this back to the imaging model in equation (5.13), we have to assume that next to the independence of the coordinate samples, also the mapping function $f$ and the noise $\varepsilon$ are context-free. The noise is therefore restricted to a spatial extent of one pixel and artifacts like blurring, that affect a local neighborhood, cannot be considered. Finding an alternative to the limiting i.i.d. assumption is the subject of the following section.

## 5.3 Contextual Probabilistic Registration Framework

The introduction of a new probabilistic framework for registration is motivated by the increasing importance of the consideration of *context information* for image registration and by the enhanced descriptiveness. The intensity value of a single pixel shows limited significance and is corrupted by noise. A local neighborhood provides more consistent information. In [Shechtman and Irani, 2007], for instance, impressive results are obtained by correctly matching objects in different images, where only their shape is similar, but the appearance is very different. They achieve it by extracting the local self-similarity around the pixels. In [Belongie et al., 2002], the shape context is introduced, which describes the distribution of the rest of the shape with respect to a given point on the shape. Recently, the contextual flow was proposed [Wu and Fan, 2009], which replaces the core assumption of intensity constancy of optical flow with a constancy of context. These are just a few examples to show the increasing importance of context in image registration.

Especially in the medical imaging literature, the consideration of *spatial information* motivates the usage of neighborhood systems. The standard example is that the entropy does not change when shuffling pixels within an image. For image registration, interpolation and multi-resolution approaches lead to an implicit integration of spatial information. However, an explicit consideration of spatial information may improve the registration. In [Rueckert et al., 2000], neighboring intensities are added to the calculation of mutual information by estimating higher-order densities. This approach is extended to larger neighborhoods and addresses the high dimensionality of the problem by utilizing uniformly distributed random lines [Bardera et al., 2006]. In [Zheng and Zhang, 2006, Zheng, 2008], a Markov random field is proposed to integrate contextual information. A $3 \times 3$ neighborhood is used around each pixel in [Russakoff et al., 2004]. The problem of the high dimensionality is tackled by introducing simplifying assumptions which reduce the estimation of a $D$-dimensional distribution to $D$ independent 1-dimensional distributions. [Legg et al., 2009] extend this by not only considering intensity values but also multi-scale derivatives. In [Loeckx et al., 2010], conditional mutual information is proposed, where the joint




(a) Graphical model as plate

(b) Contextual graphical model for registration

Figure 5.6: Observed random variables are filled blue, links indicate dependency. Left: Descriptors $d_i$ and $e_i$ are dependent on a local neighborhood $\mathcal{N}_i$ of the original images. Further, $e_i$ is dependent on $d_i$ and the transformation $T$. Right: Graphical modal illustrated for multiple descriptors.

distribution is augmented by a spatial dimension, expressing the location of the joint intensity pair. In [Zhuang et al., 2009], the densities for mutual information based similarity measures are estimated on local volumes. This large number of approaches building upon the introduction of context shows the importance of a unified theoretical model that integrates context information. The i.i.d. assumption of current frameworks prohibits, however, their consideration.

### 5.3.1 Contextual Probabilistic Graphical Model

The key component of the novel probabilistic framework is to replace the assumption of independence of coordinate samples in equation (5.14), by the Markov property, so making each pixel position dependent on a local pixel neighborhood. One could think of a variety of possibilities for modeling the local neighborhood in a maximum likelihood framework. We decided to introduce two additional layers $d$ and $e$, because it facilitates the representation of the neighborhood dependency. Each of the layers, we refer to as *description layers*, consists of latent random variables $d_i$ and $e_i$, respectively, with $1 \leq i \leq N$ and $N = |\Omega|$. The layers $d$ and $e$ are lying on the same grid as the images do, so we have a dense set of descriptors. Each descriptor $d_i$ is dependent on a local neighborhood $\mathcal{N}_i$ of the image $u(\mathcal{N}_i)$, analogously, $e_i$ is dependent on $v(\mathcal{N}'_i)$. In our model, we let each descriptor $d_i$ be dependent on a local neighborhood $\mathcal{N}_i$ of the image $u(\mathcal{N}_i)$, analogously, $e_i$ is dependent on $v(\mathcal{N}'_i)$. The relationship between descriptors $d_i$ and $e_i$ is one-to-one. The creation of the layer $e$ is dependent on the transformation $T$.

As mentioned previously, we utilize *probabilistic graphical models* [Bishop, 2006] for establishing the relationship of the random variables because they are advantageous in representing the structure and dependency for a multitude of variables. Further,

we choose a *directed* graphical model, where nodes (vertices) represent random variables and directed links (edges) express probabilistic dependency between them. The graphical model for our framework is shown in figure 5.6(a) as plate with an exemplary 4- and 3-neighborhood, for $u$ and $v$, respectively. Another illustration, not as plate, is shown in figure 5.6(b) without the consideration of $T$ due to clarity of presentation. The presented graphical model factorizes to

$$p(u, v, d, e, T) = p(T) \cdot \prod_{i=1}^{N} p(u(x_i)) \cdot p(v(T(x_i))) \cdot p(d_i|u(\mathcal{N}_i)) \cdot p(e_i|d_i, v(\mathcal{N}'_i), T). \tag{5.15}$$

We deduce the term $p(e_i|d_i, v(\mathcal{N}'_i), T)$ further by applying the product rule and Bayes' theorem. Therefore, we assume the conditional independence of $d_i$ and $v(\mathcal{N}'_i)$ given $e_i$. Moreover, we incorporate the independence of $d_i$ and $v(\mathcal{N}'_i))$, which is a consequence of the model. This leads to

$$p(e_i \mid d_i, v(\mathcal{N}'_i), T) = \frac{p(d_i \mid e_i) \cdot p(v(\mathcal{N}'_i), T \mid e_i) \cdot p(e_i)}{p(d_i) \cdot p(v(\mathcal{N}'_i), T)} \tag{5.16}$$

$$= \frac{p(e_i \mid d_i) \cdot p(e_i \mid v(\mathcal{N}'_i), T)}{p(e_i)}. \tag{5.17}$$

Setting this in the right-hand side of equation (5.15) results in

$$p(T) \cdot \prod_{i=1}^{N} p(u(x_i)) \cdot p(v(T(x_i))) \cdot p(d_i \mid u(\mathcal{N}_i)) \cdot \frac{p(e_i \mid d_i) \cdot p(e_i \mid v(\mathcal{N}'_i), T)}{p(e_i)}. \tag{5.18}$$

Therein, the marginal terms $p(T)$, $p(u(x_i))$, $p(v(T(x_i)))$, and $p(e_i)^{-1}$ represent the probabilities for the transformation, the images, and the description layer. The reason that only the descriptor layer $e$ appears in the formulation is rooted in the asymmetric formulation of the registration, by only transforming the image $v$. This can be changed with a symmetric formulation, by transforming all images, which is shown in section 5.6 for the general case of groupwise registration.

The marginal terms are used to incorporate prior information into the registration, with the purpose of improving the robustness and capture range [Zöllei, 2006]. This will be discussed for deformable registration in section 5.5.1. In most cases, we do not have any a priori knowledge about the probability distribution of these terms, so that we presume a uniform distribution, leading to

$$p(u, v, d, e, T) = \prod_{i=1}^{N} \underbrace{p(e_i \mid d_i)}_{\text{similarity}} \cdot \underbrace{p(d_i \mid u(\mathcal{N}_i)) \cdot p(e_i \mid v(\mathcal{N}'_i), T)}_{\text{coupling}}. \tag{5.19}$$

It is mainly the interplay of these three probabilities that determines the functionality of our model. The *similarity term* $p(e_i \mid d_i)$ is the standard likelihood function as used in previous unifying frameworks [Roche et al., 2000, Viola, 1995, Zöllei et al.,

2003]. However, instead of comparing the images $u$ and $v$, it compares the description layers $d$ and $e$. If we could arbitrarily modify the description layers and just optimize $p(e \mid d)$, we would simply change the layers to be totally exact. The *coupling terms* $p(d_i \mid u(\mathcal{N}_i))$ and $p(e_i \mid v(\mathcal{N}'_i), T)$ prevent this simplistic solution by expressing how well the description layers fit the original images. In the optimization, they are counterbalancing the influence of the similarity term like a regularizer.

The joint distribution, $p(u, v, e, d, T)$, we finally end up with, is different from the one used in maximum likelihood frameworks, $p(u, v, T)$. This can, however, be obtained by marginalizing with respect to the descriptors

$$p(u, v, T) = \sum_{d,e} p(u, v, d, e, T). \tag{5.20}$$

Practically, it is not possible to sum over all possible descriptors. Thus, the alignment is only optimal with respect to a specific descriptor or a small set of descriptors, which is discussed further in section 5.4.3 on hybrid approaches.

### 5.3.2 Similarity Term

The probability $p(e \mid d)$ models the calculation of the similarity on the description layers. This is exactly the part of registration where the reviewed probabilistic approaches in section 5.2 focus on. An analysis of certain similarity measures with respect to the underlying intensity mapping is provided in [Roche et al., 2000]. Commonly used measures are SSD, sum of absolute differences, correlation coefficient, and correlation ratio. Further, entropy-based metrics are mutual information [Collignon et al., 1995,Wells et al., 1996], joint entropy [Studholme et al., 1999], and the entropy of the difference image (EDI) [Penney et al., 1998]. Recently, the residual complexity (RD) [Myronenko and Song, 2009b] was proposed, which is similar to EDI. First, the difference image is calculated. Then, the complexity or information content of this image is quantified. For correctly aligned mono-modal images the difference should be around 0. A mis-alignment leads to large differences and therefore a large spread around 0. In EDI, the entropy is used to quantify the spread. For residual complexity, the compressibility is used, where the difference image of correctly aligned images contains less information and can therefore be better compressed. Also in this line of argumentation fits the minimum descriptor length criterion [Twining et al., 2004], however, with a focus on groupwise registration, further described in section 5.6. Finally, the mapping complexity [Myronenko and Song, 2009a] presents an extension of the residual complexity, applicable for multi-modal registration

### 5.3.3 Transformation

The final objective of registration is to recover the transformation, indicating the spatial relationship of the images. In this chapter, we do not focus on the details of the transformations; however, we have to differentiate between two categories. First, global transformations including rigid, affine, and projective transformation models,


Figure 5.7: Continuum between feature- and intensity-based registration, augmented with exemplary approaches. Arranged by the uniqueness of descriptors.

with the purpose of arranging the images in a common reference frame. And second, local transformations, which deform the images and are meant to represent changes in the images. An overview of non-rigid transformations is presented in [Holden, 2008]. Since rigid transformations apply the same algorithm to each pixel position, the coupling term $p(e_i \mid v(\mathcal{N}_i'), T)$ is constant, and can consequently be neglected. For affine, projective, and local transformations, however, this term becomes interesting, because it constrains together with the prior on the transformation $p(T)$ the changes in the images. These terms are generally referred to as regularization terms. The modeling of deformable registration within this framework is further detailed in section 5.5.1.

## 5.4 A Continuum of Registration Approaches

In this section, we discuss several approaches for geometry- and intensity-based registration and show how they fit into the proposed framework. These methods form, in fact, a continuum of registration approaches, going all the way from pure geometric to pure iconic registration. On the one end, we identify landmark-based registration, where users manually pick salient points in the image. The description is optimal because we exactly know about the correspondence of points. On the other end, we identify intensity-based registration, with single intensity values as minimalistic descriptors. The number of approaches in between can be arranged by the uniqueness of their descriptors, as illustrates figure 5.7.

On the right-hand side of the spectrum, we consider SIFT and GLOH with comparatively high uniqueness of the descriptors. SIFT/GLOH correspondence hypotheses are created without location information, therefore descriptors must uniquely characterize the position they are extracted from. For DAISY [Tola et al., 2008], the dense arrangement of descriptors relaxes this requirement, equally for self-similarity [Shechtman and Irani, 2007]. Entropy images (cf. chapter 6), extract structural information of images for multi-modal registration. These images resemble gradient images that are commonly applied for registration. The images in scale-space are close to the original ones, because locally weighted averages are created with an emphasis on the center location.

In figure 5.8, we show a diagram that illustrates the new registration process, which is in comparison to the previous one augmented with description layers. The layers are the output of $D(u, \Phi)$ and $D(v, \Phi)$, with $\Phi$ being the parameters that

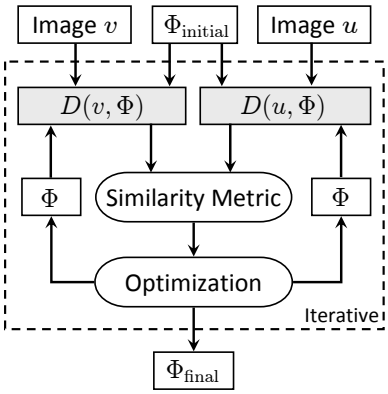



Figure 5.8: Diagram illustrating the registration process with description layers (gray shaded).

are to be calculated. As an example, the parameters for a joint registration and restoration could be $\Phi = (T, \mu, \sigma^2)$ [Woods et al., 2006], with the transformation $T$ and the Gaußian noise parameters $\mu, \sigma^2$. The iterative process starts with the initial parameters $\Phi_{\text{initial}}$ and the output is $\Phi_{\text{final}}$.

### 5.4.1 Intensity-Based Registration

Existing probabilistic frameworks for intensity-based registration focus on similarity measures and do not model common processing steps on the images. We demonstrate in the following how they can be integrated in the new framework. The proposed framework is a true extension of previous maximum likelihood frameworks, which can be obtained by setting $\mathcal{N}_i = (x_i)$, $d_i = u(x_i)$, $\mathcal{N}_i' = (T(x_i))$, and $e_i = v(T(x_i))$.

#### 5.4.1.1 Image Filtering

Image filtering is a common pre-processing step for image registration. One application of filtering is image enhancement through operations such as sharpening, noise reduction, and contrast adjustment. Another application is the creation of a scale-space [Koenderink, 1984]. Although these processing steps are very popular, it has not yet been described under which conditions they are optimal choices.

With the proposed framework, it is possible to deduce optimal filters under the incorporation of certain assumptions, similar to the derivation of similarity measures. For this, we focus on the maximization of the coupling term $p(d \mid u)$ with all considerations analogously for $p(e \mid v, T)$. Incorporating the assumption of white Gaußian noise and conditional independence we obtain

$$\max_d p(d \mid u) = \max_d \prod_{i=1}^{N} \prod_{j \in \mathcal{N}_i} p(d_i \mid u_j) = \max_d \prod_{i=1}^{N} \prod_{j \in \mathcal{N}_i} \exp\left(-\omega_j (d_i - u_j)^2\right). \quad (5.21)$$

Following the maximum likelihood estimation by calculating the log-likelihood function [Bishop, 2006] leads to the optimal solution for $d$. This estimation was extended to the usage of various norms, considering for instance least absolute values, instead of least squares. Further extensions resulted in M-estimators, and later, generalized M-estimators [Hoaglin et al., 1983]. We consider in the following the minimization problem

$$\min_{d} \sum_{i=1}^{N} ||\Lambda \cdot (d_i \mathbf{1} - u(\mathcal{N}_i))|| \quad \text{and} \quad \Lambda = \text{diag}(\omega_1, \ldots, \omega_{|\mathcal{N}_i|}) \tag{5.22}$$

with a vector norm $||.||$, weights $\Lambda$, and the one vector $\mathbf{1} = [1, 1, \ldots]^\top$. The sum in equation (5.22) is minimal if each summand is minimal, so that we can optimize for each descriptor $d_i$ separately

$$\min_{d_i} ||\Lambda \cdot (d_i \mathbf{1} - u(\mathcal{N}_i))||. \tag{5.23}$$

Calculating the derivative with respect to $d_i$ and setting it to zero leads to optimal descriptors. For different norms and $\Lambda = \mathbf{I}$, this results in the following descriptors

- $||.||_2^2$: $d_i$ is the mean of $u(\mathcal{N}_i)$
- $||.||_1$: $d_i$ is the median of $u(\mathcal{N}_i)$
- $||.||_\infty$: $d_i$ is $[\max(u(\mathcal{N}_i)) - \min(u(\mathcal{N}_i))]/2$.

As an example for least squares and arbitrary weights, we obtain

$$d_i = \frac{1}{\Pi} \sum_{j \in \mathcal{N}_i} \omega_j u_j \qquad \text{with} \qquad \Pi = \sum_{j \in \mathcal{N}_i} \omega_j. \tag{5.24}$$

Modifying the weights in this case allows for modeling arbitrary linear filters. For the creation of scale-space, the weights follow the entries of a Gaußian filter mask, with the typical decay from the center.

#### 5.4.1.2 Gradient-Based Similarity Measures

Gradient-based similarity measures are, for instance, of interest in 2D-3D registration [Penney et al., 1998]. Example metrics are gradient correlation and gradient difference. The gradients are calculated with the Sobel operator represented as $3 \times 3$ filter mask. Subsequently, the correlation coefficient or difference is evaluated between the gradients of the images. For modeling the Sobel operator in the maximum likelihood framework, as described in section 5.4.1.1, we have to adapt equation (5.24), because the weights for differential operators sum up to zero. Consequently, we do not consider the normalization factor $\Pi$ and set the weights $\omega_j$ according to the Sobel mask. The description layers of our framework represent the gradient images, which are successively matched. In a more recent article, [Shams et al., 2007] propose gradient intensity-based registration, where mutual information between the gradient

images is calculated. The description layers for both registration approaches [Penney et al., 1998, Shams et al., 2007] are the same, it is only the metric that is changing. This shows the increased modularity provided by our framework due to the explicit consideration of description layers.

#### 5.4.1.3 Local Similarity Measures

The term local similarity measure is ambiguous. It can on the one hand refer to similarity measures that calculate a global score based on locally adapted properties, such as local normalized cross correlation (LNCC) [Larose, 2001], and on the other hand, to the local evaluation of the similarity measure, as it is necessary for block matching and deformable registration [Hermosillo et al., 2002, Glocker et al., 2008]. For LNCC, the NCC is calculated within windows of $7 \times 7$ or $11 \times 11$ pixels for all spatial locations. In our framework, the descriptors $d_i$ and $e_i$ would be vectors with all the normalized intensity values of the local neighborhoods. The objective function is then the accumulated inner product of the vectors $\sum_{i=1}^{N} < d_i, e_i >$.

#### 5.4.1.4 Interpolation

For the comparison of images, the transformed image has to be interpolated on the grid of the source image. The simplest technique is to use a nearest neighbor interpolation, where the neighborhood system consists only on the nearest neighbor. The frequently applied bi-linear interpolation in 2D and tri-linear interpolation in 3D need a 4- and 8-neighborhood, respectively. More complex methods like cubic interpolation need accordingly a larger neighborhood to determine the interpolator [Gonzalez and Woods, 2002]. Except the nearest neighbor strategy, a drawback of these methods is that they may lead to the creation of artificial intensity values. The partial volume (PV) distribution interpolation addresses this issue [Maes et al., 1997, Chen and Varshney, 2003]. However, grid-aligning transformations cause problems for PV interpolation [Pluim et al., 2000b] because no interpolation results in higher joint entropy. A recent discussion of interpolation artifacts together with possible solutions is presented in [Rohde et al., 2009]. An approach that is not part of this study is the nonparametric (NP) windows technique [Kadir and Brady, 2005, Dowson et al., 2008], which is equivalent to sampling the images at an infinite resolution, leading to more robust estimates. We conclude that whatever interpolation method is applied, it bases upon a local context.

#### 5.4.1.5 Multi-Modal Registration

Intensity-based registration of images from different modalities can be solved in numerous ways. Possibilities are to apply similarity measures that can deal with complex intensity relationships, to simulate one modality from the other, or to transfer both images to a common representation. For the last two approaches the multi-modal alignment problem is reduced to a mono-modal one. Popular examples for the simulation are the creation of digitally reconstructed radiographs (DRRs)

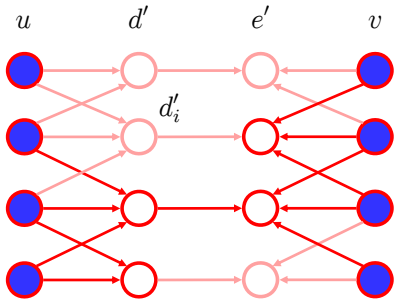



Figure 5.9: Extension of the graph of figure 5.6(b). Keypoint locations in the original images are marked red, while non-keypoint locations are marked light red. Only those descriptor nodes are marked active (red) that correspond to keypoint locations. Further, only those pairs of nodes are considered for the similarity calculation, where both are active.

for the registration of X-ray fluoroscopy images to CT [Penney et al., 1998] and the simulation of ultrasound images from CT for US-CT registration [Wein et al., 2008]. For the first example, the description layers consist of the X-ray image and DRR, and for the second example, they consist of the original and simulated ultrasound image. For the simulation, each pixel is dependent on the locations the beam penetrated before to account for attenuation. These locations constitute the local neighborhood of the descriptors. For the last approach to multi-modal registration by transferring both images to a common representation, we introduce a novel method in chapter 6 by calculating structural representations. The structural representations form the description layers, which are registered with simple similarity measures.

### 5.4.2 Geometry-Based Registration

The integration of geometric registration in our framework corresponds to embedding the feature points on a dense grid. This is similar to the dense set of descriptors used for matching in stereo reconstruction [Tola et al., 2008]. Once the descriptors are calculated for each image, the next step is the comparison between the images. Looking at the approaches for geometry-based registration, we observe that typically SSD is evaluated between the descriptors, which is derived from the similarity term $p(e|d)$. The difference to intensity-based registration is, however, the focus on certain keypoint locations. To account for this change, we extend the description layers with keypoint information, $d'_i = [d_i, k_i]$, with $k_i = 1$ for a keypoint and $k_i = 0$ for no keypoint. Analogously for $e'_i = [e_i, l_i]$. Assuming the independence of descriptor and keypoint, we obtain for the coupling terms

$$p(d_i, k_i \mid u(\mathcal{N}_i)) = p(d_i \mid u(\mathcal{N}_i)) \cdot p(k_i \mid u(\mathcal{N}_i)). \tag{5.25}$$

The term $p(k_i = 1 \mid u(\mathcal{N}_i))$ expresses the likelihood that location $x_i$ is a keypoint location in image $u$. The location $x_i$ is a keypoint if it is a keypoint in both images,

$p(k_i = 1 \mid u(\mathcal{N}_i)) \cdot p(l_i = 1 \mid v(\mathcal{N}'_i))$. Commonly in geometry-based registration, hard assignments are made, $p(k_i = 1 \mid u(\mathcal{N}_i)) \in \{0, 1\}$, meaning that the similarity of the descriptors is only considered if both locations are keypoints, as illustrated in figure 5.9. Soft assignments may, however, be more practical in noisy environments, leading to $p(k_i = 1 \mid u(\mathcal{N}_i)) \in [0, 1]$. Considering these probabilities on the entire image domain would lead to images that resemble distance maps, with the appearance being dependent on the calculation of $k_i$ and the selected distribution.

Next to the coupling terms, we obtain for the similarity term

$$p(e_i, l_i \mid d_i, k_i) = p(e_i \mid d_i) \cdot p(l_i \mid k_i) \cdot \frac{p(d_i) \cdot p(k_i) \cdot p(k_i \mid e_i) \cdot p(d_i \mid l_i)}{p(d_i, k_i)^2} \tag{5.26}$$

$$= p(e_i \mid d_i) \cdot p(l_i \mid k_i), \tag{5.27}$$

by applying Bayes' theorem and incorporating independence assumptions. These are the independence between the descriptor $d_i$ and keypoint $k_i$, as well as, the independences between keypoints and descriptors of different images ($p(k_i \mid e_i) = p(k_i)$ and $p(d_i \mid l_i) = p(d_i)$). The probability $p(l_i \mid k_i)$ models if two locations have the same keypoint value. This is comparable to the similarity term $p(e_i \mid d_i)$, but instead of expressing the similarity between descriptors, it expresses the similarity between keypoints. The extension of the joint probability (cf. equation (5.19)) with keypoint information is

$$p(u, v, d', e', T) = \prod_{i=1}^{N} p(e_i, l_i | d_i, k_i) \cdot p(d_i, k_i = 1 | u(\mathcal{N}_i)) \cdot p(e_i, l_i = 1 | v(\mathcal{N}'_i)) \tag{5.28}$$

$$= \prod_{i=1}^{N} \underbrace{p(e_i \mid d_i) \cdot p(l_i \mid k_i)}_{\text{similarity}} \cdot \underbrace{p(d_i \mid u(\mathcal{N}_i)) \cdot p(e_i \mid v(\mathcal{N}'_i))}_{\text{coupling}} \cdot \tag{5.29}$$

$$\underbrace{p(k_i = 1 \mid u(\mathcal{N}_i)) \cdot p(l_i = 1 \mid v(\mathcal{N}'_i))}_{\text{keypoint}}, \tag{5.30}$$

by incorporating the results of equations (5.25) and (5.27). Geometric registration can be classified into aligning landmarks, point clouds, surfaces, and features. We describe in the following the probabilities that are involved in the different types of registration, summarized in table 5.1.

#### 5.4.2.1 Landmarks

The term landmark-based registration is ambiguously used in the literature, where we consider it in the sense that experts identify the location of the keyoint and also provide a distinctive description. An example is a physician who identifies anatomical landmarks such as the nose, eyes, or ears on the scans. Most important is the probability $p(e_i \mid d_i)$, which evaluates the similarity that locations with the same labels overlap. The other terms can be used to model the confidence in the assignment of the label and the keypoint location. $p(d_i \mid u(\mathcal{N}_i))$ and $p(e_i \mid v(\mathcal{N}'_i))$

Table 5.1: Probability terms for various types of registration approaches. Simplified notation is used.

| Registration Type | Probability Terms |
|---|---|
| Landmark | $p(e\|d) \cdot p(d\|u) \cdot p(e\|v) \cdot p(k=1\|u) \cdot p(l=1\|v)$ |
| Point | $p(l\|k) \cdot p(k=1\|u) \cdot p(l=1\|v)$ |
| Feature | $p(e\|d) \cdot p(d\|u) \cdot p(e\|v) \cdot p(k=1\|u) \cdot p(l=1\|v)$ |
| Intensity | $p(e\|d) \cdot p(d\|u) \cdot p(e\|v)$ |

expresses the certitude in labeling, *e.g.* if the point clicked on is really the tip of the nose. With $p(k_i = 1 \mid u(\mathcal{N}_i))$ and $p(l_i = 1 \mid v(\mathcal{N}'_i))$ it is possible to model the confidence in the geometric location. The physician knows, for instance, that this is the nose but he is not confident if the tip is several pixels to the left or right.

#### 5.4.2.2 Point Clouds and Surfaces

Point clouds or surfaces are mainly extracted from the images through segmentation. Alternative sources are 3D scanners and fiducial markers [Hajnal et al., 2001]. A common drawback of these approaches is the dependency on a meaningful segmentation and the propagation of the segmentation errors to the registration. In contrast to all other approaches, we do not have descriptive information in point-based registration. We therefore use $p(l_i \mid k_i)$, instead of $p(e_i \mid d_i)$, to calculate how well points are matching. We do not consider the coupling terms $p(d_i \mid u(\mathcal{N}_i))$ and $p(e_i \mid v(\mathcal{N}'_i))$, because we do not have descriptors. Finally, the terms $p(k_i = 1 \mid u(\mathcal{N}_i))$ and $p(l_i = 1 \mid v(\mathcal{N}'_i))$ quantify the reliability of points being detected and therefore weight their influence on the overall optimization.

#### 5.4.2.3 Features

While in landmark-based registration, the localization and description of the keypoints takes place manually, and for point clouds, the localization is automatic but no description is provided, feature-based registration performs the extraction as well as the description automatically. The first task, the *keypoint localization*, identifies locations that can repeatedly be assigned under different views of the same object. Popular methods include the difference-of-Gaußian (DoG) [Lowe, 2004], Harris detector, Harris-affine, and Hessian-affine detector [Mikolajczyk and Schmid, 2004]. The output of these detectors, after normalization, is set to the variables $k_i$ and $l_i$. The second step, the *feature description*, has to represent the characteristics of the point within its local neighborhood. Frequently used image descriptors are *e.g.* Scale-Invariant Feature Transform (SIFT) [Lowe, 2004], Speeded-Up Robust Features (SURF) [Bay et al., 2008], and Gradient Location and Orientation Histogram (GLOH) [Mikolajczyk and Schmid, 2005]. The descriptors are assigned to the corresponding locations on the description layers. The last step is the *feature*

*matching*, resulting in the image transformation. For this, descriptors of both images at the corresponding locations are compared.

In our framework, $p(k_i = 1 \mid u(\mathcal{N}_i))$ and $p(l_i = 1 \mid v(\mathcal{N}'_i))$ emphasize the influence of keypoint locations or restrict it to them. The terms $p(d_i \mid u(\mathcal{N}_i))$ and $p(e_i \mid v(\mathcal{N}'_i))$ are applied for the deduction and calculation of the descriptors from the images. They ensure that the descriptors well characterize the local image context. $p(e_i \mid d_i)$ expresses the similarity of descriptors.

Looking at feature-based approaches, we clearly see the local nature of these techniques. Choosing SIFT as an example, the keypoint localization with DoG searches the local maximum in scale-space. The DoG can be modeled by setting the appropriate weights in the linear filtering in equation (5.24). The maximum search only considers the direct neighbors. The SIFT descriptor uses $4 \times 4$ blocks around the keypoint, where each block consists of $4 \times 4$ pixels of the corresponding scale-space level. In total, a $16 \times 16$ neighborhood of each keypoint is considered for building the descriptor, demonstrating the restriction to a local context. We are able to model them due to the neighborhood extension and integration of latent layers.

### 5.4.3 Hybrid Methods

Hybrid registration approaches combine multiple alignment techniques to achieve an improved registration result. So far, it has not been possible to describe hybrid approaches that combine techniques from geometric and iconic registration in a common framework, because there was no framework that enabled the modeling of both registration approaches. As seen in sections 5.4.1 and 5.4.2, the proposed probabilistic framework enables the description of a multitude of registration techniques by choosing different descriptors. A possible differentiation of hybrid approaches is to distinguish between the consecutive application of registration [Johnson and Christensen, 2002, Wang and Feng, 2004, Rohr et al., 2004, Teng et al., 2006, Azar et al., 2006, Georgel et al., 2008], or the coupling to a joint energy formulation [Pluim et al., 2000a, Hartkens et al., 2002, Kybic and Unser, 2003, Feldmann et al., 2006, Wörz and Rohr, 2008, Biesdorf et al., 2009, Wacker and Deinzer, 2009, Brox et al., 2009]. For the joint formulation, we consider the sets of descriptors $\mathcal{D}$ and $\mathcal{E}$, which can contain descriptors from geometric registration, such as SIFT, and from iconic registration, such as entropy and gradient information. The final marginalization is similar to equation (5.20)

$$p(u, v, T) \approx \sum_{d \in \mathcal{D}, e \in \mathcal{E}} p(u, v, d, e, T). \tag{5.31}$$

Since we marginalize only over finite sets and not all possible descriptors, we only achieve an approximation. In chapter 8, we present a hybrid similarity measure for ultrasound registration.

For the consecutive approaches, we face the issue of modeling the temporal evolution of the registration, in order to decide when to change the method. This problem, about the missing information of the progress of the registration, is also encountered in the hyperplane approximation of the Jacobian matrix for optimization [Jurie and

Dhome, 2002]. In [Gay-bellile et al., 2007] an approach is proposed for learning which matrix to choose at which time of the registration based on the current intensity error. These approaches could be extended to not just adapting the optimization during the registration, but also the similarity measure and descriptors.

### 5.4.4 Prior Knowledge

Prior knowledge is important for image registration because it can help to make the registration more robust and enlarge the capture range [Zöllei et al., 2007]. With the increasing availability of image databases more data becomes available for analysis and therefore more powerful priors can be defined. Prior knowledge can be integrated in several parts of the registration process. Basically, even the expertise of the user for selecting a specific similarity measure or optimization procedure can be considered as prior knowledge. In the literature, prior knowledge is mainly used on the joint distribution model and the transformation.

Prior knowledge on the transformation is modeled with the term $p(T)$ in equation (5.18). A common prior is to force the transformation to be smooth. For a prior on the joint distribution, [Leventon et al., 1998] evaluate the current samples under a previously constructed model distribution. In [Chung et al., 2002], the prior model distribution is compared to the current model distribution with the Kullback-Leibler distance. A multinomial model of joint intensities is used in [Zöllei et al., 2007] and the effect of uninformative, informative, and strong priors is analyzed. An uninformative prior only considers the observed data and therefore corresponds to standard maximum likelihood approaches and a strong prior focuses on the constructed model distribution like in [Leventon et al., 1998]. The informative prior presents a compromise between prior and current observations.

These approaches can directly be integrated into the probabilistic framework, however, it it also provides new possibilities for integrating prior information. One way is to give information about the structure of the local neighborhood by *e.g.* learning it from previous images [El-Baz and Gimel'farb, 2008]. Another way is to assign probabilities to the descriptors, which would enable to learn about descriptors that work well for the specific alignment scenarios [Seshamani et al., 2009]. And lastly, prior information can be introduced through the coupling terms $p(d|u)$ and $p(e|v,T)$ to assure the description layers to resemble the original images.

## 5.5 Dynamic Adaptation of Description Layers

In the last section, we showed how registration techniques can be modeled with the proposed framework. Further, we illustrated a continuum of registration approaches, classified by the uniqueness of their descriptors. We achieve this increased flexibility in modeling by introducing layers consisting of latent random variables. For the approaches in the last section, these layers were calculated with various deterministic algorithms and did not change during the registration. In this section, we illustrate the second advantage of our model, the dynamic adaptation of the description layers.

Instead of mainly reducing the optimization to the similarity term $p(e|d)$, we now rely on the interaction of coupling and similarity terms.

### 5.5.1 Deformable Registration

Deformable registration is frequently formulated in the context of calculus of variations [Hermosillo et al., 2002], with the energy

$$\mathcal{I}(u, v, T) = \mathcal{J}(u, v, T) + \mathcal{R}(T) \tag{5.32}$$

to maximize. The term $\mathcal{J}$ measures the similarity between the warped images, and $\mathcal{R}$ the regularity of the displacement field. The regularization is commonly performed by putting smoothness constraints on the transformation model. This formulation naturally embeds into our framework by taking the original images as $u$ and $v$, the deformed ones as $e$ and $d$, and applying the logarithm to equation (5.19)

$$\mathcal{E} = \mathcal{E}(T) + \sum_{i=1}^{N} \mathcal{E}(e_i, d_i) + \mathcal{E}(d_i, u(\mathcal{N}_i)) + \mathcal{E}(e_i, v(\mathcal{N}_i')). \tag{5.33}$$

The energy $\mathcal{E}(e_i, d_i)$ corresponds to $\mathcal{J}$. The energies $\mathcal{E}(d_i, u(\mathcal{N}_i)), \mathcal{E}(e_i, v(\mathcal{N}_i'))$, and $\mathcal{E}(T)$ to the regularizer $\mathcal{R}$, where we define the energies as $\mathcal{E}(.) = \log(p(.))$ [Mumford, 1994]. The regularization in our model is not only performed by analyzing the transformation field but also by comparing the deformed image to the original one, which we have not yet seen in the literature. This provides a probabilistic rationale for the variational formulation of deformable registration, similar to the energy functionals for segmentation of Mumford [Mumford, 1994].

We already referred to several articles that integrate context into deformable registration throughout the article. Here we just want to mention two recent approaches. [Wu and Fan, 2009] replace the core assumption of optical flow - the constancy of intensity - with constancy of context. [Shekhovtsov et al., 2008] apply Markov random fields (MRFs) for the deformable registration of images. The likelihood term $p(e \mid d)$ is modeled with univariate potentials and the regularization with pairwise potentials.

### 5.5.2 Joint Registration and Segmentation

Fundamental operations in image analysis include the segmentation and registration of images. Although they are most times solved separately, there are applications where they can benefit mutually and accordingly a joint formulation is expedient [Ashburner and Friston, 2005, D'Agostino et al., 2006, Pohl et al., 2006, Wyatt and Noble, 2003, Xiaohua et al., 2004]. The performance of any segmentation approach is primarily dependent on the discriminative power of the underlying likelihood model for the data [Wyatt and Noble, 2003]. Multiple measurements with different imaging modalities or viewpoints could therefore improve segmentation. On the other hand, the alignment of segmented images, instead of the original ones, significantly reduces

the influence of noise and consequently facilitates the registration. In our framework, the description layers represent the segmented images. The similarity term $p(e \mid d)$ drives the correct global alignment and also provides the combination of both image segmentations. The coupling terms $p(d \mid u)$ and $p(e \mid v, T)$ counterbalance the effect of letting both segmentations looking as similar as possible, by ensuring the segmentations to be close to the underlying data.

We show how the MAP MRF approach in [Wyatt and Noble, 2003] naturally integrates into our framework. The MAP problem is stated using Bayes rule

$$p(d, e, T \mid u, v) = \frac{p(u, v \mid d, e, T) \cdot p(d, e, T)}{p(u, v)} \tag{5.34}$$

with the images $u, v$ given and the transformation $T$ and segmentations $d, e$ to calculate. The likelihood term $p(u, v \mid d, e, T)$ is represented with a Gaußian mixture model (GMM), and the prior $p(d, e, T)$ with an MRF using the Ising model. At the beginning of the registration, when the images are far from being correctly aligned, the joint modeling of both images is not meaningful. Therefore, the independence of the images and consequently the labels is assumed, leading to

$$p(d, e, T \mid u, v) = \frac{p(u \mid d) \cdot p(d) \cdot p(v \mid e) \cdot p(e) \cdot p(T)}{p(u) \cdot p(v)}. \tag{5.35}$$

For the joint optimization, an alternation is done between solving for the optimal labeling with iterated conditional modes and the alignment with the Powell method.

## 5.6 Groupwise Registration

The analysis of a group or population of images necessitates their alignment to a common reference frame, referred to as *groupwise* registration. A variety of groupwise registration approaches exist, including the sequential pairwise alignment with a consecutive estimation of the global transformations [Joshi et al., 2004, Vercauteren et al., 2006, Wachinger et al., 2007] or the simultaneous registration of all images [Zöllei et al., 2003, Cootes et al., 2004, Zöllei et al., 2005, Twining et al., 2005, Learned-Miller, 2006, Wachinger et al., 2007, Sidorov et al., 2009]. The first approach is based on pairwise registration and is modeled in the framework that has been described so far. For the simultaneous registration, we have to generalize the framework, as it is presented in the following.

Assume $n$ images $\mathcal{U} = (u^1, \ldots, u^n)$ and the corresponding global transformations $\mathcal{T} = (T^1, \ldots, T^n)$. Analogously to the pairwise case, we assign a description layer $d^j$ to each image $u^j$, with $\mathcal{D} = (d^1, \ldots, d^n)$ the list of description layers. We apply once again a graphical model to deduce the probabilities, where we only use the plate for illustration because of the complexity. We show the graphical model in figure 5.10, where we let each descriptor be dependent on a local neighborhood, and let each descriptor depend on each other descriptor, leading to a complete graph in

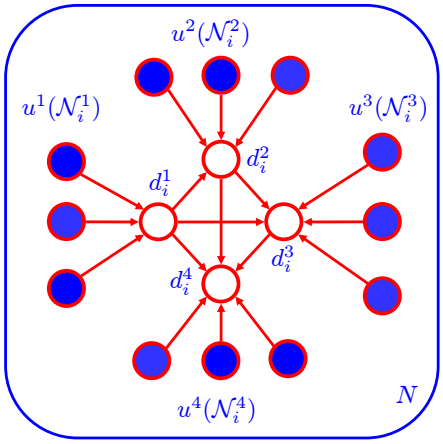



Figure 5.10: Probabilistic graphical model of the groupwise, contextual registration framework as a plate illustrated for four images $u^1, \ldots, u^4$ and exemplary 3-neighborhood. The transformations are not displayed for clarity of presentation, but they are analogously to figure 5.6(a) connected to the descriptor nodes.

the middle. Alternatively, hyper-edges [Zass and Shashua, 2008] could be used. This graphical model factorizes to

$$p(\mathcal{U}, \mathcal{D}, \mathcal{T}) = p(\mathcal{T}) \cdot p(\mathcal{U}) \cdot p(d^1|u^1, T^1) \cdots p(d^n|d^{n-1}, \ldots, d^1, u^n, T^n) \tag{5.36}$$

with $p(\mathcal{T}) = \prod_{j=1}^{n} p(T^j)$ and $p(\mathcal{U}) = \prod_{j=1}^{n} p(u^j)$. We further deduce

$$\begin{aligned}
p(d^n|d^{n-1}, \ldots, d^1, u^n, T^n) &= \frac{p(d^{n-1}, \ldots, d^1, u^n, T^n|d^n) \cdot p(d^n)}{p(d^{n-1}, \ldots, d^1, u^n, T^n)} && (5.37)\\
&= \frac{p(d^{n-1}, \ldots, d^1|d^n) \cdot p(u^n, T^n|d^n) \cdot p(d^n)}{p(d^{n-1}, \ldots, d^1) \cdot p(u^n, T^n)} && (5.38)\\
&= \frac{p(d^n, \ldots, d^1)}{p(d^{n-1}, \ldots, d^1)} \cdot \frac{p(u^n, T^n|d^n)}{p(u^n, T^n)} && (5.39)\\
&= \frac{p(d^n, \ldots, d^1)}{p(d^{n-1}, \ldots, d^1)} \cdot \frac{p(d^n|u^n, T^n) \cdot p(u^n, T^n)}{p(d^n) \cdot p(u^n, T^n)} && (5.40)\\
&= \frac{p(d^n, \ldots, d^1)}{p(d^{n-1}, \ldots, d^1) \cdot p(d^n)} \cdot p(d^n|u^n, T^n). && (5.41)
\end{aligned}$$

Two assumptions are incorporated to reach equation (5.38). First, the independence $p(d^1, \ldots, d^{n-1}, u^n, T^n) = p(d^1, \ldots, d^{n-1}) \cdot p(u^n, T^n)$, which is justified because the descriptor layers of the $n-1$ images do not depend on the $n$th image and transformation. Second, the conditional independence $p(d^{n-1}, \ldots, d^1, u^n, T^n|d^n) = p(d^{n-1}, \ldots, d^1|d^n) \cdot p(u^n, T^n|d^n)$, which is also justified with a similar argument considering the conditioning on $d^n$.

Applying this to all conditional probabilities in equation (5.36), the probabilities

nicely cancel out, and we end up with

$$p(\mathcal{U}, \mathcal{D}, \mathcal{T}) = \underbrace{p(d^1, \ldots, d^n)}_{\text{Similarity Term}} \cdot \prod_{j=1}^{n} \underbrace{\frac{p(T^j) \cdot p(u^j)}{p(d^j)}}_{\text{Prior Terms}} \cdot \underbrace{p(d^j \mid u^j, T^j)}_{\text{Coupling Term}}. \tag{5.42}$$

Identically to the pairwise framework, we apply the Markov property to make a descriptor $d_i^j$ from image $j$ and location $x_i$ dependent on a local neighborhood $\mathcal{N}_i^j$

$$p(d^j \mid u^j, T^j) = \prod_{i=1}^{N} p(d_i^j \mid u^j(\mathcal{N}_i^j)). \tag{5.43}$$

The marginal distributions $p(T^j), p(u^j), p(d^j)^{-1}$ are used to encode prior information. The transformations are integrated in the creation of the description layers, meaning that for a location $x_i$ of the global reference frame, descriptors $d_i^1, \ldots, d_i^n$ are provided. We obtain local image coordinates $x_i^j$ by applying the transformation $T^j$ to the global coordinate $x_i$, leading to $x_i^j = T^j(x_i)$.

The multivariate probability $p(d^1, \ldots, d^n)$ models the similarity between the descriptors. We consider multivariate similarity measures in more details in chapter 7 and apply them for motion modeling in chapter 10. Moreover, we employ a groupwise registration approach similar to [Joshi et al., 2004] for deformable mosaicing in appendix B. Here we just present a concise description for completeness. [Twining et al., 2004] propose the minimum descriptor length as similarity criterion for group-wise registration. The rationale is that correctly aligned images can be transmitted with a minimal amount of bits. [Learned-Miller, 2006] propose the congealing framework for the alignment of a large number of binary images from a database of handwritten digits and for the removal of unwanted bias fields in magnetic resonance images. For the deduction of congealing, independent but *not* identical distributions of the coordinate samples $d_i$ and i.i.d. description layers $d^j$ are considered, leading to

$$p(d^1, \ldots, d^n) = \prod_{i=1}^{N} \prod_{j=1}^{n} p^i(d_i^j). \tag{5.44}$$

In [Studholme and Cardenas, 2004], a joint density function for multivariate similarity estimation is constructed, which has a problem for larger image sets. We present an own approximation to the multivariate density function (cf. chapter 7), which is based on the accumulation of pairwise estimates. This factorizes as

$$p(d^1, \ldots, d^n) = \prod_{j \neq k} p(d^j | d^k). \tag{5.45}$$

We conclude this section by stating that the extension of the contextual framework to groupwise registration is possible. It allows for modeling existing groupwise registration approaches and holds the advantage of integrating local neighborhood information.

## 5.7 Conclusion and Future Work

We presented a novel probabilistic framework for image registration, which is general enough to describe intensity-based, as well as geometry-based registration. The proposed framework allows us to move from just modeling the similarity function towards a modeling of larger parts of the registration process. The key extension with respect to previous frameworks is the consideration of local neighborhood information, so replacing the assumption of independent coordinate samples by the Markov property. We reviewed various registration approaches and showed how they can be modeled within our framework. This gave us the motivation to introduce a continuum of registration approaches, limited by pure geometric and iconic registration. We showed how to use the models to derive optimal descriptors, as well as integrating the dynamic adaptation of descriptors during the registration. Finally, we presented an extension for groupwise registration. The proposed framework provides further insights about the relationship of various registration techniques, and moreover, helps to understand and classify them.

# Chapter 6

# Structural Representations for Registration

*The standard approach for multi-modal registration is to apply sophisticated similarity metrics such as mutual information. The disadvantage of these metrics, in comparison to measuring the intensity difference with* e.g. *L1 or L2 distance, is the increase in computational complexity and consequently the increase in runtime of the registration. An alternative approach, which has not yet gained much attention in the literature, is to find image representations, so called structural representations that allow for the application of the L1 and L2 distance for multi-modal images. This has not only the advantage of a faster similarity calculation but enables also the application of more sophisticated optimization strategies. In this chapter, we theoretically analyze the requirements for structural representations. Further, we introduce two approaches to create such representations, which are based on the calculation of patch entropy and manifold learning, respectively. While the application of entropy has practical advantages in terms of computational complexity, the usage of manifold learning has theoretical advantages, by presenting an optimal approximation to one of the theoretical requirements. We perform experiments on multiple datasets for rigid and deformable registration with good results with respect to both runtime and quality of alignment.*

## 6.1 Introduction

The objective of image registration is to find the correct spatial alignment between corresponding structures in images. This task is made difficult by intensity variations between images. Such variations can originate from a multitude of sources, such as illumination changes in optical images, field inhomogeneities in magnetic resonance images, and simply, different imaging modalities. In chapter 5, we have already discussed that a common approach in iconic registration is to integrate similarity metrics that are robust to those intensity variations, assuming a functional or statistical intensity relationship, instead of an identical one. On the other hand, geometric registration approaches that build upon an automatic keypoint extraction and de-

scription have to apply methods that are robust to intensity variations. Widespread descriptors such as SIFT [Lowe, 2004] and GLOH [Mikolajczyk and Schmid, 2005] achieve such robustness by building upon intensity differences, rather than absolute intensity values, by calculating histograms of image gradients. The registration of images from different modalities is, however, affected by more substantial intensity variations.

In this chapter, we introduce a representation of images that is only dependent on the depicted structures and not on the intensities used to encode them. Such a *structural representation* can assist several image processing tasks, while we focus on registration in this chapter. We obtain structural representations by calculating a dense set of descriptors that capture the structural information of each of the local patches. Subsequently, the input images are replaced by the dense set of descriptors, on which a regular intensity-based registration is performed. This guarantees a seamless integration into existing registration frameworks. Referring back to chapter 5, we see that the structural representations correspond to the description layers $d$ and $e$. Since we do not allow for a dynamic adaptation of the layers, the coupling terms are not relevant for the optimization, which is consequently reduced to the similarity term $p(e \mid d)$. The advantage of the structural representation is that the simple L1 or L2 distance can be used for the registration of multi-modal images. These metrics are computationally less expensive than mutual information and enable therefore a faster registration. This is even more important for groupwise registration, where the speed improvement becomes quadratic, as we discuss in section 6.3.2. Additionally, more efficient optimization schemes can be applied for multi-modal registration. Finally, L1 and L2 distances are better suited for parallelization than complex multi-modal similarity measures, which is important with respect to transferring the computation to GPUs for further speed-up.

In sections 6.3 and 6.4, we describe the integration of structural images for registration and theoretically analyze the properties of a structural representation. In section 6.5, we show that the minimal coding length for transferring a patch over a channel, calculated with the Shannon entropy, properly captures the information content of a patch invariant to the intensity. The created *entropy images* present an exemplary structural representation. There is, however, a risk of ambiguities, *i.e.* several patches can lead to the same entropy value. In order to address this issue, we propose to integrate spatial information to the density estimation. In section 6.6, we introduce an alternative structural representation, based on the application of manifold learning. The representation created with Laplacian eigenmaps has superior theoretical properties, because it optimally fulfills one of the requirements for a structural representation, the preservation of locality. This means that patches, which are close in high-dimensional patch space, are mapped to a close structural representation. Moreover, this technique exploits the internal similarities across modalities, similar to [Penney et al., 2008], for the structural embedding. We refer to the images created with Laplacian eigenmaps shortly as *Laplacian images*. In sections 6.7 and 6.8, we evaluate the performance of entropy and Laplacian images.

## 6.2 Related Work

There are two groups of related work, first, methods that are related because they deal with transforming a multi-modal registration to a mono-modal one, and second, articles that are related to our proposed representations (see sections 6.2.1 and 6.2.2).

Techniques that reduce a multi-modal to a mono-modal registration can be differentiated into two classes. The first ones try to simulate one modality from the other. Examples are X-Ray to CT registration with the creation of digitally reconstructed radiographs [Penney et al., 1998] and ultrasound to CT registration with the simulation of ultrasound images [Wein et al., 2008]. In our case, we are interested in a general structural representation, so that the application of these specific approaches is not suitable.

The second group consists of methods that transfer both images into a third, artificial modality. Examples are (i) the application of morphological tools [Maintz et al., 1997], (ii) recoloring images depending on the variances of the image regions [Andronache et al., 2008], (iii) the usage of edge- and ridge-information [Maintz et al., 1996, Van den Elsen et al., 1995], (iv) cross-correlating gradient directions [Haber and Modersitzki, 2007], and (v) the creation of shadow-invariant optical images [Pizarro and Bartoli, 2007]. Approaches (ii) and (iv) use cross-correlation for the comparison, indicating that the description is not truly identical. The morphological approach [Maintz et al., 1997] mainly leads to a surface extraction, and although it employs gray values instead of only binary values, much internal information is lost. Finally, edge, ridge, and gradient estimation is problematic for points where more than two regions are meeting, *e.g.* T-junctions, as discussed in section 6.7.

Following the discussion in chapter 3, the analytic signal may be a suitable candidate for a structural representation. Having the property of split of identity by separating quantitative and qualitative information, the qualitative part expressed with the local phase could be of interest. In our own tests, we have not achieved convincing results, supposedly rooted in the high intensity variations of multi-modal data. The local phase was applied for multi-modal registration in [Mellor and Brady, 2005, Zhang et al., 2007], but with mutual information as similarity measure. This confirms our experience that the local phase is not directly suited for multi-modal registration with L1 or L2 distance.

### 6.2.1 Entropy Images

Concepts from information theory, specifically the entropy, have a significant influence on image registration. The widely utilized mutual information [Wells et al., 1996, Collignon et al., 1995] is building upon the entropy calculation of joint and marginal probability distributions. Similarly, the congealing framework [Learned-Miller, 2006, Zöllei et al., 2005], which is commonly used for the simultaneous alignment of multiple images, evaluates the entropy of a pixel stack. In [Buzug et al., 1997], the entropy of the difference image is calculated to align mono-modal images. Also based on the calculation of the entropy is the scale saliency algorithm [Kadir and Brady, 2001].

Salient regions are identified in images with the criterion of unpredictability in feature and scale space, where the local entropy is used for quantifying the unpredictability.

### 6.2.2 Laplacian Images

As we will describe in more details in section 6.6, the application of Laplacian images for multi-modal registration relies on the assumption of similar internal similarities in images across modalities. This was previously exploited in a novel framework for multi-modal registration based on internal similarity [Penney et al., 2008]. The two approaches are similar because small image patches are compared to find local similarities inside images. The consecutive utilization of this information is, however, entirely different. While in [Penney et al., 2008] a few internal similarity structures are identified and then transferred to the image from the other modality, we use all the patch information to build a neighborhood graph, approximating the patch manifold embedded in high dimensions. The internal similarity in images is also exploited in [Shechtman and Irani, 2007], however, only locally for constructing image descriptors.

Finally related is the approach for learning similarity measures for multi-modal registration [Lee et al., 2009]. For the supervised learning a max-margin structured output learning is used. The approach seems related because also learning is considered, however, we are interested in finding structural representations, instead of learning the similarity measure, and our unsupervised learning with Laplacian eigenmaps does not need any training. The training is not simple because correctly aligned images from the target modalities must be available. This is even more challenging for magnetic resonance images, which vary significantly in their appearance for different echo and repetition times (TE/TR).

## 6.3 Structural Image Registration

Consider two images $I, J : \Omega \to \mathcal{I}$ defined on the image grid $\Omega$ with intensity values $\mathcal{I} = \{1, \ldots, \lambda\}$. The registration is formulated as

$$\hat{T} = \arg\max_{T \in \mathcal{T}} \mathcal{S}(I, J(T)), \tag{6.1}$$

with the space of transformations $\mathcal{T}$ and the similarity measure $\mathcal{S}$. We have already seen in chapter 5, that for a more complex intensity relationship than the identity, such as an affine, a functional, or a statistical one, typical choices for $\mathcal{S}$ are the correlation coefficient, correlation ratio, and mutual information, respectively. These are, however, more computationally expensive than the L2 distance. Our goal is therefore to find structural representations $D^I$ and $D^J$ that replace $I$ and $J$ in the optimization of equation (6.1) and for which we can apply mono-modal metrics. The registration is then formulated as follows

$$\hat{T} = \arg\max_{T \in \mathcal{T}} \mathcal{S}(D^I, D^J(T)), \tag{6.2}$$

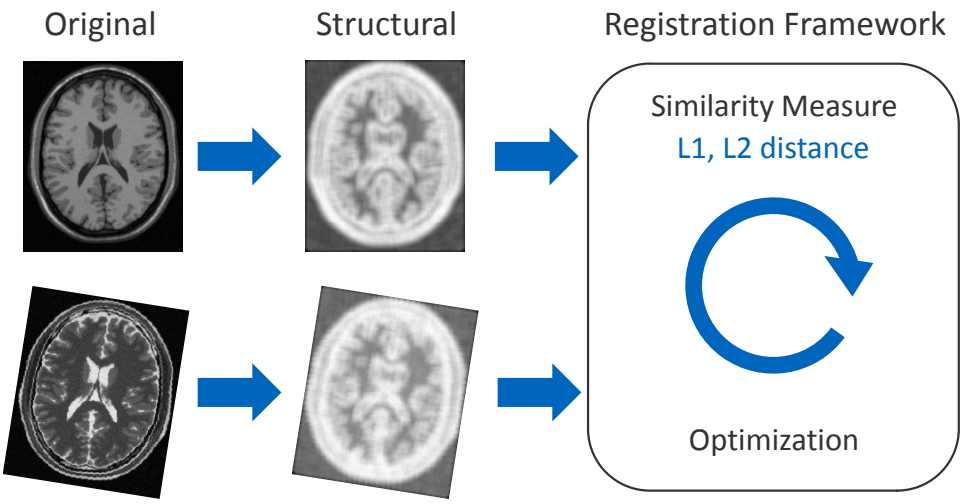



Figure 6.1: Schematic illustration of structural registration. From the original images, structural representations are calculated. In this diagram, entropy images are shown. Subsequently, these images are used in the standard intensity-based registration framework, with L1 or L2 distance as possible similarity measures.

with $\mathcal{S}$ being the L1 or L2 distance, even for images from different modalities. We illustrate the entire process schematically in figure 6.1.

### 6.3.1 Efficient Optimization

Besides the discussed similarity measure, an efficient optimization scheme is required for a fast registration. With structural representations, we transfer a multi-modal to a mono-modal registration problem; making a larger number of optimization methods applicable. Such optimization schemes are based on the Fourier transform [Reddy and Chatterji, 1996, Padfield, 2010] or the batch alignment by sparse and low-rank decomposition [Peng et al., 2010]. Fourier-based approaches calculate the correlation between images and assume linearly correlated images, analogously to [Peng et al., 2010]. This assumption is generally too limiting for multi-modal registration. The application of structural representations has the potential to make those approaches even applicable to multi-modal registration. Moreover, the efficient second-order minimization (ESM, [Benhimane and Malis, 2004]) is applicable in this case, which is further discussed in chapter 7. ESM is an extension of Gauß-Newton and was shown to converge faster than other gradient-based optimizers [Benhimane and Malis, 2004, Vercauteren et al., 2007]. ESM builds a second-order approximation of the cost function without the explicit calculation of the second derivative. The structural representation of images enables the application of ESM for multi-modal registration. Especially for deformable registration, the optimization of MI-based similarity measure with Gauß-Newton schemes is problematic, because the appearing non-sparse matrices lead to a high computational complexity [Modersitzki, 2009, Ch.9].

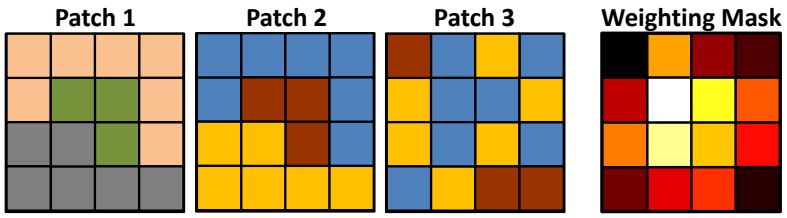



Figure 6.2: Patch 1 and 2 show the same structure but encoded with different intensities. All three patches have the same entropy of 2.0749 using an identical spatial weighting. Integrating a location dependent weighting with the modified Gaußian weighting mask shown on the right, we can differentiate patch 3 from the others.

### 6.3.2 Efficient Groupwise Registration

For groupwise registration it is even more important to have an efficient registration process, since the computational cost increases significantly when moving to the alignment of multiple images. In section 5.6, we have already mentioned that there exist various techniques to measure the similarity of a group of images, such as the congealing framework or the accumulation of pairwise estimates (APE). In chapter 7, the derivation of APE is presented in details. As we will see, $N \cdot (N-1)$ pairwise similarity measures have to be calculated to estimate the similarity of the whole group of images. Consequently, the influence of the speed-up of simple matching functions is quadratic. To conclude, structural representations for multi-modal registration enable the usage of an efficient optimizer, needing less steps, and further, they permit a faster calculation of each update step.

## 6.4 Structural Representation

In this section, we analyze the theoretical requirements on a structural representation. We break the problem of finding a structural representation for images down to the simpler problem of finding a structural representation for image patches. We denote patches that are defined on the local neighborhood $\mathcal{N}_x$ around $x$ as $P_x : \mathcal{N}_x \to \mathcal{I}$. Our objective is to find a function $f : P_x \mapsto D_x$ that assigns each patch a descriptor $D_x$ such that the descriptor captures the structural information of the patch. Since we calculate a descriptor for each location $x$, we obtain a new image with the original intensities replaced by the descriptors. Moreover, we differentiate between patches of different images, with $P_x$ being part of $I$ and $Q_x$ being part of $J$, and further $f' : Q_x \mapsto D'_x$.

We define two patches $P_x, P_y$ to be *structurally equivalent* $P_x \sim P_y$, if there exists a bijective function $g : \mathcal{I} \to \mathcal{I}$ such that

$$\forall z \in \mathcal{N}_x : P_x(z) = g(P_y(z)).$$

For an illustration, the first two patches in figure 6.2 are structurally equivalent, in contrast to the third one. The requirements $f$ and $f'$ have to fulfill are

**(R1) Locality preservation:**

$$||P_x - P_y|| < \varepsilon \quad \Longrightarrow \quad ||f(P_x) - f(P_y)|| < \varepsilon' \tag{6.3}$$

**(R2) Structural equivalence:**

$$P_x \sim Q_x \quad \Longleftrightarrow \quad f(P_x) = f'(Q_x) \tag{6.4}$$

with reasonable $\varepsilon$ and $\varepsilon'$ depending on the chosen norm. The motivation behind the first property is to ensure that similar patches are mapped to similar descriptors, which is important for the robustness to noise and the capture range of the registration. The second property states that descriptors are identical, if and only if the patches are structurally equivalent. This ensures, on the one hand '$\Rightarrow$', the desired structural representation, and on the other hand '$\Leftarrow$', a perfect discrimination of patches. Generally, the discrimination for a dense descriptor is less critical than for a sparse descriptor.

Note that in (R1) only patches from the same image are compared, because the calculation of the norm $||.||$ between patches from different modalities is not meaningful. In contrast, we require the structural equivalence (R2) only for patches of different images, because no intensity mapping is required in the same image. This model is no longer satisfiable by a global function $f$, so that we have to employ a local function for each modality, indicated with $f$ and $f'$.

### 6.4.1 Structural Equivalence vs. Modality Invariance

We would like to point out that the structural representation is different from a *modality invariant* representation, which would be the final goal. Each imaging device has its own characteristics, leading to images with specific artifacts and noise. Also, structures visible in one of the images may not be observable in the second one. As an example, compare the CT and MR images in figure 6.5, whose appearances are significantly different. It is also clear that structural images cannot detect structures, where there are none. However, we can expect to be robust to such changes, so that those structures that are present in all images can guide the registration. This problem is not specific to structural images, but is rooted in the multi-modal registration scenario, and therefore also affects multi-modal measures like MI. The application of robust metrics for comparing structural images, such as the robust M-estimation [Li et al., 1998], can limit the influence of those outliers.

## 6.5 Entropy Images

A possible interpretation of the similarity between images is to consider whether intensity changes occur at the same locations. An example to quantify the intensity

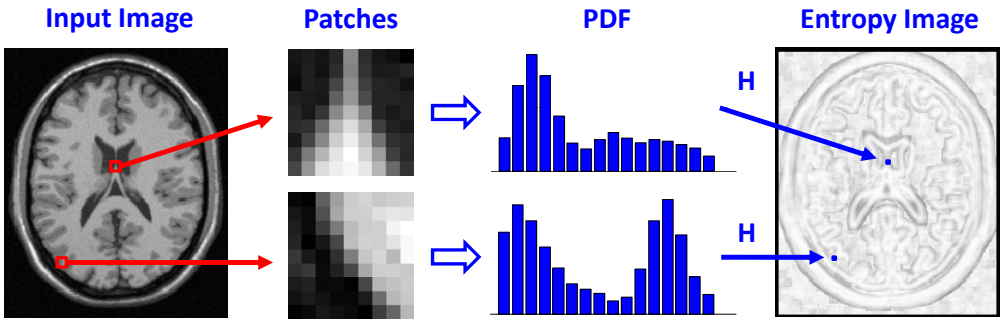



Figure 6.3: Illustration of the process for calculating entropy images. For each pixel in the image, the local neighborhood patch is selected. For these patches, the PDF of the intensity is estimated, in this example with the Parzen-window method. Finally, the entropy is computed and the result is stored in the corresponding location in order to create the entropy image.

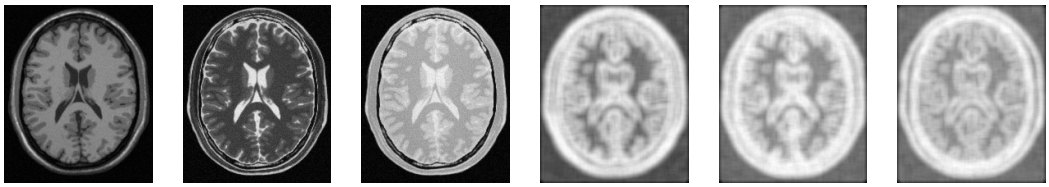


Figure 6.4: Multi-modal images (T1, T2, PD) from BrainWeb dataset together with entropy images used for rigid registration.

change is the calculation of the image gradient [Haber and Modersitzki, 2007]. This is, however, not suitable for a structural representation because of its dependency on the similarity values. A more general concept is to quantify the information content or, analogously, the bound for a lossless compression, as stated by Shannon's theorem, which is both achieved with the entropy. The Shannon entropy of a random variable $Y$ with possible values $\mathcal{I}$ is

$$\mathrm{H}(Y) = -\sum_{i \in \mathcal{I}} p(Y = i) \cdot \log p(Y = i), \tag{6.5}$$

assuming $p$ to be the probability density function of $Y$. Calculating the entropy on a dense image grid leads to

$$D_x^I = \mathrm{H}(I|_{\mathcal{N}_x}). \tag{6.6}$$

The construction of entropy images is illustrated in figure 6.3. We see that the entropy images are directly fitting in the contextual probabilistic framework of chapter 5. $D^I$ and $D^J$ are the description layers, where each descriptor $D_x^I$ is dependent on a local neighborhood $\mathcal{N}_x$ in the images.

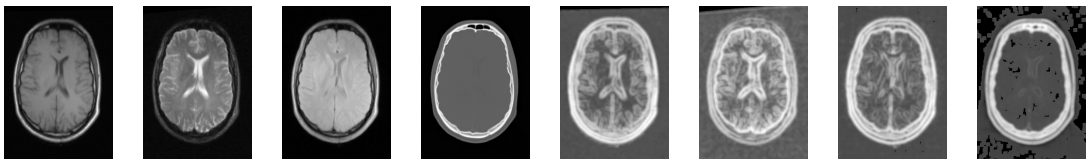


Figure 6.5: Images from RIRE dataset (T1, T2, PD, CT) together with entropy images used for rigid registration.

### 6.5.1 Verification of Structural Properties

In this section, we verify whether the entropy images fulfill the theoretical requirements on a structural representation. Since the entropy is calculated on images from both modalities, we do not have a local function $f$ for every modality. We could therefore only fulfill (R1) and (R2) if we relax the requirements and allow patches $P_x, Q_x$ to be from $I$ as well as $J$. Verifying these relaxed requirements, we see that (R1) is fulfilled because small changes in the patches also lead to small changes in the entropy. This holds for changes in the distribution, where the probability is not close to zero, because the derivative of the entropy is $-(1 + \log p(y))$. The structural equivalence, "$\Rightarrow$" of (R2), is also fulfilled because the value of the entropy is invariant to the permutation of the bins in a histogram, which is the effect of the intensity mapping $g$.

In order to be able to assess the discrimination ability of the descriptors,"$\Leftarrow$" of (R2), we quantify the number of structurally different patches. Let $\lambda = |\mathcal{I}|$ be the number of intensity levels, and $k = |\mathcal{N}_x|$ be the cardinality of the patch. We assume $\lambda \geq k$, with typical values for $\lambda = 256$ and $k = 10 \times 10$. For these numbers, we indicate the order of magnitude of the number of patches in the subsequent analysis. The total number of different patches $\eta_1$ is calculated with

$$\eta_1 = \lambda^k \approx 10^{240}. \tag{6.7}$$

The number of patches that vary in structure is equivalent to the Bell number $B$

$$\eta_2 = B(k) = \frac{1}{e} \sum_{l=0}^{\infty} \binom{l^k}{l!} \approx 10^{115}. \tag{6.8}$$

This corresponds to the number of equivalence classes of the structural equivalence relation $\sim$. Patch 1 and 2 in Fig. 6.2 are in the same class, and are therefore counted only once. The Bell numbers generally indicate the number of ways a set with $k$ elements can be partitioned into nonempty subsets. This is also the number of patches an optimal function $f$ would be able to differentiate. From a practical point of view, however, it would require more than 47 bytes per pixel to store up to $10^{115}$ different values, which could exceed the memory limit for volumetric data and decelerate the registration.

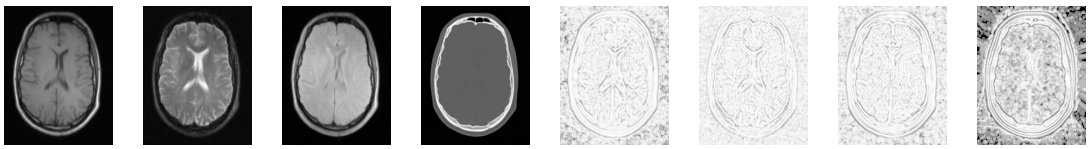


Figure 6.6: Multi-modal images (T1, T2, PD, CT) from RIRE dataset together with entropy images used for deformable registration.

The number of different distributions is

$$\eta_3 = \binom{\lambda + k - 1}{k} \approx 10^{90}, \tag{6.9}$$

which corresponds to ball picking of unordered samples with replacement.

In order to make distributions fulfill the structural equivalence, we have to sort the entries of the distribution

$$p' = \mathrm{sort}(p). \tag{6.10}$$

The number of sorted distributions $p'$ is

$$\eta_4 = \mathcal{P}(k) \approx \frac{1}{4k\sqrt{3}} e^{\pi\sqrt{2k/3}} \approx 10^8 \tag{6.11}$$

with the partition function $\mathcal{P}$. This represents the number of ways of writing an integer as a sum of positive integers, where the order of addends is not considered significant.

The final step, the mapping from ordered histograms to real values, is performed with the entropy formula in equation (6.5). For $k = 2$, the entropy uniquely assigns each ordered histogram a scalar. However, for $k \geq 3$ the mapping is no longer injective and consequently $\eta_4$ presents an upper bound to the number of different entropy values for patches. Generally, the desired number $\eta_2$ is much higher than the maximally achievable $\eta_4$, so that "$\Leftarrow$" of property (R2) is not fulfilled. Although the discrimination is more critical for a sparse than for a dense descriptor, we try to improve it by adding a spatial weighting in the density estimation in section 6.5.2.2.

### 6.5.2 Details on Entropy Estimation

There are several processing steps in the entropy estimation that influence the appearance of the entropy image, compare figures 6.5 and 6.6, where different parameters are used to create the entropy images. We evaluate the effect of those parameters for rigid and deformable registration. We present details about the experimental setup of the evaluation in section 6.7.

#### 6.5.2.1 Local Neighborhood

The size of the local neighborhood is important because it determines the cardinality of the samples for the density estimation. Consequently, there is a trade-off between

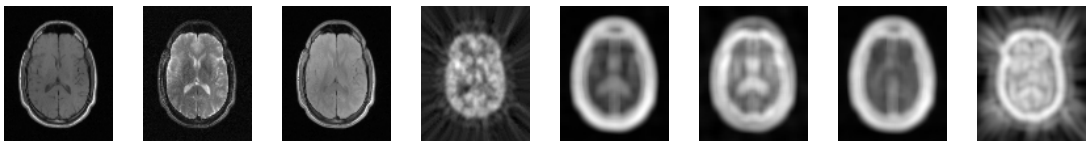


Figure 6.7: Images from RIRE dataset with PET image (PET of lower resolution, $128 \times 128$) and corresponding entropy images.

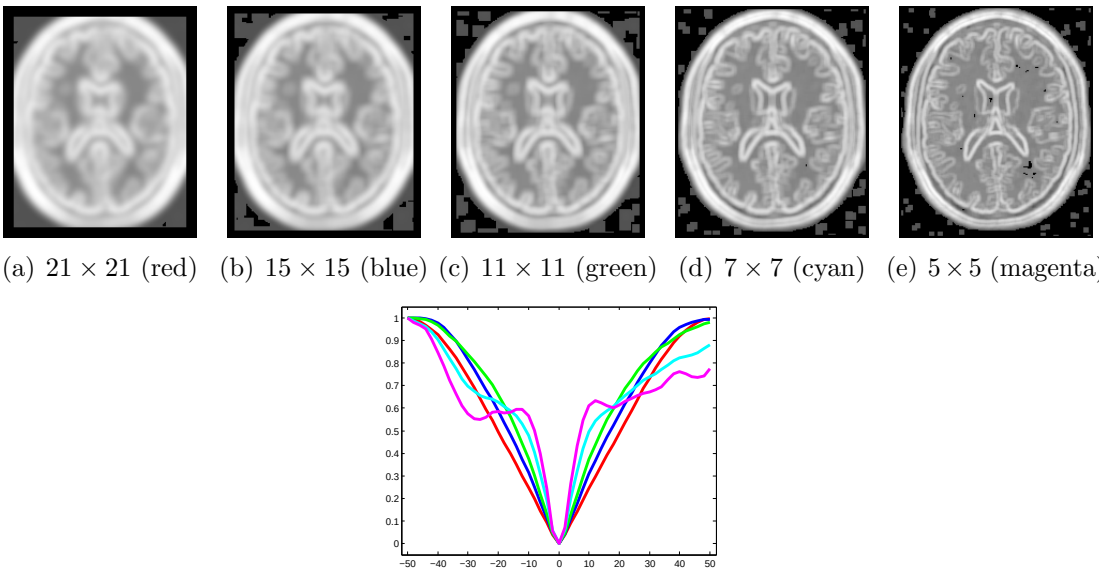
a small local neighborhood, in order to keep the estimation local, and a large neighborhood, to have sufficient statistics for an accurate estimation. For rigid registration, we evaluate local neighborhoods ranging from $5 \times 5$ to $21 \times 21$ pixels, where we found patches of size $11 \times 11$ to be a good compromise. For deformable registration, we prefer smaller patch sizes to have a more local description, and therefore test them in the range from $5 \times 5$ to $13 \times 13$, with good results for $7 \times 7$ patches. In 3D, we achieved good results with $9 \times 9 \times 9$ patches, where we adapt the neighborhood accordingly for anisotropic spacing.

An interesting effect of changing neighborhood sizes is the smoothness of the entropy images. We show images that are created with patch sizes ranging from $5 \times 5$ to $21 \times 21$ in figure 6.8. We observe smoother entropy images for larger patch sizes. This is useful for registration, because a common approach is to perform registration on several resolution and smoothness levels. However, instead of filtering with a Gaußian kernel, we change the neighborhood size. The similarity graphs for various neighborhood sizes for T1-PD registration in figure 6.8 show wider cost functions for larger patch sizes, as we would expect them for Gaußian filtering with different variances.

#### 6.5.2.2 Spatially-weighted Density Estimation

We use histogramming and the kernel-based Parzen window method for the non-parametric PDF estimation, with the latter yielding more robust results for a small number of samples. For both, the bin size has to be specified. A large number of bins makes the entropy image more sensitive to noise, while a low number deteriorates the unique patch description.

As discussed previously, we are interested in increasing the discrimination of the entropy estimation, because of the large difference between $\eta_2$ and $\eta_4$. Consider, for instance, the three patches in Fig. 6.2. While it is desirable to assign patch 1 and 2 the same value, this does not hold for patch 3. However, the PDF is the same under permutation of pixels in the patch, so that all three patches have the same entropy. To address this issue, we propose to modify the density estimation with plug-in estimators, which are entropy estimators that explicitly estimate the PDF. When constructing the image histogram $h$, each pixel's intensity has the same contribution to the respective bin. We introduce a spatial weighting function $\omega : \mathcal{N}_x \to \mathbb{R}$,



(a) $21 \times 21$ (red) (b) $15 \times 15$ (blue) (c) $11 \times 11$ (green) (d) $7 \times 7$ (cyan) (e) $5 \times 5$ (magenta)

(f) Similarity plot

Figure 6.8: Entropy images for varying patch sizes of BrainWeb T1. Similarity plot for rotation between entropy images of T1 and PD. Color of the curves as specified below the images.

assigning a weight to each patch location. The histogram update changes to

$$\forall y \in \mathcal{N}_x \quad : \quad h_x[I(y)] \;\leftarrow\; h_x[I(y)] + \omega(y). \tag{6.12}$$

This is illustrated in Fig. 6.9. We obtain the Parzen window density estimation by convolution with a Gaußian kernel and corresponding normalization.

In our experiments, we use a Gaußian, a modified Gaußian, and the identity as weighting functions $\omega$. The identity corresponds to the usual density estimation. For the Gaußian we set $\omega(y) = G_\sigma(\|y - c\|)$ with $c$ the patch center, see Fig. 6.10 for an illustration. The discrimination between patches is not optimal because the Gaußian is rotational symmetric. We therefore modify the Gaußian weighting function, see Fig. 6.2, giving it a unique weight at each patch location. We assign similar values to neighboring locations to ensure the locality preservation.

Although each location has a different weight for the modified Gaußian, the sum of several values, as it is done in the histogram calculation, can lead to the same value, and therefore ambiguities. An optimal assignment of weights to the $|\mathcal{N}_x|$ patch locations in the weighting mask, so that they are unique with respect to addition and that the dynamic range is minimal, is to use $2^1, 2^2, \ldots, 2^{|\mathcal{N}_x|}$. However, even then the dynamic range is too high, considering $2^{49} \approx 10^{15}$, leading to locations that become negligible in the entropy calculation.

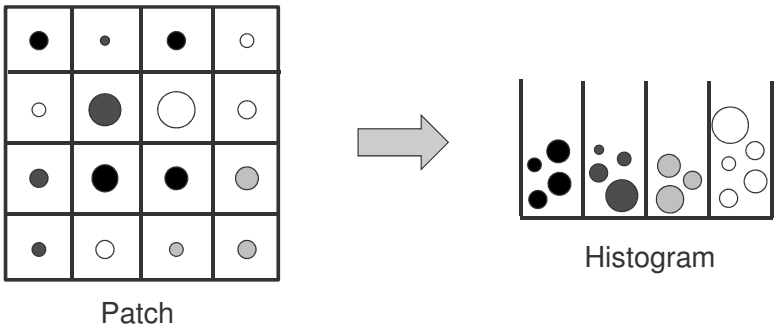



Figure 6.9: The spatially varying weight for the histogram calculation is illustrated by the size of the balls. The discrimination of patches is improved, because two patches with the same intensity values, but at different locations, may lead to different histograms. The discrimination is dependent on the selected spatial weighting.

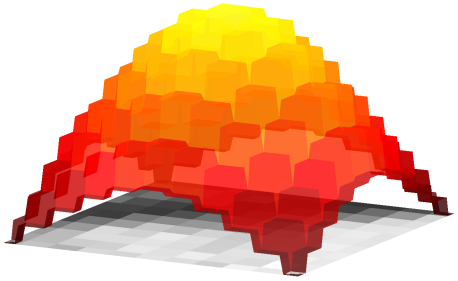


Figure 6.10: Image patch (gray) with the corresponding spatial weighting (color) used for the histogram calculation. Illustrated is a Gaußian weighting.

#### 6.5.2.3 Intensity Normalization

In order to use the whole range of the histogram, we normalize the intensity values of patches $P_x$. For this, we can either use the global minimum and maximum

$$\min_x = \inf_{y \in \Omega} I(y) \qquad \max_x = \sup_{y \in \Omega} I(y) \tag{6.13}$$

or the local extrema of the patch

$$\min_x = \inf_{y \in \mathcal{N}_x} I(y) \qquad \max_x = \sup_{y \in \mathcal{N}_x} I(y). \tag{6.14}$$

Our experiments confirm that a global approach is better suited for rigid, while the local approach is better suited for deformable registration. We show entropy images with the global approach in figures 6.4, 6.5, and 6.7 whereas the local approach is illustrated in figure 6.6.

#### 6.5.2.4 Entropy Measure

The Shannon entropy is one in a group of measures to calculate the entropy of a random variable, so that we evaluate its influence on registration. A whole class of entropy measures is deduced from the Rényi entropy [Rényi, 1961]

$$\mathrm{H}_\alpha(Y) = \frac{1}{1-\alpha} \log \left( \sum_{i \in \mathcal{I}} p(Y=i)^\alpha \right) \tag{6.15}$$

defined for $\alpha \geq 0$ and $\alpha \neq 1$. For $\alpha \to 1$, it converges to the Shannon entropy, $\mathrm{H}_1 = \mathrm{H}$. Commonly used in the group of Rényi entropies is the case for $\alpha = 2$. An alternative entropy measure, not in the group of Rényi measures is the Burg entropy [Mansoury and Pasha, 2008]

$$\mathrm{H}_{\mathrm{Burg}}(Y) = \sum_{i \in \mathcal{I}} \log p(Y=i). \tag{6.16}$$

## 6.6 Laplacian Images

As an alternative structural representation, we propose the application of manifold learning. Laplacian eigenmaps present an optimal solution to one of the requirements for a structural representation. This is the preservation of locality, meaning that patches that are close in high-dimensional patch space are mapped to a close structural representation. The requirement of structural equivalence across modalities is guaranteed by the comparable internal similarity.

Manifold learning is an approach applied for non-linear dimensionality reduction and data representation. The task of dimensionality reduction is to find the underlying structure in a large set of points embedded in a high-dimensional space and to map these points to a low-dimensional space preserving the structure. For manifold learning, being a specific type of non-linear dimensionality reduction, it is assumed that data points are sampled from a low-dimensional manifold living in a high-dimensional space. This implies that the data is really low dimensional and that the data points change smoothly and continuously. Manifold learning has recently gained much attention to assist image processing tasks such as segmentation [Zhang et al., 2006], registration [Rohde et al., 2008, Hamm et al., 2009], tracking [Lee and Elgammal, 2007], recognition [Arandjelovic and Cipolla, 2007], and computational anatomy [Gerber et al., 2009]. Common techniques for manifold learning are Isomap [Tenenbaum et al., 2000], local linear embedding [Roweis and Saul, 2000], and Laplacian eigenmaps [Belkin and Niyogi, 2003]. We focus on Laplacian eigenmaps because the technique is well founded on mathematical concepts and computationally efficient. We apply manifold learning also at later parts of this dissertation for breathing gating (cf. chapter 9) and patient position detection (cf. appendix A).

An overview of the calculation of Laplacian images is presented in figure 6.11. We use all the patches in an image to build a neighborhood graph, approximating the manifold embedded in high dimensional patch space. Subsequently, the graph

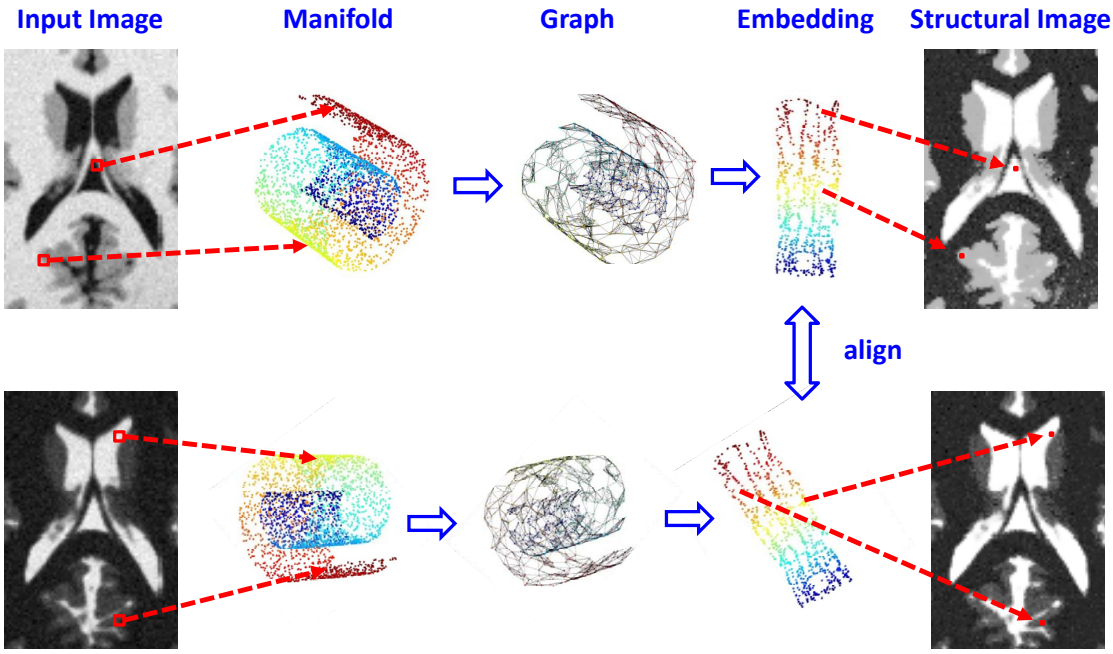



Figure 6.11: Structural representation with Laplacian eigenmaps. Patches of images lie on a manifold in high-dimensional patch space. The manifold is approximated by the neighborhood graph. The low-dimensional embedding is calculated with the graph Laplacian. Embeddings from different modalities have to be aligned to obtain the final representation.

Laplacian is calculated to find an optimal mapping to low-dimensional space. Since the embedding in low-dimensional space is arbitrary, as long as it preserves the locality, we have to align embeddings from different modalities with an affine point-based registration. This finally leads to the structural representation that is used in the intensity-based registration

### 6.6.1 Laplacian Eigenmaps

Laplacian eigenmaps build upon the construction of a neighborhood graph that approximates the manifold, on which the data points lie. Subsequently, the graph Laplacian is applied to calculate a low-dimensional representation of the data that preserves locality. Considering $k$ points $\mathbf{a}_1, \ldots, \mathbf{a}_k$ in $\mathbb{R}^N$ lying on a manifold $\mathcal{M}$, we want to find a set of corresponding points $\mathbf{b}_1, \ldots, \mathbf{b}_k$ in the low-dimensional space $\mathbb{R}^n$ ($n \ll N$). We define weights $\mathbf{W}$ between all pairs of input points, which reflect the locality of points. One possibility for the weights proposed in [Belkin and Niyogi, 2003] is derived from the heat kernel

$$\mathbf{W}_{i,j} = \exp(-||\mathbf{a}_i - \mathbf{a}_j||_2^2/t). \tag{6.17}$$

The parameter $t$ weights the influence of neighboring points. The optimization in Laplacian eigenmaps that tries to preserve the locality of points in low-dimensional

space is

$$\underset{\{\mathbf{b}_1,\ldots,\mathbf{b}_k\}}{\arg\min} \sum_{i,j} \mathbf{W}_{i,j} \cdot ||\mathbf{b}_j - \mathbf{b}_i||_2^2. \tag{6.18}$$

We see that points that are close in high-dimensional space should be arranged close in low-dimensional space to avoid a high cost caused by the high weight. Calculating the diagonal matrix $\mathbf{D}_{i,i} = \sum_j \mathbf{W}_{i,j}$, we can construct the Laplacian matrix $\mathbf{L}$ of the graph

$$\mathbf{L} = \mathbf{D} - \mathbf{W}. \tag{6.19}$$

With the relationship $\sum_{i,j} \mathbf{W}_{i,j} \cdot ||\mathbf{b}_j - \mathbf{b}_i||_2^2 = 2 \cdot \text{trace}(\mathbf{B}^\top \mathbf{D}\mathbf{B} - \mathbf{B}^\top \mathbf{W}\mathbf{B})$, the optimization can be formulated as trace minimization

$$\min_{\substack{\mathbf{B} \in \mathbb{R}^{k \times n} \\ \mathbf{B}^\top \mathbf{D}\mathbf{B} = \mathbf{I} \\ \mathbf{b^{(l)}}^\top \mathbf{D}\mathbf{1} = \mathbf{0}}} \text{trace}(\mathbf{B}^\top \mathbf{L}\mathbf{B}) \tag{6.20}$$

with $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k]^\top$ and further considering column vectors $\mathbf{b}^{(l)}$, enabling to write $\mathbf{B} = [\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(n)}]$. The constraints guarantee, among others, the orthogonality of the different dimensions, preventing a collapse onto a subspace of dimension less than $n$. Many dimensionality reduction techniques can be formulated as trace optimization problems [Kokiopoulou et al., 2011a]. The solution of the optimization problem is given by the first $n$ eigenvectors corresponding to the lowest, non-zero eigenvalues of the generalized eigenvalue problem $(\mathbf{D} - \mathbf{W})\mathbf{v} = \lambda \mathbf{D}\mathbf{v}$.

In [Belkin and Niyogi, 2003], a further derivation is presented for the Laplace Beltrami operator, which is the analogon of the Laplacian of a graph on manifolds. We assume a twice differentiable function $m : \mathcal{M} \to \mathbb{R}^n, \mathbf{a}_i \mapsto \mathbf{b}_i$. It is shown that the following holds for the relationship between the distances on the manifold and embedding space

$$||m(\mathbf{a}_i) - m(\mathbf{a}_j)|| \leq d_{\mathcal{M}} \cdot ||\nabla m(\mathbf{a}_i)|| + o(d_{\mathcal{M}}), \tag{6.21}$$

with $\mathbf{a}_i, \mathbf{a}_j \in \mathcal{M}$ and $d_{\mathcal{M}} = \text{dist}_{\mathcal{M}}(\mathbf{a}_i, \mathbf{a}_j)$ the manifold distance. We see that $||\nabla m(\mathbf{a}_i)||$ indicates how close nearby points are mapped. Consequently, a map that best preserves locality on average is found with the following minimization

$$\underset{||m||_{L^2(\mathcal{M})}=1}{\arg\min} \int_{\mathcal{M}} ||\nabla m(\mathbf{a}_i)||^2 \, \mathrm{d}\mathbf{a}_i. \tag{6.22}$$

It is this optimization problem, for which Laplacian eigenmaps provide an optimal solution, by calculating eigenfunctions of the Laplace Beltrami operator.

### 6.6.2 Verification of Structural Properties

The optimal locality preservation as is provided by the Laplacian eigenmaps, see equation (6.22), is exactly what was required for the structural representation in (R1). We only have to identify the points $\mathbf{a}_i, \mathbf{a}_j$ with the patches $P_x, P_y$. The mapping

$m$ is therefore a suitable candidate for the function that provides the structural representation $f$.

For the second property (R2), we consider manifolds $\mathcal{M}$ and $\mathcal{M}'$ for two different modalities with patches $P_x \in \mathcal{M}$ and $Q_x \in \mathcal{M}'$. The mappings $m$ and $m'$ provide $m : P_x \mapsto p_x$ and $m' : Q_x \mapsto q_x$, with $p_x$ and $q_x$ being the low-dimensional representations. Since the intensity, with which objects are depicted in the images, varies with the modality, the two manifolds $\mathcal{M}$ and $\mathcal{M}'$ are not directly comparable. Considering, however, the assumption that the internal similarity of both modalities is equivalent, as in [Penney et al., 2008], we conclude that the structure or shape of both manifolds is similar. Since Laplacian eigenmaps preserve locality when embedding the manifold in a low-dimensional space, this structure is preserved in low dimensions. We could then directly use the coordinates of $p_x$ as descriptor for the corresponding location $D_x$. This is, however, not possible because the embedding of the structure in low-dimensional space is arbitrary, as long as it preserves the locality. The embeddings of both manifolds $\mathcal{M}$ and $\mathcal{M}'$ are therefore only similar when correcting for rotation, translation, and scale. Consequently, an affine registration between the point sets $\mathcal{P} = \{p_x : x \in \Omega\}$ and $\mathcal{Q} = \{q_x : x \in \Omega\}$ has to be performed. Recently, an affine ICP for point set alignment was proposed in [Du et al., 2010] that performs a fast registration. The coordinates of the registered embeddings finally provide the structural descriptors.

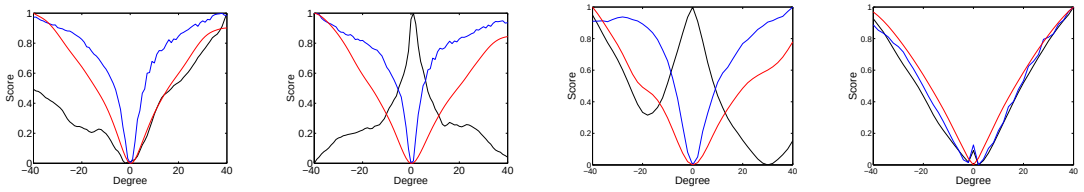
We conclude that $m$ fulfills the requirements (R1) and (R2) for a structural representation. It has therefore theoretical advantages in comparison to entropy images, since they only fulfill the relaxed properties, and further, the preservation of locality is optimal for $m$.

### 6.6.3 Application of Laplacian Eigenmaps

In our scenario, we consider one dimension of the ambient space for each image pixel of the patches. We use patches of size $15 \times 15$, so that $N = 225$. This size proved to be a good compromise between too small patches that do not contain enough structural information, and too large patches that contradict the required locality and further lead to a higher computational burden.

We construct a graph with a node for each point $P_x$ and with edges connecting neighboring nodes. The neighborhood can be defined with an $\delta$-neighborhood around each point, so $P_y$ is in the $\delta$-neighborhood of $P_x$ if $||P_x - P_y||_2^2 < \delta$. Although this is geometrically motivated, a disadvantage is the selection of the parameter $\delta$. In our implementation, we search instead for the $l = 500$ nearest neighbors and add edges between them in the adjacency graph. Further, heat kernel-based weights (cf. equation (6.17)) are assigned to the edges with $\mathbf{W}_{xy} = e^{-||P_x - P_y||_2^2/t}$.

We select $n = 1$ as dimension for the low-dimensional space. The reasons for not increasing this value are, first, that we obtain good results, and second, that for $n > 1$ we would have to store a vector in each pixel position instead of a scalar. This increases the computational complexity and memory consumption of the registration, and moreover, makes the visualization more challenging. Additionally,




(a) T1-T2 (b) T1-PD (c) CT-T2 (d) PET-T2

Figure 6.12: Similarity plots for RIRE dataset (black: SSD, blue: MI, red: eSSD).

the low dimensionality facilitates the affine alignment to a correction of scale and shift. In fact, a robust normalization accounting for outliers and flipping, makes the point-based registration in this case even superfluous. This holds the advantage of making the Laplacian images more comparable to entropy images, because they, as well, do not need any intermediary step.

Although the Laplacian images have superior theoretical properties than entropy images, this comes at a much higher computational cost. For the construction of the neighborhood graph, all patches have to be compared among each other. This means that the computational complexity increases quadratically with the number of patches. Thinking about its application for the alignment of volumetric data, this becomes even more challenging for the comparison of all subvolumes. This is different for entropy images, where the complexity increases only linearly. The Laplacian images present a global approach to the creation of structural images, while the entropy images work entirely locally.

## 6.7 Experiments with Entropy Images

For entropy images, we conduct experiments on T1, T2, and PD-weighted MR images from the BrainWeb database[1] and CT, T1, T2, PD, and PET (Positron Emission Tomography) images from the Retrospective Image Registration Evaluation (RIRE) database[2]. We work with BrainWeb images containing 3% noise and 20% intensity non-uniformity, in order to achieve realistic results. For both databases the ground truth alignment is provided. We depict axial slices of the original and entropy images in figures 6.4, 6.5, 6.6, and 6.7. The average time for the creation of an entropy image in C++ is $0.078s$ for a slice of $256 \times 256$, and $11.01s$ for a volume of $512 \times 512 \times 29$.

For our deformable experiments we use the freely available deformable registration software DROP[3], which applies free-form deformations based on cubic B-splines and an efficient linear programming with a primal-dual scheme. To emphasize that we consider deformable registration as an available tool, we perform all experiments with the standard settings of DROP, without any optimization towards the application of entropy images. This guarantees that similar results are obtainable with alternative

[1] http://www.bic.mni.mcgill.ca/brainweb/
[2] http://www.insight-journal.org/rire/
[3] http://www.mrf-registration.net

Table 6.1: Pairwise random registration study for BrainWeb with translational error $t$ in mm, rotational error $r$ in degree.

| Data Set | Sim | $t_x$ | $t_y$ | $t_z$ | $r_x$ | $r_y$ | $r_z$ | $t_{\text{std}}$ | $r_{\text{std}}$ | $t_{\text{total}}$ | $r_{\text{total}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T1-T2 | MI | 0.189 | 0.247 | 0.162 | 0.430 | 0.331 | 0.122 | 1.570 | 2.849 | **0.376** | 0.579 |
| T1-T2 | eSSD | 0.165 | 0.209 | 0.622 | 0.051 | 0.056 | 0.282 | 0.077 | 0.170 | 0.695 | **0.317** |
| T1-PD | MI | 0.251 | 0.319 | 0.257 | 0.359 | 0.417 | 0.165 | 1.906 | 2.303 | **0.506** | 0.621 |
| T1-PD | eSSD | 0.155 | 0.119 | 0.538 | 0.047 | 0.064 | 0.334 | 0.077 | 0.152 | 0.5929 | **0.365** |
| T2-PD | MI | 0.171 | 0.112 | 0.143 | 0.192 | 0.156 | 0.119 | 0.792 | 0.861 | 0.274 | 0.290 |
| T2-PD | eSSD | 0.046 | 0.041 | 0.141 | 0.045 | 0.040 | 0.044 | 0.039 | 0.025 | **0.161** | **0.085** |

registration approaches.

### 6.7.1 Rigid Registration

For rigid registration, the standard configuration for the entropy images is: $11 \times 11$ patches (in 2D) and $9 \times 9 \times 9$ patches (in 3D), 64 bins, Gaußian weighting, global normalization, Parzen-window estimation, and Shannon entropy. We evaluate the usage of SSD, MI, and SSD on entropy images (eSSD) for the rigid alignment by analyzing the similarity plots for the various mutli-modal combinations of both datasets, cf. Fig. 6.12. The plots are created by rotating the images around the image center for the range of [-40°, 40°], with 0° corresponding to the correct alignment. Generally, we observe that SSD fails, which was to be expected, whereas MI and eSSD indicate the correct alignment. In most cases eSSD provides smoother curves with a wider basin than MI, which is advantageous for registration, because it enables a larger capture range. Remarkable is also the problem of MI for the registration of PET images.

We further run a pairwise registration study for the various combinations of the multi-modal volumes, with an initial random deviation of maximal ±20 mm in translation and ±20° in rotation from the correct pose, to compare the performance of eSSD and MI. For each configuration 100 registrations are performed using the best neighbor optimizer. In tables 6.1 and 6.2, we show the absolute mean error for each pose parameter and the total mean error for translation $t_{\text{total}}$ and rotation $r_{\text{total}}$ for the two datasets. Additionally, we indicate the standard deviation of the error for translation $t_{\text{std}}$ and rotation $r_{\text{std}}$. Lowest errors are indicated in bold face in the tables. On the BrainWeb dataset eSSD and MI lead to comparable results. The large values of the standard deviation for MI indicate that some of the registrations did not converge correctly. Compared to the low standard deviations for eSSD, we conclude that the registration is more robust with eSSD. On the MR volumes of the RIRE dataset, eSSD performs significantly better than MI, matching with our observations from the similarity plots. For the alignment with CT volumes the registration error increases for both eSSD and MI. The experiments for the PET registration are performed on volumes with a lower resolution to match the resolution of the PET volume ($128 \times 128 \times 29$). The registration with eSSD achieves excellent results, in contrast to MI. The good registration results for eSSD on the RIRE

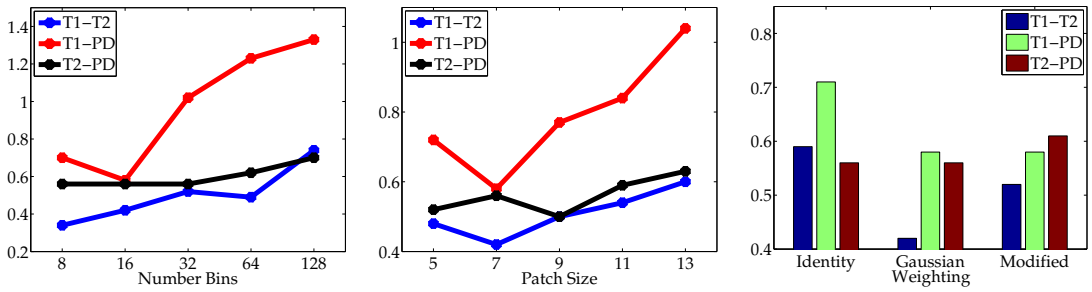



Figure 6.13: Evaluation of error $\tau$ for deformable registration for varying bin number, patch size, and spatial weighting on Brainweb.

Table 6.2: Pairwise random registration study for RIRE with translational error $t$ in mm, rotational error $r$ in degree.

| Data Set | Sim | $t_x$ | $t_y$ | $t_z$ | $r_x$ | $r_y$ | $r_z$ | $t_{\text{std}}$ | $r_{\text{std}}$ | $t_{\text{total}}$ | $r_{\text{total}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T1-T2 | MI | 0.719 | 0.395 | 1.531 | 1.594 | 2.252 | 1.467 | 0.235 | 0.177 | 1.754 | 3.139 |
| T1-T2 | eSSD | 0.042 | 0.224 | 0.396 | 1.120 | 1.582 | 0.538 | 0.046 | 0.057 | **0.461** | **2.013** |
| T1-PD | MI | 0.190 | 0.251 | 0.856 | 0.635 | 0.877 | 0.546 | 0.290 | 0.180 | 0.944 | 1.231 |
| T1-PD | eSSD | 0.061 | 0.048 | 0.431 | 0.344 | 0.648 | 0.558 | 0.050 | 0.048 | **0.442** | **0.926** |
| T2-PD | MI | 0.196 | 0.344 | 2.068 | 2.365 | 1.196 | 0.952 | 0.718 | 2.186 | 2.141 | 2.976 |
| T2-PD | eSSD | 0.093 | 0.110 | 1.065 | 1.984 | 1.196 | 0.308 | 0.279 | 0.493 | **1.080** | **2.355** |
| CT-T1 | MI | 1.925 | 1.004 | 1.312 | 1.718 | 2.951 | 0.763 | 2.394 | 5.529 | 2.710 | 3.951 |
| CT-T1 | eSSD | 0.963 | 1.269 | 0.702 | 2.433 | 0.728 | 0.169 | 2.990 | 9.577 | **1.997** | **3.089** |
| CT-T2 | MI | 4.567 | 1.501 | 2.314 | 3.896 | 9.738 | 1.332 | 4.543 | 9.783 | **5.806** | **11.634** |
| CT-T2 | eSSD | 1.288 | 8.488 | 6.025 | 35.905 | 1.903 | 0.885 | 7.932 | 28.631 | 11.063 | 36.896 |
| CT-PD | MI | 2.348 | 0.613 | 1.394 | 1.395 | 4.056 | 0.805 | 2.677 | 5.622 | 2.943 | 4.784 |
| CT-PD | eSSD | 0.770 | 0.988 | 0.475 | 1.487 | 0.248 | 0.248 | 1.824 | 5.971 | **1.442** | **1.619** |
| PET-T1 | MI | 9.071 | 7.730 | 13.409 | 29.226 | 23.578 | 4.945 | 13.373 | 22.343 | 20.869 | 46.234 |
| PET-T1 | eSSD | 0.053 | 0.057 | 0.089 | 0.040 | 0.038 | 0.042 | 0.051 | 0.023 | **0.135** | **0.078** |
| PET-T2 | MI | 9.027 | 8.566 | 11.631 | 35.444 | 22.239 | 5.522 | 11.545 | 20.217 | 20.705 | 52.131 |
| PET-T2 | eSSD | 0.043 | 0.038 | 0.057 | 0.040 | 0.040 | 0.042 | 0.034 | 0.024 | **0.093** | **0.079** |
| PET-PD | MI | 10.131 | 7.685 | 13.354 | 27.693 | 24.451 | 4.565 | 12.853 | 20.023 | 21.165 | 45.606 |
| PET-PD | eSSD | 0.048 | 0.056 | 0.194 | 0.043 | 0.056 | 0.043 | 0.065 | 0.033 | **0.217** | **0.094** |

data are confirmed by the low standard deviations. Comparing the errors for the different pose parameters, we observe that the translational error along the z-axis $t_z$ is commonly larger than along the other axes. This is due to the lower resolution along the z-axis. The low resolution is also the reason for the lower rotational error around the z-axis $r_z$ in comparison to $r_x$ and $r_y$.

We measure an average computation time for the evaluation of SSD being a factor of 15 faster than MI. This leads to a quick amortization of the additional processing time needed for the creation of the entropy images.

### 6.7.2 Deformable Registration

We first evaluate the application of gradient images [Haber and Modersitzki, 2007] for deformable registration. In figure 6.14 we show two synthetic images that model

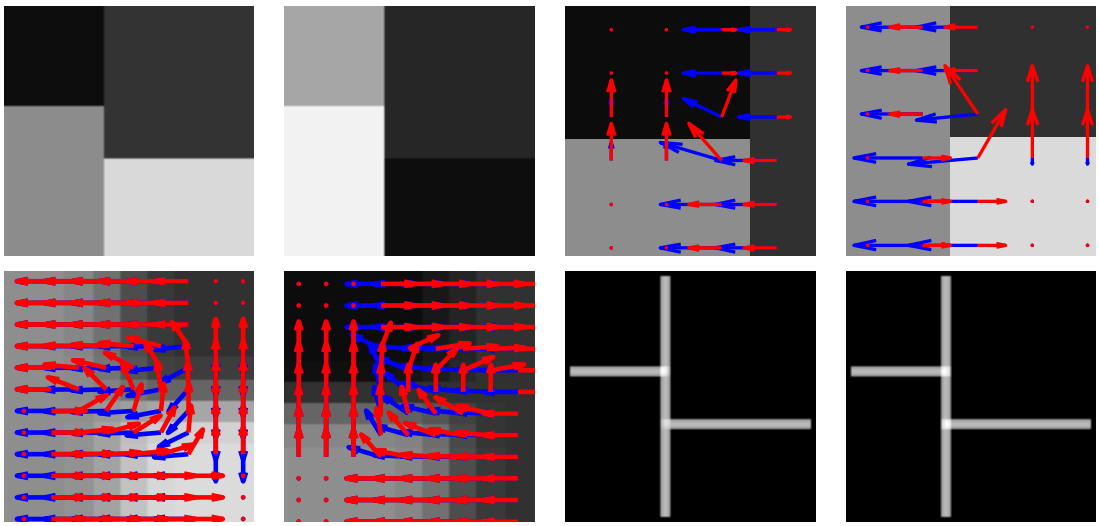


Figure 6.14: Illustration of two synthetic multi-modal images ($1^{st}$ & $2^{nd}$) together with a zoom on gradient fields ($3^{rd}$ & $4^{th}$), gradient fields of smoothed images ($5^{th}$ & $6^{th}$) and entropy images ($7^{th}$ & $8^{th}$). The best structural representation is achieved with entropy images.

a possible multi-modal scenario, with image regions being depicted in different colors. We calculate the gradient fields of both images and overlay them, once for the upper and once for the lower T-junction. Further, we smooth the synthetic images to create a more realistic scenario and show the gradient fields again. We observe that the gradients are pointing in different directions, where the smoothing propagates this effect to a larger region. In contrast, the entropy images consistently represent the structure of the images, leading to good registration results.

On the medical databases, we deform one of the two images with a deformation $d_g$ serving as ground truth. Next, we run the registration with the deformed image as target and the image from the other modality as source to calculate the estimated deformation $d_c$. We calculate the average Euclidean difference of the

Table 6.3: Registration errors $\tau$ in mm for various configurations for calculating the entropy images. (B: Brainweb, R: RIRE dataset)

| Technique | T1-T2$_B$ | T1-PD$_B$ | T2-PD$_B$ | T1-T2$_R$ | T1-PD$_R$ | T2-PD$_R$ | CT-T1$_R$ | PET-T2$_R$ |
|---|---|---|---|---|---|---|---|---|
| NMI (reference) | 0.63 | 0.79 | 0.66 | 0.94 | 1.04 | 1.33 | 1.84 | 3.42 |
| Local Norm. | 0.42 | 0.58 | 0.56 | 0.48 | 0.44 | 0.41 | 5.86 | 0.64 |
| Global Norm. | 0.99 | 2.08 | 0.87 | 2.76 | 4.30 | 4.11 | 6.87 | 1.83 |
| Parzen Window | 0.42 | 0.58 | 0.56 | 0.48 | 0.44 | 0.41 | 5.86 | 0.64 |
| Histogramming | 0.54 | 0.91 | 0.66 | 0.94 | 1.12 | 1.42 | 6.02 | 0.97 |
| Shannon | 0.42 | 0.58 | 0.56 | 0.48 | 0.44 | 0.41 | 5.86 | 0.64 |
| Rényi, $H_2$ | 0.47 | 1.11 | 0.64 | 0.49 | 0.54 | 0.64 | 6.39 | 1.25 |
| Burg | 1.82 | 4.61 | 2.43 | 2.81 | 2.37 | 2.68 | 6.71 | 3.46 |

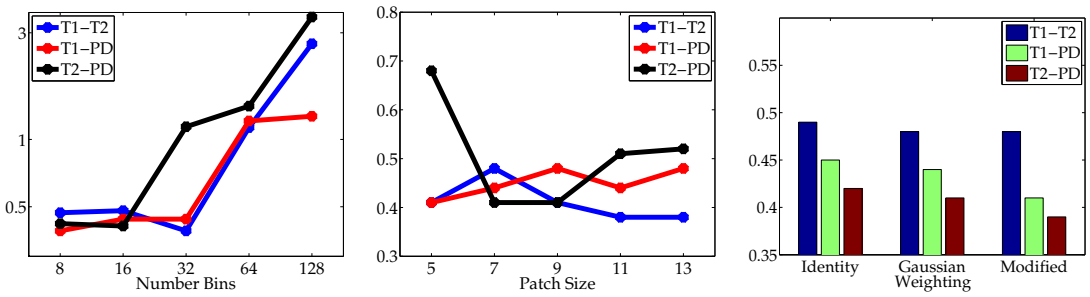



Figure 6.15: Evaluation of error $\tau$ for deformable registration for varying bin number, patch size, and spatial weighting on RIRE.

deformation fields $\tau = \frac{1}{|\Omega|} \sum_{x \in \Omega} \|d_c(x) - d_g(x)\|$ for quantifying the residual error of the registration.

The results for the experiments are shown in figures 6.13 and 6.15, and table 6.3. The error of the registration with the original images using normalized mutual information is stated in the table as reference. The standard configuration for the entropy image for deformable registration is: $7 \times 7$ patches, 16 bins, Gaußian weighting, local normalization, Parzen-window estimation, and Shannon entropy. In our experiments, we evaluate each of the parameters by changing one of them from the standard configuration and letting the others constant.

From figures 6.13 and 6.15, we see that best results are achieved around 16 bins. While reducing it further to 8 bins also leads to good results, increasing it further to 32 bins leads to an increase in error. A good compromise in the patch size for the different datasets and modalities is approximately $7 \times 7$. Larger patch sizes still lead to good results on the RIRE images, but on the Brainweb images we observe a significant increase of the error. Smaller patches lead to an inaccurate density estimation because of the small number of samples. For the weighting, we observe a general reduction of the error when using a more advanced weighting than the standard identical one.

From table 6.3, we see that a local normalization of the intensity values and the Parzen window method for the density estimation are essential for good deformable registration results. For the entropy estimation, we obtain best results for the Shannon entropy. The results for the special Rényi entropy $H_2$ are comparable, while the ones for the Burg entropy are not good.

The results of the deformable registration on T1, T2, and PD images show a slight advantage for normalized mutual informaiton (NMI) on the Brainweb dataset, while for the RIRE dataset the entropy images lead to a significant improvement. The registration with CT and PET is more challenging, because of the significant differences in the images. The registration of CT with entropy images is inferior to NMI. For the registration of PET, entropy images are superior to NMI.

The registration with eSSD is on average 6.6 times faster than with NMI. This includes the time for the creation of the entropy images.

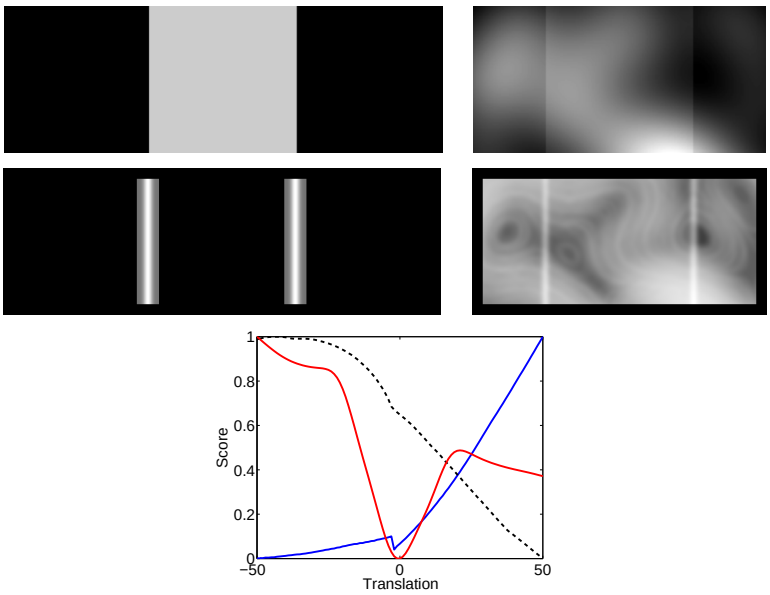



Figure 6.16: First line, synthetic images with bias field [Myronenko and Song, 2009b]. Second line, entropy images. Third line, similarity plot with SSD (blue), MI (black, dashed), and eSSD (red).

### 6.7.3 Bias field

A typical situation that challenges the application of mutual information is the registration of images, which contain high intensity non-uniformity due to the bias field. Recently, the residual complexity [Myronenko and Song, 2009b] was presented as a technique that can cope with such situations. The approach was motivated with synthetic images, similar to those in figure 6.16. We create plots of SSD and MI by translating the images, analogously to [Myronenko and Song, 2009b]. As expected, they fail in indicating the correct alignment. In contrast, eSSD correctly indicates the perfect alignment. There is an interesting parallel, because it was shown that the usage of local statistics for the MI calculation [Yi and Soatto, 2009] is more robust to intensity non-uniformity. For the calculation of entropy images we also use local statistics. However, we have the advantage of estimating only marginal instead of joint distributions, which is more robust for few samples from small local neighborhoods.

## 6.8 Experiments with Laplacian Images

We have already mentioned in section 6.6.3 that the computational complexity of the Laplacian images prohibits its practical usage on a large scale at the moment. We nevertheless want to validate if the theoretical advantages also lead to better registration results. We limit the analysis to 2D regions of the BrainWeb images (3% noise, 20% intensity non-uniformity) and RIRE images. In figure 6.17, we show the original images, the entropy images, and the Laplacian images. We can clearly

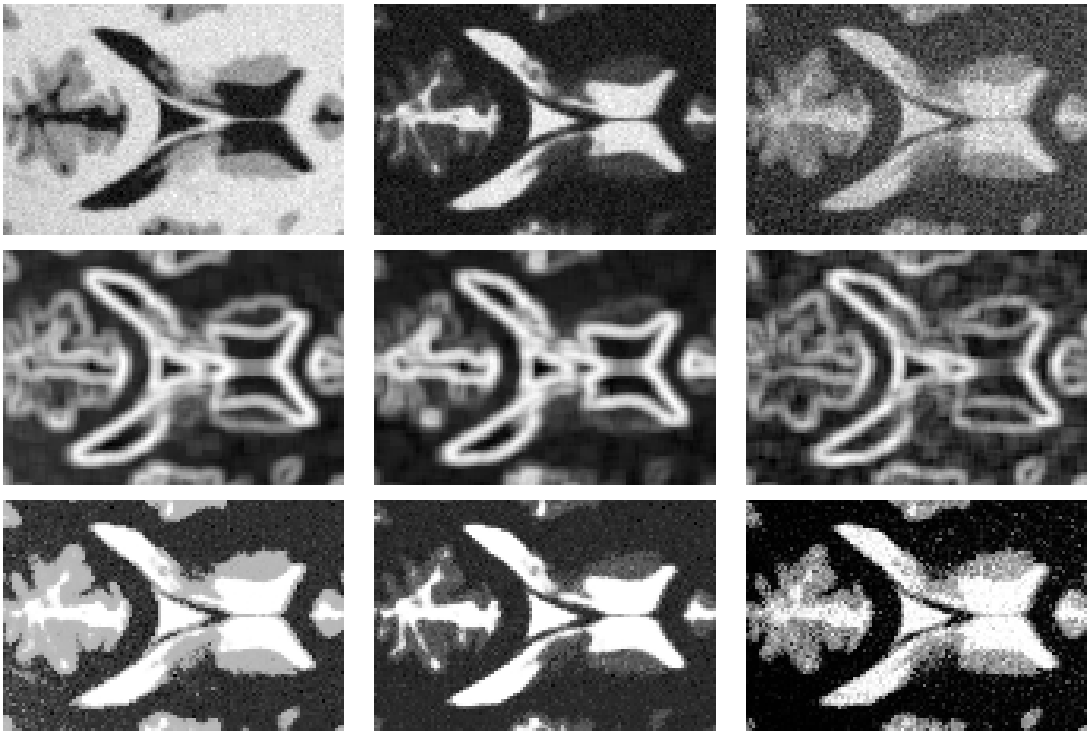


Figure 6.17: T1 (left), T2 (middle), and PD (right) images. First line: original images. Second line: entropy images. Third line: Laplacian images.

observe the different nature of the entropy and Laplacian images. Entropy images resemble gradient images, pronouncing boundaries and changes in the images. The Laplacian images, however, look like the original images, but with a different coloring. We can further observe that the assumption of comparable internal similarities in the images is justified, because the appearance of the Laplacian images across the modalities is very similar. In figure 6.19, we show the results for the RIRE images with T1, T2, and CT. The assumption of comparable internal similarity between MR and CT brain is more challenging, because of the low contrast of internal brain structures in CT.

In order to quantify the promising visual appearance for image registration, we show surface plots of the similarity measures for rotation and translation in figures 6.18 and 6.20, respectively. We compare the usage of the L2 distance on the original images, MI on the original images, and L2 distance on entropy and Laplacian images for all combinations of multi-modal alignment. The maxima indicate the best alignment. MI shows a very sharp peak at the correct position, but seems to have a limited capture range. Entropy images also indicate the correct position, but the cost functions contain several local maxima. We observe the cost function with the largest capture range for the Laplacian images.

Although similarity plots give a good intuition about the performance of different similarity measures, it is only a registration study that indicates the final quality. We

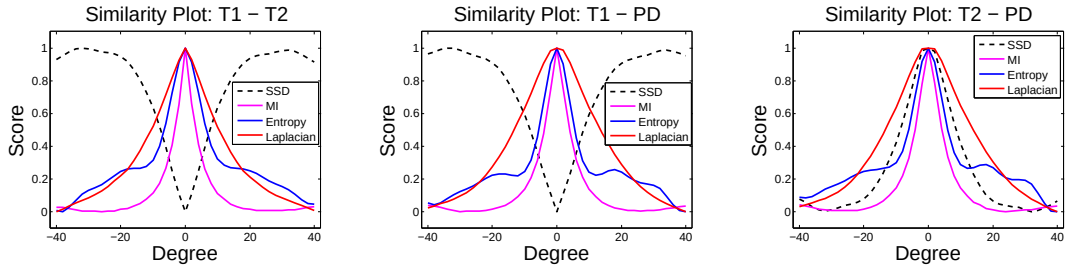



Figure 6.18: Plot of similarity measures with respect to rotation of images in figure 6.17. Maxima indicate best alignment.

Table 6.4: Registration study for Laplacian images. RMS errors for translation $t$ in mm and rotation $r$ in degree.

| Datasets | Similarity | $r$ | $t_x$ | $t_y$ | RMS |
|---|---|---|---|---|---|
| T1-T2 | L2 | 4.879 | 9.019 | 6.471 | 7.000 |
| | MI | 2.325 | 3.768 | 5.226 | 3.954 |
| | Entropy | 2.084 | 4.539 | 5.231 | 4.180 |
| | Laplacian | 2.584 | 2.061 | 2.168 | **2.271** |
| T1-PD | L2 | 2.760 | 6.422 | 5.755 | 5.227 |
| | MI | 2.304 | 4.138 | 4.907 | 3.937 |
| | Entropy | 2.283 | 4.782 | 4.750 | 4.108 |
| | Laplacian | 1.750 | 3.007 | 1.929 | **2.297** |
| T2-PD | L2 | 1.784 | 2.947 | 2.916 | 3.942 |
| | MI | 2.161 | 4.628 | 3.812 | 3.680 |
| | Entropy | 1.723 | 4.296 | 3.780 | 3.450 |
| | Laplacian | 1.171 | 2.350 | 1.984 | **1.900** |

perform a registration study for all multi-modal image combinations. The random starting position deviates up to ±15 mm in translation and ±15° in rotation from the correct pose. We show the average absolute error for translation and rotation, together with the overall root mean squared error (RMS), for 100 registration runs for each configuration in table 6.4. We weight 1 mm equal to 1° to quantify translational and angular displacement from the ground truth in one single value. We see that the positive impression of the Laplacian images from the similarity plots is confirmed by the registration results. We obtain a significantly lower error in comparison to MI and entropy images. Our experiments therefore confirm the theoretical advantages of Laplacian images in comparison to entropy images in practice.

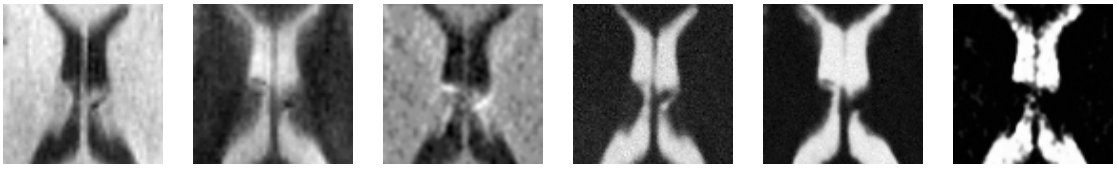


Figure 6.19: Left to right: T1, T2, and CT images. Next, corresponding Laplacian images in same order.

## 6.9 Discussion

We extensively evaluated the performance of entropy images for rigid, deformable, and groupwise registration. The results on the BrainWeb data are comparable to the application of MI. The registration results on the RIRE database are dependent on the modalities. We obtain very good results for the alignment of PET images, and in most other cases we are comparable or slightly better than MI. Finally, we indicated that entropy images are robust to intensity non-uniformity.

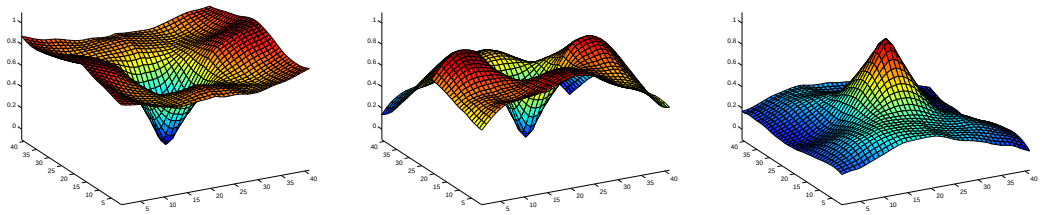
For the Laplacian images, we were able to validate that their superior theoretical properties also lead to better registration results. The computational complexity for the creation of these representations is, however, limiting. We think that the juxtaposition of entropy and Laplacian images is interesting because they present two completely different approaches for obtaining structural representations. While entropy images could be seen as non-linear filtering of an image, the Laplacian images identify self-similarities in an image. Entropy images present a local approach, and Laplacian images a global approach to the calculation of structural representations.

Thinking along the line of entropy images, one could imagine several alternative techniques to quantify the structure of a patch. We experimented, for instance, with the compressibility of patches or the calculation of the entropy from gradient instead of intensity histograms. These approaches, however, bear no obvious advantages for our applications and are more complex to compute. We are convinced that the simplicity of the entropy images together with the fast calculation and good performance leads to a high practical value. Further, there are many articles in the computer vision literature [Sizintsev et al., 2008, Wei and Tao, 2010] that deal with a fast local histogram calculation, such as integral histograms, distributive histograms, and the efficient histogram-based sliding window, that allow for a reduction of the processing time for certain configurations of entropy images.

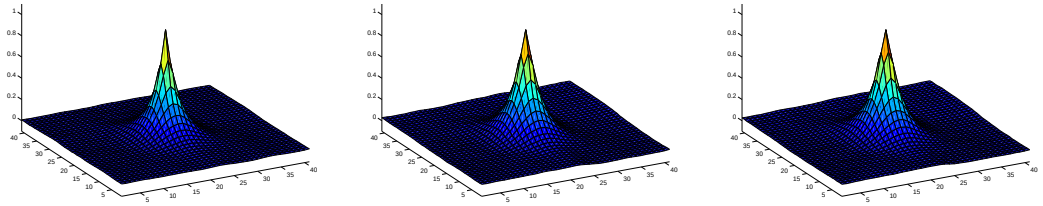
## 6.10 Conclusion

In this chapter, we analyzed structural image representations for multi-modal registration. We formulated theoretical properties that such a structural representation should fulfill. Generally, the multi-modal registration with such representations has the advantage that mono-modal metrics are applicable, as well as, more efficient

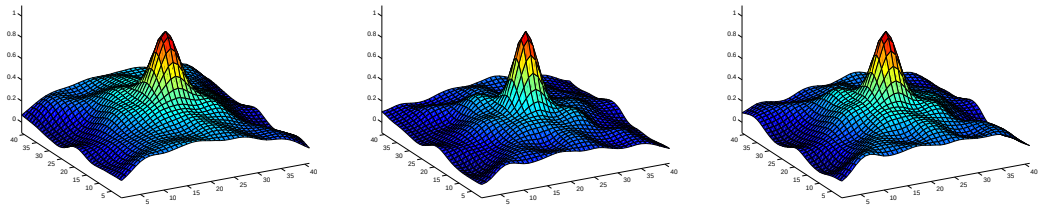
optimization schemes. Throughout the chapter we introduced two possible structural representations, the entropy and Laplacian images. The entropy images only fulfill certain requirements of a relaxed version of the theoretical properties, however, they are fast to compute and lead to good alignments, making them a very practical solution. The Laplacian images fulfill all the theoretical requirements, the preservation of locality even optimally. These superior theoretical properties also lead to better registration results; however, these advantages come at a significantly higher computational cost. Finally, we think that the application of structural representations for multi-modal registration is an interesting research direction, which presents an alternative to the usage of sophisticated similarity metrics.



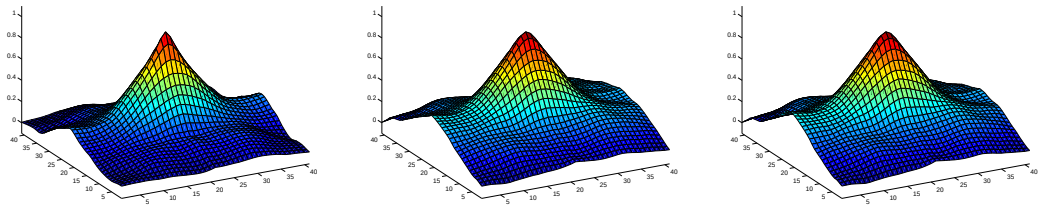
(a) L2 Original, T1-T2 (b) L2 Original, T1-PD (c) L2 Original, T2-PD



(d) MI Original, T1-T2 (e) MI Original, T1-PD (f) MI Original, T2-PD



(g) L2 Entropy, T1-T2 (h) L2 Entropy, T1-PD (i) L2 Entropy, T2-PD



(j) L2 Laplacian, T1-T2 (k) L2 Laplacian, T1-PD (l) L2 Laplacian, T2-PD

Figure 6.20: Plot of similarity measures with respect to translation in x and y direction. Maxima indicate best alignment.

# Part III

# ULTRASOUND MOSAICING

This part describes the first application: ultrasound mosaicing. We propose to apply simultaneous registration for the alignment of a group of ultrasound images. We further adapt the registration to ultrasound by introducing specific similarity measures. The first contribution considers locally adaptive Nakagami-based metrics and the second one separates reflectivity and scattering regions.

# Chapter 7

# Simultaneous Registration

*We address the alignment of a group of images for ultrasound mosaicing with simultaneous registration. We introduce a new class of multivariate similarity measures, referred to as* accumulated pair-wise estimates *(APE), and derive efficient optimization methods for it. A strict mathematical deduction of APE is presented and a connection to the congealing framework is established. Moreover, we address the increased computational complexity of simultaneous registration by deriving efficient gradient-based optimization strategies for APE: Gauß-Newton and the efficient second-order minimization (ESM). The application of structural images is very interesting in this context, because they enable the usage of ESM in multi-modal registration. Finally, we evaluate the performance of the optimization strategies with respect to the similarity measures, leading to very good results for ESM.*

## 7.1 Introduction

In chapter 2, we stated typical disadvantages of ultrasound imaging, such as the limited field of view and the viewing angle dependent artifacts. We also mentioned that the combination of several images helps to alleviate such problems. Several fusion approaches and their clinical importance were discussed in chapter 4, together with the description of the acoustic impedance estimation. Essential for the combination is, however, the identification of the correct alignment of the images. While this is difficult for 2D data, because all images should lie on the same plane, the prevalence of 3D imaging in ultrasound leads to a wide applicability of mosaicing.

The rigid intensity-based registration that we use for the alignment is not trivial to compute because of the limited amount of overlap between the images [Wachinger et al., 2007]. This limited overlap puts a special interest on the overlap invariance of similarity measures. An additional difficulty lies in the interface enhancing nature of ultrasound images, making acquisitions of the same object but from varying viewing angles not necessarily look the same. The standard approach in mosaicing is to find the global alignment of multiple images from a sequence of pairwise ones [Gee et al., 2003, Poon and Rohling, 2005, Ni et al., 2008]. Drawbacks of such approaches are that registration errors are accumulated and that only a fraction of the

available information is taken into account. In our previous work [Wachinger et al., 2007, Wachinger, 2007], we evaluated several registration strategies for mosaicing. Next to the sequential pairwise approach, we also investigated the combination of multiple pairwise registrations with a Lie group based normalization and the application of simultaneous registration. The conclusion of this comparison was that the best alignment is achieved with simultaneous registration. The reasons are that the consideration of all images alleviates the overlap issue and that the optimization on global transformations impedes the accumulation of errors.

### 7.1.1 Simultaneous Registration

Simultaneous registration is generally applied for aligning a group or population of images to a canonical pose. Examples are the alignment of handwritten digits or face images for their later identification [Learned-Miller, 2006, Huang et al., 2007, Cox et al., 2008, Cox et al., 2009] and the alignment of 3D tomographic images for the creation of an atlas [Zöllei et al., 2005]. First approaches to this groupwise registration problem identified one image as template, and registered all other images to it with a pair-wise approach. While this is a valid strategy for certain applications where such a template exists, in most cases it leads to an undesired introduction of bias with respect to the *a priori* chosen template. *Simultaneous registration* presents a method to circumvent this problem, however, it necessitates *multivariate similarity measures* and an optimization in a higher-dimensional space.

The direct estimation of multivariate measures with high-order joint density functions is prohibitive, because for a reliable estimation of the joint density, the number of samples would have to grow exponentially with the number of images; however, it only grows linearly. Approximations are therefore necessary, like the *congealing* framework presented by [Learned-Miller, 2006]. Another approach was presented in [Wachinger et al., 2007], which *accumulates pair-wise estimates* (APE). The derivation of APE was mainly based on analogies. Moreover, the relationship between congealing and APE has not yet been investigated.

When aligning multiple data sets simultaneously, instead of successively, one has to consider two consequences for the optimization method. First, the registration scenario becomes more complex because the parameter space increases linearly with the number of images. And second, the evaluation of the multivariate similarity measure is more expensive. One is therefore interested in an efficient optimization procedure, which finds the optima robustly and with a minimal amount of evaluations of the objective function. We focus on gradient-based methods because they promise a fast convergence rate due to the guidance of the process by the gradient.

In this chapter, we address the previously mentioned problems of simultaneous registration. First, we present a strict mathematical deduction of APE from a maximum likelihood framework. Second, we describe an extended version of congealing, enriched with neighborhood information, which allows us to show the connection between APE and congealing. Third, we derive efficient gradient-based optimization strategies for simultaneous registration with APE as multivariate similarity frame-

work. And finally, we use structural representations (cf. chapter 6) to apply ESM to multi-modal registration.

### 7.1.2 Related Work

Simultaneous registration has many applications in computer vision, pattern recognition, and medical imaging when it comes to the alignment of multiple images. [Learned-Miller, 2006] propose the congealing framework for the alignment of a large number of binary images from a database of handwritten digits and for the removal of unwanted bias fields in magnetic resonance images. Congealing sums up the entropy of pixel stacks over the image. [Huang et al., 2007] apply congealing to align 2D face images, essential for their later identification. In [Learned-Miller, 2006, Huang et al., 2007, Cox et al., 2008, Cox et al., 2009], a sequential update of the registration parameters is performed. [Zöllei et al., 2005] use congealing for the simultaneous alignment of a population of brain images for brain atlas construction. [Studholme and Cardenas, 2004] construct a joint density function for multivariate similarity estimation, which has the afore mentioned problem for larger image sets. [Cootes et al., 2004] use the minimum description length for the alignment of a group of images in order to create statistical shape models. This criterion demands a great deal of memory so that it only works for a limited number of volumes [Zöllei et al., 2005]. In [Sidorov et al., 2009] a stochastic optimization approach for groupwise registration of face images is proposed. One image is selected at a time and aligned to the remaining images using a similarity term that is close the voxel-wise variances, see section 7.2.2.1. It is argued that by randomly selecting the image to update the warp, an approximation to a fully simultaneous registration is achieved. [Wang et al., 2009b] propose an attribute vector-guided groupwise registration algorithm and show that including more attributes can improve the robustness of the algorithm. [Wachinger et al., 2007] propose simultaneous registration for volumetric mosaicing. This poses slightly different requirements on the multivariate similarity measure, because the number of overlapping images varies and can be rather small on specific locations. The therein introduced APE is flexible enough to deal with such situations. APE is a general framework to extend pairwise to multivariate similarity measures. The specific case of APE with sum of squared differences was used by [Cox et al., 2008, Cox et al., 2009], referring to it as least squares congealing.

A good overview of gradient-based optimization methods is provided in [Baker and Matthews, 2004] and [Madsen et al., 2004]. Based on their results, we do not consider the Levenberg-Marquardt algorithm because of its very similar behavior to Gauß-Newton. A new method, which is not covered in these articles, comes from the field of vision-based control. It is an efficient-second order optimization method introduced by [Benhimane and Malis, 2004]. They showed that ESM has striking advantages in convergence rate and convergence frequency in comparison to Gauß-Newton (GN) and steepest-descent (SD). [Vercauteren et al., 2007] achieve good results for the pairwise alignment of 2D images with ESM.

Once the update is calculated, either an additive or a compositional scheme for

updating the current transformation can be applied. In several articles [Chefd'hotel et al., 2002, Baker and Matthews, 2004, Stefanescu et al., 2004, Vercauteren et al., 2009] the advantages of a compositional update are noted, which we consequently focus on in our work.

## 7.2 Multivariate Similarity Metrics

In this section, we present a deduction of APE from the maximum likelihood framework in chapter 5 and show its connection to congealing. Considering $n$ images $\mathcal{I} = \{I_1, \ldots, I_n\}$ and the transformation parameters $\mathbf{x}$, the ML estimation for registration is formulated as:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \log p(I_1, \ldots, I_n; \mathbf{x}). \tag{7.1}$$

with the joint density function $p$, and the estimated alignment $\hat{\mathbf{x}}$. This corresponds to the similarity term in equation (5.42) and is similar to the formulation in [Zöllei, 2006]. For notational ease, we will no longer consider $\mathbf{x}$ explicitly in the density function.

### 7.2.1 Accumulated Pair-Wise Estimates

APE approximates the joint likelihood function with pair-wise estimates:

$$\log p(I_1, \ldots, I_n) \approx \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j | I_i). \tag{7.2}$$

Assuming a Gaußian distribution of the density $p$, i.i.d. coordinate samples, and various intensity mappings between the images, popular similarity measures such as SSD, NCC, CR, MI can be derived from the log-likelihood term $\log p(I_j|I_i)$ as discussed in chapter 5. APE therefore presents a framework for a class of similarity measures.

In order to deduce APE, we first derive a pair-wise approximation with respect to image $I_n$ using the product rule and conditional independence:

$$p(I_1, \ldots, I_n) \overset{\text{Prod.Rule}}{=} p(I_1, \ldots, I_{n-1} | I_n) \cdot p(I_n) \tag{7.3}$$

$$\overset{\text{Cond.Indep.}}{=} \prod_{i=1}^{n-1} p(I_i | I_n) \cdot p(I_n). \tag{7.4}$$

Second, we take the $n$-th power of the joint density function and perform the derivation of equation (7.4) with respect to each of the images, leading to:

$$p(I_1, \ldots, I_n)^n = \prod_{i=1}^{n} p(I_i) \cdot \prod_{i=1}^{n} \prod_{j \neq i} p(I_j | I_i). \tag{7.5}$$

Third the logarithm is applied:

$$\log p(I_1, \ldots, I_n)^n = \sum_{i=1}^{n} \log p(I_i) + \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j | I_i)$$

leading to the desired approximation of the high dimensional density:

$$\log p(I_1, \ldots, I_n) = \frac{1}{n} \sum_{i=1}^{n} \log p(I_i) + \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j | I_i) \tag{7.6}$$

$$\approx \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j | I_i) \tag{7.7}$$

wherein we no longer consider the multiplicative constant $\frac{1}{n}$ and the prior term $\sum_{i=1}^{n} \log p(I_i)$. The prior may, however, be used in future applications to incorporate further knowledge about the registration problem. The presented deduction is not limited to similarity measures and presents a general approximation of higher order densities by pairwise ones.

### 7.2.2 Congealing

In the congealing framework [Learned-Miller, 2006], independent but *not* identical distributions of the coordinate samples $s_k \in \Omega$ on the grid $\Omega$ are assumed:

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} p^k(I_1(s_k), \ldots, I_n(s_k)). \tag{7.8}$$

Assuming further i.i.d. input images $I_i$ leads to:

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} p^k(I_i(s_k)). \tag{7.9}$$

The idea of congealing is schematically illustrated in figure 7.1. In the following, we derive a more general form of congealing that applies, instead of the assumption of independent images, the Markov property. This means that images are independent, if we know a certain local neighborhood of images around the current one. While the consideration of neighboring pixels, surrounding a sample $s_k$, was already discussed in [Learned-Miller, 2006], referred to as pixel cylinder, the consideration of neighboring images has not yet been proposed. So, instead of independent images, we assume that each image $I_i$ depends on a certain neighborhood $\mathcal{N}_i$ of images:

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} p^k(I_i(s_k) | I_{\mathcal{N}_i}(s_k)). \tag{7.10}$$

We refer to this approximation as *Markov-congealing*. The size of the neighborhood depends on the structure in the image stack. If there is no further information about the images, considering a total neighborhood seems reasonable. If there is, however, a certain order or evolution in the stack (camera parameters, motion, etc.), the neighborhood can be chosen appropriately to reflect this structure.

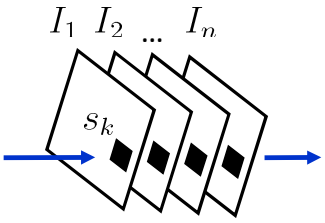



Figure 7.1: The idea of congealing is to focus the similarity estimation at a specific location $s_k$ along all the images.

#### 7.2.2.1 Voxel-wise Variances

The Markov-congealing allows us to derive the voxel-wise variances as proposed in [Wachinger et al., 2007] and applied in [Sidorov et al., 2009]. The term *voxel-wise* estimation [Zöllei et al., 2005] is used, since the approach taken in the congealing framework focuses on certain pixel or voxel locations at a time. Voxel-wise variances combine the approach of a voxel-wise similarity estimation with the assumptions underlying SSD, which are Gaußian distributed intensity values and the identity as intensity mapping.

We incorporate the neighborhood information by estimating the mean $\mu_k$ for each voxel location $s_k$ with:

$$\mu_k = \frac{1}{n} \sum_{l=1}^{n} I_l(s_k). \tag{7.11}$$

Following the formal definition of a local neighborhood $\mathcal{N}_i$ in the Markov sense, the calculation of the mean should not include the image $I_i$ itself [Li, 2009]. This leads, however, to higher computational costs because for each image and for each voxel location a different mean has to be calculated. We therefore go ahead with the more practical approximation, leading to:

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left( -\frac{(I_i(s_k) - \mu_k)^2}{2\sigma^2} \right). \tag{7.12}$$

This leads to the formula for voxel-wise SSD:

$$\log p(I_1, \ldots, I_n) \approx - \sum_{s_k \in \Omega} \sum_{i=1}^{n} (I_i(s_k) - \mu_k)^2 \tag{7.13}$$

$$= - \sum_{s_k \in \Omega} \mathrm{Var}[\mathcal{I}(s_k)] \tag{7.14}$$

with $\mathcal{I}(s_k)$ the intensities at location $s_k$ across all images. Looking at equation (7.13), we see that voxel-wise SSD leads to the calculation of the variance at each location and subsequently accumulates the values. The variance is one of the measures to

express the *statistical dispersion* of a random variable [Bishop, 2006]. In contrast to entropy, which measures the structuredness of a variable, it can only deal with mono-modal matchings. An interesting equality exists between voxel-wise SSD and APE SSD:

$$\sum_{s_k \in \Omega} \sum_{i=1}^{n} (I_i(s_k) - \mu_k)^2 \quad \stackrel{!}{=} \quad \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{s_k \in \Omega} (I_i(s_k) - I_j(s_k))^2. \tag{7.15}$$

We show the key steps of the proof of the equality, starting from the left-hand side in equation (7.15) and neglecting the summation over the spatial locations:

$$\sum_{i=1}^{n} (I_i(s_k) - \mu_k)^2 = \sum_{i=1}^{n} \left[ I_i^2(s_k) - 2 \cdot I_i(s_k) \cdot \left( \frac{1}{n} \sum_{j=1}^{n} I_j(s_k) \right) + \frac{1}{n^2} \left( \sum_{j=1}^{n} I_j(s_k) \right)^2 \right] \tag{7.16}$$

$$= \sum_{i=1}^{n} I_i^2(s_k) - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} I_i(s_k) I_j(s_k) + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} I_i(s_k) I_j(s_k). \tag{7.17}$$

The right-hand term deduces to:

$$\frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} (I_i(s_k) - I_j(s_k))^2 = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ I_i^2(s_k) + I_j^2(s_k) - 2 \cdot I_i(s_k) I_j(s_k) \right] \tag{7.18}$$

$$= \sum_{i=1}^{n} I_i^2(s_k) - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} I_i(s_k) I_j(s_k). \tag{7.19}$$

This equality between APE SSD and voxel-wise SSD is a reason and motivation to investigate the general relationship between APE and Congealing.

### 7.2.3 Comparison of APE and Congealing

In the last sections, we discussed APE and congealing as separate approximations to the high dimensional density and demonstrated the equality for SSD. In this section, we investigate if there is a direct theoretical relationship between these two approaches. It is in fact possible to deduce a connection between APE and Markov-congealing. We show the connection between the two approximations, by starting with the Markov-congealing, equation (7.10), and derive the formula of APE,

equation (7.6), from it:

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} p^k(I_i(s_k)|I_{\mathcal{N}_i}(s_k)) \tag{7.20}$$

$$\overset{\text{Bayes}}{=} \prod_{s_k \in \Omega} \prod_{i=1}^{n} p^k(I_{\mathcal{N}_i}(s_k)|I_i(s_k)) \frac{p^k(I_i(s_k))}{p^k(I_{\mathcal{N}_i}(s_k))} \tag{7.21}$$

$$\overset{\text{C.Idp.}}{=} \prod_{s_k \in \Omega} \prod_{i=1}^{n} \left[ \prod_{j \in \mathcal{N}_i} p^k(I_j(s_k)|I_i(s_k)) \right] \frac{p^k(I_i(s_k))}{p^k(I_{\mathcal{N}_i}(s_k))} \tag{7.22}$$

$$\overset{\text{Idp.}}{=} \prod_{s_k \in \Omega} \prod_{i=1}^{n} \left[ \prod_{j \in \mathcal{N}_i} p^k(I_j(s_k)|I_i(s_k)) \right] \frac{p^k(I_i(s_k))}{\prod_{j \in \mathcal{N}_i} p^k(I_j(s_k))} \tag{7.23}$$

Applying the logarithm and assuming a total neighborhood leads to:

$$\log p(I_1, \ldots, I_n) = \sum_{s_k \in \Omega} \sum_{i=1}^{n} \sum_{j \neq i} \left( \log p^k(I_j(s_k)|I_i(s_k)) - \log p^k(I_j(s_k)) \right) \tag{7.24}$$

$$+ \sum_{s_k \in \Omega} \sum_{i=1}^{n} \log p^k(I_i(s_k)). \tag{7.25}$$

An assumption that is different between the pair-wise and voxel-wise approaches, per design, is that the voxel-wise coordinate samples are not identically distributed. To relate the two approaches, we set the distribution of the coordinate samples identical:

$$\log p(I_1, \ldots, I_n) = \sum_{i=1}^{n} \sum_{j \neq i} (\log p(I_j|I_i) - \log p(I_j)) + \sum_{i=1}^{n} \log p(I_i) \tag{7.26}$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j|I_i) + \sum_{i=1}^{n} \log p(I_i) - \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j) \tag{7.27}$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j|I_i) + \sum_{i=1}^{n} \log p(I_i) - (n-1) \sum_{j=1}^{n} \log p(I_j) \tag{7.28}$$

Comparing this result to APE in equation (7.6), we observe that both are equivalent up to the term $-(n-1)\sum_{j=1}^{n} \log p(I_j)$. Again assuming that no prior information is available, we conclude that the approximations with APE and Markov-congealing are equal, under the consideration of (i) a total neighborhood, (ii) conditional independent images, and (iii) an identical distribution of coordinate samples. While (iii) was explicitly chosen by the design of congealing and (ii) by the deduction of APE, the novel part is the neighborhood (i), which relates these two approaches. The Markov-congealing in equation (7.10) presents therefore an intermediate between APE and congealing.

To conclude, for congealing no specific distribution has to be selected, because the similarity can directly be calculated with the sample entropy. Markov-congealing

and APE do not present actual similarity measures, but frameworks, where further information about the distribution has to be provided to derive similarity measures. Incorporating *e.g.* a Gaußian distribution and an identity intensity mapping leads to SSD like extensions. APE, in contrast to congealing, assumes an identical distribution of coordinate samples, which makes a reliable estimation for a small number of overlapping images possible. For congealing, a larger number is necessary, because the estimation is done with the information at one location at a time. Consequently, the choice, which multivariate similarity approximation to choose, is application dependent. We will focus on APE because of its versatility.

## 7.3 Efficient Optimization Methods

In this section, we present efficient gradient-based optimization methods for simultaneous registration. More precisely, we focus on APE as similarity measure and 3D rigid transformations as transformation model, where the parameterization can be easily adapted to different types of alignments. In contrast to [Learned-Miller, 2006, Huang et al., 2007, Cox et al., 2008, Cox et al., 2009], we do not update one parameter at a time, but update all parameters at once. Problems with the sequential update are illustrated in [Cox et al., 2009].

### 7.3.1 Transformation Parameterization

We parameterize the spatial transformations with Lie groups because 3D rigid transformations do not form a vector space. We perform a geometric optimization using local canonical coordinates. It has the advantage that the geometric structure of the group is taken care of intrinsically [Lee and Moore, 2005, Mahony and Manton, 2002]. This enables us to use an unconstrained optimization. Alternatively, one could embed them into the Euclidean space and perform a constrained optimization with Lagrange multipliers.

Each rigid 3D transformation $\mathbf{x}$ is an element of $\mathbb{SE}(3)$, the special Euclidean group. It is possible to describe them with a $4 \times 4$ matrix having the following structure:

$$\mathbf{x} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{7.29}$$

with the rotational part $\mathbf{R}$, element of the special orthogonal group $\mathbf{R} \in \mathbb{SO}(3)$, and the translational part $\mathbf{t} \in \mathbb{R}^3$.

$\mathbb{SE}(3)$ forms a manifold and is a group under standard matrix multiplication, therefore it is a Lie group. On Lie groups, the tangent space at the group identity defines a Lie algebra. The Lie algebra captures the local structure of the Lie group. The Lie algebra of $\mathbb{SE}(3)$ is denoted by $\mathfrak{se}(3)$, and is defined by:

$$\mathfrak{se}(3) = \left\{ \begin{bmatrix} \boldsymbol{\Upsilon} & \mathbf{v} \\ \mathbf{0} & 0 \end{bmatrix} | \boldsymbol{\Upsilon} \in \mathbb{R}^{3\times3}, \mathbf{v} \in \mathbb{R}^3, \boldsymbol{\Upsilon}^\top = -\boldsymbol{\Upsilon} \right\}.$$

The standard basis of $\mathfrak{se}(3)$ is $\mathcal{L} = \{\mathbf{l}_1, \ldots, \mathbf{l}_6\}$ with:

$$\mathbf{l}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{l}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{l}_3 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{7.30}$$

$$\mathbf{l}_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{l}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{l}_6 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{7.31}$$

Each element $\mathcal{H} \in \mathfrak{se}(3)$ can be expressed as a linear combination of matrices $\mathcal{H} = \sum_{i=1}^{6} h_i \mathbf{l}_i$ with $h_i$ varying over the manifold [Zefran et al., 1998] and $\mathbf{h} = [h_1, \ldots, h_6]^\top$. The exponential map relates the Lie algebra to the Lie group:

$$\exp : \mathfrak{se}(3) \to \mathbb{SE}(3) \tag{7.32}$$

$$\mathcal{H} \mapsto \exp\left(\sum_{i=1}^{6} h_i \mathbf{l}_i\right) = \sum_{j=0}^{\infty} \frac{1}{j!} \left(\sum_{i=1}^{6} h_i \mathbf{l}_i\right)^j .$$

It exists an open cube $V$ around $\mathbf{0}$ in $\mathfrak{se}(3)$ and an open neighborhood $U$ of the identity matrix $\mathbf{I} \in \mathbb{SE}(3)$ such that the group exponential is smooth and one-to-one onto, with a smooth inverse, therefore a diffeomorphism. An explicit expression for the calculation of the exponential for elements in $\mathbb{SE}(3)$ exists, as shown in [Murray et al., 1994, pp.413]. For the restriction to $\mathbb{SO}(3)$, the explicit formula is known as Rodrigues' formula.

Using the local coordinate charts, there exists for any $\mathbf{y} \in \mathbb{SE}(3)$ in some neighborhood of $\mathbf{x}$ a vector in the tangent space $\mathcal{H} \in \mathfrak{se}(3)$, such that:

$$\mathbf{y} = \mathbf{x} \circ \exp(\mathcal{H}) = \mathbf{x} \circ \exp\left(\sum_{i=1}^{6} h_i \mathbf{l}_i\right). \tag{7.33}$$

Let us further denote the transformation of a point $\mathbf{p} = [x, y, z, 1]^\top \in \mathbb{R}^4$ in homogeneous coordinates through the mapping $\mathbf{y} \in \mathbb{SE}(3)$ with $w(\mathbf{y}, \mathbf{p})$:

$$w : \mathbb{SE}(3) \times \mathbb{R}^4 \to \mathbb{R}^4 \tag{7.34}$$

$$(\mathbf{y}, \mathbf{p}) \mapsto w(\mathbf{y}, \mathbf{p}) = \mathbf{p}'. \tag{7.35}$$

Finally, for ease of notation we define an extension of the exponential, enabling the direct application of the parameter vector $\exp(\mathbf{h}) := \exp(\mathcal{H})$.

### 7.3.2 Optimization Methods

The global transformation $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, with $\mathbf{x}_i \in \mathbb{SE}(3)$, maps the points from each of the image spaces to the joint image space, $\mathbb{R}^4 \to \mathbb{R}^4, \mathbf{p} \mapsto w(\mathbf{x}_i, \mathbf{p})$. The cost function $E$ that we want to optimize is a sum of squared smooth functions:

$$E(\mathbf{x}) = \sum_{i \neq j} F_{i,j}(\mathbf{x}) = \sum_{i \neq j} \frac{1}{2} ||\mathbf{f}_{i,j}(\mathbf{x})||^2 \tag{7.36}$$

with $F_{i,j}$ representing the pair-wise similarity measure.

Regarding equation (7.36), we see that we deal with a non-linear least-squares problem. Therefore, efficient optimization methods were proposed that achieve in many cases linear, or even quadratic, convergence without the explicit calculation of the second derivatives.

The starting point of all the following optimization methods is a Taylor expansion of the cost function around the current transformation $\mathbf{x}$ along the gradient direction $\mathbf{h}$:

$$E(\mathbf{x} \circ \exp(\mathbf{h})) \approx E(\mathbf{x}) + \mathbf{J}_E(\mathbf{x}) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^\top \cdot \mathbf{H}_E(\mathbf{x}) \cdot \mathbf{h} \tag{7.37}$$

with $\mathbf{J}_E(\mathbf{x}) = \left.\frac{\partial E(\mathbf{x}\circ\exp(\mathbf{h}))}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{0}}$ and $\mathbf{H}_E(\mathbf{x}) = \left.\frac{\partial^2 E(\mathbf{x}\circ\exp(\mathbf{h}))}{\partial \mathbf{h}^2}\right|_{\mathbf{h}=\mathbf{0}}$ the Jacobian and Hessian, respectively, of $E$ at the point $\mathbf{x}$. The global gradient direction $\mathbf{h}$ is a combination of local elements $\mathbf{h}_i$, resulting in $\mathbf{h} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$. The Newton (NT) method then has the following compositional update:

$$\mathbf{H}_{F_{i,j}} \mathbf{h}_{i,j}^{\mathrm{NT}} = -\mathbf{J}_{F_{i,j}}^\top \qquad \mathbf{x} \leftarrow \mathbf{x} \circ \exp(\mathbf{h}^{\mathrm{NT}}). \tag{7.38}$$

The global update $\mathbf{h}^{\mathrm{NT}}$ is obtained by summing up the pairwise updates, following the structure of the cost function $E$ in equation (7.36), leading to

$$\mathbf{h}^{\mathrm{NT}} = \left[\sum_i \mathbf{h}_{i,1}^{\mathrm{NT}}, \ldots, \sum_i \mathbf{h}_{i,n}^{\mathrm{NT}}\right]. \tag{7.39}$$

Unfortunately, the explicit calculation of the Hessian causes problems because it is numerically not well-behaved and computationally expensive, so that its usage is not recommended [Baker and Matthews, 2004]. In the field of non-linear least squares optimization most of the methods use an approximation of the Hessian [Madsen et al., 2004]. In the following we present different possibilities for approximating the Hessian by a positive definite matrix $\hat{\mathbf{H}}$.

#### 7.3.2.1 Steepest-Descent

For SD, the Hessian is approximated by the identity $\hat{\mathbf{H}} = \alpha \cdot \mathbf{I}$, leading to the update:

$$\alpha \cdot \mathbf{h}^{\mathrm{SD}} = -\mathbf{J}_E^\top(\mathbf{x}) \qquad \mathbf{x} \leftarrow \mathbf{x} \circ \exp(\mathbf{h}^{\mathrm{SD}})$$

with $\alpha$ the step length. Consequently, SD only considers a first-order Taylor expansion of $E$ and has linear convergence.

#### 7.3.2.2 Gauß-Newton

The approximation of the Hessian for Gauß-Newton is based on a linear approximation of the components of $\mathbf{f}$ in a neighborhood of $\mathbf{x}$. For small $||\mathbf{h}||$ we obtain from the Taylor expansion:

$$\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h})) \approx \mathbf{f}(\mathbf{x}) + \mathbf{J}_{\mathbf{f}}(\mathbf{x}) \cdot \mathbf{h}. \tag{7.40}$$

For notational ease, we often write $\mathbf{f}$ instead of $\mathbf{f}_{i,j}$ when no reference to the images $i$ and $j$ is necessary. Setting this linear approximation in our cost function $E$, as defined in equation (7.36), gives:

$$E(\mathbf{x} \circ \exp(\mathbf{h})) \approx \sum_{i \neq j} \frac{1}{2} ||\mathbf{f}_{i,j}(\mathbf{x} \circ \exp(\mathbf{h}))||^2 \tag{7.41}$$

$$= \sum_{i \neq j} \frac{1}{2} \mathbf{f}_{i,j}(\mathbf{x} \circ \exp(\mathbf{h}))^\top \mathbf{f}_{i,j}(\mathbf{x} \circ \exp(\mathbf{h})) \tag{7.42}$$

$$= \sum_{i \neq j} \left( F_{i,j}(\mathbf{x}) + \mathbf{h}^\top \mathbf{J}_{\mathbf{f}_{i,j}}^\top \mathbf{f}_{i,j} + \frac{1}{2} \mathbf{h}^\top \mathbf{J}_{\mathbf{f}_{i,j}}^\top \mathbf{J}_{\mathbf{f}_{i,j}} \mathbf{h} \right). \tag{7.43}$$

By comparison with equation (7.37), and considering the gradient $\mathbf{J}_F = \mathbf{J}_\mathbf{f}^\top \mathbf{f}$, we can see that the Hessian is approximated by $\hat{\mathbf{H}} = \mathbf{J}_\mathbf{f}^\top \mathbf{J}_\mathbf{f}$.

We get the global Gauß-Newton step $\mathbf{h}^{\mathrm{GN}}$ by the pairwise optimal steps $\mathbf{h}_{i,j}^{\mathrm{GN}}$, analogously to the Newton method, see equation (7.39). This leads to the update:

$$(\mathbf{J}_{\mathbf{f}_{i,j}}^\top \mathbf{J}_{\mathbf{f}_{i,j}}) \mathbf{h}_{i,j}^{\mathrm{GN}} = -\mathbf{J}_{\mathbf{f}_{i,j}}^\top \mathbf{f}_{i,j} \qquad \mathbf{x} \leftarrow \mathbf{x} \circ \exp(\mathbf{h}^{\mathrm{GN}})$$

with $\mathbf{h}^{\mathrm{GN}} = \left[ \sum_i \mathbf{h}_{i,1}^{\mathrm{GN}}, \ldots, \sum_i \mathbf{h}_{i,n}^{\mathrm{GN}} \right]$. Gauß-Newton has only in specific cases quadratic convergence [Madsen et al., 2004, Benhimane and Malis, 2004].

#### 7.3.2.3 ESM

The efficient second-order minimization procedure originally comes from the field of vision-based control [Benhimane and Malis, 2004]. It is an extension of GN and incorporates further knowledge about the specificity of the optimization problem in order to achieve better results.

More precisely, ESM uses the fact, that if the images are aligned with the optimal transformation, the images and therefore also their gradients should be very similar to each other. This can be used to ameliorate the search direction of the Newton methods. For the standard Newton method, the first- and second-order derivatives around $\mathbf{x}$ are used to build a second-order approximation, see equation (7.38). The Gauß-Newton method neglects the second derivative and can therefore only build a first-order approximation. For ESM, the first-order derivatives around $\mathbf{x}$ and $\mathbf{x} \circ \exp(\mathbf{h})$ are used to build a second-order approximation without the necessity of an explicit calculation of a second-order derivative.

To deduce the ESM, we start with a second-order Taylor approximation of the function $\mathbf{f}$:

$$\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h})) \approx \mathbf{f}(\mathbf{x}) + \mathbf{J}_\mathbf{f}(\mathbf{x}) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^\top \cdot \mathbf{H}_\mathbf{f}(\mathbf{x}) \cdot \mathbf{h}. \tag{7.44}$$

Subsequently, we do a second Taylor expansion around $\mathbf{x}$, but this time of the Jacobian of $\mathbf{f}$:

$$\mathbf{J}_\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h})) \approx \mathbf{J}_\mathbf{f}(\mathbf{x}) + \mathbf{H}_\mathbf{f}(\mathbf{x}) \cdot \mathbf{h}. \tag{7.45}$$

Plugging this first-order series in the approximation shown in equation (7.44), we get a second-order approximation without second-order derivatives:

$$\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h})) \approx \mathbf{f}(\mathbf{x}) + \frac{1}{2}[\mathbf{J}_\mathbf{f}(\mathbf{x}) + \mathbf{J}_\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h}))] \cdot \mathbf{h}. \tag{7.46}$$

The problem about this approximation is the calculation of the Jacobian $\mathbf{J}_\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h}))$, which is dependent on the update $\mathbf{h}$ that we want to solve for. We illustrate a solution for this problem in section 7.3.3.3.

Comparing equations (7.46) and (7.40) shows the similarity between the Gauß-Newton and ESM procedure. For the development of the update rule we proceed therefore analogously to Gauß-Newton. The only difference is the usage of $\mathbf{J}_\mathbf{f}^{\text{ESM}} = \frac{1}{2}[\mathbf{J}_\mathbf{f}(\mathbf{x}) + \mathbf{J}_\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h}))]$ instead of only $\mathbf{J}_\mathbf{f}(\mathbf{x})$. This leads to an approximation of the Hessian by $\hat{\mathbf{H}} = \mathbf{J}_\mathbf{f}^{\text{ESM}^\top} \mathbf{J}_\mathbf{f}^{\text{ESM}}$. The compositional update is:

$$\left(\mathbf{J}_{\mathbf{f}_{i,j}}^{\text{ESM}^\top} \mathbf{J}_{\mathbf{f}_{i,j}}^{\text{ESM}}\right) \mathbf{h}_{\mathbf{f}_{i,j}}^{\text{ESM}} = -\mathbf{J}_{\mathbf{f}_{i,j}}^{\text{ESM}^\top} \mathbf{f}_{i,j} \qquad \mathbf{x} \leftarrow \mathbf{x} \circ \exp(\mathbf{h}^{\text{ESM}})$$

with $\mathbf{h}^{\text{ESM}} = \left[\sum_i \mathbf{h}_{i,1}^{\text{ESM}}, \ldots, \sum_i \mathbf{h}_{i,n}^{\text{ESM}}\right]$. ESM has a quadratic convergence rate [Benhimane and Malis, 2004].

## 7.3.3 Gradient Calculation

In the last section, we introduced the gradients $\mathbf{J}_E$, $\mathbf{J}_\mathbf{f}$, and $\mathbf{J}_\mathbf{f}^{\text{ESM}}$ without further explaining their calculation. This will be the subject of this part, together with an analysis on how the gradient calculation changes for different similarity measures.

### 7.3.3.1 Steepest-Descent

We begin with the gradient for the general cost function $E$ by considering one moving image at a time. W.l.o.g., we assume $I_i$ as fixed and $I_j$ as moving image leading to $F_{i,j}(\mathbf{x} \circ \exp(\mathbf{h})) = \text{SM}(I_i(\mathbf{x}), I_j(\mathbf{x} \circ \exp(\mathbf{h}))$, with SM a pair-wise similarity measure. The point-wise gradient has the form:

$$\begin{aligned}
[\mathbf{J}_E(\mathbf{x})]_\mathbf{p} &= \left[ \frac{\partial E(\mathbf{x} \circ \exp(\mathbf{h}))}{\partial \mathbf{h}} \bigg|_{\mathbf{h}=\mathbf{0}} \right]_\mathbf{p} \\
&= \sum_{i \neq j} \frac{\partial \text{SM}(I_i(\mathbf{x}), I_j(\mathbf{x} \circ \exp(\mathbf{h})); \mathbf{p})}{\partial \mathbf{h}} \bigg|_{\mathbf{h}=\mathbf{0}} \\
&= \sum_{i \neq j} [\mathbf{J}_{\text{SM}_{i,j}}(\mathbf{x})]_\mathbf{p} \cdot [\mathbf{J}_{I_j}(\mathbf{x})]_\mathbf{p} \cdot [\mathbf{J}_w(\mathbf{x})]_\mathbf{p} \cdot \mathbf{J}_e(\mathbf{x}).
\end{aligned} \tag{7.47}$$

Note that in order to avoid working with the large update vector $\mathbf{h}$ we calculate local updates $\mathbf{h}_{i,j}$. These are combined cf. equation (7.39) to form the global update.

The Jacobian $[\mathbf{J}_{\mathrm{SM}}(\mathbf{x})]_{\mathbf{p}}$ is a scalar value, corresponding to the derivative of the similarity measure:

$$\begin{aligned}
[\mathbf{J}_{\mathrm{SM}_{i,j}}(\mathbf{x})]_{\mathbf{p}} &= \left.\frac{\partial \mathrm{SM}(I_i(\mathbf{x}), I_j(\mathbf{x} \circ \exp(\mathbf{h})); \mathbf{p})}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{0}} \\
&= \left.\frac{\partial \mathrm{SM}(I_i(\mathbf{x}), I; \mathbf{p})}{\partial I}\right|_{I=I_j(\mathbf{x}\circ\exp(\mathbf{0}))=I_j(\mathbf{x})} \\
&= \nabla \mathrm{SM}(I_i(\mathbf{x}), I_j(\mathbf{x}); \mathbf{p}). \qquad (7.48)
\end{aligned}$$

The Jacobian $[\mathbf{J}_{I_j}(\mathbf{x})]_{\mathbf{p}}$ is a matrix of dimension $(1 \times 3)$, corresponding to the spatial derivative of the moving image under the current transformation $\mathbf{x}$:

$$\begin{aligned}
[\mathbf{J}_{I_j}(\mathbf{x})]_{\mathbf{p}} &= \left.\frac{\partial I_j(w(\mathbf{x} \circ \exp(\mathbf{h}), \mathbf{p}))}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{0}} \qquad (7.49) \\
&= \left.\frac{\partial I_j(w(\mathbf{x}, w(\exp(\mathbf{h}), \mathbf{p})))}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{0}} \qquad (7.50) \\
&= \left.\frac{\partial I_j(w(\mathbf{x}, \mathbf{z}))}{\partial \mathbf{z}}\right|_{\mathbf{z}=w(\exp(\mathbf{0}),\mathbf{p})=\mathbf{p}} \qquad (7.51) \\
&= \nabla I_j(w(\mathbf{x}, \mathbf{p})). \qquad (7.52)
\end{aligned}$$

The Jacobian $[\mathbf{J}_w(\mathbf{x})]_{\mathbf{p}}$ is of dimension $(3 \times 16)$, corresponding to the derivative of the vector $w(\mathbf{Z}, \mathbf{p})$ with respect to the elements of the matrix $\mathbf{Z}$:

$$\begin{aligned}
[\mathbf{J}_w(\mathbf{x})]_{\mathbf{p}} &= \left.\frac{\partial w(\mathbf{x} \circ \exp(\mathbf{h}), \mathbf{p})}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{0}} \qquad (7.53) \\
&= \left.\frac{\partial w(\mathbf{Z}, \mathbf{p})}{\partial \mathbf{Z}}\right|_{\mathbf{Z}=\mathbf{x}\circ\exp(\mathbf{0})=\mathbf{x}} \qquad (7.54) \\
&= \begin{bmatrix} \mathbf{p}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{p}^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{p}^\top & \mathbf{0} \end{bmatrix}. \qquad (7.55)
\end{aligned}$$

The Jacobian $\mathbf{J}_e(\mathbf{x})$ is of dimension $(16 \times 6)$, corresponding to the derivative of the exponential mapping with respect to each of the transformation parameters $h_i$:

$$\begin{aligned}
\mathbf{J}_e(\mathbf{x})_i &= \left.\frac{\partial \exp(\mathbf{h})}{\partial h_i}\right|_{\mathbf{h}=\mathbf{0}} = \left.\frac{\partial \exp(\sum_{i=1}^{6} h_i \mathbf{l}_i)}{\partial h_i}\right|_{\mathbf{h}=\mathbf{0}} \qquad (7.56) \\
&= \exp\left(\sum_{i=1}^{6} h_i \mathbf{l}_i\right)\Bigg|_{\mathbf{h}=\mathbf{0}} \cdot \mathbf{l}_i = \mathbf{l}_i \qquad (7.57)
\end{aligned}$$

Stacking the vectorized basis vectors $[\mathbf{l}_i]_v$ of $\mathfrak{se}(3)$ leads to:

$$\mathbf{J}_e(\mathbf{x}) = \left[[\mathbf{l}_1]_v, \ldots, [\mathbf{l}_6]_v\right]. \qquad (7.58)$$

Analyzing the Jacobian matrices further, not focusing on a specific point but in their general form, we see that their layout is:

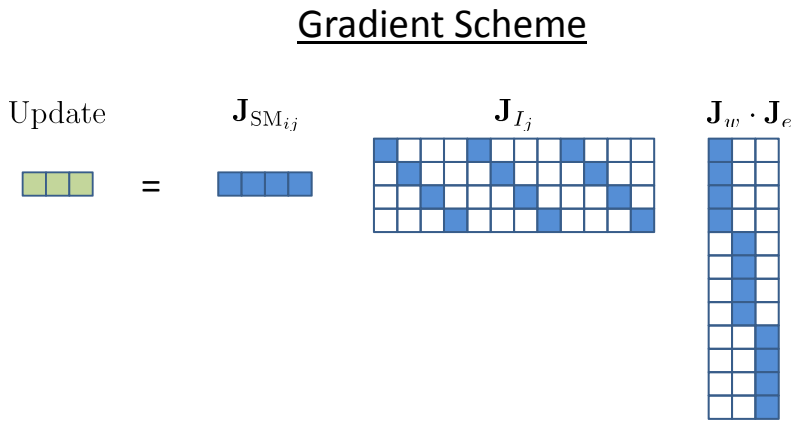



Figure 7.2: Schematic illustration of the Jacobian matrices and the resulting update. Only the colored boxes are unequal zero.

- $\mathbf{J}_{\mathrm{SM}_{i,j}}$ is a $1 \times N$ vector containing the derivatives of the similarity metric
- $\mathbf{J}_{I_j}$ is an $N \times 3N$ matrix with the image gradients in all 3 directions on the diagonals
- $\mathbf{J}_w$ is a $3N \times 16$ matrix
- $\mathbf{J}_e$ is a $16 \times 6$ matrix

with $N = |\Omega|$ the number of points in the grid. The product of the Jacobians $\mathbf{J}_w \cdot \mathbf{J}_e$ for one specific point $\mathbf{p}$ has the form

$$[\mathbf{J}_w]_{\mathbf{p}} \cdot \mathbf{J}_e = \begin{pmatrix} 0 & p_z & -p_y & 1 & 0 & 0 \\ -p_z & 0 & p_x & 0 & 1 & 0 \\ p_y & -p_x & 0 & 0 & 0 & 1. \end{pmatrix} \tag{7.59}$$

The extension of the product to $N$ points leads to a matrix with the following structure:

$$\mathbf{J}_w \cdot \mathbf{J}_e = \begin{pmatrix} 0 & p_z^1 & -p_y^1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & p_z^N & -p_y^N & 1 & 0 & 0 \\ -p_z^1 & 0 & p_x^1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -p_z^N & 0 & p_x^N & 0 & 1 & 0 \\ p_y^1 & -p_x^1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_y^N & -p_x^N & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7.60}$$

We schematically depict the scheme of the Jacobian matrices in figure 7.2. The illustration is corresponding to an image consisting of $N = 4$ pixels, 3 dimensions,

and 3 transformation parameters (3 translations). The calculation of the Jacobian of the cost function corresponds to an accumulation of all point-wise Jacobians:

$$\mathbf{J}_E(\mathbf{x}) = \sum_{\mathbf{p}\in\Omega} [\mathbf{J}_E(\mathbf{x})]_{\mathbf{p}}. \tag{7.61}$$

#### 7.3.3.2 Gauß-Newton

For the derivation of the gradient $\mathbf{J}_\mathbf{f}$, which is part of the Gauß-Newton optimization, we have to guarantee that the cost function fulfills further presumptions. The Gauß-Newton procedure was deduced by starting at a least-squares problem $E(\mathbf{x}) = \sum_{i\neq j} \frac{1}{2}||\mathbf{f}_{i,j}(\mathbf{x})||^2$, see equation (7.36). When considering SSD we can simply set $E(\mathbf{x}) = \sum_{i\neq j} \mathrm{SSD}_{i,j}(\mathbf{x})$, since SSD is intrinsically a least-squares problem.

This is not the case for other similarity measures like correlation ratio or mutual information. In order to ensure the least-squares nature, we square the similarity measures, leading to $E(\mathbf{x}) = \sum_{i\neq j} ||\mathrm{SM}_{i,j}(\mathbf{x})||^2$. Obviously, optimizing the squared similarity measure has far-ranging consequences, which we investigate further in section 7.3.3.4. The gradient $\mathbf{J}_\mathbf{f}$ is then calculated as:

$$[\mathbf{J}_{\mathbf{f}_{i,j}}(\mathbf{x})]_{\mathbf{p}} = \left[ \frac{\partial \mathbf{f}_{i,j}(\mathbf{x} \circ \exp(\mathbf{h}))}{\partial \mathbf{h}} \bigg|_{\mathbf{h}=0} \right]_{\mathbf{p}} \tag{7.62}$$

$$= \frac{\partial \mathrm{SM}(I_i(\mathbf{x}), I_j(\mathbf{x} \circ \exp(\mathbf{h})); \mathbf{p})}{\partial \mathbf{h}} \bigg|_{\mathbf{h}=0} \tag{7.63}$$

$$= [\mathbf{J}_{\mathrm{SM}_{i,j}}(\mathbf{x})]_{\mathbf{p}} \cdot [\mathbf{J}_{I_j}(\mathbf{x})]_{\mathbf{p}} \cdot [\mathbf{J}_w(\mathbf{x})]_{\mathbf{p}} \cdot \mathbf{J}_e(\mathbf{x}). \tag{7.64}$$

Stacking all the point-wise derivatives leads to the Jacobian of $\mathbf{f}$:

$$\mathbf{J}_{\mathbf{f}_{i,j}}(\mathbf{x}) = \begin{pmatrix} [\mathbf{J}_{\mathbf{f}_{i,j}}(\mathbf{x})]_{\mathbf{p}_1} \\ \vdots \\ [\mathbf{J}_{\mathbf{f}_{i,j}}(\mathbf{x})]_{\mathbf{p}_{|\Omega|}} \end{pmatrix}. \tag{7.65}$$

#### 7.3.3.3 ESM

The last gradient that remains is $\mathbf{J}_\mathbf{f}^{\mathrm{ESM}}$ for the ESM. Here we also consider the squared similarity measures like for GN. The calculation of $\mathbf{J}_\mathbf{f}^{\mathrm{ESM}}$ is difficult because part of its calculation is $\mathbf{J}_\mathbf{f}(\mathbf{x} \circ \exp(\mathbf{h}))$, which depends on $\mathbf{h}$ that we want to solve for. In order to address this issue, [Benhimane and Malis, 2004] consider the optimal update step $\mathbf{h}^{\mathrm{opt}}$ for the current location $\mathbf{x}$, leading to the perfect alignment $\mathbf{x}^{\mathrm{opt}} = \mathbf{x} \circ \exp(\mathbf{h}^{\mathrm{opt}})$. To consider the influence of this optimal update step for the product:

$$\mathbf{J}_{\mathbf{f}_{i,j}}(\mathbf{x} \circ \exp(\mathbf{h}^{\mathrm{opt}})) \cdot \mathbf{h}^{\mathrm{opt}} = \frac{\partial \mathbf{f}_{i,j}(\mathbf{x} \circ \exp(\mathbf{h}))}{\partial \mathbf{h}} \bigg|_{\mathbf{h}=\mathbf{h}^{\mathrm{opt}}} \cdot \mathbf{h}^{\mathrm{opt}},$$

we have to analyze each of the four factors resulting from the derivation, see equation (7.64).

We proceed from right to left, starting with $\mathbf{J}_e$. [Benhimane, 2006, pp. 157] present a proof that:

$$\mathbf{J}_e(\mathbf{x}) \cdot \mathbf{h}^{\text{opt}} = \mathbf{J}_e(\mathbf{x}^{\text{opt}}) \cdot \mathbf{h}^{\text{opt}}$$

utilizing the properties of the Lie algebra and the exponential map. Next, the derivative of the transformation $\mathbf{J}_w$ is the same for $\mathbf{x}$ and $\mathbf{x} \circ \exp(\mathbf{h}^{\text{opt}})$ [Benhimane and Malis, 2004]. In order to have an approximation of the third term, the main assumption of ESM is incorporated. The gradient of the perfectly aligned image $\nabla I_j(\mathbf{x} \circ \exp(\mathbf{h}^{\text{opt}}))$ can be approximated by the gradient of the fixed image $\nabla I_i(\mathbf{x})$, leading to:

$$\mathbf{J}_{I_j}(\mathbf{x} \circ \exp(\mathbf{h}^{\text{opt}})) \approx \mathbf{J}_{I_i}(\mathbf{x}). \tag{7.66}$$

This takes the specificity of our optimization problem into account, because for image registration the possibility exists to approximate this gradient. Naturally, this is only feasible for images of the same modality. The last term is the derivative of the similarity measure, which we approximate by $\mathbf{J}_{\text{SM}}(\mathbf{x})$. This finally leads to the overall approximation:

$$\mathbf{J}_{\mathbf{f}_{i,j}}(\mathbf{x} \circ \exp(\mathbf{h}^{\text{opt}})) \approx \mathbf{J}_{\text{SM}_{i,j}}(\mathbf{x}) \cdot \mathbf{J}_{I_i}(\mathbf{x}) \cdot \mathbf{J}_w(\mathbf{x}) \cdot \mathbf{J}_e(\mathbf{x}). \tag{7.67}$$

Considering the definition of the gradient $\mathbf{J}_{\mathbf{f}}^{\text{ESM}} = \frac{1}{2}(\mathbf{J}_{\mathbf{f}}(\mathbf{x}) + \mathbf{J}_{\mathbf{f}}(\mathbf{x} \circ \exp(\mathbf{h})))$, and equations (7.64) and (7.67), we finally get:

$$\mathbf{J}_{\mathbf{f}_{i,j}}^{\text{ESM}} = \frac{1}{2} \cdot \mathbf{J}_{\text{SM}_{i,j}} \cdot [\mathbf{J}_{I_i} + \mathbf{J}_{I_j}] \cdot \mathbf{J}_w \cdot \mathbf{J}_e. \tag{7.68}$$

evaluated at the current alignment $\mathbf{x}$.

#### 7.3.3.4 Gradient of Similarity Measures

As mentioned in the last section, we optimize the squared similarity measures for NCC, CR, and MI to ensure the least-squares nature of the optimization problem. For sum of squared differences this is not necessary. The interesting question is about the consequences of optimizing the squared function instead. Assuming a function $\phi$ and its squared version $\Phi = \phi^2$. The first and second derivatives of $\Phi$ are $\Phi' = 2 \cdot \phi \cdot \phi'$ and $\Phi'' = 2 \cdot (\phi')^2 + 2 \cdot \phi \cdot \phi''$. These indicate that there are new extrema for $\phi = 0$ and that they change their type for $\phi < 0$, which is problematic. NCC, CR, and MI have a lower bound, which is -1 and 0, respectively. This enables to avoid these optimization problems, by adding a constant $\nu$ to the similarity measures $\text{SM}_{i,j} + \nu$, guaranteeing that they are in the positive range.

We list the actual derivatives of the similarity measures in [Wachinger and Navab, 2009b], but they can also be found in *e.g.* [Hermosillo et al., 2002]. Note that for the calculation of the update $\mathbf{h}$ of the least-squares problems, either an LU- or Cholesky-decomposition could be used on the normal equations $(\mathbf{J}_{\mathbf{f}}^{\top}\mathbf{J}_{\mathbf{f}})\mathbf{h} = -\mathbf{J}_{\mathbf{f}}\mathbf{f}$, or a QR-decomposition on $\mathbf{J}_{\mathbf{f}}\mathbf{h} = -\mathbf{f}$. Since the normal equations worsen the numerical condition of the problem, the QR-decomposition presents the stabler choice.

#### 7.3.3.5 Relationship to Forward/Inverse Compositional Update

In [Baker and Matthews, 2004], the differences between forward and inverse compositional updates are discussed. The inverse update scheme has computational advantages, because the image gradient can be pre-computed. In our case of simultaneous registration, where all images move, we are no longer able to pre-compute the gradient image, and consequently the difference vanishes. It is, however, interesting to look at ESM from this perspective, because instead of either considering the gradient of the fixed or moving image, both gradients are combined. Hence, ESM presents a combination of forward and inverse compositional update.

## 7.4 Multi-Modal Registration with ESM

The fundamental assumption of ESM, the approximation $\mathbf{J}_{I_j}(\mathbf{x} \circ \exp(\mathbf{h}^{\text{opt}})) \approx \mathbf{J}_{I_i}(\mathbf{x})$, prevents its direct application to multi-modal registration. The reason is that for multi-modal images, the gradient directions and orientations are not comparable. A solution to address this issues has, however, been presented in chapter 6, with the creation of structural images. Since they convert a multi-modal to a mono-modal registration problem, they make the approximation in equation (7.66) meaningful and ESM therefore applicable.

Another positive aspect of structural representations is a faster similarity evaluation, due to the reduced computationally complexity of the metrics. This is even more important for simultaneous registration, where the multivariate similarity metrics are more demanding to calculate. For the case of APE, the influence of the faster similarity evaluation is quadratic, since all combinations of pair-wise estimates are calculated in equation (7.36).

A final advantage of structural representations in combination with ESM, that we want to mention, concerns the validation of registration results. The validation of rigid registration is generally easier than the validation of deformable registration, because under the assumption of a rigid object, it is possible to measure the camera pose to obtain ground truth data. The drawback is, however, that there are rarely volumetric acquisitions of group of images from a static object. One example concerns volumetric ultrasound acquisitions with a tracked ultrasound transducer. Another interesting application is the alignment of multi-modal volumes for neurosurgery. Effort has been taken to exactly measure the location with bone-implanted fiducial markers in order to provide ground truth data [West et al., 1997]. This data is well suited for rigid registration experiments because the acquisitions are performed within a short time frame, and further, the skull provides a rigid frame limiting deformations. Since this is a multi-modal scenario, the validation is only possible in combination with structural representations. The structural representation that we employ in combination with ESM are entropy images.

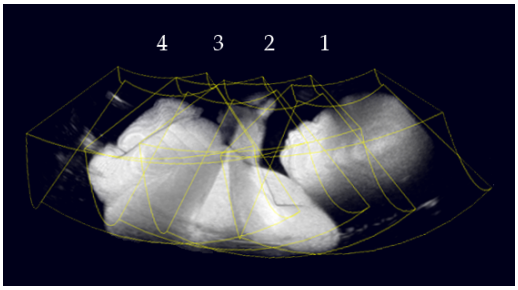



Figure 7.3: Mosaic of baby phantom from 4 acquisitions.

## 7.5 Experiments

We perform experiments for two different applications to test APE in combination with the described optimization procedures. One application is the registration of multiple ultrasound volumes for volumetric ultrasound mosaicing and the second application is the alignment of a group of multi-modal volumes.

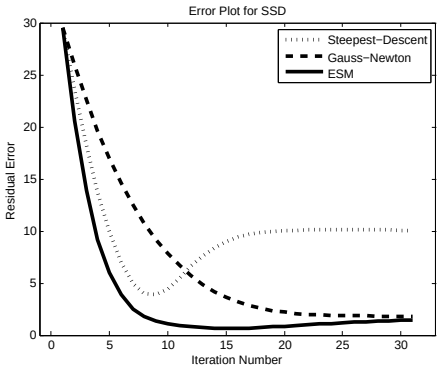
### 7.5.1 Ultrasound Mosaicing

For ultrasound mosaicing, the experiments were conducted on four 3D ultrasound acquisitions from a baby phantom, having a resolution of $64 \times 64 \times 64$ voxels, see figure 7.3. We displaced the volumes randomly from the correct position, guaranteeing an accumulated residual error of 30 over all the volumes. We weight 1mm equal to 1° to make translational and angular displacement from the ground truth comparable. Starting from the random initial position we run 100 simultaneous registrations for each configuration to assess its performance.

In figures 7.4 and 7.5, the averaged residual error is plotted with respect to the iteration number. For SSD, see figure 7.4(a), we only have one plot because we do not have to consider the squared variant of it, as already mentioned. The best performance is obtained with ESM, leading to the fastest convergence. But also the Gauß-Newton method leads to a robust convergence. Steepest-descent does not perform well. Although it seems to approach the correct alignment nicely at the beginning, it diverges into another optimum. In the table in figure 7.4(b), the number of registrations that diverged are listed. We consider a registration diverged, if the residual error after 30 steps is larger than half the initial error.

For CR, see figure 7.4(c), the results for GN and ESM are not good. All of the 100 runs diverged. Steepest-descent, although slower, performs better. The situation changes a lot, when optimizing the squared function, see figure 7.4(d). The ESM quickly approaches the correct alignment. Also GN improves, but the result is still not convincing. We also plot the curve for SD as reference, although it is the one of CR, because we do not use the squared variant for SD.

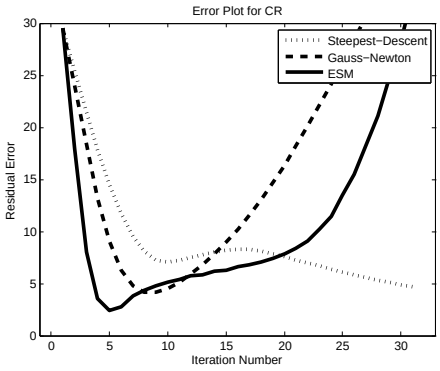
For NCC and MI, see figure 7.5, the situation is pretty similar to CR. The performance of GN and ESM when using the non-squared similarity measures is



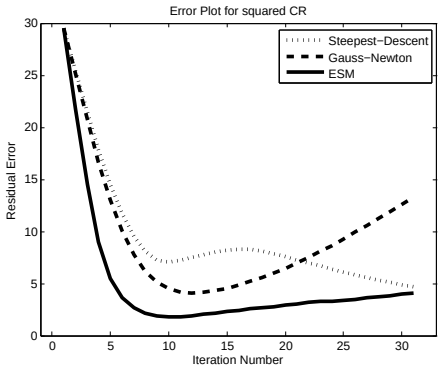

(a) SSD

| | SD | GN | ESM |
|---|---|---|---|
| SSD | 6 | 0 | 0 |
| CR | 0 | 100 | 100 |
| $CR^2$ | - | 25 | 0 |
| NCC | 0 | 88 | 0 |
| $NCC^2$ | - | 7 | 0 |
| MI | 0 | 38 | 12 |
| $MI^2$ | - | 31 | 2 |

(b) Number of diverged registrations from 100




(c) CR




(d) Squared CR

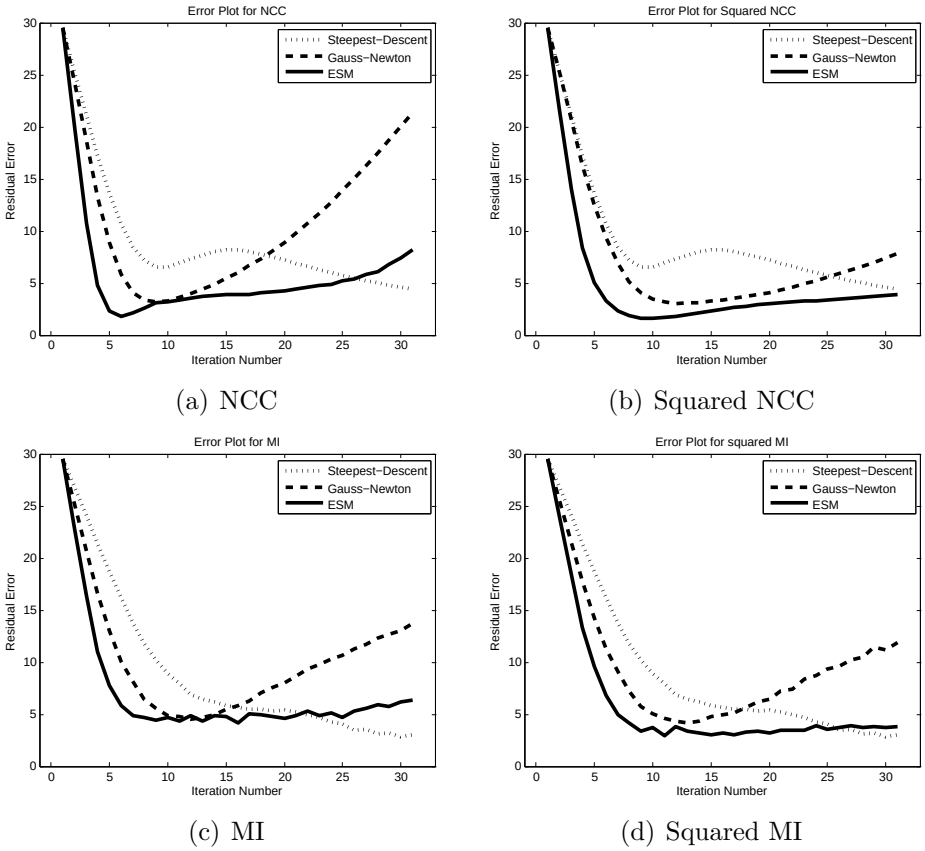
Figure 7.4: Plot of the average residual error for each iteration step for SSD, CR, and squared CR. Comparing CR and squared CR shows the improved performance of GN and ESM. ESM converges the fastest and leads to the smallest residual error.

insufficient, leading to a high divergence rate. The situation improves enormously when optimizing the squared function instead. ESM always performs better than GN, both, with respect to speed and robustness. Furthermore, the performance of SD is interesting. Although the convergence is slower, compared to the others, it is in most cases robust. All the registrations are performed on an Intel dual-core 2.4 GHz processor having 2 GB of RAM. The time for one registration, where we allowed for 30 iterations, is below one minute.

### 7.5.2 Multi-Modal Registration

For multi-modal registration, we conduct experiments on T1, T2, and PD-weighted MR images from the BrainWeb database[1] and CT, T1, T2, and PD images from the Retrospective Image Registration Evaluation (RIRE) database[2]. These are the same

[1] http://www.bic.mni.mcgill.ca/brainweb/
[2] http://www.insight-journal.org/rire/




(a) NCC (b) Squared NCC

(c) MI (d) Squared MI

Figure 7.5: Plot of the average residual error for each iteration step for NCC, squared NCC, MI, and squared MI. The convergence of GN and ESM is significantly improved for the squared similarity measures. ESM is converging the fastest.

images that we used in chapter 6. Cross-sectional slices of the original volumes and entropy volumes for Brainweb are shown in figure 6.4 and for RIRE in figure 6.5. For the calculation of the entropy images, we use $9 \times 9 \times 9$ sub-volumes in the case of isotropic voxel spacing. For anisotropic spacing we adapt the neighborhood accordingly. Further, we perform a spatially weighted density estimation using a Gaußian weighting scheme together with a kernel-based Parzen window method. We select 64 bins and a global normalization of the intensity values. The Shannon entropy is chosen to measure the entropy.

In the simultaneous registration study, we compare ESM on entropy volumes to Gauß-Newton on entropy and original volumes. For Gauß-Newton on the original volumes, we select mutual information as similarity metric, which is corresponding to the sate-of-the-art configuration. Further, we use SSD as similarity measure for the registration with entropy volumes. We run 50 registrations, each starting from a random initial position. Each initial position has an accumulated RMS error of 45 over all volumes from the correct alignment, again weighting 1mm equal to 1°. The average residual error for each step is shown in figure 7.6. We observe that for

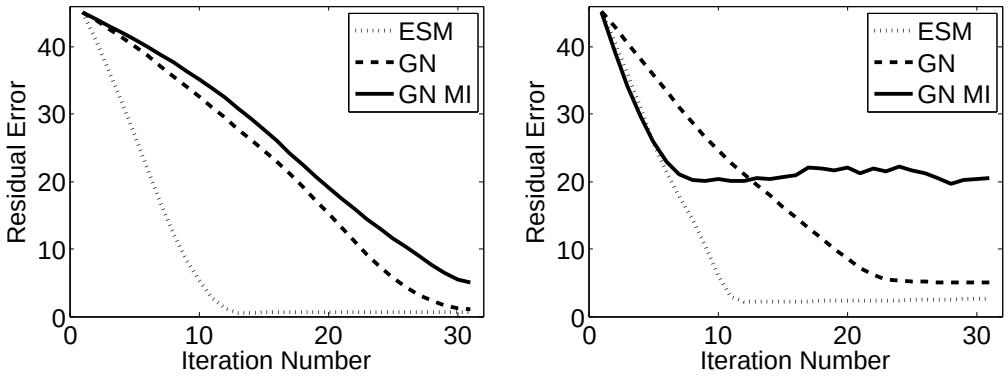



Figure 7.6: Multi-modal simultaneous registration for BrainWeb (left) and RIRE (right) volumes. For GN and ESM SSD on the entropy volumes is used, for GN MI mutual information on the original volumes.

the BrainWeb dataset the convergence of GN on the original volumes with MI and GN on the entropy images with SSD is identical, as to be expected. ESM converges, however, significantly faster than GN. On the RIRE data, most registrations do not converge for GN on the original volumes. GN with entropy images leads to good registration results. The convergence of ESM is, however, once again significantly faster than the one of GN.

## 7.6 Discussion

The experiments show the good performance of simultaneous registration using the APE framework and gradient-based optimization. The performance of the optimization methods, however, depends on the chosen similarity measure. In our experiments, the squared versions of NCC, CR, and MI perform better for GN and ESM. For all measures, the fastest approximation to the correct results is obtained with ESM. In most cases GN was faster than SD.

The convergence graphs are not all monotonic, as one would expect; approaching the ground truth further with each iteration until the convergence is achieved. The reasons for the increase lie, one the one hand, in the averaging over the 100 registrations, thus diverging trials lead to a large residual error that is averaged over. And on the other hand, we see the reasons in the complex registration scenario. For ultrasound mosaicing, the volumes are inherently contaminated by speckle patterns, making it a difficult registration problem. Analogously, the registration of multi-modal volumes is challenging. The performance of ESM indicates that it is more robust in such noisy scenarios because the gradient information of both images is considered. Finally, our results show that structural representations and ESM nicely complement each other.

## 7.7 Conclusion

We presented further insights into multivariate similarity measures and optimization methods for simultaneous registration of multiple images. First, we deduced APE from a ML framework and showed its relation to the congealing framework. This required an extension of the congealing framework with neighborhood information. Second, we focused on efficient optimization methods for APE. We started the deduction of the optimization methods from the same Taylor expansion, to provide the reader a good overview of the methods and further insights into the relatively unknown ESM. We further presented the optimization of intrinsically non-squared similarity metrics in a least-squares optimization framework. Finally, we illustrated the application of ESM for multi-modal registration with structural representations. Our experiments showed a superior performance of ESM with respect to speed and accuracy for the case of ultrasound mosaicing and multi-modal registration.

## 7.8 Future Work

In recent years, other techniques than gradient-based optimization schemes were used for image registration. Next to the already mentioned Fourier-based techniques [Reddy and Chatterji, 1996, Padfield, 2010] or the batch alignment by sparse and low-rank decomposition [Peng et al., 2010], alternative approaches include the usage of stochastic optimization [Wells et al., 1996, Cole-Rhodes et al., 2003], discrete optimization [Zikic et al., 2010] and convex optimization [Kokiopoulou and Frossard, 2009, Kokiopoulou et al., 2011b]. For discrete optimization, it is possible to converge to strong optima, however, this is dependent on the generally sparse discretization of the search space. Optimization methods that try to combine both approaches, in a discrete-continuous optimization were proposed [Trobin et al., 2008, Lempitsky et al., 2008]. Convex optimization gained much popularity in computer vision and pattern recognition. Typical tasks such as segmentation [Chan et al., 2006] and reconstruction [Kolev et al., 2009] are formulated as convex optimization problems. For registration, only limited work is available for affine 2D [Kokiopoulou and Frossard, 2009] and recently for affine 3D [Kokiopoulou et al., 2011b] alignment. The principle is to approximate the images with convex atoms. The application of affine transformations does not change the convexity of the approximation. Comparing two images, which is done by calculating the difference, leads to a difference of convex functions (DC). For DC cost functions, DC programming methods can be applied the find the global optimum, *e.g.* the cutting plane algorithm [Kokiopoulou and Frossard, 2009]. The registration with such registration algorithms is so far limited to pairwise, mono-modal, affine alignment. It would be interesting to extend such approaches to the alignment of images from different modalities, to groups of images, and to more general transformation models.

# Chapter 8

# Ultrasound Specific Similarity Measures

*In this chapter, we combine the maximum likelihood framework for registration of chapter 5 with the theory of speckle statistics and sound propagation of chapter 2. After presenting an overview of imaging models and similarity metrics introduced in the literature, we propose a locally adaptive Nakagami-based similarity measure. Moreover, we illustrate that the viewing angle dependency of ultrasound applies an implicit weighting on the images, negatively affecting the registration process. To address this issue, we additionally introduce a new matching function separating reflectivity and scattering regions, which are the results of two different types of physical interactions of the ultrasound beam with tissue.*

## 8.1 Introduction

In [Wachinger et al., 2007], it was shown that the pairwise registration of ultrasound images is challenging due to the missing overlap invariance of similarity measures, favoring a total overlap of the images. The application of simultaneous registration is a possible solution to this problem, as discussed in chapter 7. Faced with the registration of only two images, this approach is, however, not applicable. Hence, we try to adapt similarity measures to the particularities of ultrasound, as stated in chapter 2, to improve their performance. Especially important is to take speckle characteristics into account by selecting appropriate noise distributions. The developed measures can directly be applied for simultaneous registration, where more accurate pairwise estimates yield a more accurate groupwise estimate with APE.

In section 8.3, we will present an overview of imaging models and similarity measures presented in the literature. Further, we introduce a new spatially adaptive metric for envelope data based on the Nakagami distribution. In section 8.4, we focus on the effects of the viewing angle dependency of ultrasound for registration. A first attempt is to apply a model of ultrasound reflection to reweight ultrasound images, trying to make them viewing angle independent. Due to a lack of accuracy of the model, we propose a new matching function, separating reflection and scattering.

## 8.2 Related Work

A large number of articles address the registration of ultrasound images [Strintzis and Kokkinidis, 1997, Krücker et al., 2002, Cohen and Dinstein, 2002, Xiao et al., 2002, Boukerroui et al., 2003, Revell et al., 2004, Zikic et al., 2006, Poon and Rohling, 2005, Wachinger et al., 2007, Grau et al., 2007, Basarab et al., 2007, Elen et al., 2008, Esther Leung et al., 2008, Myronenko et al., 2009]. Typical application areas are (i) motion measurements in echocardiography for detecting and characterizing abnormalities, (ii) breast deformation analysis to assess the elastic properties of tissues, (iii) assessment of tissue strain with elastography, and (iv) multi-view compounding. In [Strintzis and Kokkinidis, 1997], likelihood functions are presented assuming a multiplicative Rayleigh noise. This is further extended in [Cohen and Dinstein, 2002], assuming that both, the moving and the fixed image, are affected by multiplicative speckle noise. Moreover, the log-compression is incorporated into the imaging model to achieve better results on B-mode data. These similarity metrics are successfully applied in [Boukerroui et al., 2003, Revell et al., 2004] for motion estimation with a block matching approach. In [Basarab et al., 2007], block matching with normalized cross-correlation is performed for ultrasound flow and elasticity imaging. [Krücker et al., 2002, Poon and Rohling, 2005] apply a block matching approach to improve the quality of the compounded images. While [Krücker et al., 2002] evaluate several similarity measures, with the conclusion that SSD is better suited for low noise levels, [Poon and Rohling, 2005] entirely focus on NCC. In [Myronenko et al., 2009], similarity measures based on the bivariate Rayleigh and Nakagami distribution are presented. Common to all the presented methods is that the parameters of the distributions are set heuristically on a global basis, which is at odds with the underlying local data variation.

One of the contributions of this chapter is a locally adaptive similarity measure. There are several approaches that are related in terms of general registration not focusing on ultrasound. A typical situation that challenges the application of mutual information is the registration of images, which contain high intensity non-uniformity due to the bias field, see section 6.7.3. Approaches that address this issue try to estimate the joint density in local regions or do a combination of local and global estimation [Loeckx et al., 2010, Zhuang et al., 2009]. This local adaptation is different from our contribution, because no noise estimates are performed. Further, the estimation of joint densities is complicated for small regions, as required for local MI. Recently, learning-based similarity measure were proposed for multi-modal registration [Lee et al., 2009, Bronstein et al., 2010]. Supervised learning is performed on previously registered data to learn an appropriate similarity function. A major drawback of these methods is the necessity of having access to correctly aligned data for training. The adaptation of the distribution parameters in our approach could be regarded as learning, with the learning being directly performed on the actual images.

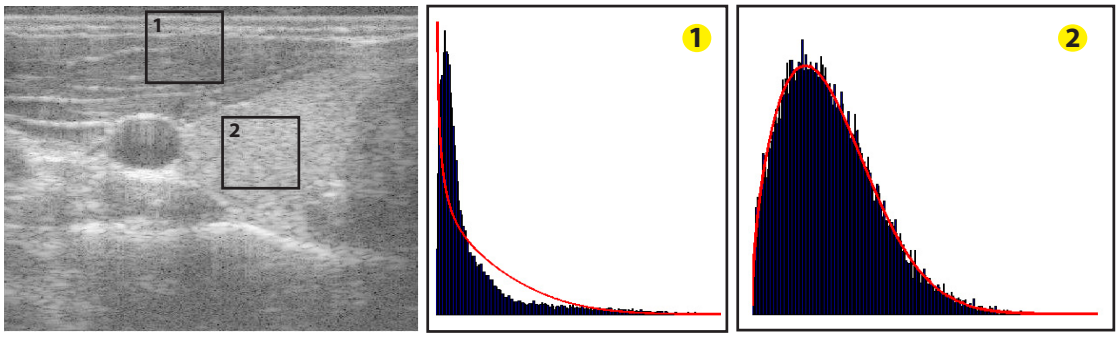



Figure 8.1: Left: Ultrasound image with two example regions. Middle and right: histograms of the regions together with Nakagami MLE fits (red).

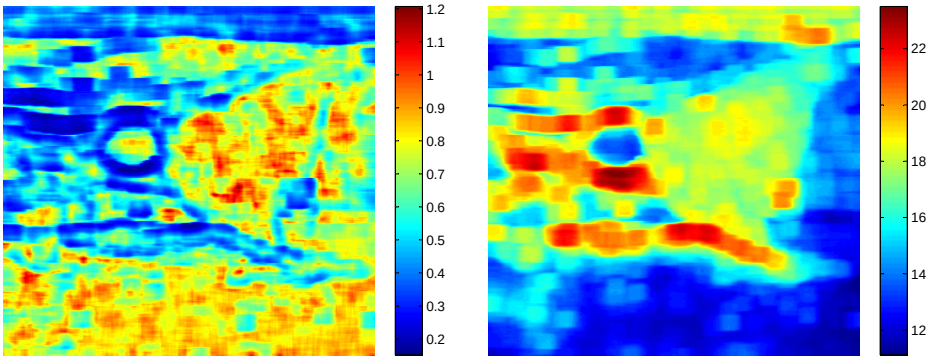
## 8.3 Ultrasound Specific Likelihood Functions

In the field of segmentation and classification, research on various distributions for modeling ultrasound scattering has been performed over the last years. In section 2.4, we illustrated that the commonly used Rayleigh distribution only models a specific scattering scenario. Alternative distributions were introduced that deal with varying numbers of scatterers per resolution cell and with the presence of coherent structures. Especially interesting is the Nakagami distribution, because it allows for modeling a large number of scattering conditions, while being computationally efficient [Shankar, 2000]. Adaptation to various noise conditions is achieved by setting the shape $m$ and scale $\omega$ parameter accordingly. In figure 8.1, we show the histogram of two regions of the ultrasound image together with the corresponding Nakagami MLE fit. We observe a large disparity of the histograms. The Nakagami parameters estimated for each patch on the same RF image are visualized in figure 8.2. The large variation of the distribution parameters highlights the necessity of spatial adaptation of noise models to accommodate for the speckle locality.

A Nakagami-based similarity measure has recently been proposed in [Myronenko et al., 2009]. It is problematic that the parameters of the distribution are, first, set globally on the entire image, and second, determined heuristically. As mentioned above, the advantage of the Nakagami distribution is to model various scattering scenarios; however, this requires specifying the parameters adequately. Since multiple scattering scenarios are usually present within an ultrasound image, as illustrated in figures 8.1 and 8.2, we argue that a local adaptation increases the descriptiveness. While this analysis focuses on ultrasound registration, the local adaptation of similarity measures is of general interest and is also applicable to other domains.

### 8.3.1 Ultrasound Likelihood Functions

In the following, we present an overview of ultrasound similarity measures proposed in the literature and derive a new locally adaptive Nakagami-based metric. The likelihood functions are incorporated in the maximum likelihood estimation for




(a) $m$ map (b) $\omega$ map

Figure 8.2: Illustration of Nakagami MLE parameters shape $m$ and scale $\omega$, calculated densely on the ultrasound image in figure 8.3.

registration as discussed in chapter 5. Considering ultrasound envelope images $I$ and $J$ together with the transformation $T$, registration is formulated as

$$\hat{T} = \arg\max_{T} \log p(I \mid J, T, \varepsilon) \tag{8.1}$$

with $\hat{T}$ the estimated transformation and noise $\varepsilon$. Below, we list several imaging models and denote them with the initials of the authors, prevailing in the literature.

### $SK_1$: Multiplicative Rayleigh Noise

In [Strintzis and Kokkinidis, 1997], an imaging model based on a multiplicative Rayleigh noise is presented

$$I = J \cdot \varepsilon \tag{8.2}$$

with $\varepsilon$ Rayleigh distributed. This noise model considers only one image to be degraded by noise, while the other one has to be noiseless. Knowing the distribution $p(\varepsilon)$, we have to find the distribution for the likelihood $p(I \mid J, T)$. Assuming that we know the distribution of a random variable $x$ and we want to calculate the distribution of a random variable y, with both being related by the function $g$, so $y = g(x)$. The fundamental theorem [Papoulis and Pillai, 1991, p.93] states that the distribution of the random variable $y$ is calculated with

$$p(y) = \frac{p(x)}{|g'(x)|}. \tag{8.3}$$

In our case, we set $g(x) = x \cdot J$ and obtain $\frac{dg(x)}{dx} = J$ so that

$$p(I \mid J, T) = \frac{1}{J} \cdot p(\varepsilon). \tag{8.4}$$

Setting the variance of the Rayleigh distribution to $\frac{2}{\pi}$ leads to the log-likelihood function of $SK_1$

$$\log p(I \mid J,T) = \log \frac{1}{J} \cdot p\left(\frac{I}{J}\right) \tag{8.5}$$

$$\approx \log\left(\frac{I}{J^2}\right) - \frac{\pi}{4}\frac{I^2}{J^2}. \tag{8.6}$$

A second noise model is proposed in [Strintzis and Kokkinidis, 1997] with a signal dependent Gaußian noise, not further considered here.

### $CD_1$: Division of Rayleigh Noises

A more realistic and refined model is proposed in [Cohen and Dinstein, 2002], assuming that each image is contaminated by a multiplicative Rayleigh noise $\varepsilon_1$ and $\varepsilon_2$, respectively. Considering the underlying, noise-free scene $S$, the imaging model is formulated as $I = S \cdot \varepsilon_1$ and $J = S \cdot \varepsilon_2$, leading to

$$I = J \cdot \frac{\varepsilon_1}{\varepsilon_2} = J \cdot \eta \tag{8.7}$$

with the division of probabilities $\eta = \frac{\varepsilon_1}{\varepsilon_2}$. The distribution of the divisional noise $p(\eta)$ is calculated with [Papoulis and Pillai, 1991, p.138]

$$p(\eta) = \int_{-\infty}^{\infty} \varepsilon_2 \cdot p(\eta\varepsilon_2, \varepsilon_2)\, \mathrm{d}\varepsilon_2. \tag{8.8}$$

Considering the noise in the images to be independent $p(\varepsilon_1, \varepsilon_2) = p(\varepsilon_1)p(\varepsilon_2)$, of equal variance, and Rayleigh distributed, the integration results in

$$p(\eta) = \frac{2 \cdot \eta}{(\eta^2 + 1)^2}. \tag{8.9}$$

The log-likelihood function of $CD_1$ is

$$\log p(I \mid J,T) = \log \frac{1}{J} \cdot p\left(\frac{I}{J}\right) \tag{8.10}$$

$$= \log \frac{1}{J} \frac{2 \cdot \frac{I}{J}}{\left(\left(\frac{I}{J}\right)^2 + 1\right)^2} \tag{8.11}$$

$$\approx \log I - 2\log J - 2\log\left[\left(\frac{I}{J}\right)^2 + 1\right]. \tag{8.12}$$

### $CD_2$: Logarithm of Division of Rayleigh Noises

The second model in [Cohen and Dinstein, 2002] considers next to the noise contamination of both images also the log-compressed nature of ultrasound images

$$\log I = \log(J \cdot \eta) = \log J + \log \eta. \tag{8.13}$$

With setting $\tilde{I} = \log I$ and $\tilde{J} = \log J$

$$\eta = \exp(\tilde{I} - \tilde{J}). \tag{8.14}$$

The likelihood function, applying the fundamental theorem again, but this time with respect to the log-compressed images, is

$$p(\tilde{I} \mid \tilde{J}, T) = \eta \cdot p(\eta). \tag{8.15}$$

This is obtained with $g(\eta) = \log J + \log \eta$ and the derivative $g'(\eta) = \frac{1}{\eta}$. The log-likelihood function for $CD_2$ is

$$\begin{aligned}
\log p(\tilde{I} \mid \tilde{J}, T) &= \log \frac{\exp(\tilde{I})}{\exp(\tilde{J})} \cdot p(\exp(\tilde{I} - \tilde{J})) && (8.16)\\
&= \log \frac{\exp(\tilde{I})}{\exp(\tilde{J})} \cdot \frac{2 \cdot \exp(\tilde{I} - \tilde{J})}{\left[\exp(\tilde{I} - \tilde{J})^2 + 1\right]^2} && (8.17)\\
&= \log \frac{2 \cdot \exp(2(\tilde{I} - \tilde{J}))}{\left[\exp(2(\tilde{I} - \tilde{J})) + 1\right]^2} && (8.18)\\
&\approx \tilde{I} - \tilde{J} - \log[\exp(2(\tilde{I} - \tilde{J})) + 1]. && (8.19)
\end{aligned}$$

The application of the presented ultrasound specific likelihood terms for simultaneous registration is presented in [Wachinger and Navab, 2008]. Working with B-mode images, we achieved a superior performance with $CD_2$ than with SSD and NCC.

### Locally Adaptive Nakagami-Based

As explained in section 2.4, the Nakagami distribution is well suited for modeling speckle statistics in ultrasound. We also discussed that many of the theoretical properties are lost by a conversion to B-mode, see section 2.4.6. Since we are interested in adapting the similarity measure to various scattering scenarios, we directly work with envelope data in order to achieve meaningful results. This is different to [Strintzis and Kokkinidis, 1997, Cohen and Dinstein, 2002, Myronenko et al., 2009] that work on B-mode.

For the calculation of divisional noises in equation (8.8), a bivariate distribution is required, if no assumptions about independence are incorporated as in [Cohen and Dinstein, 2002]. Since speckle is due to physical interaction of the beam with the tissue, the same speckle patterns appear under constant acquisition conditions. In modern ultrasound systems with high frame rates and for acquisitions from the same viewing angle, it is therefore reasonable to assume the random variables $\varepsilon_1$ and $\varepsilon_2$ to be correlated [Cobbold, 2007]. The bivariate Nakagami distribution is [Myronenko et al., 2009]

$$p(\varepsilon_1, \varepsilon_2) = \frac{m^{m+1}(\varepsilon_1 \varepsilon_2)^m}{2^{m-1}\sigma^{2(m+1)}(1-\rho)\rho^{(m-1)/2}\Gamma(m)} \exp\left(-\frac{m(\varepsilon_1^2 + \varepsilon_2^2)}{2(1-\rho)\sigma^2}\right) I_{m-1}\left(-\frac{m\sqrt{\rho}\varepsilon_1\varepsilon_2}{(1-\rho)\sigma^2}\right) \tag{8.20}$$

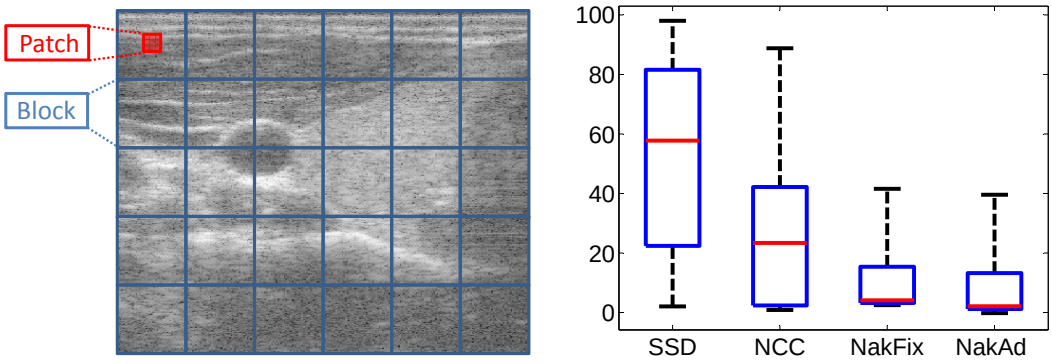



Figure 8.3: Left: Illustration of the subdivision of the image domain into blocks and patches. Right: Boxplot of errors from rigid registration study.

with $\rho$ being a squared correlation coefficient, the distribution widths $\sigma^2$ being equal, and a modified Bessel function of the first kind of order $m-1$

$$I_{m-1}(z) = \sum_{k=0}^{\infty} \frac{z^{2k+m-1}}{2^{2k+m-1}\Gamma(k+m)k!}. \tag{8.21}$$

Incorporating the bivariate Nakagami distribution in the conditional density leads to

$$p(I|J,T) = \frac{1}{J}p(\eta) = \frac{1}{J}\int_{-\infty}^{\infty} \varepsilon_2 \cdot p(\eta\varepsilon_2, \varepsilon_2)\,\mathrm{d}\varepsilon_2 \tag{8.22}$$

$$= \frac{1}{J}2(1-\rho)^m \frac{\Gamma(2m)}{\Gamma(m)^2}\frac{\eta^{2m-1}}{(\eta^2+1)^{2m}}\left(1-\frac{4\rho\eta^2}{(\eta^2+1)^2}\right)^{-\frac{2m+1}{2}} \tag{8.23}$$

$$= \frac{1}{J}\frac{2(1-\rho)^m}{\beta(m)}\cdot\frac{1}{\eta}\cdot\frac{\eta^{2m}}{(\eta^2+1)^{2m}}\cdot\left(1-\frac{4\rho\eta^2}{(\eta^2+1)^2}\right)^{-\frac{2m+1}{2}} \tag{8.24}$$

with $\beta(m) = \frac{\Gamma(m)^2}{\Gamma(2m)}$. More details on the deduction, however, assuming log-compressed envelope data and therefore a different imaging model, are presented in [Myronenko, 2010]. Finally, computing the log-likelihood yields the Nakagami-based similarity measure

$$\log p(I|J,T)$$

$$= -\log J + \log\frac{2(1-\rho)^m}{\beta(m)} - \log\eta + 2m\log\left(\frac{\eta}{\eta^2+1}\right) - \frac{2m+1}{2}\log\left(1-\frac{4\rho\eta^2}{(\eta^2+1)^2}\right) \tag{8.25}$$

$$\approx -\log J - \log\left(\frac{I}{J}\right) + 2m\log\left(\frac{\left(\frac{I}{J}\right)}{\left(\frac{I}{J}\right)^2+1}\right) - \frac{2m+1}{2}\log\left(1-\frac{4\rho\left(\frac{I}{J}\right)^2}{\left(\left(\frac{I}{J}\right)^2+1\right)^2}\right). \tag{8.26}$$

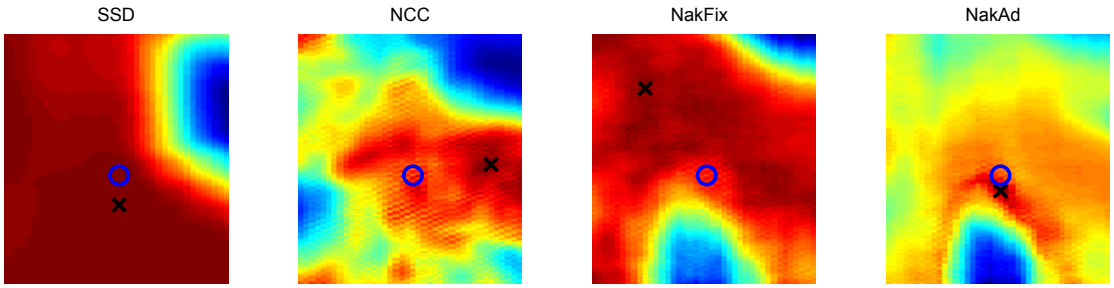



Figure 8.4: Similarity plots for various measures. Blue circle: correct alignment, black cross: similarity maximum.

### 8.3.2 Parameter estimation

During each iteration of the similarity computation, the bivariate Nakagami model has to be instantiated. This requires the specification of a shape parameter $m$ as well as a correlation coefficient $\rho$, see equation (8.26). Considering the need for high locality, the patch size has to be kept at minimum. However, this sparsity constraint is detrimental for the computation of the $m$ parameter of the bivariate Nakagami distribution. Additionally, because of the high frequency of this computation, complexity has to be kept at bare minimum. Therefore, we decided to use a fast approximation scheme. The distribution parameters are calculated separately for the moving $(m_I, \omega_I)$ and the fixed patches $(m_J, \omega_J)$ with maximum likelihood estimation. Then, given the individual MLE parameters, the joint shape parameter $m$ of the bivariate Nakagami is approximated by $m = (m_I + m_J)/2$.

For computing the correlation coefficient $\rho$ for the bivariate Nakagami, probably the most obvious way is to directly use the corresponding intensities of the patches. However, due to the noise-susceptibility of ultrasound, a probabilistic correlation score was chosen in order to increase the overall reliability. In this respect, we employ the Bhattacharyya coefficient $(BC)$

$$BC(p, q) = \int \sqrt{p(x)q(x)}\, dx \tag{8.27}$$

with $0 \leq BC \leq 1$, which is a measure of the relative overlap between two probability distributions $p$ and $q$. In our case, the distributions correspond to the Nakagami MLE estimates of the moving and the fixed image, respectively.

### 8.3.3 Experiments

Experiments are performed on several RF datasets, acquired with a linear transducer at 3.3 MHz. The RF data is sampled with 40 MHz. We calculate the RF envelope with the 2D analytic signal, as presented in chapter 3. It is important to remark that we directly benefit from the improved statistical properties of the envelope resulting from the 2D analytic signal in this application. Depending on the depth setting, the

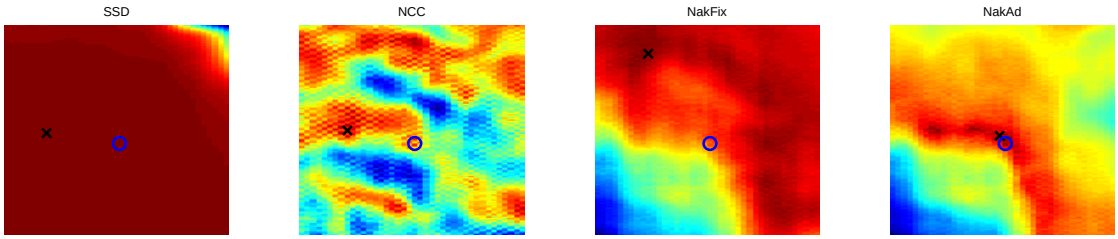



Figure 8.5: Similarity plots for various measures. Blue circle: correct alignment, black cross: similarity maximum.

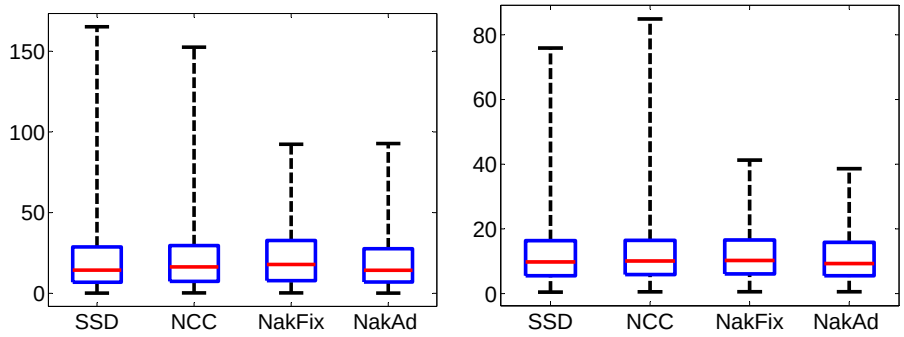
| | SSD | NCC | NakFix | NakAd |
|---|---|---|---|---|
| Dataset 1 | 9.3 | 9.9 | 10.4 | **9.1** |
| Dataset 2 | 9.8 | 10.1 | 10.2 | **9.2** |
| Dataset 3 | 9.9 | 10.4 | 11.1 | **9.5** |
| Dataset 4 | 15.7 | 15.7 | 17.0 | **13.8** |

Table 8.1: Median of errors of random registration study for various datasets and similarity metrics. Lowest errors are indicated in boldface.

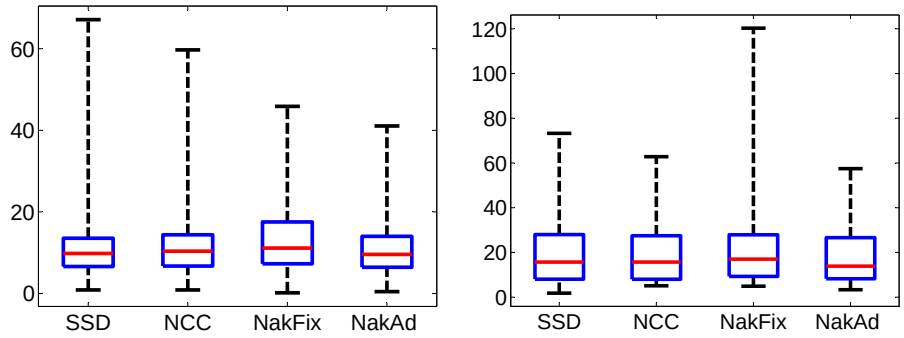
images have a resolution between 1157 and 2080 pixels in axial, and 256 pixels in lateral direction. Each dataset consists of a moving and a fixed image. We perform block matching on 100 blocks in the images, which are equally distributed across the image. This is schematically illustrated in figure 8.3. For each block, we have the coordinates from a manual alignment, serving as ground truth data. We compare SSD and NCC with the similarity measure presented in equation (8.26). Once we use heuristic values for the distribution parameters, referred to as NakFix, and once we estimate them on the images, referred to as NakAd. For the heuristic case, we choose the parameters used in [Myronenko et al., 2009] $m = 0.5$ and $\rho = 0.8$. NakFix is therefore the analogon of the similarity measure presented in [Myronenko et al., 2009] for not log-compressed envelope data.

To compare the different similarity measures, we extract a patch of $91 \times 11$ pixels in the moving image and shift it over the block in the fixed image. The distribution parameters are estimated on these patches. We illustrate similarity plots for two different blocks in figures 8.4 and 8.5. In both cases, SSD and NCC are not able to correctly indicate the correct alignment. Moreover, the similarity plot of NCC shows several local minima. Also the maximum of NakFix is far off the correct alignment. Interestingly, the adaptation of the parameters, as it is done in NakAd, significantly changes the similarity function, leading to good results in these cases.

The similarity plots provide a first impression of the performance, which we further evaluate by performing block matching. As discussed in section 8.2, block matching is commonly applied in ultrasound registration. We extract patches from the moving image and try to find the corresponding patch in the fixed image. For each image, this is done for one patch per block, so on 100 patches. Since we know




(a) Dataset 1 (b) Dataset 2




(c) Dataset 3 (d) Dataset 4

Figure 8.6: Boxplots of errors from random registration study for block matching.

the alignment of the blocks, we can calculate the Mahalanobis distance with respect to the ground truth position, serving as error measure. We select the Mahalanobis distance to compensate for the significantly higher resolution in axial direction. To be able to perform statistics on the results of the registration, we perform a random registration study. For this we randomly displace the patch 100 times from the ground truth position, with maximal initial deviation of $\pm 40$ pixels in axial and $\pm 10$ pixels in lateral direction. The errors over all patches and all runs are shown in the boxplots in figure 8.6. We list the median errors in table 8.1. We observe that the median, the box, and the whiskers are lowest for NakAd in all cases. The performance of SSD and NCC is comparable, with slight advantages for NCC with respect to the whiskers. The performance of NakFix for datasets 1 - 3 is slightly worse than NakAd. For dataset 4, NakFix is not leading to good results.

Next to the analysis of the presented similarity measure for block matching, we also perform experiments for global rigid registration. Therefore, the image is separated into blocks, as already discussed previously. This time, we estimate the parameters for each block and evaluate the similarity measure on the block. Accumulating the similarity estimates from all blocks leads to the global similarity estimate. We also perform a random registration study with 100 runs and initial deviations up to $\pm 240$ pixels in axial and $\pm 70$ pixels in lateral direction. The results are shown in figure 8.3. The median errors are SSD: 57.8, NCC: 23.4, NakFix:

4.1, and NackAd: 2.1. NCC performs better than SSD, but the best results are obtained for the Nakagami-based similarity measures, with the proposed adaptive version outperforming the fixed version. Our results therefore confirm the theoretical advantages of applying a more adequate noise model and the local adaptation of the similarity measure.

## 8.4 Viewing Angle Dependency

In the last section, we presented various models to account for speckle noise. The proposed similarity measure assumes correlated speckle in the images, which is likely the case for images acquired from a similar pose. For volumetric mosaicing, we commonly have a limited number of acquisitions from largely varying poses in order to gain field of view. Especially when scanning parts that show strong reflections such as bone, diaphragm, or kidney surface, the viewing angle dependency has to be taken into account. Already in our previous experiments, we noted problems with the overlap invariance of the similarity measures, favoring a total overlap of the images [Wachinger et al., 2007]. The reasons for this could either be rooted in the similarity measures themselves, as [Cahill et al., 2008] addressed, or in the ultrasound images. We show that the viewing angle dependency of ultrasound images not only causes angle dependent artifacts like shadow, but also puts an implicit weighting on the images, favoring a total overlap.

To illustrate the problems that affect the registration process when working with viewing angle dependent ultrasound images, we create two artificial images acquired with a curved linear probe, see figure 8.7. The images show a region of high reflectivity, having a cosine intensity profile, and a scattering region. Multiplicative Rayleigh distributed noise is added to the images (log-compressed) to simulate the speckle noise of the ultrasound images. The correct alignment of the two images, which can be acquired with lateral displacement, corresponds to an overlap of the scattering regions. However, as we show in the similarity plot of figure 8.7, similarity measures like $CD_2$ favor a total overlap of the images (0 displacement). The reason is the dominance of the reflectivity structure, leaving only a local minimum at the desired position.

To address this issue, we first evaluate methods for reweighting ultrasound images, making them viewing angle independent. However, currently used ultrasonic imaging models are not accurate enough, as we will show, to allow such a reweighting. We instead propose a new ultrasound matching function, consisting of a first term, measuring the alignment of reflectivity structures, and a second term, measuring the alignment of the remaining texture. This new matching function has advantages for US-US, as well as, US-CT registration.

### 8.4.1 Reweighting Ultrasound Images

A first attempt, when thinking about the problem, is to identify structures of high reflectivity, calculate the incidence angle of the beam, and use the ultrasonic imaging

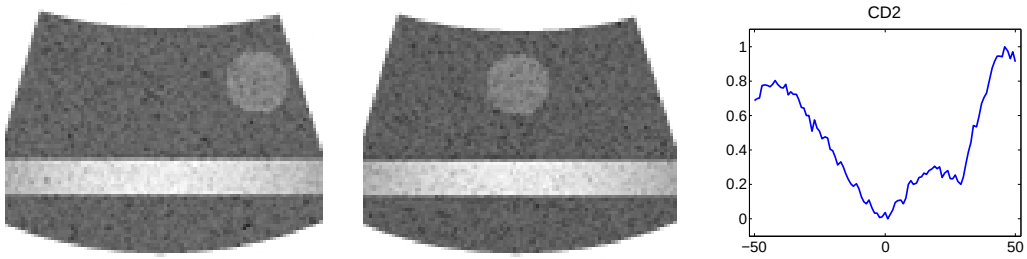



Figure 8.7: Artificial ultrasound images showing reflection, scatter, and speckle. Similarity plot using $CD_2$ along lateral direction.

model as shown in equation (2.8)

$$I_r(\mathbf{x}) = I_i(\mathbf{x}) \cdot \cos^m \theta_i(\mathbf{x}) \cdot \zeta_I(\mathbf{x}), \tag{8.28}$$

to reweight the image, with $\zeta_I$ the reflection coefficient. The result would be the creation of normalized ultrasound images being less or not at all viewing angle dependent. In the case of our example, the reflectivity region would result in a constant similarity value for translations in lateral direction, so that the scattering region would dominate the registration process.

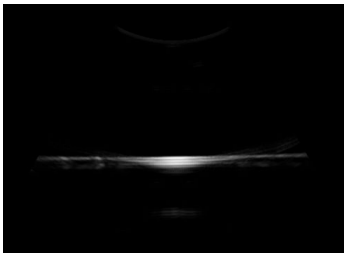
In the following, we will investigate whether the presented ultrasound model is accurate enough for reweighting the images. We acquired ultrasound B-mode and RF images with a curved linear transducer in a water bath with just one object in it, having a diffuse surface, usually used for single wall calibration, see figure 8.8. The reflection coefficient $\zeta_I(\mathbf{x})$ is the same for every beam, as we have the common water-object boundary. The incident intensity $I_i(\mathbf{x})$ is dependent on the attenuation of the beam along its way to the boundary. The attenuation coefficient, which includes absorption and scattering effects, is $2.2 \cdot 10^{-3} \frac{\text{dB}}{\text{cm·MHz}}$ for water. In comparison, the attenuation coefficients for fat ($0.6 \frac{\text{dB}}{\text{cm·MHz}}$), liver ($0.9 \frac{\text{dB}}{\text{cm·MHz}}$), and kidney ($1.0 \frac{\text{dB}}{\text{cm·MHz}}$), are about three orders of magnitude larger. Further, the length of the beam in the middle and the one on the side vary about 2 cm (1.93 cm exactly). This leads to an attenuation of $22.8 \cdot 10^{-3}$ dB by using an acquisition frequency of 2.6 MHz, considering also the way back to the transducer. Since the ratio of change of intensity $\frac{I_1}{I_0}$ is close to 1 for values close to 0 dB, we can neglect the effect of attenuation for our experiment.

The received signal, therefore only depends on the incident angle of the beam. The subsequent cost function has to be minimized in order to find the optimal exponent $m$, characterizing best the type of the interface

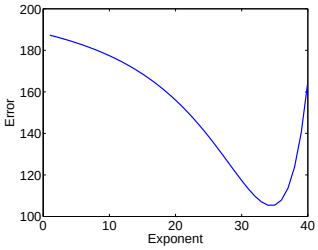
$$\mathcal{C}(m) = \int \left( \frac{|I_r(\mathbf{x})|}{\cos^m \theta_i(\mathbf{x})} - \zeta_I(\mathbf{x}) \cdot I_i(\mathbf{x}) \right)^2 \mathrm{d}\mathbf{x} \tag{8.29}$$

$$\frac{d\,\mathcal{C}(m)}{d\,m} = \int 2 \left( \frac{|I_r(\mathbf{x})|}{\cos^m \theta_i(\mathbf{x})} - \zeta_I(\mathbf{x}) I_i(\mathbf{x}) \right) |I_r(\mathbf{x})| \cdot \ln(\cos \theta_i(\mathbf{x})) \cdot \cos^{-m} \theta_i(\mathbf{x})\, \mathrm{d}\mathbf{x}$$

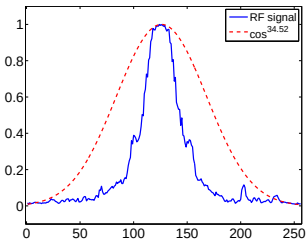
with $\zeta_I(\mathbf{x}) \cdot I_i(\mathbf{x})$ being constant for all $\mathbf{x}$ on the surface and corresponding to the measured reflectivity of the middle beam, and $|I_r(\mathbf{x})|$ the absolute value of the received signal.



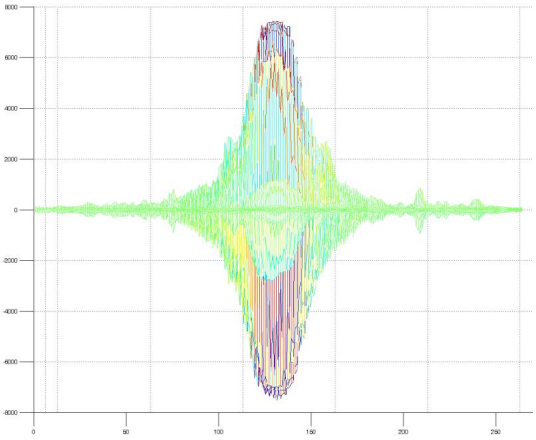
(a) B-Mode US image



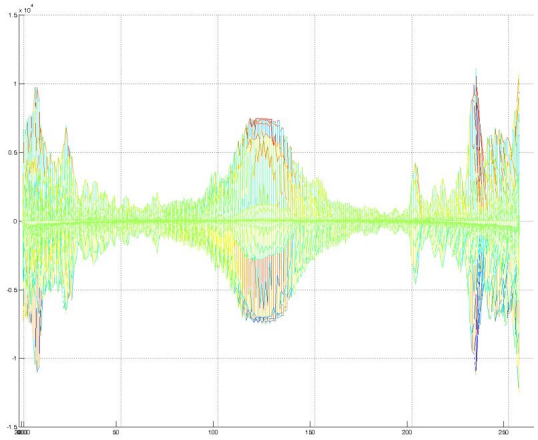

(b) Cost Function




(c) Intensity Profile and Best Approximation



(d) RF Profile (x-axis: lateral)



(e) Reweighted RF Profile

Figure 8.8: Reweighting experiment on ultrasound RF data.

We plot the cost function in figure 8.8(b), where a gradient descent optimization with the above stated gradient of the cost function leads to $m = 34.52$. In figure 8.8(c), we show the intensity profile of the RF data, together with the optimal cosine variant. The advantage of working with RF data is that no further processing steps or filters are applied on the data, making a meaningful evaluation of the model possible. In figure 8.8(e), we show the reweighted RF signal with $\cos^{34.52}$. As can be seen, the result is not satisfying because instead of having a constant intensity profile, we have a "W" shaped one. For other interfaces it may be possible to find an exponent $m$ so that reweighting works, however, this would only characterize this specific type of interface and would not be generally applicable. We conclude that the ultrasound model in equation (8.28) is usually not accurate enough to allow for reweighting ultrasound images, in order to make them viewing angle independent.

This finding also affects an algorithm for US-CT registration [Wein et al., 2008], which bases upon the simulation of ultrasound images, applying this ultrasonic imaging model. In the next section, we introduce a new ultrasound matching scheme, usable for US-US and US-CT registration, to address this issue.

### 8.4.2 A New Ultrasound Matching Scheme

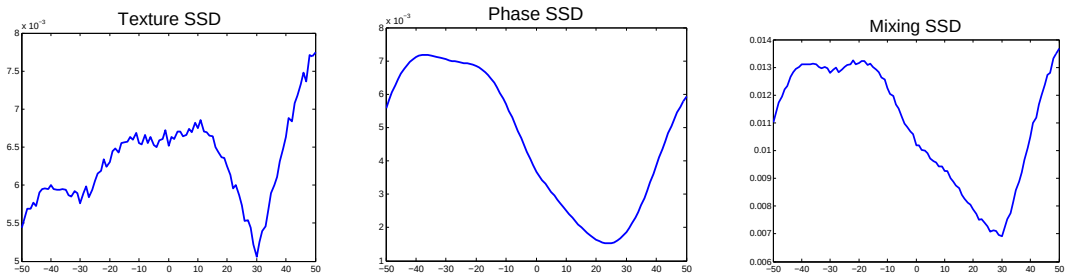
As we concluded, a correction of the angle dependency of the images with the current model is not feasible. Our approach is therefore to separate reflection and scattering, which are the results of two different physical interactions of the beam with the tissue, combined in one image. Reflection occurs at large scale tissue boundaries and is viewing angle dependent. Scattering is caused by microscopic tissue inhomogeneities and provides the internal texture of the organs. Scattering is to a certain extent dependent on the direction of insonification because the ultrasound point spread functions are not spherically symmetric. However, since we are not able to match single scattering responses from the resolution cells, it is more appropriate to consider entire scattering regions, and the echogeneity of these regions does not change with the viewing angle. In the following, we first present the method that we use for the crucial identification of regions of reflectivity in the ultrasound images, and then we describe the new matching function.

#### Reflectivity Regions

We use the local phase $\phi$, as described in chapter 3, for identifying regions of reflectivity because it provides us with structural information independent of the brightness and contrast [Grau et al., 2007]. This independence is very important for the extraction of reflectivity regions, since the absolute intensity of the reflection varies with the incident angle. The local phase was already used by [Mulet-Parada and Noble, 2000] for boundary detection in echocardiography. Recently, [Hacihaliloglu et al., 2008] achieved good results in segmenting bone in ultrasound images employing the local phase. [Mellor and Brady, 2005] apply mutual information on local phase images for multi-modal image registration. [Grau et al., 2007, Zhang et al., 2007] use the local phase for aligning 3D echocardiographic sequences. We threshold the phase image with a value $\tau$ to obtain a binary mask, indicating reflectivity regions. Analogously to the identification in chapter 4, we use $\tau = 0.7$ for all experiments, showing that this is not a crucial parameter. Alternatively, the local phase values could directly be used as weights, resulting in a fuzzy like mask.

#### Matching Function

Since the local phase image provides us with structural information, building a matching function upon this like in [Grau et al., 2007] would already significantly reduce the effects of the viewing angle dependency on the registration. However, this also leads to the following drawbacks: First, we would be completely dependent on the outcome of the algorithm calculating the local phase. And second, texture information in the form of scattering regions would not be integrated. Focusing on boundaries seems totally fine for echcardiographic applications, where not much texturing exists, but for other applications, like abdominal ones, this is not optimal. Consequently, we propose a hybrid matching function which considers both, feature- and intensity-based aspects by using reflection and texture.




(a) $\mathcal{M}^T$ (b) $\mathcal{M}^R$ (c) $\mathcal{M}$

Figure 8.9: Similarity plots for images from Figure 8.7 along lateral direction using SSD.

Considering two images $I$ and $J$, an image grid $\Omega$, regions of reflectivity $\Lambda_I = \{x \in \Omega | \phi_I(x) > \tau\}$ and $\Lambda_J = \{x \in \Omega | \phi_J(x) > \tau\}$, further $\Lambda = \Lambda_I \cup \Lambda_J$, the matching function is

$$\mathcal{M}(I, J) = \mathcal{M}^T_{\Omega \setminus \Lambda}(I, J) + \lambda \cdot \mathcal{M}^R_{\Omega}(\phi_I, \phi_J) \tag{8.30}$$

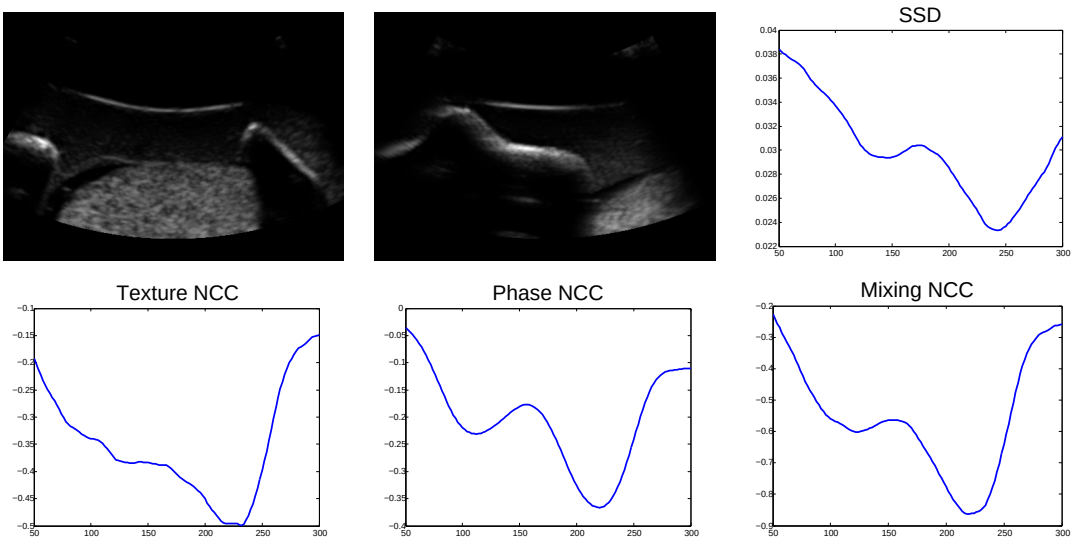
with $\mathcal{M}^T_{\Omega \setminus \Lambda}$ measuring the similarity of the texture regions, excluding reflectivity regions $\Omega \setminus \Lambda$, and with $\mathcal{M}^R_{\Omega}(\phi_I, \phi_J)$ measuring the similarity of the phase images on the entire grid $\Omega$. The matching function can be easily extended to the multivariate case with APE (cf. chapter 7).

Suitable similarity measures for $\mathcal{M}^T$ are, next to standard measures like SSD, NCC, and mutual information, the previously discussed ultrasound specific similarity measures. For them it is important to consider whether RF or B-mode data is used. Suitable similarity measures for $\mathcal{M}^R$ are once again the standard measures, and the similarity measure proposed in [Grau et al., 2007], taking the cosine of the phase difference. The disadvantage, when choosing two different similarity measures, is that the correct selection of the weighting term $\lambda$ is important, since it has to map the similarity score into a comparable range. Even when working with normalized images, the outcomes of the similarity measures are not directly comparable, so that we perform the experiments with the same similarity measures for both terms and simply set $\lambda = 1$.

With regard to the probabilistic framework of chapter 5, the presented metric integrates as hybrid approach. The description layers are once the phase images and once the original images. The restriction of the similarity evaluation on the original images to non-reflectivity regions, $\Omega \setminus \Lambda$, is modeled with the keypoint mechanism for geometric registration.

### 8.4.3 Experiments

Figure 8.9 shows the results of the new matching function with SSD as similarity metric for texture and reflection for the data set illustrated in figure 8.7. Comparing the similarity plots for $CD_2$ and $\mathcal{M}^T$, which both use the texture, we can see the improvement of not considering the reflectivity region $\Lambda$ in the calculation. We




(d) $\mathcal{M}^T$ (e) $\mathcal{M}^R$ (f) $\mathcal{M}$

Figure 8.10: Ultrasound images acquired from laterally displace positions. Similarity plots for SSD and the different parts from our new matching scheme.

selected $CD_2$ because the images are log-compressed. The plot of the term $\mathcal{M}^R$, measuring the similarity of the phase images is very smooth, however, does not indicate the correct alignment at 30. Combining both curves, $\mathcal{M}$, leads to an accurate cost function with wider capture range than $\mathcal{M}^T$.

We also conducted experiments for US-US registration on ultrasound images acquired with a curved linear transducer, see figure 8.10. The images are acquired from an ultrasound phantom, with a true displacement of 88.0 mm. The bony structure on the left side of the first image is depicted in the middle of the second image. The similarity plots for SSD, NCC, and $CD_2$ are comparable, with SSD indicating the correct alignment at 96.8 mm. The combined cost function, with NCC as similarity measure for both terms, indicates the correct alignment at 88.8 mm, which is close to the true displacement.

## 8.5 Conclusion

The alignment of ultrasound images is complicated by the viewing angle dependent nature of the images and their inherent contamination by speckle noise. We introduced a locally-adaptive similarity measure based on the bivariate Nakagami distribution. As discussed above, the adaptation of the similarity measure to various scattering scenarios is necessary to model the noise correctly. Moreover, we analyzed the effects of the viewing angle dependency on the registration process. This is especially important for acquisitions at different poses and of regions with strong reflections. Due to a lack of accuracy of reweighting, we introduced a new hybrid matching scheme for ultrasound images, incorporating feature- and intensity-based aspects.

# Part IV

# MOTION MODELING

This part is about the second clinical application of this dissertation: motion modeling. First, we address the creation of 4D data by retrospective gating, proposing a purely image-based system with manifold learning. Second, we introduce a new registration technique for time-resolved data, considering the spatial and temporal components simultaneously to guarantee a smooth deformation field along the temporal direction.

# Chapter 9

# Image-Based Breathing Gating

*Respiratory motion is a challenging factor for image acquisition and image-guided procedures in the abdominal and thoracic region. In order to address the issues arising from respiratory motion, it is often necessary to detect the respiratory signal. In this chapter, we propose a novel, purely image-based retrospective respiratory gating method for ultrasound and MRI. Further, we use this technique to provide a solution for breathing-affected 4D ultrasound acquisitions with a wobbler probe, and similarly, the creation of 4D MR with a slice stacking approach. We achieve the gating with Laplacian eigenmaps, which assign to each image frame a coordinate in low-dimensional space by respecting the neighborhood relationship, making them well suited for analyzing the breathing cycle. We validate the image-based gating on several ultrasound and MR datasets by comparing to alternative gating approaches.*

## 9.1 Introduction

Respiration is a cyclic, irregular motion that leads to deformations in the abdominal and thoracic region. The respiratory signal monitors the current breathing phase of the patient. For numerous applications, it is necessary to assign to each image the corresponding respiratory phase in which it was acquired. One example is the previously discussed image mosaicing, where only the fusion of images from the same breathing state leads to consistent panorama images. An alternative to breathing gating are breath-hold acquisitions, but they further complicate the procedure and are dependent on the patient's ability for breath-hold. Another application is radiation therapy, where organ motion due to respiration can lead to inaccuracies during target localization [Flampouri et al., 2006]. These inaccuracies require the treatment margins to be much larger than the tumor size, resulting in exposure of the healthy tissue in the surroundings to a high radiation dose [Colgan et al., 2008]. It is therefore necessary to study the respiratory motion in the planning process to decrease treatment margins. For an accurate analysis, 4D imaging techniques are necessary in order to handle the motion of the organs over time [Remmert et al., 2007, Li et al., 2008]. Prospective and retrospective gating methods exist, with advantages of retrospective approaches being discussed in [Rohlfing et al., 2001].

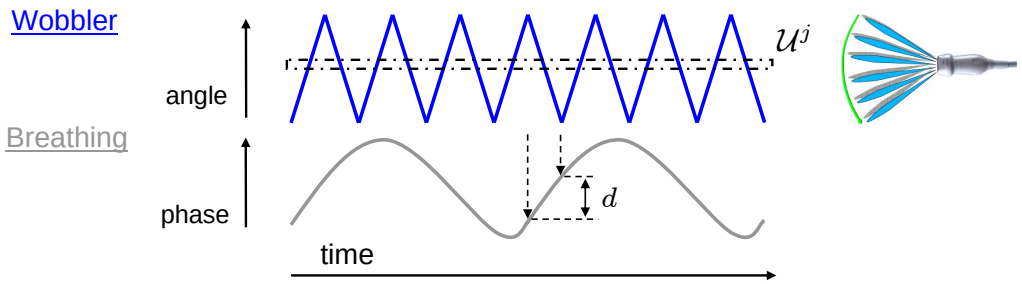



Figure 9.1: Wobbler angle (blue) and respiratory phase (gray) over time. Dashed lines indicate respiratory change $d$ within one sweep. Dash dotted line indicates frames from same angle over several breathing cycles.

We describe the creation of 4D ultrasound and 4D MR data with a retrospective approach in further details throughout the chapter. 4D data is a prerequisite for the motion modeling with deformable registration in chapter 10.

The measurement of the breathing phase is generally achieved with external gating systems, attached to the patient. The disadvantages of the usage of such systems are long setup times, the prolongation of the overall acquisition, and its high costs. Additionally, the image data stream and the respiratory signal have to be synchronized, which is not trivial. For certain imaging devices, such as CT and MR scanners, solutions for the synchronization exist, however, we are not aware of such a possibility for ultrasound; leaving the synchronization to the user. The consequence is that gating systems are rarely used in practice. We propose a purely image-based retrospective respiratory gating system using manifold learning. The proposed method is fully automatic and does not need any prior information about the anatomy, training data, or user interaction. The basic performance of our algorithm is to assign to a stream of images acquired from the same position over time the corresponding respiratory signal. In the following, we explain how we us this basic technique for creating 4D data.

### 9.1.1 4D Ultrasound with Wobbler

One application that we investigate in more details, and for which we have not yet found a solution proposed in the literature, is the acquisition of breathing-affected 4D ultrasound with a mechanically steered transducer, also referred to as wobbler. The problem for using a wobbler in such a scenario is that images in one sweep do not contain consistent information, but represent the anatomy in different breathing states. We illustrate this in figure 9.1, where we schematically plot the deviation angle of the wobbler together with the respiratory signal over time. The phase difference $d$ indicates the range of breathing phases accumulated in one sweep. We propose to select all frames acquired from the same angle (dash dotted line) and to apply the image-based gating on each of these sets of images. Having the respiratory signal estimated for each angle, we align these local curves and apply a robust spline curve fitting to create a globally consistent respiratory signal. This, consequently,

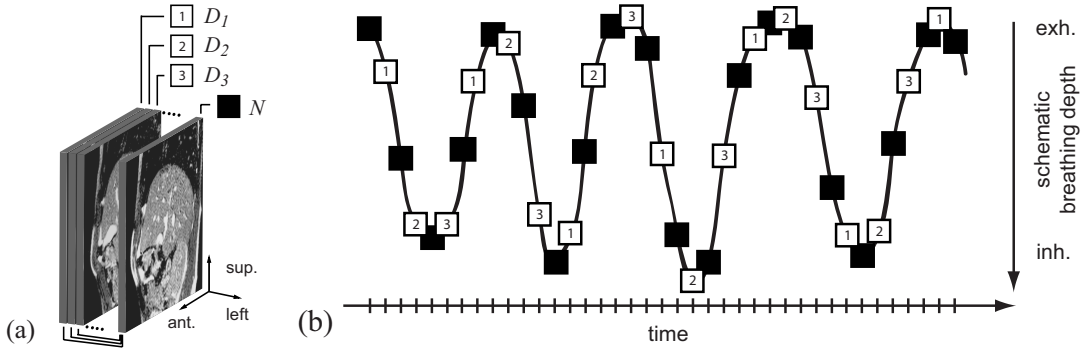



Figure 9.2: (a) Sagittal slices from the volume of interest. Data slices $D^1, D^2, D^3$ and the dedicated navigator slice, $N$. (b) The interleaved acquisition of data and navigator slices. Solid squares indicate navigator slices while the others indicate data slices with the position number displayed in the boxes. In this case only 3 positions are illustrated. Courtesy of Martin von Siebenthal [von Siebenthal et al., 2007].

allows us to reconstruct volumes for specific breathing stages.

An alternative to the application of a wobbler to obtain 4D ultrasound would be a native 3D transducer with elements arranged on a 2D array. Such systems, however, are still expensive and the access to data streaming and radio frequency data is very restricted. Our proposed method is, nevertheless, also interesting in conjunction with 2D array transducers, because we can perform the breathing-gating on the acquired volumetric data.

### 9.1.2 4D MRI with Navigator Slices

Several techniques based on MRI have been proposed in the literature to handle the respiratory motion. Among these are breath-hold and slice stacking techniques. The problem with breath-hold is that patients may not be able to hold their breath during the acquisition. [von Siebenthal et al., 2007] use the slice stacking method where 2D slices from different locations having the same breathing state are stacked together to reconstruct a 3D image for that state. They acquire dedicated high quality slices, called *navigator slices*, at a fixed location to determine a similarity criterion for sorting the data slices.

The acquisition process for the slice stacking approach is illustrated in figure 9.2. It is differentiated between *data slices* $D_i^p$ with $p$ indicating the position and $i$ indicating time and *navigator slices* $N_i$. As already mentioned, the navigator slices are acquired at a fixed location and are used to estimate the breathing phase. An alternating acquisition scheme,

$$\{\ldots, N_i, D_{i+1}^p, N_{i+2}, D_{i+3}^{p+1}, N_{i+4}, D_{i+5}^{p+2}, \ldots\}, \tag{9.1}$$

is applied to interleave the data slices and navigator slices, as shown in figure 9.2. Further details about the acquisition process are presented in [von Siebenthal, 2008].

In order to find data slices that are acquired in the same breathing state, its surrounding navigator slices are compared. The rationale is to assume that if the enclosing navigator slices are similar, then the data slices are in the same breathing state. For comparing navigator slices, non-rigid registrations are performed [von Siebenthal et al., 2007]. Although this produces the desired results, the processing time, ranging from 10 to 100 hours, is the major drawback of this method for usage in clinical practice. Our proposed method is able to recover the breathing signal from the navigator slices within seconds. It can either be applied on its own or as a pre-processing step for the non-rigid registration by defining a search window. This significantly reduces the processing time for the non-rigid registration because of the pre-selection of navigator slices to be registered.

## 9.2 Related work

There are many articles on image-based gating in ultrasound for detecting the cardiac motion [Treece et al., 2002, Zhu et al., 2003, de Winter et al., 2003, Karadayi et al., 2006]. These approaches apply techniques that are either (i) specific to detecting the cardiac signal *e.g.* centroid algorithm [Karadayi et al., 2006], (ii) based on user interaction [Treece et al., 2002], or (iii) designed for intravascular ultrasound [Zhu et al., 2003, de Winter et al., 2003]. In [Sundar et al., 2009a], a general technique for breathing gating is proposed and applied to ultrasound data. It bases on the phase correlation technique to estimate the motion between successive frames. The breathing phase is estimated from the energy change between consecutive frames. The inherent limitation of the phase correlation algorithm is that it finds the *global translation* in the image plane. Considering that ultrasound images and MR slices are 2D cross sections of the body, the organ motion is not necessarily in-plane, and consequently, there is no uniform global translation. This issue is further discussed in section 9.4. In [von Siebenthal, 2008], the non-rigid registration of navigator slices is applied for image-based breathing gating in MRI.

We already presented an overview of manifold learning approaches and applications in section 6.6.1. Interesting in regard to our approach is the 4D CT reconstruction of the lung [Georg et al., 2008]. Manifold learning is performed on slabs, where a manual inspection is necessary to crop the slabs to contain only lung tissue. Further, Isomap is used to create the low-dimensional embedding. In our work, we deal with the specific challenges of the integration of 4D ultrasound wobbler data and 4D MRI slice stacking, because each 4D imaging technique has its own acquisition protocol. Especially in 4D MRI, there is always a trade-off between acquisition speed, image resolution, and signal-to-noise ratio (SNR) which makes it more challenging than 4D CT. To this end, we focus on Laplacian eigenmaps, since it provided us with better results in comparison to Isomap.

## 9.3 Manifold Learning for Gating

The general idea of manifold learning is to project a manifold in high-dimensional space $\mathbb{R}^{\mathcal{N}}$ to a low-dimensional space $\mathbb{R}^n$, while preserving the local neighborhood, as described in section 6.6.1. For our applications, we consider one dimension of the ambient space for each image pixel, so $\mathcal{N}$ is corresponding to the resolution of the images. For the low-dimensional space, we investigated $n = \{1, 3\}$ as possible dimensions; experiments with $n = 2$ led to similar results than $n = 1$. Considering $k$ images $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ that are acquired over several breathing cycles, the manifold learning $m$ assigns to each image a coordinate in the low-dimensional space $\phi_i$

$$m : \mathbb{R}^{\mathcal{N}} \rightarrow \mathbb{R}^n \tag{9.2}$$

$$\mathbf{u}_i \mapsto \phi_i, \tag{9.3}$$

with $1 \leq i \leq k$. The suggestion that images during free breathing lie on a low-dimensional manifold in the ambient space seems to be justified because variations between neighboring slices are smooth, and further, slices from the same respiratory phase but different acquisition times are similar. Each image in the respiratory cycle corresponds to a point in high-dimensional space. During breathing, we move back and forth along this manifold or trajectory in high-dimensional space. The underlying optimization problem of manifold learning tries to optimally preserve local information, meaning that similar images are mapped to similar positions in the low-dimensional space. With manifold learning, we are therefore able to project the manifold in high-dimensional space, with the images lying on it, to low dimensions. Consequently, it is reasonable to use $\phi_i$ as an estimate for the respiratory phase. Important to notice is that we do not use a parameterization of the expected motion, as it is the case in [Sundar et al., 2009a] with global translation. The proposed method can therefore deal with complex motion patterns, *e.g.* local translation, rotation, deformation, out-of-plane motion, because the low-dimensional embedding is only based on the similarity of images.

### 9.3.1 Laplacian Eigenmaps

We propose the application of Laplacian eigenmaps [Belkin and Niyogi, 2003], for the respiratory phase estimation because the technique is well founded on mathematical concepts (Laplace Beltrami operator) and computationally efficient. Laplacian eigenmaps build upon the construction of a graph, which represents the neighborhood information of the dataset. Subsequently, the graph Laplacian is applied to calculate a low-dimensional representation.

We construct a graph with a node for each point $\mathbf{u}_i$ and with edges connecting neighboring nodes. We select for each image $\mathbf{u}_i$ the $l$ nearest neighbors, by evaluating the norm $||\mathbf{u}_i - \mathbf{u}_\iota||_2^2$. Further, heat kernel-based weights are assigned to the edges with $\mathbf{W}_{i\iota} = e^{-||\mathbf{u}_i - \mathbf{u}_\iota||_2^2/t}$. The similarity measure is important for the neighborhood selection and weighting, where the calculation of the Euclidean norm between the points is equivalent to calculating SSD between the images. We also experiment

with the correlation coefficient as similarity measure, which is up to additive and multiplicative constants equivalent to the calculation of SSD on normalized input images [Viola, 1995]. We therefore only have to normalize the input images to achieve the performance of the correlation coefficient. The normalization is done by subtracting the mean and by dividing by the standard deviation. Once the neighborhood graph is constructed, the eigenvectors of the graph Laplacian provide the embedding map.

### 9.3.2 Global Consistency in 4D US

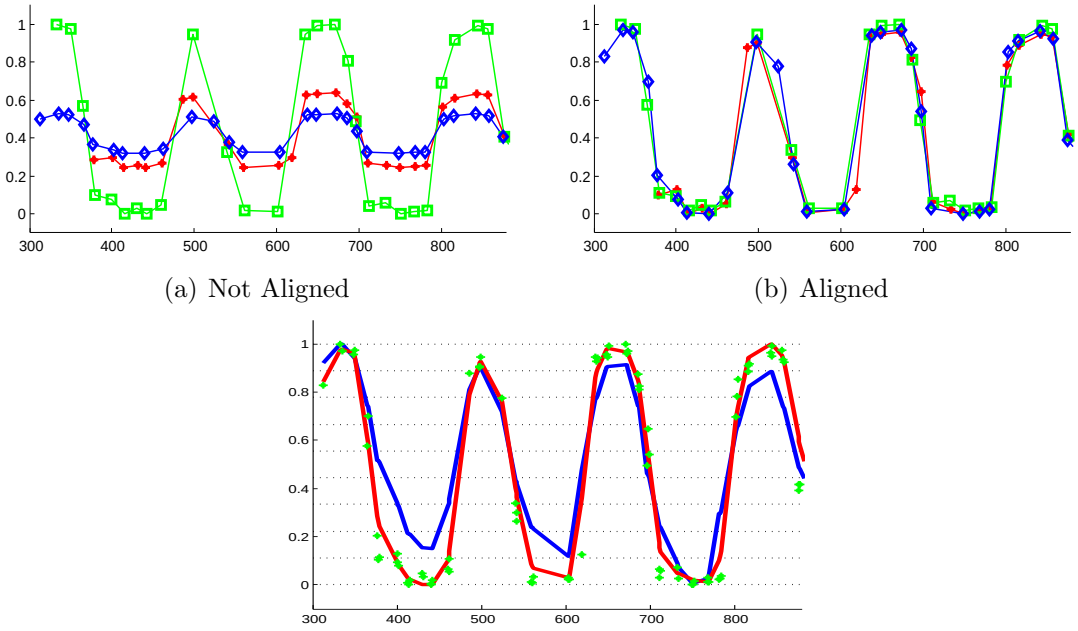
The image acquisition processes for US and MR bear certain similarities, however, the existence of navigator slices for MR facilitates the application. For MR, we only perform the manifold learning on navigator slices, which are acquired at the same position, leading directly to a global respiratory signal. For ultrasound, we do not have those specific slices and therefore have to perform the manifold learning on the data slices. The idea is to estimate a local breathing signal for each deflection angle and, subsequently, calculate the global respiratory signal from the local ones.

Given $\mathcal{U}$ the set of all acquired images, we partition the set in disjunct subsets $\mathcal{U}^1, \ldots, \mathcal{U}^\alpha$, corresponding to the number of different deflection angles $\alpha$ of the wobbler (dash dotted region in figure 9.1). We perform the manifold learning for each of the subsets separately $m^j(\mathbf{u}_i) = \phi_i^j$, with $1 \leq j \leq \alpha$. So depending on the acquisition angle of the ultrasound image $\mathbf{u}_i$, the corresponding manifold learning $m^j$ is performed. Considering all the phases estimated from one angle, we have the local respiratory signals $\Phi^j = \{\phi_1^j, \ldots, \phi_v^j\}$, with $v$ the number of frames per angle. Each local signal contains a consistent estimation of the breathing signal. It is, however, not possible to directly compare local signals, because the 1D projection of the manifold learning can be in an arbitrary range. This is illustrated in figure 9.3(a) with exemplary three local signals corresponding to three angular positions. A simple normalization of each of the local signals $\Phi^j$ is not sufficient because the extreme positions of the breathing cycle may not be reached within them. Consequently, we affinely register local signals in order to retrieve the best scaling $s^j$ and translation $t^j$

$$\Phi^j \mapsto s^j \cdot \Phi^j + t^j. \tag{9.4}$$

Note that this is a 1D affine registration and that scaling and translation are only performed in breathing phase direction (y-axis). We do not have to register in temporal direction, because the acquisition time of the images is provided by the ultrasound system. This is, in fact, a groupwise registration scenario, where we choose to align each pair of neighboring curves with a pairwise registration, starting from the middle one. As cost function serves the distance between curves. The result of the alignment is shown in figure 9.3(b).

The values of the partial signals $\Phi^j$ are now comparable, but may still contain outliers. Consequently, we apply a robust curve fitting to all the sample points to retrieve the global breathing signal. We experimented with various curve models, including Fourier, sum of sine waves, and splines. We achieved best results with




(a) Not Aligned

(b) Aligned

(c) Reconstructed

Figure 9.3: (a) Local breathing signals from manifold learning before alignment. Illustrated is the case for 3 angles (3 colors). X-axis indicates ultrasound frame number. (b) Local breathing signals after alignment. (c) Reconstructed global breathing signal (red) is calculated by robustly fitting a spline curve through the aligned local signals (green crosses). The ground truth signal is shown in blue. Dotted lines indicate the separation of the breathing cycle into several stages. For each stage a volume is compounded.

fitting a spline curve because it allows for the most flexibility, which is important due to irregularity of free breathing. The value of the fitted curve represents the breathing phase of the ultrasound frames, see figure 9.3(c). The proposed method is not limited to analyzing breathing motion, but could also be applied for studying cardiac motion. For the cardiac application, the motion curve was reported to be more regular [Brant and Helms, 2007], so that a Fourier-based curve model may be more appropriate [Ionasec et al., 2010].

In a final step, the breathing cycle is classified into several breathing stages. For each of the breathing stages, the ultrasound frames along the various angles are gathered, and compounded into a final volume, see figures 9.3(c) and 9.4.

## 9.4 Experiments for 4D Ultrasound

For our experiments, we use an ultrasound system from Ultrasonix (Richmond, Canada) and an optical tracking system from A.R.T. (Weilheim, Germany). Both systems are connected to a workstation PC. For the synchronization, we time stamp

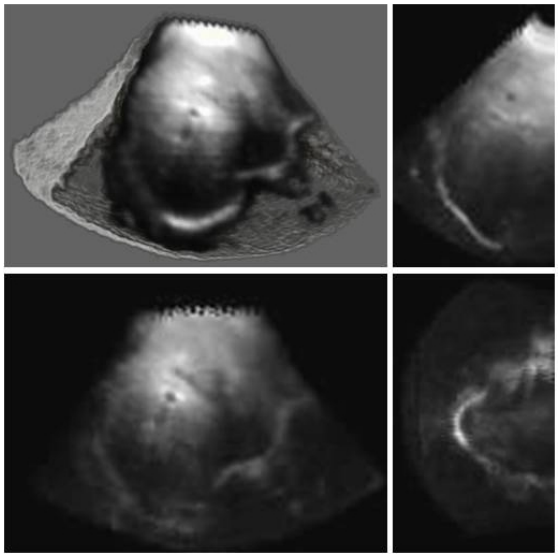


Figure 9.4: 3D rendering of compounded volume and cross-sectional views for one of the breathing stages.

the data on the tracking system and use a network time server to calculate the time offset. For the ultrasound data, we use the direct streaming of B-mode images over the network. We perform tests on multiple patient datasets acquired from different positions, focusing on the liver and kidney.

In order to validate our results, we compare them to the measurements of an external gating system. In [Martinez-Möller et al., 2007], four different gating systems are compared with the best results for an elastic belt and an optical tracking system. We use the optical tracking and place a tracking target, consisting of 7 retro-reflective marker balls, on the chest of the patient. Our setup is more accurate than the one in [Martinez-Möller et al., 2007] for PET tracking, because our field of view is not hindered by the tube, enabling the tracking with four cameras from different views. Since we are only interested in the relative motion of the tracking target, and not in the absolute pose, constant target offsets and calibration errors do not influence the result; leading to a precise monitoring of the respiratory motion. The tracking system provides the pose of the tracking target in 3D space, which consists of 3 translational and 3 rotational components. We apply a principal component analysis of the 6D tracking data to find the principal component along which direction we measure the breathing motion. This is, for instance, done analogously for motion modeling in lung radiotherapy [McClelland et al., 2006]. Further, we low-pass filter the signal to remove cardiac motion and extract the respiratory signal. We refer to the tracked signal as ground truth, which is not completely correct because it contains tracking errors. However, it is the best that can currently be achieved [Martinez-Möller et al., 2007] and is sufficient to validate the performance of our image-based approach.

We compare our approach to the phase correlation technique applied in [Sundar

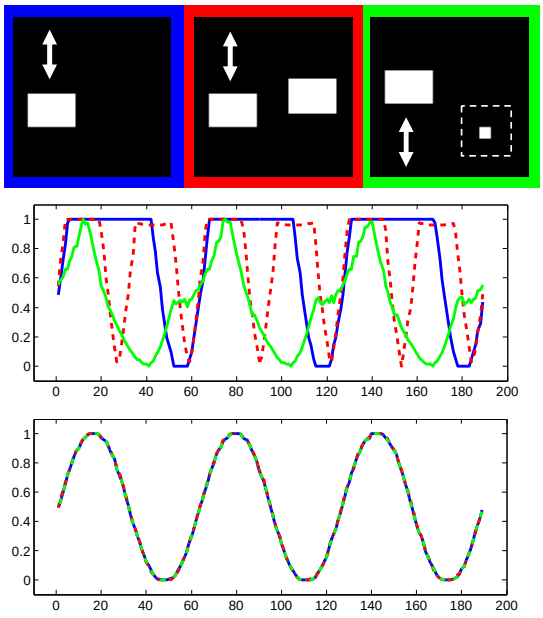



Figure 9.5: Analysis of the gating techniques for synthetic images. Three different motion scenarios are illustrated in the top row. The corresponding gating curves for the phase correlation are shown in the middle row. The last row shows results for manifold learning. Note that in this case the gating results for all three scenarios are identical and corresponding to the surrogate signal.

et al., 2009a] [1]. Unfortunately, with this technique, we do not achieve meaningful results for our datasets. We think that this is due to the limitation of the approach to approximate the 3D motion with a global translation in 2D. In order to illustrate this limitation, we produced synthetic images that show periodic motion. The first scenario consists of a rectangle moving up and down, see figure 9.5. For the second, we add a fixed rectangle, and for the third we add a rectangle that grows and shrinks. We plot the corresponding energy curves of the phase correlation technique. We further show the gating results for our manifold learning approach in figure 9.5, which are identical for all three scenarios and corresponding to the ground truth signal. The result with the phase correlation technique for the first scenario (blue) is approximating the true signal. The addition of a fixed object (red) leads to a signal with double frequency and the addition of the shrinking/growing object (green), leads to a further distortion of the motion signal. Since the results are already not optimal for this easy synthetic case, it is comprehensible that this approach is not best suited for breathing estimation in a noisy ultrasound or MR environment with 3D anatomy moving in and out of plane. Additional results of the phase correlation technique on real data are discussed in the following paragraphs.

The resolution of our ultrasound images is $640 \times 480$ pixels. We downsample the images in each direction by a factor of 2, leading to $\mathcal{N} = \frac{1}{4} \cdot 640 \cdot 480$. This enables a

[1] We want to thank the authors of [Sundar et al., 2009a] for sharing source code.

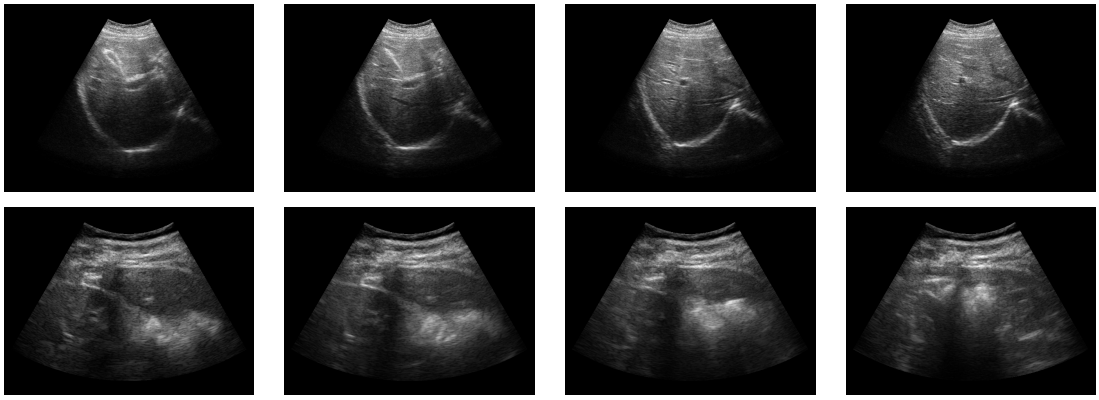


Figure 9.6: 2D Ultrasound images over time from liver (abdomen, right upper quadrant, oblique section) and kidney (left lateral decubitus position, right intercostal flank section).

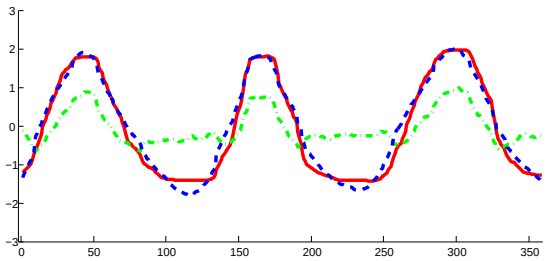



Figure 9.7: Breathing gating results for 2D (red solid: estimated signal, blue dashed: ground truth, green dash-dotted: phase correlation).

faster processing and leads to no noticeable degradation of the manifold learning. We show excerpts of two datasets in figure 9.6. We perform all our experiments with a graph neighborhood of $l = 14$. The number of images for manifold learning varies between 100 and 300, where we did not notice a dependency of the performance on the number of input samples. For the low-dimensional space, we perform experiments with embeddings to 1D and 3D. In the case of ultrasound images, the results are very similar, so that we concentrate on the 1D case.

In figure 9.7, we show the result of the respiratory gating for one of the 2D datasets together with the ground truth signal. For comparison, we also plot the result of the phase correlation technique, which confirms the findings of the synthetic experiments. We calculate the correlation coefficient (CC) between the detected and ground truth signal for multiple 2D datasets, cf. table 9.1. It is remarkable that the ground truth signal is almost perfectly detected. All peaks in the ground truth signal also appear in the detection. Further, the calculation of the correlation, which is in the range of 95%, confirms the visual similarity of the graphs. We also experimented with normalizing the images before passing them to the manifold learning, noticed however no significant improvement.

Table 9.1: 2D+t results

| 2D Correlation Coeff | |
|---|---|
| liver1 | 95.4 % |
| liver2 | 94.4 % |
| liver3 | 93.6 % |
| kidney | 97.3 % |

Table 9.2: 3D+t results

| 3D Correlation Coeff | |
|---|---|
| liver 30° | 94.3 % |
| liver 45° | 95.8 % |
| liver 60° | 96.8 % |
| kidney 45° | 94.4 % |

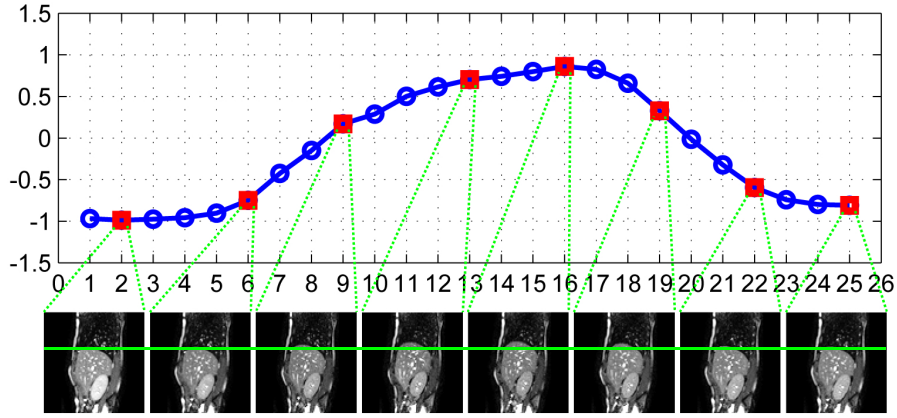



Figure 9.8: Sample slices selected from one breathing cycle. Red-square markers indicate the slice locations in the signal. The auxiliary line assists in observing the liver movement.

For the 4D experiments, we show the result of a fitted curve in figure 9.3. We also calculate the correlation coefficient between the fitted curves and ground truth for four datasets, see table 9.2. We experimented with three different angular ranges, 30°, 45°, and 60° (maximum of probe), for which the probe steers to 15, 21, and 29 different angular positions. We split the breathing signal into 9 different breathing stages, and compound a 3D volume for each of the stages. A volume rendering of one of the volumes is shown in figure 9.4.

We perform statical tests to analyze the significance of our results. Under the null-hypothesis $H_0$ that the correlation between tracking and gating is lower than 93.6%, we obtain a p-value lower than $p < 0.005$. The null-hypothesis can therefore be rejected and the result is considered to be statistically significant.

All image-based approaches rely on ultrasound acquisitions from the same position, because otherwise it is not possible to differentiate between probe motion and breathing motion. To investigate this assumption, we attached a tracking target to the transducer and analyzed its trajectory. This analysis showed only a negligible deviation. The still position therefore does not limit the applicability of our method, which is also confirmed by our good gating results.

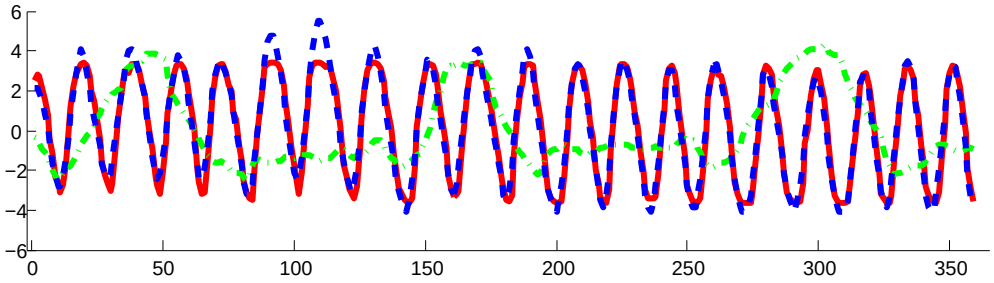



Figure 9.9: Breathing gating results for MRI navigator slices from ETH dataset (red solid: estimated signal, blue dashed: ground truth, green dash-dotted: phase correlation). The correlation between the manifold learning and ground truth is 98%, between phase correlation and ground truth is 2%.

## 9.5 Experiments for 4D MRI

We perform experiments on four different MRI datasets to evaluate the proposed method. Two of them originate from ETH and two from UCL. For each dataset, diaphragm tracking is performed to have a ground truth (GT) signal to compare to [Timinger et al., 2005, Nguyen et al., 2009, King et al., 2009, Rijkhorst et al., 2010]. The general disadvantage of diaphragm tracking is its limitation to certain slice positions and orientations. For our experiments this is fine because we only need a reference signal and we can select the slice with the best visibility of the diaphragm for tracking. A further advantage of the proposed method is that it is applicable to almost any slice orientation and position, as is shown in following. Moreover, diaphragm tracking necessitates the manual placement of a tracking window, while the proposed method is totally automatic.

### 9.5.1 ETH Datasets

The first dataset is the navigator slices acquired at ETH [von Siebenthal et al., 2007], as described in section 9.1.2. These slices have a spatial resolution of $255 \times 255$, a temporal resolution of about 2.7 Hz, and are acquired with a Philips 1.5T Achieva. The sagittal plane was chosen as the imaging plane due to its ease of tracking vascular structure with minimal out-of-plane motion, since the dominant motion of the liver is in the superior-inferior direction [Rohlfing et al., 2004]. A reconstructed breathing cycle with sample MR images is shown in figure 9.8. We further show a comparison between the reconstructed signal of all navigator slices and the ground truth respiratory signal in figure 9.9. The visual similarity of the curves is confirmed by a correlation of 98%. Also illustrated is the signal obtained from the phase correlation technique, leading to a correlation of 2%.

The proposed approach is computationally much more efficient than [von Siebenthal et al., 2007]. In our case, most costly is the calculation of the neighborhood graph with the weights. Assuming $k$ images, SSD has to be calculated $\mathcal{O}(k^2)$ times. In [von

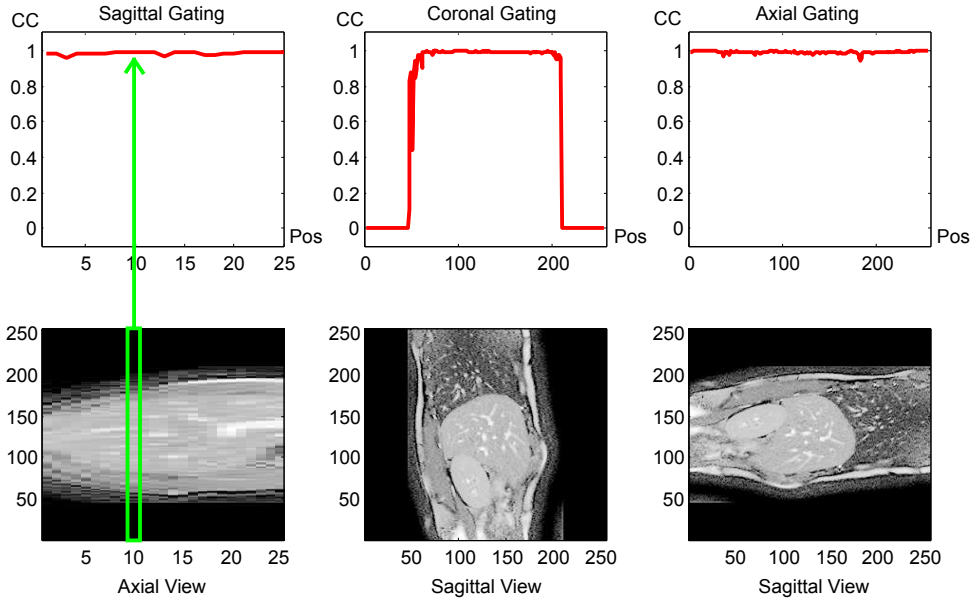



Figure 9.10: The three graphs show the correlation coefficient of the estimated signal with the ground truth for the different orientations of the ETH dataset. For each slice position and orientation, a separate gating is performed, leading to 537 separate gatings in this experiment. The images in the second row are cross-sections, with the x-axis of the graph and the image corresponding. To provide an overview of sagittal slices, for instance, we show an axial view.

Siebenthal et al., 2007], a deformable registration between all image combinations is calculated, leading to $\mathcal{O}(k^2)$ deformable registrations. Since SSD is evaluated in each iteration of the registration, we clearly see the computational benefit of our approach.

The second dataset is the publicly available 4D data from ETH[2], which is created following the slice stacking principle. Each volume consists of $256 \times 256 \times 25$ voxels with a spatial resolution of $1.37 \times 1.37 \times 4\text{mm}^3$. 14 volumes are available for one breathing cycle. We perform manifold learning on 2D slices at all possible positions and orientations, leading to 537 separately estimated respiratory signals with manifold learning in this experiment. This provides further insights whether certain regions or orientations are better suited for gating, which is further discussed in section 9.6. We plot the correlation coefficient with respect to the GT signal for all orientations and positions in figure 9.10. Also in this figure, we show cross-sectional views, to have a better overview to which slice position and orientation a specific result corresponds. We further perform a statistical analysis of these results, which is summarized in table 9.3. We exclude the results of coronal slices showing only background from the statistics. Shown is the mean correlation coefficient, the standard deviation, and the correlation coefficient of the null-hypothesis of the t-test such that we obtain

[2]http://www.vision.ee.ethz.ch/4dmri/

Table 9.3: Statistical analysis of results for gating on MRI. Correlation coefficients are calculated for all orientations and all positions for 3 datasets. Mean CC and standard deviations for each orientation are indicated. We further calculated the CC for the null hypothesis such that the p-value of a one-tailed t-test is $p < 0.005$.

| Datasets | Orientation | # of Pos. | Mean CC | STD CC | CC s.t. ($p < 0.005$) |
|---|---|---|---|---|---|
| ETH | Sagittal | 25 | 98.5 % | 0.90 | 98.0 % |
| | Coronal | 256 | 99.2 % | 0.31 | 99.1 % |
| | Axial | 256 | 99.1 % | 0.66 | 99.0 % |
| UCL | Sagittal | 78 | 97.1 % | 1.37 | 96.7 % |
| | Coronal | 143 | 95.4 % | 2.61 | 94.8 % |
| | Axial | 288 | 94.2 % | 4.46 | 93.5 % |
| UCL (filtered) | Sagittal | 78 | 97.7 % | 0.94 | 97.4 % |
| | Coronal | 143 | 95.7 % | 2.60 | 95.1 % |
| | Axial | 288 | 94.4 % | 3.19 | 93.9 % |
| UCL Mot. Mod. | Sagittal | 160 | 98.9 % | 0.54 | 98.8 % |
| | Coronal | 190 | 98.7 % | 0.98 | 98.5 % |
| | Axial | 200 | 98.5 % | 1.13 | 98.3 % |

significant results ($p < 0.005$). Consider the coronal ETH slices, the hypothesis that the correlation between the ground truth and tracking signal is above 99.1% is statistically significant ($p < 0.005$). The CC in the test is lower for the sagittal direction because of the lower number of slices, which is determining the degrees of freedom of the student's $t$ distribution function. But nevertheless, the CC is in the range of 98.0% to 99.1%.

### 9.5.2 UCL Datasets

The third dataset is from UCL and consists of free-breathing MR scans with a field of view covering the whole abdomen. A balanced-SSFP sequence (TR/TE=4.3/1.46 ms, 30° flip angle) was used to obtain high-resolution 4D dynamic scans during free breathing using a 1.5 T MR scanner (Philips Achieva, Best, The Netherlands) at Guy's Hospital, London. Parallel imaging with a 32-channel coil array using a SENSE acceleration factor of 4.6 resulted in scan times of approximately one second per dynamic volume. Since the largest liver motion normally occurs in the sagittal plane [von Siebenthal et al., 2007, Rijkhorst et al., 2010], the highest reconstruction resolution of $1.4 \times 1.4$ mm was chosen in this plane, resulting in a slice thickness of 4 mm. 25 volumes were acquired over 4 breathing cycles. Once again, manifold learning is performed on 2D slices for all possible positions and orientations. We plot the correlation coefficient with respect to the GT signal for all orientations in figure 9.11. The statistical analysis is summarized in table 9.3. In contrast to the ETH dataset, the UCL dataset contains more noise. This is comprehensible because the ETH dataset is the result of a sophisticated and time consuming slice stacking approach, while the volumes of the UCL dataset are acquired in real-time with a

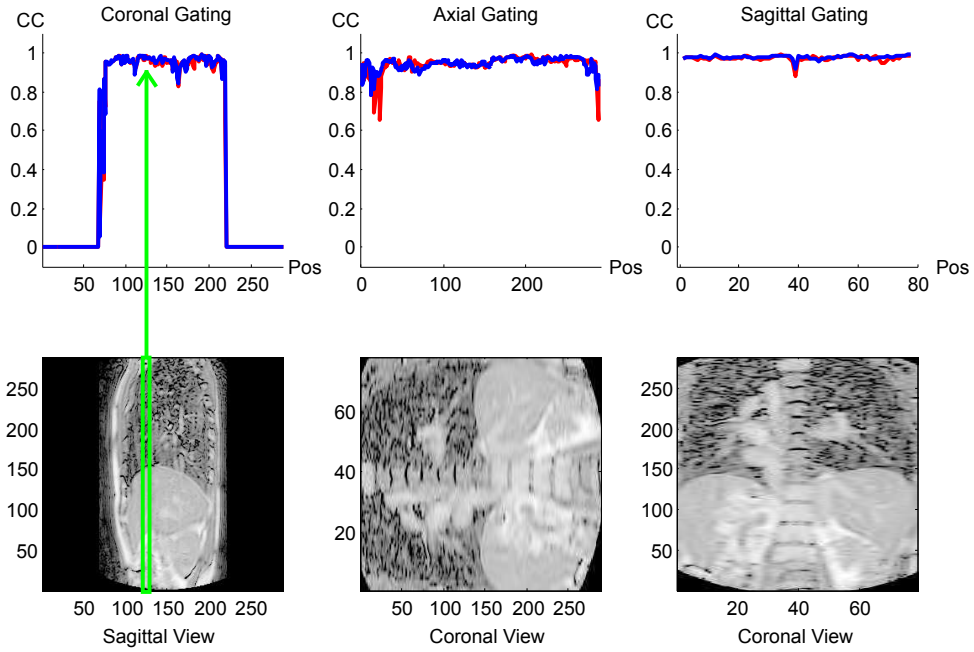



Figure 9.11: The three graphs show the correlation coefficient of the estimated signal with the ground truth for the different orientations of the UCL dataset. The red solid line shows the results of gating on original data, the blue dashed line shows the results on filtered data. For each slice position and orientation, a separate gating with only those slices is performed, leading to 509 separate gatings in this experiment. The images in the second row are cross-sections, with the x-axis of the graph and the image corresponding. To provide an overview of coronal slices, for instance, we show a sagittal view.

fast imaging protocol. This leads to slightly worse gating results. We are, however, able to improve the result by pre-processing the data with median filtering. The results of the gating on the original slices and noise reduced slices are shown in figure 9.11. The results of the statistical analysis are summarized in table 9.3. The mean CC before filtering ranges between 94.2% and 97.1%, while after the filtering it is between 94.4% and 97.7%. The standard deviation is decreased by about 40% in sagittal and axial orientations after filtering.

The fourth 4D dataset is created by using a motion model similar to the one presented in [Rijkhorst et al., 2010]. A set of dynamic 4D MR data was acquired as described above, and tissue displacements throughout the liver were computed by registering each volume within the set to a reference volume using a non-rigid fluid registration method [Crum et al., 2005]. A ground truth surrogate respiratory signal is computed by positioning a sector inside each MR volume, and using a navigator window at the location of the diaphragm. By combining the signal with the registration results, a second order polynomial is fitted at each spatial location, allowing for the creation of motion fields at arbitrary time points. We apply the

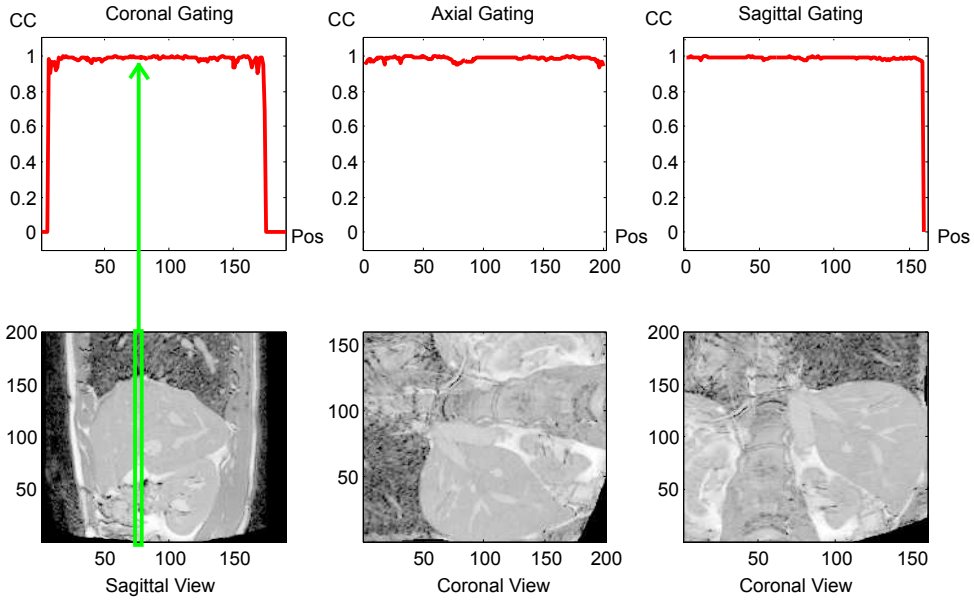



Figure 9.12: The three graphs show the correlation coefficient of the estimated signal with the ground truth for the different orientations of UCL dataset created with the motion model. For each slice position and orientation, a separate gating with only those slices is performed, leading to 550 gatings in this experiment. The images in the second row are cross-sections, with the x-axis of the graph and the image corresponding. To provide an overview of coronal slices, for instance, we show a sagittal view.

motion model to a separately obtained breath-hold 3D MRI to create 4D MRI. Once again, manifold learning is performed on 2D slices at all possible positions and orientations. We plot the correlation coefficient with respect to the GT signal for all orientations in figure 9.12. The last position on sagittal slices is constant, due to boundary effects of the motion model, and does consequently not lead to usable gating results. The statistical analysis is summarized in table 9.3, with the mean CC ranging between 98.5% and 98.9%.

### 9.5.3 Embedding Dimensionality

In this section, we want to further analyze the dimensionality of the embedding space and perform experiments for 3D and 1D. The projection of the ETH navigator slices to 1D space is shown in figure 9.13, and the projection to 3D space in figure 9.15. We arbitrarily select one of the slices as reference slice. We then look for the closest points in 1D and 3D space with respect to the reference image, calling them the best match in 1D and 3D, respectively. As distance measure we use the Euclidean distance. The image corresponding to the 1D best match is shown in the left column of figure 9.14, and the 3D best match is shown in the right column. Considering the auxiliary lines, circles, and arrows, we observe that the breathing state of the reference image

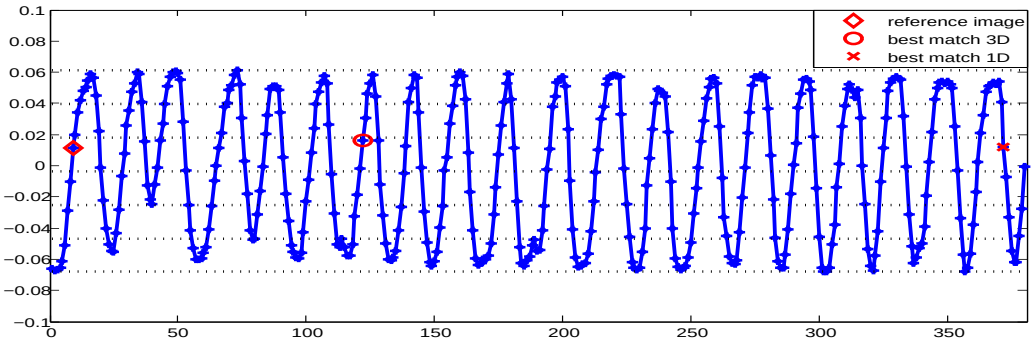



Figure 9.13: Calculated respiratory signal from 1D. X axis shows the slice numbers. Y axis shows corresponding 1D values from manifold learning. We highlight the reference slice and its best matches using 1D and 3D signals.

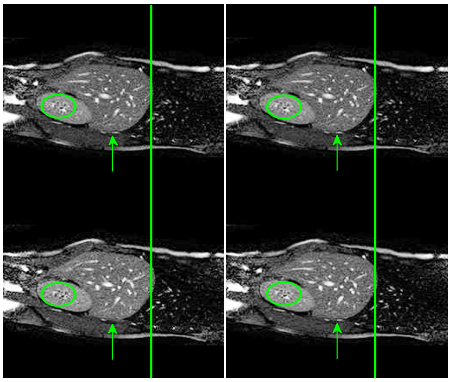


Figure 9.14: Results using 1D and 3D signals. Top row: reference slice, twice. Bottom row left: best match using 1D signal. Bottom row right: best match using 3D signal. Red line, circle and arrow assist comparison.

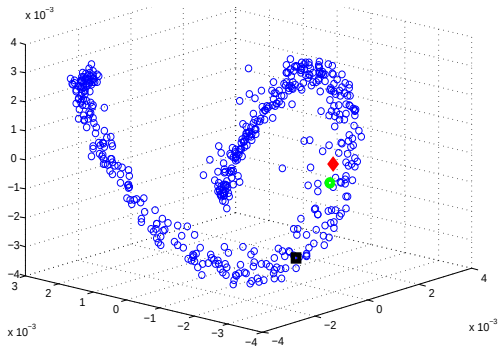
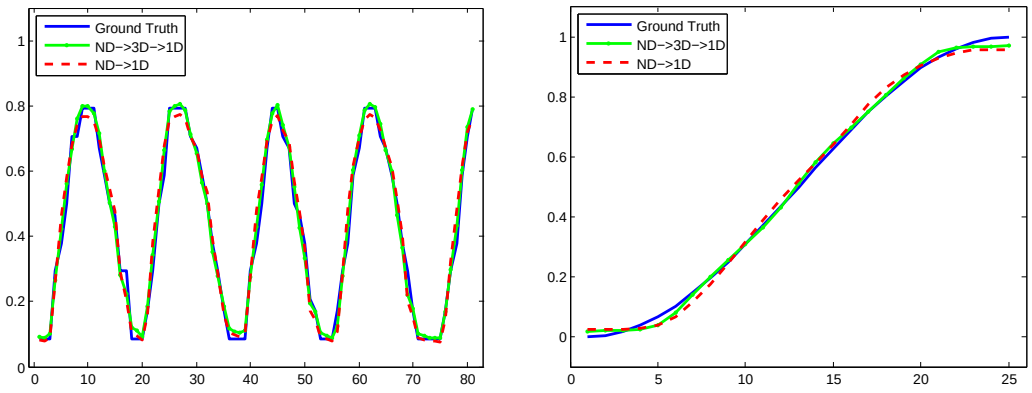


Figure 9.15: Scatter plot of manifold learning results using 3D. Red-diamond: reference slice, green-circle: 3D best match, black-square: 1D best match

and 1D best match is not equivalent. In contrast, the slice corresponding to the 3D best match is almost completely identical to the reference image. Showing that the embedding to 3D has advantages in finding similar slices. In order to get further insights, we also show the best match in 3D on the 1D signal in figure 9.13 and the best match in 1D on the 3D plot in figure 9.15. We observe that the 1D best match is pretty far away from the reference image in the 3D plot.

For some applications, like finding the closest slice, the embedding into 3D is fine. For other purposes, such as visualization, instantiation of a motion model [Rijkhorst et al., 2010], or comparison to a 1D ground truth, a 1D signal is more appropriate. If we want to reduce the 3D signal to 1D, we have to perform another dimensionality reduction. The successive application of multiple manifold learning causes no problems. In figure 9.16(a), we compare the direct embedding to 1D and the combination




(a) ETH navigators (b) UCL

Figure 9.16: Gating experiment performed on the ETH navigator slices and UCL data to evaluate the difference between a direct 1D embedding and a 3D embedding with a successive 1D reduction. For the ETH data, the CC between the direct embedding to 1D and the GT is 98.2% and the CC between the 3D→1D embedding and GT is 98.8%. For the UCL data, the CC between the direct embedding to 1D and the GT is 99.8% and the CC between the 3D→1D embedding and GT is 99.9%.

of first an embedding to 3D and a successive reduction to 1D on navigator slices. We observe that the plots are very similar, which is confirmed by correlations of 98.2% and 98.8%, respectively. We perform the same experiment on the UCL motion model data, with the results shown in figure 9.16(b). The correlations are 99.8% and 99.9%, respectively. These results show that two successive dimensionality reductions do not significantly improve the result and that a direct embedding to 1D is reasonable, if a 1D signal is needed.

## 9.6 Discussion

For ultrasound, we achieved correlations between the proposed method and external gating of around 95%. For the ETH and UCL motion model data, the correlation is in the range of 98%. For the UCL free-breathing MR data it is in the range of 95%, with a slight improvement after filtering. The very low standard deviation is noticeable. This shows that the proposed method is very versatile and leads in almost all scenarios to excellent gating results. For the two UCL datasets, we achieved the lowest performance on axial slices, and the best results on sagittal slices. This is in line with previous observations [von Siebenthal et al., 2007]. The slightly lower correlation for the sagittal planes on the ETH data is a bit surprising and may be due to the large slice thickness.

Our experiments have further shown that the phase correlation technique was not able to extract the respiratory signal from the data. Diaphragm tracking is a valid alternative, does however request user interaction by placing the window and is further limited to views showing the diaphragm. Our results show that the

proposed method can be applied to arbitrary orientations and that even gating on axial planes leads to good results. For ultrasound, diaphragm tracking is even more limited, because it is difficult to have a nice view of the diaphragm during the entire breathing cycle; showing a clear advantage of our method.

The respiratory signal is displaying the current respiratory state of the patient. It is, however, not entirely clear, how this state is defined. One possibility would be to measure the amount of air in the lungs; another possibility would be to measure the displacement of the diaphragm. While there is definitely a high correlation between such quantifications, they are not identical. Since we want to automatically extract the respiratory signal from the data, it is important for our evaluation to relate to a ground truth respiratory signal. Due to the lack of a global consensus, we compared our results to the tracking results from an external tracking system and the tracking of the diaphragm, which are both commonly used procedures. We also want to point out that these respiratory signals are prone to errors during acquisition; however, the error is generally low enough to make the comparison in our scenario meaningful.

The discussion about the definition of the respiratory signal is continued when it comes to the dimensionality. Our experiments show, that we achieve a better discrimination for MR with an embedding to 3D than 1D. The 3D signal over time can definitely not be as nicely visualized as the 1D signal, however, it may as well be considered as an adequate representation. The question whether we first have to extract a 1D signal out of the 3D signal to have a usable respiratory signal cannot be generally answered, but is dependent on the specific application. If the interest lies in finding the closest slice, as it is the case for 4D imaging, this search can directly be performed in 3D space. For an easy visualization and comparison to an alternative gating signal, we may however be interested in a 1D signal. We showed in our experiments that a 1D signal can be obtained from a 3D embedding through the subsequent application of a second manifold learning, the performance is, however, similar to a direct reduction to 1D.

## 9.7 Conclusion

We presented an automatic, image-based respiratory gating method for ultrasound and MR using manifold learning. Moreover, we proposed a solution for acquiring 4D breathing data with a wobbler probe and also acquiring 4D MR with the slice stacking approach. Our method has the advantage that it is fully automatic and does not require a training phase or prior information about the underlying anatomy, nor the interaction of the user. To analyze the performance of our algorithm for ultrasound, we performed experiments on various datasets showing different organs and sections. The results of these experiments were very good, for both, 2D and 3D ultrasound data over time. For MRI, we worked on four different datasets and performed manifold learning on all positions and orientations. A comparison to a tracking-based gating approach is performed, leading to almost similar results and very high correlation. Finally, our approach presents an attractive alternative to external tracking and gating systems with their various setup issues and synchronization problems.

# Chapter 10

# Temporal Groupwise Registration for Motion Modeling

*We present a novel method for the registration of time-resolved image sequences for motion modeling, called Spatio-Temporal grOupwise Registration using free-form deforMations (STORM). It is a* simultaneous *registration method, applying APE as similarity measure and FFD B-splines as transformation model. Furthermore, STORM is a* spatio-temporal *registration method meaning that both, the spatial and the temporal information, are utilized during the registration. This ensures the smoothness and consistency of the resulting deformation fields, which is especially important for motion modeling. Experiments are conducted on synthetic and medical images. Results show the good performance and the robustness of the proposed approach with respect to artifacts, as well as, its ability to correct for larger deformation in comparison to standard pairwise techniques.*

## 10.1 Introduction

In radiation therapy, target localization is one of the biggest challenges, since the exposition of healthy tissue to ionizing radiation should be kept as low as possible [Colgan et al., 2008]. Organ motion due to respiration, however, can lead to inaccuracies during target localization [Flampouri et al., 2006]. Modeling and analyzing organ motion is therefore important for such applications. Thanks to dynamic imaging modalities, which provide time-resolved images of the anatomy, the modeling of organ motion becomes feasible. Time-resolved images are extensively used to study cardiac [King et al., 2010, Klein and Huesman, 2002, Peyrat et al., 2010], lung [Bystrov et al., 2009, Castillo et al., 2010, Flampouri et al., 2006] and liver [Rohlfing et al., 2004, von Siebenthal, 2008] motion.

The creation of motion models necessitates the non-rigid alignment of time-resolved volumes. The standard approach is, first, to perform pairwise registrations, and second, to combine the deformation fields to create the motion model. This can result in inconsistent and non-smooth motion fields along the temporal direction [Sundar et al., 2009b]. In applications, such as radiotherapy, it is, however,

crucial to ensure smoothness and consistency over time to estimate the delivered dose accurately at any time point in the image sequence [McClelland et al., 2006]. Further, the pairwise registration approach has the disadvantage that either, all registrations are performed towards one target leading to large displacements fields, or, registrations between adjacent volumes are calculated causing the accumulation of errors in the image sequence [Castillo et al., 2010].

Methods that address the issue of incorporating temporal information [Bai and Brady, 2009, Ledesma-Carbayo et al., 2005] still have the problem of fixed reference image selection. Although there are spatio-temporal methods that avoid the selection of a reference image in the sequence, either their main focus is to register two different image sequences [Perperidis et al., 2005, Peyrat et al., 2010], or to register two image sequences with the second one being the replication of the reference image [Sundar et al., 2009b]. On the other hand, groupwise registration techniques are proposed that address the choice of the reference image, as discussed in chapter 7. Although, their application areas are quite different from motion modeling, the main goal of these approaches is to reduce the bias introduced by a fixed reference image [Joshi et al., 2004, Crum et al., 2004].

In this chapter, we propose a new registration method for the deformable alignment of time-resolved images. Instead of applying pairwise registrations, we perform a simultaneous registration of all images in the sequence. This eliminates the bias introduced by reference image selection and the error accumulation during the sequential registration. The second novelty of our approach addresses the need for a smooth deformation field. Instead of having a separate deformation field for each image, we create one deformation field that is one dimension higher than the image dimension. Since we work with free-form deformations (FFDs) based on B-splines, this intrinsically ensures that we obtain a smooth deformation field at each update step, also along the temporal direction. This leads to further advantages for the registration method because it is more robust to outliers, can handle large deformations, and allows for temporal interpolation, which differentiates it from existing approaches. From a technical point of view, the method proposed in this chapter is an extension of the mosaicing approach in chapter 7, where we employed APE with a rigid transformation model, to FFDs with an additional temporal dimension.

## 10.2 Related Work

Approaches aiming at the registration of time-resolved image sequences mainly focus on 4D modeling of organ motion of heart, liver or lung. [Klein and Huesman, 2002] use 4D deformable image registration to compensate for the cardiac motion in PET images, following a pairwise approach. [Rohlfing et al., 2004] apply intensity-based deformable registration on gated MR images to model liver motion during a respiratory cycle. They register each volumetric phase image to a reference image by using the displacement field computed for the previous volumetric image as the initial deformation. [Sarrut et al., 2006] simulate a 4D CT image of the thorax by

first registering end-diastole and end-systole images and then linearly interpolate images in between. These methods either do not use temporal information or fix a reference frame during the registration. [Sundar et al., 2009b] address the temporal smoothness problem by considering motion estimation as the registration of two 4D image sequences. The second sequence was, however, created by replicating the end-diastolic image. This means that although this method addresses the problem of temporal smoothness, the existence of the reference image still remains an issue.

Simultaneous registration approaches align groups of images without the necessity of reference selection. We refer to chapter 7 for related work on simultaneous registration. Most related to our approach is the method in [Balci et al., 2007], combining the congealing framework with free-form deformations for the deformable registration of a group of brain images. A drawback of simultaneous registration approaches is that they do not take temporal information into account, since they are not designed for motion estimation. [Bystrov et al., 2009] try to address both aspects, temporal smoothness and reference image selection, but the method is limited to local motion trajectories.

## 10.3 Temporal Groupwise Registration

Assuming $n$ images $I_1, \ldots, I_n$, with $I_i : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^N$, we stack the images and create an image $\mathcal{I}$ of dimension $N+1$

$$\mathcal{I} = [I_1, I_2, \ldots, I_n], \tag{10.1}$$

with $\mathcal{I} : \Omega \times \{1, \ldots, n\} \to \mathbb{R}$. The transformation $\mathbf{T} : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ is defined as

$$\mathbf{T}_{\boldsymbol{\alpha}}(\mathbf{x}) = \mathbf{x} + \mathbf{D}_{\boldsymbol{\alpha}}(\mathbf{x}) \tag{10.2}$$

with the spatial coordinate $\mathbf{x} \in \mathcal{I}$ and $\mathbf{D}_{\boldsymbol{\alpha}}$ the deformation field with deformation parameters $\boldsymbol{\alpha}$. We denote the warped images with $\mathcal{I}^{\downarrow}(\mathbf{x}) = \mathcal{I}(\mathbf{T}(\mathbf{x}))$. In the following sections, we describe the main components of our registration method: transformation model, similarity metric, and optimization method.

### 10.3.1 Transformation Model

We use FFD B-splines to model non-rigid motion [Rueckert et al., 1999], with an additional dimension along the temporal direction. For the temporal direction, we can set B-spline functions of a different order or we can adapt the spacing between the control points to model the desired smoothness along the temporal direction. We derive the formulas for $N = 2$ to make it more comprehensible, leading to $\mathbf{x} = [x, y, t]$. The extension to arbitrary dimensions is straightforward. $\boldsymbol{\alpha}$ is defined as a grid of $n_x \times n_y \times n_t$ control points on image $\mathcal{I}$ with spacings $\delta_x, \delta_y$ and $\delta_t$. Applying cubic B-splines we obtain

$$D_{\boldsymbol{\alpha}}(\mathbf{x}) = \sum_{a=0}^{3} \sum_{b=0}^{3} \sum_{c=0}^{3} B_a(r) B_b(s) B_c(u) \cdot \boldsymbol{\alpha}(\tilde{i} + a, \tilde{j} + b, \tilde{k} + c) \tag{10.3}$$

where

$$\tilde{i} = \left\lfloor \frac{x}{\delta_x} \right\rfloor - 1, \qquad \tilde{j} = \left\lfloor \frac{y}{\delta_y} \right\rfloor - 1, \qquad \tilde{k} = \left\lfloor \frac{t}{\delta_t} \right\rfloor - 1 \tag{10.4}$$

$$r = \frac{x}{\delta_x} - \left\lfloor \frac{x}{\delta_x} \right\rfloor, \qquad s = \frac{y}{\delta_y} - \left\lfloor \frac{y}{\delta_y} \right\rfloor, \qquad u = \frac{t}{\delta_t} - \left\lfloor \frac{t}{\delta_t} \right\rfloor \tag{10.5}$$

and $B_0$ to $B_3$ are cubic B-spline basis functions. For the interpolation of $2D + t$ dimensional data we have

$$\boldsymbol{\alpha}(i,j,k) = \begin{pmatrix} \alpha_x(i,j,k) \\ \alpha_y(i,j,k) \\ \alpha_t(i,j,k) \end{pmatrix}.$$

Consequently, the number of B-spline coefficients is $3 \cdot n_x \cdot n_y \cdot n_t$.

### 10.3.2 Accumulated Pairwise Estimates

We apply accumulated pairwise estimates (cf. chapter 7) as similarity measure, focusing on SSD. In order to describe it in the scope of this temporally extended framework, we use the following notation. Be $\mathbf{v} \in \Omega$ and $g, h \in \{1, \ldots, n\}$. The point $\mathbf{x}$ that is related to the spatial location $\mathbf{v}$ and the temporal location $g$ is denoted by $\mathbf{x}_\mathbf{v}^g = [\mathbf{v}, g]$. Therefore, we can access the same pixel by $I_g(\mathbf{v}) = \mathcal{I}(\mathbf{x}_\mathbf{v}^g)$, leading to the cost function

$$\mathcal{C}(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{g \neq h} ||I_g^\downarrow - I_h^\downarrow||^2 \tag{10.6}$$

$$= \frac{1}{2} \sum_{g \neq h} ||\mathbf{f}_{g,h}(\boldsymbol{\alpha})||^2 = \frac{1}{2} \sum_{g \neq h} \mathbf{f}_{g,h}(\boldsymbol{\alpha})^\top \mathbf{f}_{g,h}(\boldsymbol{\alpha}) = \sum_{g \neq h} F_{g,h}(\boldsymbol{\alpha}) \tag{10.7}$$

$$= \frac{1}{2} \sum_{g \neq h} \sum_{\mathbf{v} \in \Omega} (\mathcal{I}(\mathbf{T}_{\boldsymbol{\alpha}}(\mathbf{x}_\mathbf{v}^g)) - \mathcal{I}(\mathbf{T}_{\boldsymbol{\alpha}}(\mathbf{x}_\mathbf{v}^h)))^2. \tag{10.8}$$

with $\mathbf{f}_{g,h}(\boldsymbol{\alpha}) = I_g^\downarrow - I_h^\downarrow$ the difference image as vector and $F_{g,h}(\boldsymbol{\alpha})$ the similarity value between two images.

### 10.3.3 Optimization

Simultaneous registration leads to a computationally challenging optimization problem because of the increase of the parameter space and the more complex evaluation of the similarity metric, as discussed in chapter 7. This is even more challenging for deformable than for rigid registration. Interestingly, we have fewer parameters to optimize for the $(N+1)$D deformation field, as we would have for $n$ separate $N$D fields. To address the computational complexity, we again use gradient-based optimization procedures, this time in combination with stochastic sampling in the spatial domain [Viola and Wells, 1997, Zöllei et al., 2005, Balci et al., 2007]. The

stochastic sampling significantly reduces the computational cost. It is further noted in [Viola and Wells, 1997] that the noisy estimate of the gradient can reduce the problem of local minima. The update of the parameters is calculated by

$$\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} + \mathbf{u} \tag{10.9}$$

with $\mathbf{u}$ being the current update. For steepest-descent, we have

$$\mathbf{u}^{\mathrm{SD}} = \tau \cdot \frac{\partial \mathcal{C}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \tau \sum_{g \neq h} \frac{\partial F_{g,h}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \tau \sum_{g \neq h} \mathbf{f}_{g,h,}(\boldsymbol{\alpha})^{\top} \cdot \mathbf{J}_{\mathbf{f}_{g,h}}(\boldsymbol{\alpha}) \tag{10.10}$$

with the Jacobian $\mathbf{J}_{\mathbf{f}_{g,h}}$ and the step length $\tau$, which is determined by a line search. For Gauß-Newton, we have to solve the linear system

$$(\mathbf{J}_{\mathbf{f}_{g,h}}^{\top} \mathbf{J}_{\mathbf{f}_{g,h}}) \mathbf{u}_{g,h}^{\mathrm{GN}} = -\mathbf{J}_{\mathbf{f}_{g,h}}^{\top} \mathbf{f}_{g,h}, \tag{10.11}$$

where a QR-decomposition of $\mathbf{J}_{\mathbf{f}_{g,h}} \mathbf{h}_{g,h}^{\mathrm{GN}} = -\mathbf{f}_{g,h}$ leads to a stable and efficient solution. The global update is calculated with $\mathbf{u}^{\mathrm{GN}} = \sum_{g,h} \mathbf{u}_{g,h}^{\mathrm{GN}}$. See chapter 7 for more details on optimization techniques.

#### 10.3.3.1 Point-Wise Derivatives

Considering the moving image $h$, the partial derivative with respect to control point $\alpha^{i,j,k}$ along direction $x$, with other directions analogously, is

$$\frac{\partial \mathcal{C}(\boldsymbol{\alpha})}{\partial \alpha_x^{i,j,k}} = -\sum_{g \neq h} \sum_{\mathbf{v} \in \Omega} (I_g^{\downarrow}(\mathbf{v}) - I_h^{\downarrow}(\mathbf{v})) \cdot \nabla I_h^{\downarrow}(\mathbf{v}) \cdot \begin{pmatrix} B_{d^h}(r^h) B_{e^h}(s^h) B_{f^h}(u^h) \\ 0 \\ 0 \end{pmatrix} \tag{10.12}$$

with $\nabla I_h^{\downarrow}(\mathbf{v}) = (\nabla_x I_h^{\downarrow}(\mathbf{v}), \nabla_y I_h^{\downarrow}(\mathbf{v}), \nabla_t I_h^{\downarrow}(\mathbf{v}))$ and $r^h$, $s^h$, $u^h$ as defined in equation (10.5) with respect to the location $\mathbf{x}_{\mathbf{v}}^h$ and

$$d^h = i - \left\lfloor \frac{x}{\delta_x} \right\rfloor + 1, \quad e^h = j - \left\lfloor \frac{y}{\delta_y} \right\rfloor + 1, \quad f^h = k - \left\lfloor \frac{h}{\delta_t} \right\rfloor + 1. \tag{10.13}$$

We consider $B_{d^h}(r^h) = 0$ for $d^h < 0$ and $d^h > 3$, analogously for $B_{e^h}(s^h)$ and $B_{f^h}(u^h)$. The derivative is unequal zero only in a neighborhood of the control point. We use the shorter notation $B_{\mathbf{v},h}^{i,j,k} := B_{d^h}(r^h) B_{e^h}(s^h) B_{f^h}(u^h)$ to indicate the dependency of the B-spline product on the spatial coordinate $[\mathbf{v}, h]$ and on the grid location $i, j, k$.

Combining the derivatives along all directions leads to

$$\frac{\partial \mathcal{C}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{i,j,k}} = \sum_{g \neq h} \sum_{\mathbf{v} \in \Omega} (I_g^{\downarrow}(\mathbf{v}) - I_h^{\downarrow}(\mathbf{v})) \cdot \begin{pmatrix} \nabla_x I_h^{\downarrow}(\mathbf{v}) \\ \nabla_y I^{\downarrow}(\mathbf{v}) \\ \nabla_t I_h^{\downarrow}(\mathbf{v}) \end{pmatrix}^{\top} \cdot \begin{bmatrix} B_{\mathbf{v},h}^{i,j,k} & 0 & 0 \\ 0 & B_{\mathbf{v},h}^{i,j,k} & 0 \\ 0 & 0 & B_{\mathbf{v},h}^{i,j,k} \end{bmatrix} \tag{10.14}$$

$$= \sum_{g \neq h} \sum_{\mathbf{v} \in \Omega} (I_g^{\downarrow}(\mathbf{v}) - I_h^{\downarrow}(\mathbf{v})) \cdot \nabla I_h^{\downarrow}(\mathbf{v}) \cdot \mathbf{B}_{\mathbf{v},h}^{i,j,k}. \tag{10.15}$$

#### 10.3.3.2 Global Derivatives

With the insights and notation introduced in the point-wise case, we can now describe the global update with respect to all parameters $\boldsymbol{\alpha}$. The Jacobian $\mathbf{J}_{\mathbf{f}_{g,h}}$ with $M = |\Omega|$ is

$$\mathbf{J}_{\mathbf{f}_{g,h}}(\boldsymbol{\alpha}) = \begin{bmatrix} \frac{\partial I_g^{\downarrow}(\mathbf{v}_1) - I_h^{\downarrow}(\mathbf{v}_1)}{\partial \boldsymbol{\alpha}^{1,1,1}} & \cdots & \frac{\partial I_g^{\downarrow}(\mathbf{v}_1) - I_h^{\downarrow}(\mathbf{v}_1)}{\partial \boldsymbol{\alpha}^{n_x,n_y,n_z}} \\ \vdots & \ddots & \vdots \\ \frac{\partial I_g^{\downarrow}(\mathbf{v}_M) - I_h^{\downarrow}(\mathbf{v}_M)}{\partial \boldsymbol{\alpha}^{1,1,1}} & \cdots & \frac{\partial I_g^{\downarrow}(\mathbf{v}_M) - I_h^{\downarrow}(\mathbf{v}_M)}{\partial \boldsymbol{\alpha}^{n_x,n_y,n_z}} \end{bmatrix} \tag{10.16}$$

$$= - \begin{bmatrix} \nabla I_h^{\downarrow}(\mathbf{v}_1) \cdot \mathbf{B}_{\mathbf{v}_1,h}^{1,1,1} & \dots & \nabla I_h^{\downarrow}(\mathbf{v}_1) \cdot \mathbf{B}_{\mathbf{v}_1,h}^{n_x,n_y,n_z} \\ \vdots & \ddots & \vdots \\ \nabla I_h^{\downarrow}(\mathbf{v}_M) \cdot \mathbf{B}_{\mathbf{v}_M,h}^{1,1,1} & \dots & \nabla I_h^{\downarrow}(\mathbf{v}_M) \cdot \mathbf{B}_{\mathbf{v}_M,h}^{n_x,n_y,n_z} \end{bmatrix} \tag{10.17}$$

$$= - \begin{bmatrix} \nabla I_h^{\downarrow}(\mathbf{v}_1) & & \\ & \ddots & \\ & & \nabla I_h^{\downarrow}(\mathbf{v}_M) \end{bmatrix} \begin{bmatrix} \mathbf{B}_{\mathbf{v}_1,h}^{1,1,1} & \dots & \mathbf{B}_{\mathbf{v}_1,h}^{n_x,n_y,n_z} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{\mathbf{v}_M,h}^{1,1,1} & \dots & \mathbf{B}_{\mathbf{v}_M,h}^{n_x,n_y,n_z} \end{bmatrix} \tag{10.18}$$

$$= -\nabla I_h^{\downarrow} \cdot \mathbf{B}_h \tag{10.19}$$

where the matrix $\mathbf{B}_h$ contains the B-spline weights for all spatial locations and grid locations with respect to image $h$. This leads to the global update for steepest-descent

$$\mathbf{u}^{\mathrm{SD}} = \frac{\partial \mathcal{C}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = - \sum_{g \neq h} \mathbf{f}_{g,h,}(\boldsymbol{\alpha})^{\top} \cdot \nabla I_h^{\downarrow} \cdot \mathbf{B}_h \tag{10.20}$$

and for Gauß-Newton we solve the overdetermined system

$$-\nabla I_h^{\downarrow} \cdot \mathbf{B}_h \cdot \mathbf{u}_{g,h} = (\nabla I_h^{\downarrow} \cdot \mathbf{B}_h)^{\top} \mathbf{f}_{g,h} \tag{10.21}$$

with a QR-decomposition. Note that the Jacobian is very sparse and therefore sparse QR-solvers significantly improve the performance. The dimensionalities of the involved entities are

$$\mathbf{f}_{g,h} \in \mathbb{R}^{M \times 1} \qquad \nabla I_h^{\downarrow} \in \mathbb{R}^{M \times 3M} \tag{10.22}$$

$$\mathbf{B}_h \in \mathbb{R}^{3 \cdot M \times 3 \cdot n_x \cdot n_y \cdot n_t} \qquad \mathbf{J}_{g,h} \in \mathbb{R}^{M \times 3 \cdot n_x \cdot n_y \cdot n_t}. \tag{10.23}$$

### 10.3.4 Implementation Details

We position the images along the temporal direction, reflecting its acquisition time. Alternatively, one could decide to use different criteria to position the images along the temporal direction, such as amount of change. For some applications it may further be useful to allow for deformations along the temporal direction, if there is uncertainty about the exact acquisition time, or different parts of the image were acquired at different time points. For our application, we do not need this flexibility so that we set the update along the temporal direction to zero.

We apply a multi-resolution approach to reduce computational cost and to increase robustness to local minima. Further, we apply a multi-grid approach, with several grid refinements on each image resolution, to capture large displacements, as well as, fine details. In order to guarantee smoothness and consistency over time, we use a a single control point grid of one higher dimension and overlay this onto the $(N+1)$D image. For the temporal dimension, we use fewer control points than the number of images to increase the dependency between the images and, at the same time, to further decrease the computational cost. The software is developed using open-source libraries ITK and VTK.

### 10.3.5 Advantages of STORM for Motion Modeling

After having described the method, we want to point out the advantages of STORM for motion modeling.

(A) **Smooth and Temporally Consistent Deformation Field:** Current approaches perform first the registration of the images, and then the combination and regularization of the transformation fields. The problem is that for the regularization step, the image information is no longer considered, and therefore transformation fields may be created that do not match to the underlying images. In contrast, our single B-spline deformation field across all the images intrinsically guarantees a smooth and consistent transformation along the temporal direction in each update step. Similarity and regularization aspects are jointly optimized.

(B) **Robustness Against Outliers:** Dynamic imaging is more challenging than its static counterpart, causing more artifacts and noise in the images. Our method is robust to such effects because the $(N+1)$D deformation field interlaces neighboring images, with the number of control points determining to which extent. Images containing outliers can therefore still be correctly aligned, if neighboring images drive the registration.

(C) **Handling Large Displacements**: During the registration of time-resolved images, one of the main issues is the registration of distant images. If the displacement between the images is too large, pairwise methods fail to find the mapping. By incorporation temporal information and using groupwise registration, STORM is able to better handle large displacements in the sequence.

(D) **Time Interpolation**: In some applications, it is desirable to see the organ at a specific state of the motion [Ehrhardt et al., 2006, Schreibmann et al., 2006]. However, due to the low temporal resolution of time-resolved images, this is not always possible. Thanks to the higher dimensional deformation field of STORM, it is feasible to resample images for any time point by applying the inverse mapping.

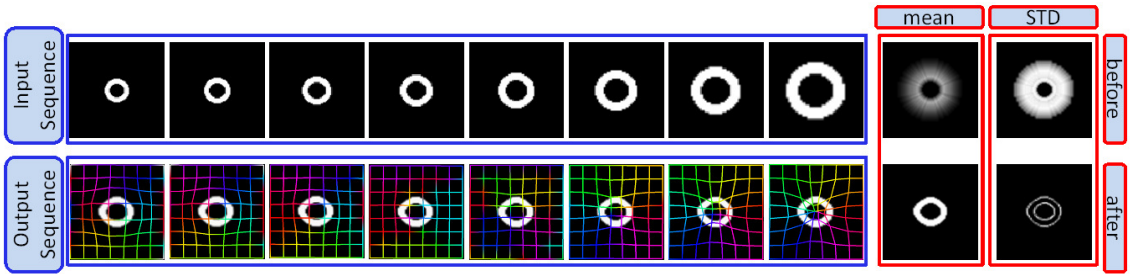
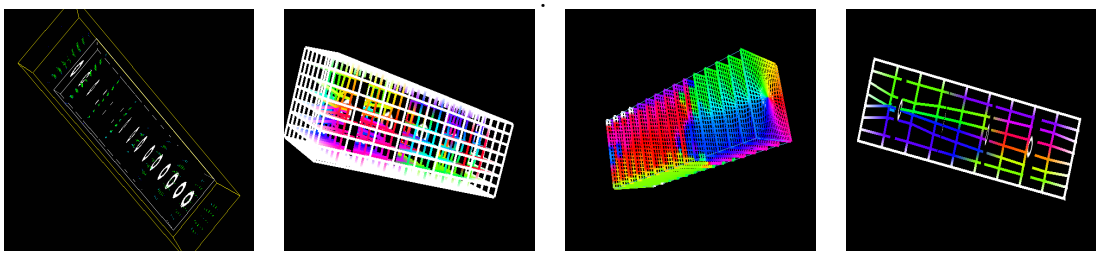



Figure 10.1: Input and output sequences together with the statistics images for growing ring experiment.



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

Figure 10.2: (a) $(N+1)$D image. (b) $(N+1)$D wireframe mesh. (c) $N$D meshes in $(N+1)$D. (d) A cut-plane from the $(N+1)$D mesh along the temporal direction.

(E) **Tracking using deformation field**: In 4D radiation therapy, it is necessary to extract motion trajectories of a tumor region, in order to perform the pre-operative dosimetry planning [Flampouri et al., 2006]. Our approach enables tracking of certain structures by clipping the $(N+1)$D deformation field along the temporal dimension.

## 10.4 Experiments and Results

In order to evaluate the performance of STORM, we experiment with temporally-related synthetic and medical data sets. In addition to using visual inspection for qualitative evaluation, mean and standard deviation (STD) along the temporal direction are computed for the quantitative validation. For clarity of presentation, every second image is shown for each sequence.

### 10.4.1 Synthetic Data

Synthetic image sequences are created by applying gradually increasing deformations to a template image to simulate a smooth deformation over time, reflecting a smooth motion. The experiments are conducted to demonstrate the advantages listed in section 10.3.5. For each experiment a multi-resolution setting with 30 iterations on the highest level that is increased by a factor of 1.2 for each lower level is used

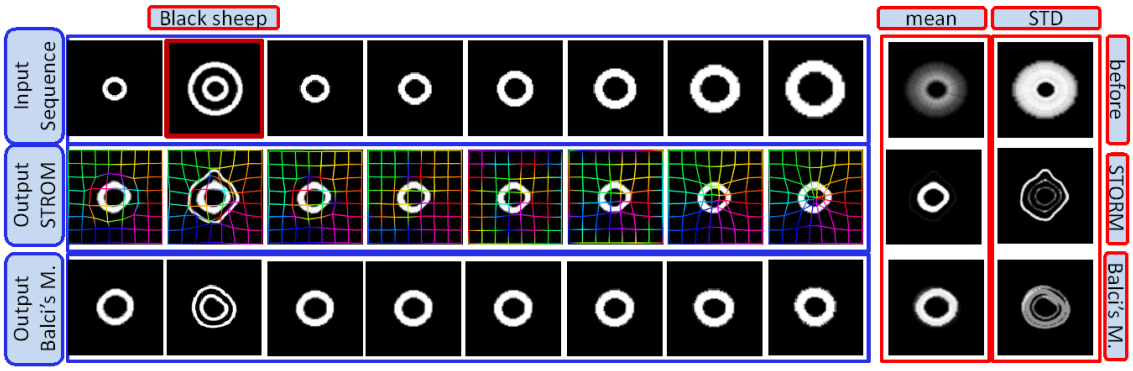



Figure 10.3: Input and output sequences together with the statistics images for the black sheep experiment using our method and Balci's method.

together with 8 control points in spatial dimensions and 7 control points in the temporal dimension.

## Growing Ring

The input images shown in Fig. 10.1 are binary images of a continuously growing ring of size $100 \times 100$ pixels. The registered images are also shown in the same figure. We observe that all images are correctly aligned to the initially unknown target. The wireframe meshes are presented to visualize the deformations. The direction of the displacements is encoded in the meshes with the hue component. Mean and STD images, calculated along the temporal direction, confirm the correct registration. The mean image clearly shows the ring, and the STD image only shows small deviations on the edges of the ring caused by interpolation artifacts. The $(N+1)$D image and the control point grid are shown in Figs. 10.2(a) and 10.2(b), respectively. The visualization of the $N$D meshes in $(N+1)$D is shown in Fig. 10.2(c). The deformation along the temporal direction can be observed by clipping the $(N+1)$D deformed mesh along the time-axis as shown in Fig. 10.2(d). The smoothness and consistency of the deformation along the temporal dimension in this figure, as well as, the smoothness of the change of colors in Fig. 10.2(c) demonstrates the advantages (A) and (E).

## Black Sheep

In this experiment, the same growing ring sequence is used with an outlier image introduced to create a 'black sheep' in the sequence. The input images are shown in Fig. 10.3. The registration result that we expect, having the smoothness of motion in mind, is the alignment of the inner ring with the others, while treating the outer ring as a separate structure. We compare our method to the groupwise registration method proposed in [Balci et al., 2007]. The results for the two methods are shown in Fig. 10.3. Looking at the final alignment in Balci's case, it is clear that it does not fulfill the expectations since the outer ring is aligned with the others instead of the inner, real one. In our case, however, the outer ring is identified correctly and

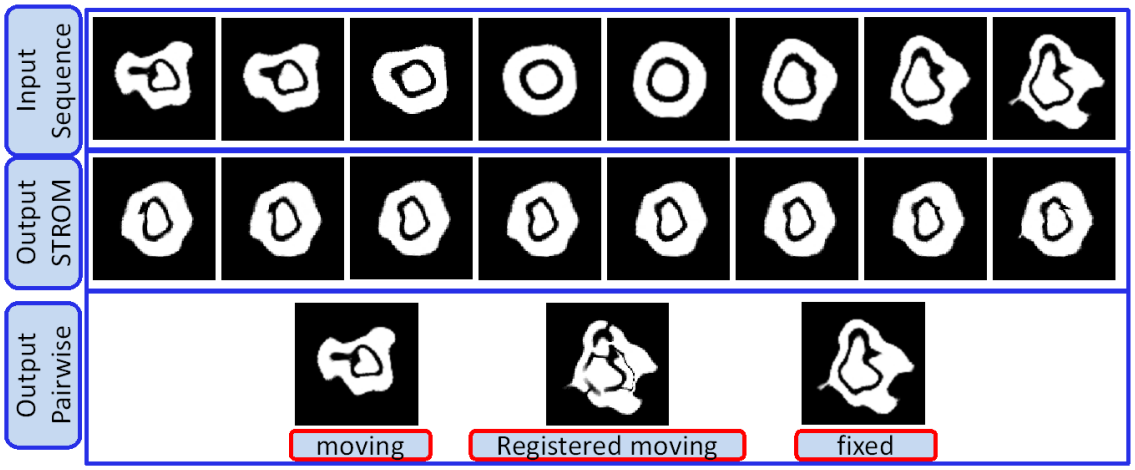



Figure 10.4: Input and output sequences together with the statistics images for first vs. last experiment. Last row: Results for experiment with pairwise registration method.

excluded from the final alignment as expected. These findings are also supported by the STD and mean images shown in the same figure. The reason for the failure of Balci's method and the success of our method is the consideration of the temporal relationship during the registration, making the method more robust against outliers, as noted in (B).

## First vs. Last

In this case, STORM is tested on a data set with large displacements between the first and last image to model large motion deformations. A randomly created B-spline deformation field is asymmetrically applied to a template image of size $100 \times 100$ pixels. The goal of this experiment is to demonstrate advantage (C) and the insufficiency of pairwise registration in such cases. In Fig. 10.4, the input sequence is shown where the middle image is used for creating the sequence. The results in the same figure show that the images are aligned up to some small structures which are difficult for many registration algorithms using FFDs. For the pairwise registration, after trying different implementations of demons algorithm available in ITK, we decided to use the diffeomorphic demons proposed in [Vercauteren et al., 2008a], since it achieved the best performance. After registering the first image (moving) to the last image (fixed) in the sequence, we show the deformed moving image in Fig. 10.4. The dissimilarity between the registration result and the fixed image shows that the large deformation is difficult to handle by pairwise methods.

## Time Interpolation

In this experiment, we want to illustrate advantage (D), the interpolation of images along the temporal dimension. For this we use the sequence in Fig. 10.5, where

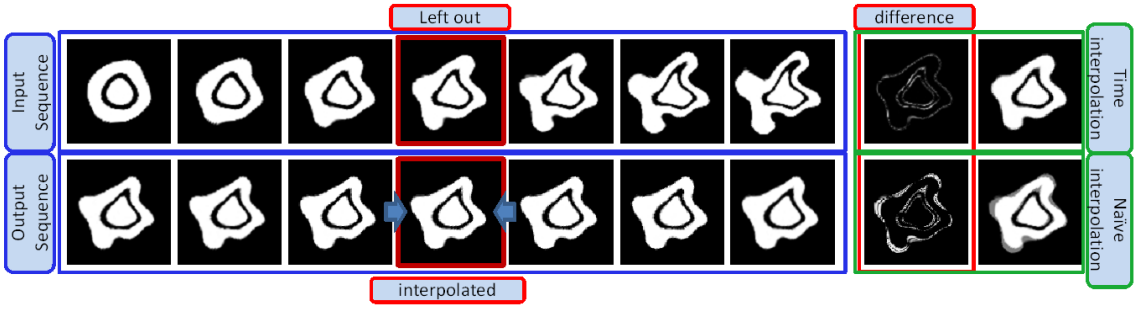



Figure 10.5: Illustration of interpolation in time using STORM. Time interpolation and naive interpolation are compared by showing the difference images to the original image.

we omit the indicated image and try to reconstruct it. First, the registration of the reduced sequence is performed. Second, the temporally adjacent images to the reconstructed time point are selected. Third, a linear intensity interpolation between the neighboring, deformed images is performed - with more emphasis to the temporally closer image. Fourth, the inverse transformation is applied to the interpolated image [Rueckert et al., 2006], creating the reconstruction. For validation, we compare the reconstructed image to the original image in the sequence by calculating the difference image. Further, we perform a naïve reconstruction by intensity interpolation of the input images. The results clearly show the good performance of the proposed time interpolation.

### 10.4.2 Medical Data

First, a time-resolved image sequence consisting of 10 MR images is used. The images are so called "navigator slices" that are used for 4D MRI reconstruction in [von Siebenthal, 2008]. The images of size $255 \times 255$ pixels and temporal resolution 2.7 Hz are acquired from the liver using the sagittal plane with Philips 1.5T Achieva during one respiratory cycle. The second experiment is done with a CT sequence extracted from the POPI-model [Vandemeulebroucke et al., 2007] which consists of 10 volumes corresponding to phases of a respiratory cycle[1]. The images are of size $482 \times 141$ pixels and correspond to the sagittal planes in the posterior thoracic region. A multi-level control point grid and image pyramids are used with 3 resolution levels for images and 2 levels for control point grids. For each level of the control point grid, registration is performed at each level of the image pyramid. 20 iterations are used for the finest level of the control point grid and image pyramid. This number is increased by a factor of 1.2 for each image level and doubled for each control point grid level.

[1]The data was obtained from the Léon Bérard Cancer Center & CREATIS lab, Lyon, France. http://www.creatis.insa-lyon.fr/rio/popi-model

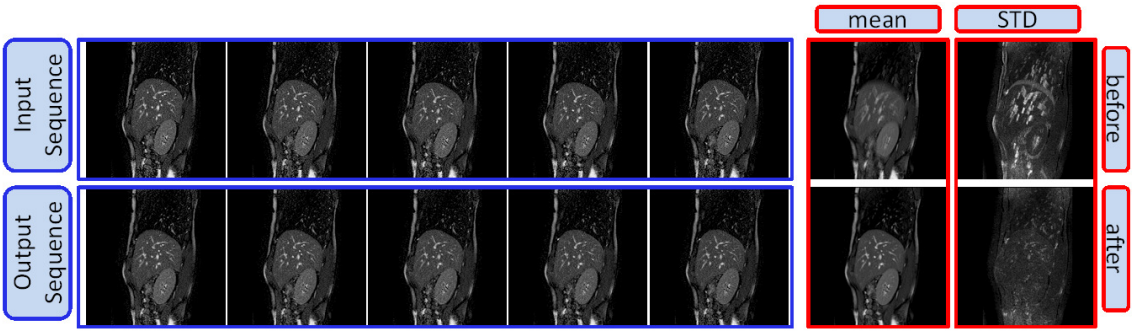
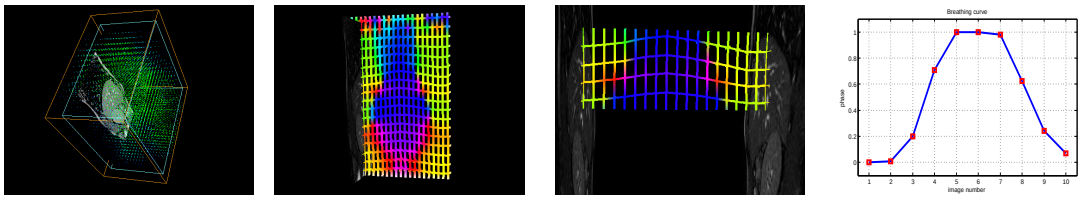



Figure 10.6: Experiment with MRI sequence. Input, output sequences and mean, STD images before and after the registration.




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

Figure 10.7: (a) $(N+1)$D deformation field. (b) A temporal slice from $(N+1)$D mesh. (c) A closer look into the mesh in (b). (d) The breathing curve for the data set. Note the similarity between the curves in (c) and the curve in (d). The color coded deformation on the mesh shows the direction of motion along the time-axis. Individual lines can be seen as the local motion paths.

## MRI Sequence

An excerpt from the 10 images in the sequence is shown in Fig. 10.6. The first and the last images in the sequence correspond to the exhale state of the respiration. We perform registration on the sequence using STORM, with the results being presented in the same figure. The visual inspection and also the calculated statistics images show the good performance of our method for aligning the sequence, and therefore its applicability for registering real medical images. We present the $(N+1)$D deformation field and a slice along the temporal dimension from the $(N+1)$D mesh in Figs. 10.7(a) and 10.7(b). Furthermore, we recover the breathing curve for this sequence, see figure 10.7(d), using the image-based gating technique (cf. chapter 9). The similarity between the breathing curve and the deformation field in temporal direction in Fig. 10.7(c) suggests the correspondence between the motion track recovered by both methods.

## CT Sequence

We perform registration using 10 images. The results are presented in Fig. 10.8. The mean and the STD images in the same figure show the very good performance of STORM on the thoracic images. The high STD around the upper edge of the

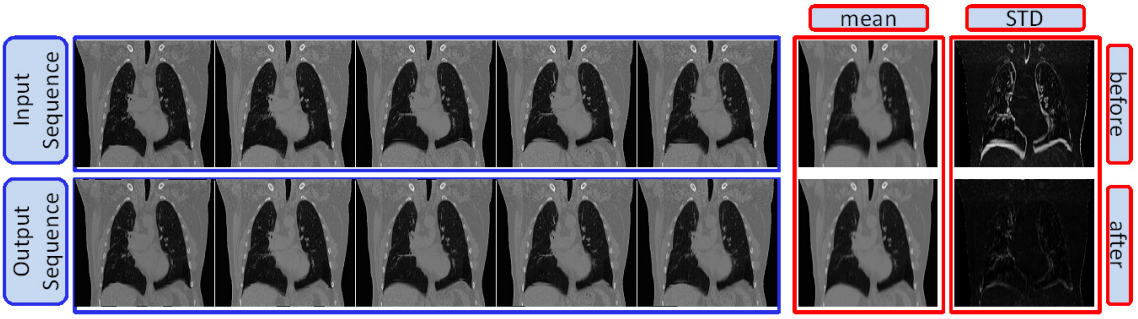
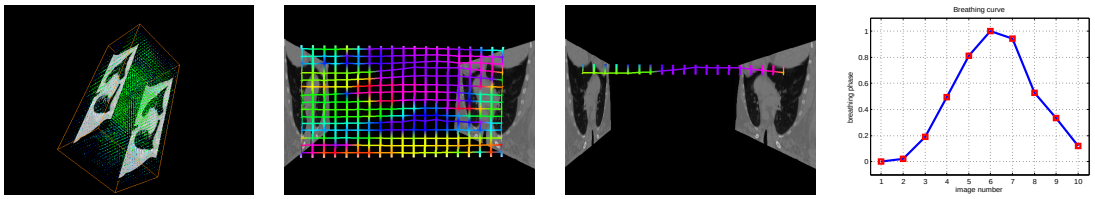



Figure 10.8: Experiment with CT sequence. Input, output sequences and mean, STD images before and after the registration.




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

Figure 10.9: (a) $(N+1)$D deformation field. (b) A temporal slice from $(N+1)$D mesh. (c) A closer look into the mesh in (b). (d) The breathing curve for the data set. Note the similarity between the curves in (c) and the curve in (d). The color coded deformation on the mesh shows the direction of motion along the time-axis. Individual lines can be seen as the local motion paths.

liver is due to the image construction artifacts, which is quite visible in the input images. We again present images of the $(N+1)$D deformation field and the mesh. Furthermore, as done for the MRI sequence, we recover the breathing curve, see Fig. 10.9(d). The breathing signal from manifold learning is once again similar to the recovered motion trajectories of STORM.

## 10.5 Motion Modeling in Ultrasound

Ultrasound is in comparison to CT and MR rarely used for motion modeling. Considering the degradation of ultrasound images by speckle noise and artifacts, this is comprehensible. Further, standard 2D+t acquisitions are not sufficient to account for complex 3D motion [Ledesma-Carbayo et al., 2005]. With the method presented in chapter 9, we are, however, able to construct 3D+t US data, which is well suited for motion modeling. Moreover, the typical advantages of ultrasound (non ionizing, portable, cheap) make it also an interesting alternative for motion modeling, if the challenging task of motion estimation could be reliably performed. STORM seems to be well suited for this task, because it was designed for delivering robust registration results. This is, on the one hand, due to the application of multivariate similarity measures, which proved advantages for ultrasound registration (cf. chapter 7). And,

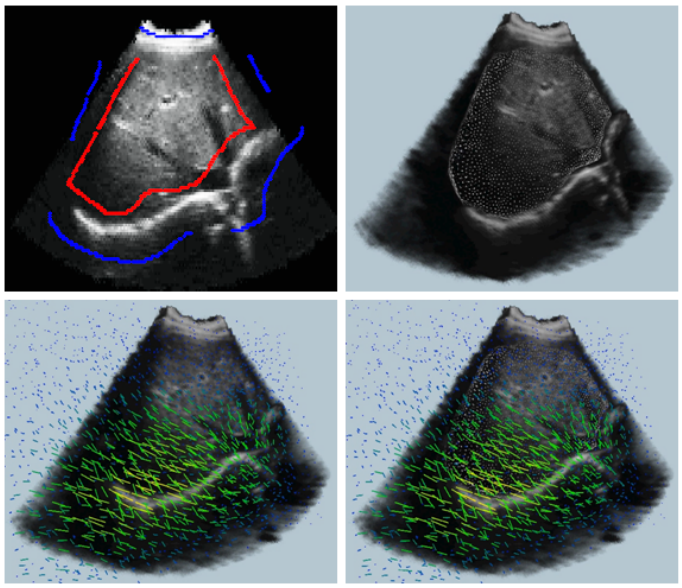


Figure 10.10: Top left: seed points for random walks segmentation. Top right: volume rendering with segmented liver mesh. Bottom left: volume rendering with deformation field. Bottom right: volume rendering with deformation field and mesh. For the illustrations the data of the first experiment is used (cf. figure 10.11).

on the other hand, due to the imposed temporal constraints across images, which potentially limit the influence of reverberation, mirror, and shadow artifacts (cf. advantage (B)). An interesting application for US motion modeling may come up in combination with high-intensity focused ultrasound (HIFU) [Kennedy et al., 2003] to more accurately heat and destroy pathogenic tissue. Currently MR guidance is performed for targeting [Tempany et al., 2003].

In our study, we concentrate on modeling the motion of the liver, for which we have not found any previous references. Existing approaches for US motion estimation mainly focus on cardiac and breast applications [Boukerroui et al., 2003, Ledesma-Carbayo et al., 2005, Elen et al., 2008, Esther Leung et al., 2008, Myronenko et al., 2009]. Our process for motion modeling is as follows. First, we acquire a sequence of ultrasound frames with the wobbler transducer during free breathing. From them we create consistent 4D ultrasound data by applying the presented image-based gating approach. The advantage is that we can directly save the data on the ultrasound system without the need of synchronization to an external gating system. Next, a groupwise deformable registration of the volumes is performed with STORM. The resulting deformation field can directly be used for analyzing the motion. We additionally apply the deformation to a segmentation of one of the volumes, which enables to propagate the segmentation over time and therefore to achieve a 4D segmentation. For the initial segmentation, we use random walks [Grady, 2006], where a manual placement of seed points in two slices of the volumes proved to be sufficient to attain good segmentation results (cf. figure 10.10).

We evaluate the motion modeling on several 4D ultrasound data sets, where we

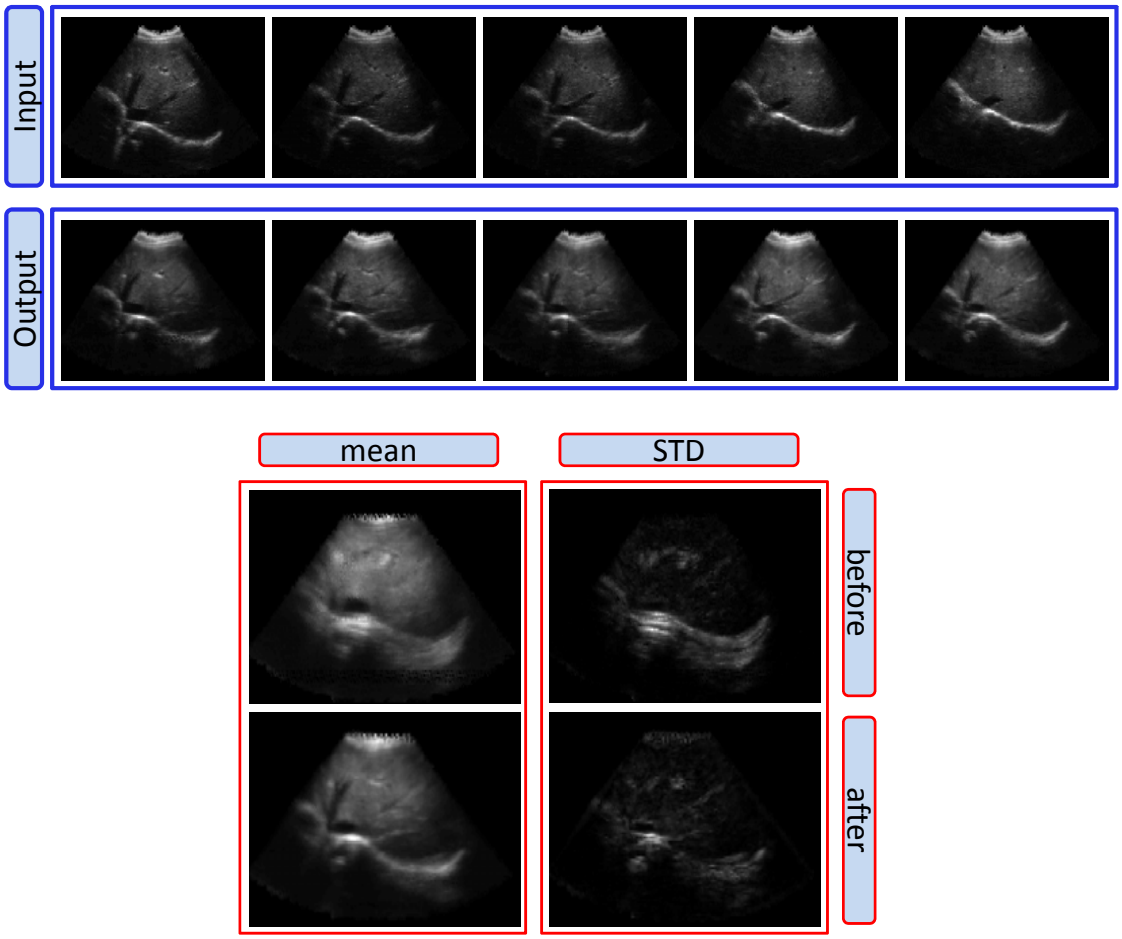



Figure 10.11: First experiment with US sequence. Input, output sequences and mean, STD images before and after the registration.

show the results of two experiments in figures 10.11 and 10.12. The first sequence consists of 5 acquisition over the breathing cycle, the second one consists of 8 acquisitions, where we show an excerpt of 5 images. The number of control points along the temporal direction is half the number of volumes plus the additional 3 control points necessary for cubic B-splines, leading to 5 and 7 control points in the experiments, respectively. The images have a resolution of $200 \times 100 \times 50$ voxels. Comparing the mean and STD images before and after the registration we observe a clear improvement. The mean images before the registration show several diaphragms, while we have one clear diaphragm afterwards, which indicates a correct alignment. The same observation holds for the STD images. In figure 10.10 we illustrate the deformation field of the fourth volume with respect to the first volume in the sequence. Moreover, we show the mesh of the segmentation of the liver which is deformed by the calculated deformation field.

The similarity measures we apply in this study are the multivariate extensions of SSD and NCC (normalization of volumes) with APE. In chapter 8, we discussed fur-

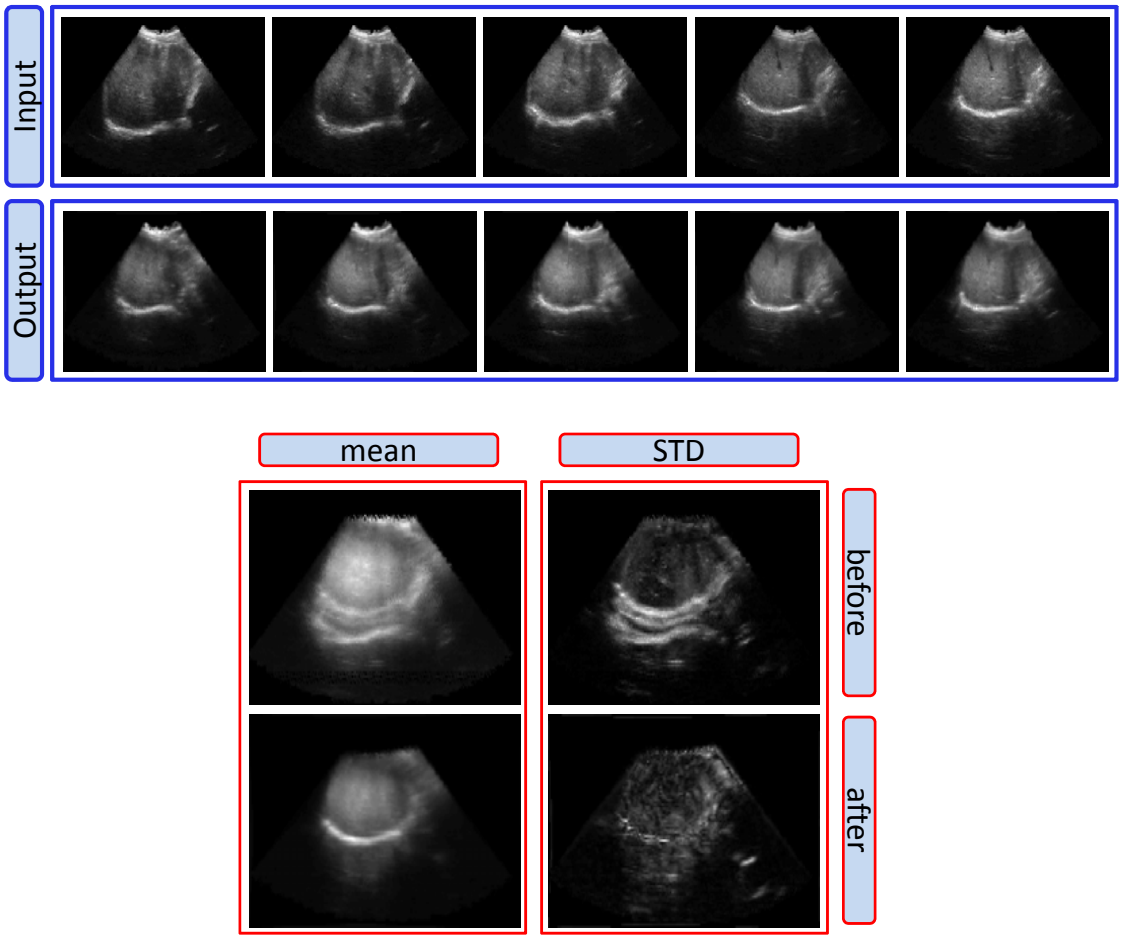



Figure 10.12: Second experiment with US sequence. Input, output sequences and mean, STD images before and after the registration.

ther possibilities for adapting the similarity evaluation to ultrasound. The separation of reflectivity and scattering does not seem to be beneficial in this scenario, although there are large reflections at the diaphragm, because we acquire images from the same position. Further, the locally adaptive Nakagami-based metric is designed for RF data, while we currently work with 4D B-mode. The acquisition of 4D RF data is possible; however, the complex scan geometry necessitates the interpolation of values for the scan conversion, which may affect the distribution. This is subject of ongoing research. Integrating the discussed $CD_2$ may lead to slight improvements, which has to be validated in future work.

## 10.6 Discussion and Conclusion

In this chapter, we have proposed a novel approach for the registration of time-resolved image sequences by taking advantage of groupwise registration. Unlike traditional methods, a single higher dimensional B-splines based transformation was

used. This provides the possibility of having a deformation field also in between the images. Thanks to this feature, we were able to interpolate images in the temporal direction, which is necessary in many applications. APE was adapted to the registration of time-resolved images sequences.

Experiments on synthetic images revealed STORM's advantages in different scenarios over the classical pairwise and groupwise registration methods. It was shown that STORM is more robust against outliers and large deformations within the sequence. The results showed that the use of temporal information ensures the consistency and smoothness in the deformation field along the temporal direction. Moreover, the integration of groupwise registration helped to handle large deformations within the sequence. We performed time interpolation by reconstructing images at any time point in the sequence, which is especially important for applications like radiotherapy treatment planning or 4D image reconstruction. Experiments on CT, MRI and US sequences showed the good performance of STORM on real medical images. It was one of the first times that ultrasound was used for modeling the motion of the liver. Finally, we obtained 4D segmentations by propagating the initial segmentation over time with the deformation field.

## 10.7 Annex: Towards a Lie Group Optimization with FFD B-splines

FFD B-spline deformation fields are by construction smooth, but they are not easily invertible. Applications like computational anatomy and the presented temporal interpolation of images rely on invertible transformations. Commonly applied are diffeomorphisms, which are differentiable deformations with differentiable inverse. In [Vercauteren et al., 2009], a method is presented to construct diffeomorphic transformations from vector fields with the exponential map. Similar to rigid transformations in chapter 7, diffeomorphisms do not form a vector space. Diffeomorphisms are a group under composition, enabling to put a Lie group structure on the space of diffeomorphisms [Vercauteren, 2008]. Having the Lie group property, it is possible to perform efficient optimization with ESM. In our scenario, we apply the exponential map on vector fields parameterized by FFD B-splines to achieve diffeomorphisms. Working with parameterized transformations reduces the number of variables to optimize, which is especially important for simultaneous temporal registration.

Denoting the Lie group of diffeomorphisms as $\mathbb{G}$. On Lie groups, the tangent space at the group identity defines a Lie algebra. The Lie algebra captures the local structure of the Lie group. The Lie algebra of $\mathbb{G}$ is denoted by $\mathfrak{g}$. Let $L = n \cdot |\Omega|$ be the number of voxels and $K = n_x \cdot n_y \cdot n_t$ the number of control points. The Lie algebra is spanned by B-spline basis functions $\mathbf{b}^l \in \mathbb{R}^{L\times 1}$ with $1 \leq l \leq K$. Each basis function $\mathbf{b}^l$ contains the B-spline weights of the control point $l$ with respect to all spatial coordinates $\mathbf{x} \in \Omega \times \{1, \ldots, n\}$. It holds therefore $b^l_m = B^{i,j,k}_{\mathbf{x}}$, which we have defined in section 10.3.3.1, with index $m$ corresponding to $\mathbf{x}$. Stacking all basis functions leads to the matrix $\mathcal{B} = (\mathbf{b}^1, \ldots, \mathbf{b}^K)$. Each element $\mathcal{U} \in \mathfrak{g}$ is expressed as

a linear combination of basis functions

$$\mathcal{U} = \sum_{i=1}^{K} \mathbf{b}^i \cdot [\alpha_x^i, \alpha_y^i, \alpha_t^i] = \mathcal{B} \cdot \boldsymbol{\alpha} \tag{10.24}$$

with local coordinates $\alpha_x^i, \alpha_y^i, \alpha_t^i$ and

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_x^1 & \alpha_y^1 & \alpha_t^1 \\ \vdots & \vdots & \vdots \\ \alpha_x^K & \alpha_y^K & \alpha_t^K \end{bmatrix}. \tag{10.25}$$

The entries of $\boldsymbol{\alpha}$ are the B-spline coefficients. The exponential map relates the Lie algebra to the Lie group:

$$\exp : \mathfrak{g} \to \mathbb{G} \tag{10.26}$$
$$\mathcal{U} \mapsto \exp(\mathcal{U}) = \exp(\mathcal{B}\boldsymbol{\alpha}) = \sum_{j=0}^{\infty} \frac{1}{j!} (\mathcal{B}\boldsymbol{\alpha})^j .$$

For the optimization, the idea is to find from a current diffeomorphism $\mathbf{T} \in \mathbb{G}$ the update $\mathcal{U}$ and to apply the intrinsic update rule on the Lie group with the exponential map

$$\mathbf{T} \leftarrow \mathbf{T} \circ \exp(\mathcal{U}). \tag{10.27}$$

Important in this case is to have a fast calculation of the exponential. Dealing with stationary speed vector fields, it is possible to apply the sum and square algorithm (SSA) to calculate the exponential [Arsigny et al., 2006, Vercauteren et al., 2009]. In [Bossa et al., 2008] it was shown that SSA achieves best results in realistic registration scenarios in comparison to alternative approximations of the exponential. It bases on the property of one-parameter subgroups that for an integer $\eta$ holds

$$\exp(\mathcal{U}) = \exp(\eta^{-1} \cdot \mathcal{U})^{\eta}. \tag{10.28}$$

The idea of SSA is to make the maximal displacement in the vector field, $\max ||\mathcal{U}||$, smaller than 0.5 times the voxel spacing. The deformation field corresponding to this vector field is therefore a diffeomorphism. Moreover, since diffeomorphisms are a group with respect to composition, the squaring of the deformation fields leads again to a diffeomorphism. In details, for the SSA as in [Arsigny et al., 2006, Vercauteren et al., 2009] we select $\eta$ such that

$$\max ||2^{-\eta} \cdot \mathcal{U}|| \leq 0.5 \cdot \text{voxel spacing}. \tag{10.29}$$

An explicit first order integration leads to $\mathbf{V} = \mathrm{Id} + 2^{-\eta} \cdot \mathcal{U}$. Finally, a recursive squaring is $\eta$ times performed $\mathbf{V} \leftarrow \mathbf{V} \circ \mathbf{V}$, leading to the transformation. A big advantage when working with FFDs is that we only have to impose constraints on the maximal displacement with respect to the grid spacing. It was shown in [Choi

and Lee, 2000, Rueckert et al., 2006] that for FFDs based on cubic B-splines it is sufficient to impose

$$\max ||2^{-\eta} \cdot \boldsymbol{\alpha}|| < 0.4 \cdot \text{grid spacing}, \tag{10.30}$$

to obtain diffeomorphisms, setting $\mathbf{V} = \mathrm{Id} + \mathcal{B}(2^{-\eta} \cdot \boldsymbol{\alpha})$. Since the grid spacing is much larger than the voxel spacing, $\eta$ will be smaller and therefore fewer squarings necessary.

It is problematic with the composition of deformations resulting from B-spline vector fields that the composed result cannot necessarily be expressed on the same basis [Hagenlocker and Fujimura, 1998, Rueckert et al., 2006], while we would like to keep the same parameterization throughout the optimization. To address this issue, we assume that the current transformation $\mathbf{T}$ is related to a B-spline vector field. Consequently, there exists an element in the Lie algebra $\mathcal{T} \in \mathfrak{g}$ such that $\mathbf{T} = \exp(\mathcal{T})$, resulting in the update $\exp(\mathcal{T}) \circ \exp(\mathcal{U})$. The task is to find a vector field $Z(\mathcal{T} + \mathcal{U})$ such that

$$\exp(\mathcal{T}) \circ \exp(\mathcal{U}) \approx \exp(Z(\mathcal{T} + \mathcal{U})). \tag{10.31}$$

Theoretically, this poses problems because we deal with an infinite dimensional space, which has a Lie group structure but is not an actual Lie group [Vercauteren et al., 2008b]. In practice, it was shown [Bossa et al., 2007] that the Baker-Campell-Hausdorff (BCH) formula can be successfully applied for diffeomorphisms. By using the first terms of the BCH, good results for the approximation were reported for brain atlas construction [Bossa et al., 2007]. The approximation advocated in [Vercauteren et al., 2008b] is

$$Z(\mathcal{T}, \mathcal{U}) := \mathcal{T} + \mathcal{U}. \tag{10.32}$$

Since both vector fields are defined on the same B-spline grid, we have

$$Z(\mathcal{T}, \mathcal{U}) = \mathcal{T} + \mathcal{U} = \mathcal{B} \cdot \boldsymbol{\alpha}_{\mathcal{T}} + \mathcal{B} \cdot \boldsymbol{\alpha}_{\mathcal{U}} = \mathcal{B}(\boldsymbol{\alpha}_{\mathcal{T}} + \boldsymbol{\alpha}_{\mathcal{U}}) \tag{10.33}$$

with $\boldsymbol{\alpha}_{\mathcal{T}}$ the B-spline coefficients for the vector field $\mathcal{T}$ and analogously for $\boldsymbol{\alpha}_{\mathcal{U}}$. This leads to the approximation of the composition by adding the B-spline coefficients and applying the exponential

$$\exp(\mathcal{T}) \circ \exp(\mathcal{U}) \approx \exp(\mathcal{B}(\boldsymbol{\alpha}_{\mathcal{T}} + \boldsymbol{\alpha}_{\mathcal{U}})). \tag{10.34}$$

The incorporated assumption that we know the B-spline coefficients for $\mathbf{T}$ is no limitation because we start the registration with the identity $\mathbf{T} = \mathbf{I}$, which corresponds to $\boldsymbol{\alpha}_T = \mathbf{0}$.

Let us further denote the transformation of a point $\mathbf{p} = [x, y, t]^\top \in \mathbb{R}^3$ through the mapping $\exp(\mathcal{T}) \in \mathbb{G}$ with $w(\exp(\mathcal{T}), \mathbf{p})$

$$w : \mathbb{G} \times \mathbb{R}^3 \to \mathbb{R}^3 \tag{10.35}$$

$$(\exp(\mathcal{T}), \mathbf{p}) \mapsto w(\exp(\mathcal{T}), \mathbf{p}) = \exp(\mathcal{T})[\mathbf{p}] = \boldsymbol{\delta}_{\mathbf{p}} \cdot \exp(\mathcal{T}), \tag{10.36}$$

which corresponds to the selection of an element in the dense field, $\exp(\mathcal{T})[\mathbf{p}]$, or expressed in terms of a matrix-vector multiplication, $\boldsymbol{\delta}_{\mathbf{p}} \cdot \exp(\mathcal{T})$, with $\boldsymbol{\delta}_{\mathbf{p}} \in \mathbb{R}^{1 \times L}$ everywhere zero except for being one at the position of $\mathbf{p}$.

### 10.7.1 Optimization on Lie Groups

With the presented update step and Lie group parameterization in the last section, we can directly apply the presented optimization methods for Lie groups of chapter 7. For simplicity, we consider the specific case of SSD, as earlier in this chapter.

#### 10.7.1.1 Gauß-Newton

The Jacobian of the difference vector $\mathbf{f}_{g,h}$ used for Gauß-Newton has the form (cf. equation (7.64))

$$[\mathbf{J}_{\mathbf{f}_{g,h}}(\exp(\mathcal{T}))]_{\mathbf{p}} = -[\mathbf{J}_{I_h}(\exp(\mathcal{T}))]_{\mathbf{p}} \cdot [\mathbf{J}_w(\exp(\mathcal{T}))]_{\mathbf{p}} \cdot \mathbf{J}_e(\exp(\mathcal{T})) \tag{10.37}$$

with

1. $[\mathbf{J}_{I_h}(\exp(\mathcal{T}))]_{\mathbf{p}} = \nabla I_h^{\downarrow}(w(\mathrm{Id}, \mathbf{p})) = \nabla I_h^{\downarrow}(\mathbf{p}) \in \mathbb{R}^{3\times 3}$
2. $[\mathbf{J}_w(\exp(\mathcal{T}))]_{\mathbf{p}} = \frac{\partial w(\mathbf{Z},\mathbf{p})}{\partial \mathbf{Z}}\Big|_{\mathbf{Z}=\exp(\mathbf{0})=\mathrm{Id}} = \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{p}} \\ \boldsymbol{\delta}_{\mathbf{p}} \\ \boldsymbol{\delta}_{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} 0\dots010\dots0 \\ 0\dots010\dots0 \\ 0\dots010\dots0 \end{pmatrix} \in \mathbb{R}^{3\times L}$
3. $\mathbf{J}_e(\exp(\mathcal{T})) = \frac{\partial \exp(\mathcal{B}\cdot\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\Big|_{\boldsymbol{\alpha}=\mathbf{0}} = \exp(\mathcal{B}\cdot\boldsymbol{\alpha})|_{\boldsymbol{\alpha}=\mathbf{0}} \cdot \mathcal{B} = \mathcal{B} \in \mathbb{R}^{L\times K}$

This is very similar to the Jacobian derived in equation (10.17), with a slight discrepancy due to different paramterization and indices.

#### 10.7.1.2 ESM

The presented parameterization of the deformation as elements of a Lie group enables the application of ESM. The gradient for ESM is

$$\mathbf{J}^{\mathrm{ESM}}_{\mathbf{f}_{g,h}}(\exp(\mathcal{T})) = \mathbf{J}_{\mathbf{f}_{g,h}}(\exp(\mathcal{T})) + \mathbf{J}_{\mathbf{f}_{g,h}}(\exp(\mathcal{T}) \circ \exp(\mathcal{U}^{\mathrm{opt}})), \tag{10.38}$$

considering the optimal alignment $\exp(\mathcal{T}^{\mathrm{opt}}) = \exp(\mathcal{T}) \circ \exp(\mathcal{U}^{\mathrm{opt}})$. We again approximate the gradient of the perfectly aligned moving image with the fixed image

$$[\mathbf{J}_{I_h}(\exp(\mathcal{T}^{\mathrm{opt}}))]_{\mathbf{p}} \approx [\mathbf{J}_{I_g}(\exp(\mathcal{T}))]_{\mathbf{p}}. \tag{10.39}$$

Further, considering the results from chapter 7 that

$$[\mathbf{J}_w(\exp(\mathcal{T}^{\mathrm{opt}}))]_{\mathbf{p}} = [\mathbf{J}_w(\exp(\mathcal{T}))]_{\mathbf{p}} \tag{10.40}$$

$$\mathbf{J}_e(\exp(\mathcal{T}^{\mathrm{opt}})) \cdot \boldsymbol{\alpha}^{\mathrm{opt}} = \mathbf{J}_e(\exp(\mathcal{T})) \cdot \boldsymbol{\alpha}^{\mathrm{opt}} \tag{10.41}$$

we obtain for the gradient of ESM

$$\begin{aligned}\left[\mathbf{J}^{\mathrm{ESM}}_{\mathbf{f}_{g,h}}\right]_{\mathbf{p}} &= -([\mathbf{J}_{I_g}(\exp(\mathcal{T}))]_{\mathbf{p}} + [\mathbf{J}_{I_h}(\exp(\mathcal{T}))]_{\mathbf{p}}) \cdot [\mathbf{J}_w(\exp(\mathcal{T}))]_{\mathbf{p}} \cdot \mathbf{J}_e(\exp(\mathcal{T})) \\ &= -(\nabla I_g^{\downarrow}(\mathbf{p}) + \nabla I_h^{\downarrow}(\mathbf{p})) \cdot \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{p}} \\ \boldsymbol{\delta}_{\mathbf{p}} \\ \boldsymbol{\delta}_{\mathbf{p}} \end{pmatrix} \cdot \mathcal{B}. \end{aligned} \tag{10.42}$$

With the Jacobian we are able to calculate the update field $\mathcal{U}$, which is composed with the current transformation $\mathcal{T}$, as described in equation (10.34).

# Part V

# EPILOGUE

# Chapter 11

# Conclusion

*In this dissertation, we discussed a variety of improvements in the field of image registration and ultrasound imaging. The main focus was set on the alignment of ultrasound images, where a good understanding of ultrasound imaging and image registration is necessary to adapt the methods accordingly. Finally, all advancements showed its potential in ultrasound mosaicing and motion modeling.*

## Ultrasound

We presented an overview of ultrasound imaging, studied models for wave propagation and analyzed scattering statistics. The described concepts served as basis for the following adaptation of methods to ultrasound. We proposed the application of the 2D analytic signal for extracting the envelope of RF data. This leads to improved statistical properties, which we illustrated with goodness-of-fit tests. We benefited from these properties for the design of ultrasound specific similarity measures, where we intended to exactly model the statistics. Moreover, the local phase was used to identify structures of high reflectivity. Next, the quality of ultrasound can further be improved by combining the information from several views. The fusion is challenging because of the viewing angle dependency of ultrasound. Consequently, we presented a new method that tries to estimate the acoustic impedance from multiple views.

## Registration

After discussing the modality that we mainly work on, we continued with an overview of the main methodological component, image registration. We devised a novel probabilistic framework that allows describing a large number of registration techniques. The framework incorporated, in contrast to previous ones, neighborhood information. The dependencies were structured by introducing an additional layer of random variables, which can be seen as descriptors. This allowed us to model registration techniques in the continuum of intensity- and feature-based approaches. Moreover, we took a closer look on current multi-modal registration approaches and proposed a new technique with structural representations. The aim of these representations is to capture the structural information in images, which should be

the same across modalities. Consequently, we were able to perform multi-modal registration in a mono-modal framework. This enabled the application of a larger variety optimization techniques and a faster similarity evaluation, which is especially interesting for groupwise registration. The specific representations we presented are entropy and Laplacian images. Laplacian images have theoretical advantages; however, the high computational cost limits their application.

## Ultrasound Mosaicing

The first clinical application, where we combined ultrasound imaging and registration, was ultrasound mosaicing. We illustrated that the application of simultaneous registration is advantageous to deal with the particularities of ultrasound. By taking all the image information into account, we achieved more robust registration results. We deduced a new class of multivariate similarity measures and presented efficient gradient-based optimization schemes for them. Most prominently, we applied the efficient second-order minimization, which achieved the fastest convergence in our experiments. In combination with structural images, we even managed to perform simultaneous, multi-modal registration with ESM. Moreover, we devised similarity measures that account for the viewing angle dependency and noise statistics in ultrasound. A new similarity measure was proposed that adapts itself to the local region, where the similarity has to be estimated. We achieved this, on the one hand, by having a distribution, which is flexible enough to well characterize several scattering scenarios, and on the other hand, by estimating the parameters of the distribution on local patches. The adaptation led to improvements for block-matching and rigid registration. In a second approach, we considered ultrasound images with strong reflectivity in the context of mosaicing in more details, which cause problems due to the viewing angle dependency. The presented approach split images into parts of high reflectivity, corresponding the large tissue interfaces, and speckle regions, corresponding to microscopic tissue inhomogeneities.

## Motion Modeling

For ultrasound mosaicing, we assumed a rigid registration scenario. In the last part of the dissertation, we extended this to cope with deformations, which is required for motion modeling. Dealing with a non-rigid scenario has far-reaching consequences, complicating the image acquisition. We proposed an image-based gating system for the creation of breathing-affected 4D data. The respiratory signal was extracted from the image information with manifold learning. The algorithm is completely automatic and not limited to a specific body part, as it is the case for diaphragm tracking. In comparison to an external gating system, we achieved very good results. After the creation of 4D images, we calculate the motion fields with deformable registration. We devised a new algorithm that takes the specificity of the problem, the smooth deformation along temporal direction, into account. This was achieved by working with a single 4D deformation field, instead of multiple 3D deformation

fields. Moreover, the simultaneous registration approach taken has advantages in regard to template selection and large displacements. As a special feature, the 4D deformation field allowed for the interpolation of volumes at arbitrary image points. Our experimental evaluation on MR, CT, and ultrasound images showed the robustness and quality of the approach. Finally, we were able to demonstrate that 4D motion modeling of the liver in ultrasound is feasible.

# Chapter 12

# Perspectives

*In this dissertation, we discussed several subjects in medical image analysis and presented extensions to the state of the art. These extensions are, however, only small steps on a long journey of discovery. We try to indicate further steps that may be worth taking.*

## Ultrasound

In the future, with the increasing distribution of 2D array transducers in the market, the access to 3D RF data may be easier. Since RF data comes with a higher resolution than B-mode images and shows the raw measured signals, we consider it the data type of choice when working with ultrasound. The increased resolution leads to a higher computational burden, but with the increasing computation power and possibilities for downsampling this is no real limitation. As indicated in the thesis, working with RF data is advantageous when incorporating ultrasound noise statistics, which is necessary to improve the results on processing ultrasound images. With this being said, future 3D mosaicing and motion modeling can benefit from the integration of RF data. In this regard, the calculation of the local amplitude with the 3D analytic signal, instead of the 2D analytic signal, would be interesting. Unfortunately, such an extension has not yet been proposed. Also challenging when working with the 2D or 3D analytic signals is the complex geometry of curved transducers.

In inhomogeneous regions of ultrasound images, *e.g.* at boundaries, the Nakagami distribution does not lead to good fits. We would require a mixture of Nakagami to well approximate the distribution in such cases. This could lead to improvements in the quantification of the statistical properties of the 2D analytic signal, and further, may allow the design of superior similarity measures. The reasons why we have not yet taken this step are the challenging derivation of the formulas and the increasing complexity of the calculation of the estimates.

The outcome of the envelope estimation with the 2D analytic signal is influenced by a set of parameters. Depending on the purpose of the image, if it is presented to a physician or if it serves as input for further computational tasks, different parameters may have to be selected. This goes in the direction of application specific imaging. Going one step further, we could directly integrate this in the image

acquisition process. Examples may be the acquisition of ultrasound images with several frequencies, with several focal zones to clearly capture the boundary, and the inclusion of beam steering to obtain responses from different insonification angles. Such images may not lead to visually pleasing results, but they may be well suited for the subsequent analysis and improve the results of segmentation, classification, and registration. The adaptation requires the close collaboration of researchers, who work on the physics of ultrasound imaging and develop acquisition protocols, as well as scientists, who work on processing these images. Such a close collaboration will also be necessary for addressing the challenging task of acoustic impedance estimation. Only the application of a precise model for ultrasound propagation, together with accurate data, will result in realistic estimates of the impedance.

## Registration

With the presented probabilistic framework, various registration techniques can be modeled. Further, certain pre-processing steps can be derived as being optimal in a maximum likelihood sense under the consideration of specific assumptions. We exemplified this for a number of approaches, while in future work, more techniques could be investigated. It would be very interesting to try to implement this theoretic model and perform registrations with it. In its most general form, it is very generic and an implementation is not directly feasible. Limiting the model to certain descriptors in a hybrid registration setting together with a flexible adaptation of the description layers, could lead to nice practical results of the presented concepts.

For multi-modal registration, we presented two exemplary structural representations, which originate from completely different fields and lead to different structural images. These are definitely not the only structural representations that exist and more appropriate ones may still be discovered. In order to make the Laplacian images applicable in practice, more efficient ways for their calculation have to be investigated. Possible directions are to move away from the dense setting and to concentrate on keypoint locations, or to apply other dimensionality reduction techniques.

## Optimization

Gradient-based registration approaches are more efficient because they use the additional information provided by the gradient of the cost function to direct the search. A remaining problem is, however, that such methods are prone to local optima. In segmentation and reconstruction, there is a growing interest in convex formulations. The optimization of a strictly convex cost function leads to the global optimum. Recently, a convex formulation of 2D [Kokiopoulou and Frossard, 2009] and 3D [Kokiopoulou et al., 2011b] registration was proposed. It is limited to an affine transformation model and SSD as cost function. With the proposed structural representations, we could apply the convex registration in multi-modal scenarios. Moreover, an extension of the convex formulation to simultaneous registration would be interesting.

## Motion Modeling

The presented image-based gating approach is retrospective. For the investigated application of creating 4D data, this is not an issue. There are, however, situations like the instantiation of a motion model that require a real-time gating signal. This is problematic with manifold learning techniques because they only provide a mapping for the points in the data set and not for the entire space. Methods exist to address this out-of-sample problem. A naive approach for our application may be as follows. We acquire data over a couple of breathing cycles and perform manifold learning on them. This data can be seen as training data. After the gating, we know for each frame in the training set the corresponding breathing phase. In the second stage, we try to perform the gating in real-time. For this, we only have to find for every incoming frame the closest frame in the training data set and assign the breathing phase. Efficient techniques in this direction are proposed in computer vision for image retrieval. An alternative approach would be the application of different dimensionality reduction techniques, which map the entire space and therefore directly allow for mapping new images to low dimensions. Such approaches may, however, be less accurate.

For motion modeling, we are interest in a smooth deformation field along the temporal direction, as previously mentioned. Having a smooth deformation within the image domain may, however, lead to inaccuracies. During breathing the liver slides along the peritoneum, which is not well characterized by a smooth motion field. The integration of an explicit regularization term in the deformable registration, which would have to be location dependent, is a possible solution. For ultrasound, this is less problematic than for MR and CT because of the limited field of view.

Finally, it is not possible to end this thesis having only quoted a French mathematician. We finish therefore with the words of Carl Friedrich Gauß [Gauß, 1808]:

> It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again.

# Part VI

# Appendix

# Appendix A

# Manifold Learning for Patient Position Detection in MRI

*Magnetic resonance imaging is performed without ionizing radiation, however, the applied radio frequency power leads to heating, which is dependent on the body part being imaged. Determining the patient position in the scanner allows to better monitor the absorbed power and therefore to optimize the image acquisition. Low-resolution images, acquired during the initial placement of the patient in the scanner, are exploited for detecting the patient position. We use Laplacian eigenmaps to learn the low-dimensional manifold embedded in the high-dimensional image space. Our experiments clearly show that the presumption of the slices lying on a low dimensional manifold is justified. We obtain very good classification results with a nearest neighbor classifier operating on the low-dimensional embedding.*

## A.1 Introduction

Current magnetic resonance scanners allow the acquisition of high resolution images. However, this comes with a higher dose of radio frequency power applied to the patient. This leads to heating in the body, which has to be monitored by measuring the specific absorption rate (SAR). The limits for the maximal SAR depend on the patient position inside the scanner. If the position is not known, the lowest SAR limit along the body, which is in the neck area, is to be set to the global limit [Keil et al., 2006]. Determining the patient position within the MR scanner enables to exploit the maximal image resolution by imposing an SAR model, which is adapted to the body region.

In our previous work [Keil et al., 2006], we proposed to use the novel *move during scan* imaging protocol to determine the patient position. This enables the acquisition of low-resolution images during the initial positioning of the patient in the scanner. It has the advantage that the workflow does not have to be altered, ensuring its seamless integration. During the acquisition, the bed moves with a relatively high but constant speed, leading to low-resolution slices of $64 \times 64$ pixels and a slice spacing of 7.5 mm to 15 mm (see Figure A.1 for examples and a coronal view).

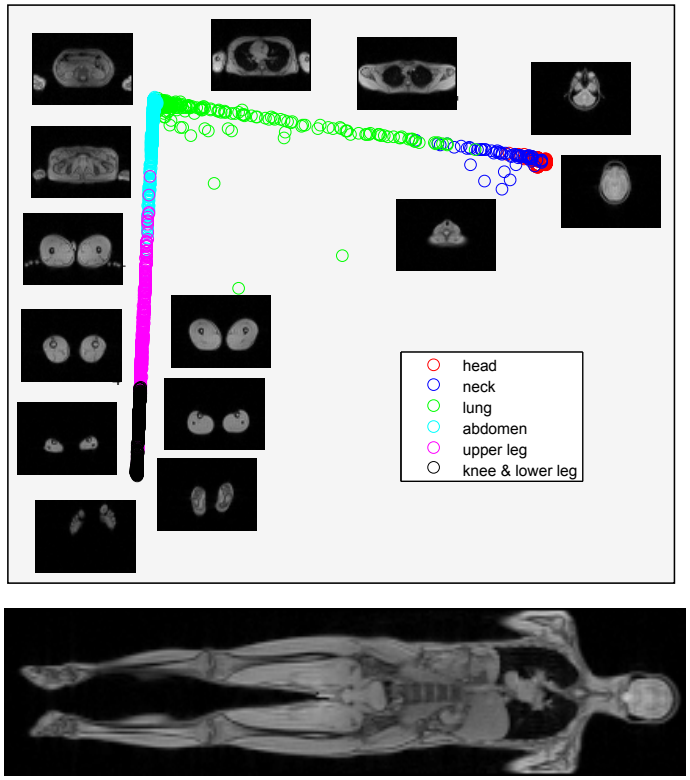
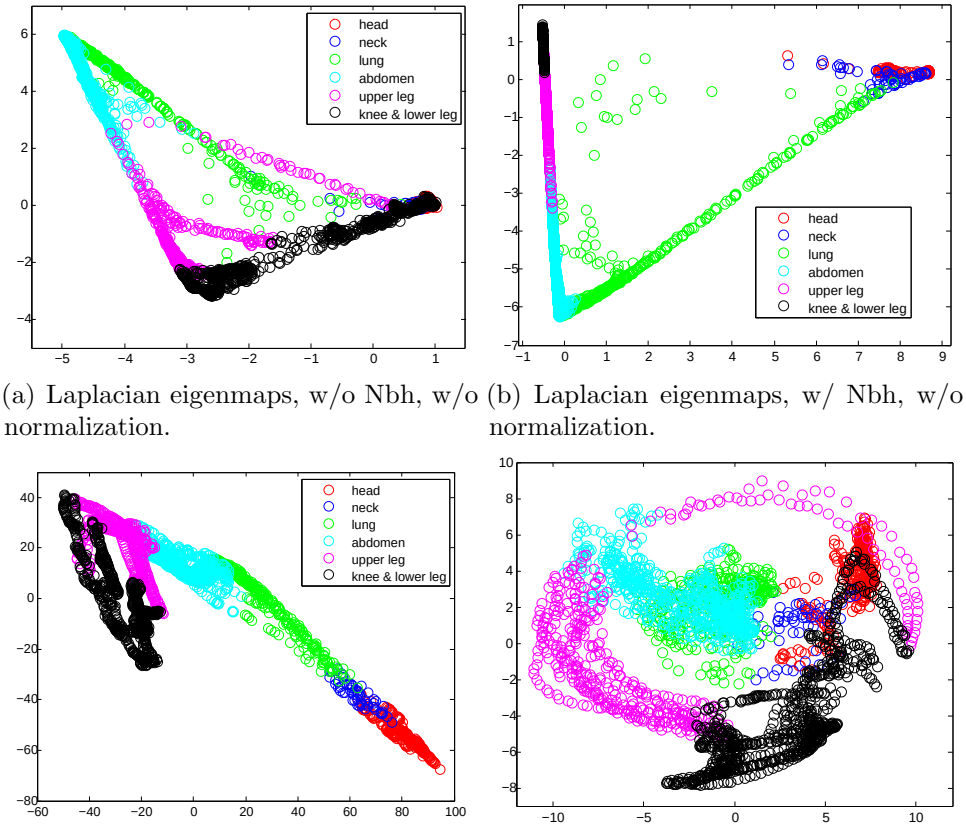



Figure A.1: Low dimensional embedding of manifold using Laplacian eigenmaps (w/ Nbh, w/ normalization) with ground truth labeling. Below, coronal plane of move during scan volume.

The objective of patient position detection is to classify these slices according to different regions of the body such as head, abdomen, and lower leg. In [Keil et al., 2006], principal component analysis (PCA) is used to reduce the dimensionality before the classification is performed. PCA [Pearson, 1901] and also the independent component analysis (ICA) [Hyvärinen et al., 2001] find a set of basis images, and represent an input image as linear combination of those. However, the representation of images as a linear combination of those basis images may require many of those for an accurate representation, and further, the internal structure may not be easy to identify by analyzing the weighting parameters. Having data that lies on a low-dimensional manifold living in a high-dimensional space, more appropriate methods were proposed that respect this structure, cf. section 6.6.1. In this chapter, we evaluate those methods for patient position detection, focusing on Laplacian eigenmaps [Belkin and Niyogi, 2003], with exemplary low-dimensional embeddings shown in Figures A.1 and A.2. Additionally, we investigate if the MR slices lie on a manifold with intrinsic low dimensionality. In order to achieve good classification results, we adapt manifold learning methods by integrating the consideration of neighboring slices and a preceding normalization. The classification is performed with a nearest neighbor classifier.




(a) Laplacian eigenmaps, w/o Nbh, w/o normalization.

(b) Laplacian eigenmaps, w/ Nbh, w/o normalization.

(c) Isomap, w/ Nbh, w/ normal.

(d) PCA, w/ Nbh, w/ normalization

Figure A.2: Low dimensional embeddings. Data labeling with ground truth segmentation.

## A.2 Background

The task of dimensionality reduction is to find the underlying structure in a large set of points embedded in a high dimensional space. The advantages of a successful dimensionality reduction are: first, it assists the classification because a direct classification in high-dimensional space does not respect the manifold structure whereas the classification on the low-dimensional embedding does. And second, the lower-dimensional embedding enables to visualize the manifold and, therefore, indicates whether a classification seems feasible.

Manifold learning for dimensionality reduction has recently gained much attention to assist image processing tasks, as described in section 6.6.1. The applications vary by using either the intensity image as input or calculating coordinate transformations, for instance diffeomorphic warps [Souvenir and Pless, 2007, Gerber et al., 2009, Hamm et al., 2009], between the images, which subsequently serve as input. The calculation of the deformation field makes sense, if all images to be analyzed show similar objects *e.g.* brain images. In our application, we have to deal with images from all body parts, for which it would not be possible to register one to the other. We therefore

Table A.1: Results of classification for Laplacian eigenmaps with neighborhood and image normalization. The overall correct classification rate is 94.0%.

| | Recognized Class Labels in % | | | | | |
|---|---|---|---|---|---|---|
| | Head | Neck | Lung | Abd. | U. Leg | L. Leg |
| Head | **95.0** | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Neck | 25.9 | **69.0** | 5.2 | 0.0 | 0.0 | 0.0 |
| Lung | 0.0 | 0.9 | **92.0** | 6.9 | 0.3 | 0.0 |
| Abd. | 0.0 | 0.0 | 2.0 | **97.2** | 0.8 | 0.0 |
| U. Leg | 0.0 | 0.0 | 0.0 | 2.1 | **84.5** | 13.5 |
| L. Leg | 0.0 | 0.0 | 0.0 | 0.0 | 1.6 | **98.4** |

directly perform the dimensionality reduction on the original images. The suggestion that the MR slices lie on a low-dimensional manifold in the ambient space seems to be justified because variations between neighboring slices are smooth, and further, slices from the same body position but different patients are similar.

## A.3 Position Detection

In this section we describe the details of the proposed manifold learning approach. Throughout, we consider one dimension of the ambient space for each image pixel. Considering $k$ points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^N$ lying on a manifold, we want to find a set of corresponding points $\mathbf{y}_1, \ldots, \mathbf{y}_k$ in the low-dimensional space $\mathbb{R}^n$ ($n \ll N$).

### A.3.1 Pre-Processing

Precedent to applying the dimensionality reduction, we normalize the intensity values in the images. Inhomogeneities and imaging artifacts lead to variation in the intensity values making the normalization necessary. Further, we integrated neighborhood information to make each slice also dependent on its locally neighboring slices. The idea is similar to the consideration of neighborhood information in many image processing tasks by looking not only at single pixel intensities but the local context, to make the processing more robust (cf. chapter 5). In our case, we create a new data point $\mathbf{x}'_i$ by concatenating it with exactly those neighbors of distance $m$ away

$$\mathbf{x}'_i = [\mathbf{x}_{i-m},\ \mathbf{x}_i,\ \mathbf{x}_{i+m}]. \tag{A.1}$$

The concatenation leads to input slices with three times higher dimensionality. In order to avoid the dimensionality growth, we downsample the data points $\mathbf{x}'_i$ by the factor 3, so that the input dimensionality remains $N$.

Table A.2: Results of classification for Laplacian eigenmaps w/o neighborhood but w/ image normalization. The overall correct classification rate is 92.6%.

| | Recognized Class Labels in % | | | | | |
|---|---|---|---|---|---|---|
| | Head | Neck | Lung | Abd. | U. Leg | L. Leg |
| Head | **95.7** | 2.8 | 0.0 | 0.0 | 0.0 | 1.4 |
| Neck | **60.3** | 34.5 | 5.2 | 0.0 | 0.0 | 0.0 |
| Lung | 0.3 | 0.3 | **91.7** | 6.6 | 0.9 | 0.3 |
| Abd. | 0.0 | 0.0 | 2.7 | **96.4** | 1.0 | 0.0 |
| U. Leg | 0.0 | 0.0 | 1.8 | 1.8 | **83.8** | 12.6 |
| L. Leg | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | **96.7** |

Table A.3: Results of classification for Laplacian eigenmaps w/o neighborhood and w/o image normalization. The overall correct classification rate is 90.0%.

| | Recognized Class Labels in % | | | | | |
|---|---|---|---|---|---|---|
| | Head | Neck | Lung | Abd. | U. Leg | L. Leg |
| Head | **90.1** | 2.6 | 0.0 | 0.0 | 0.0 | 7.4 |
| Neck | **62.1** | 6.9 | 5.2 | 0.0 | 6.9 | 19.0 |
| Lung | 0.0 | 0.0 | **92.6** | 4.6 | 1.7 | 1.1 |
| Abd. | 0.0 | 0.0 | 2.7 | **96.5** | 0.8 | 0.0 |
| U. Leg | 0.0 | 0.0 | 6.4 | 2.1 | **86.1** | 5.5 |
| L. Leg | 7.2 | 0.2 | 0.0 | 0.0 | 3.0 | **89.6** |

### A.3.2 Laplacian Eigenmaps

We construct a graph with a node for each point $\mathbf{x}_i$ and with edges connecting the $l$ nearest neighboring nodes. Heat kernel-based weights are assigned to the edges with $\mathbf{W}_{ij} = e^{-||\mathbf{x}_i - \mathbf{x}_j||_2^2/t}$. Once the neighborhood graph is constructed, the eigenvectors of the graph Laplacian provide the embedding map. In our implementation we choose to consider the 40 nearest neighbors, so $l = 40$ and, further, dimension $n = 2$ as intrinsic manifold dimensionality with $N = 64 \times 64$ the dimensionality of the ambient space. In our experiments, we found out that $n = 1$ is not sufficient for correctly distinguishing the classes and the good results for $n = 2$ make a further increase superfluous.

### A.3.3 Classification

The dimensionality reduction significantly facilitates the classification and simple classifiers, such as $k$-nearest neighbors (KNN), can be applied for performing the

classification in the low-dimensional space. Following the comparison of supervised learning algorithms [Caruana and Niculescu-Mizil, 2006], random forests and decision trees outperform KNN, and therefore could lead to further improvements of the classification results.

## A.4 Experiments

We evaluate the manifold learning embedding on 13 whole-body data sets, consisting of acquisition from male and female patients. Further, some acquisitions were done with a pillow below the legs and the position of the arms varies. A manual labeling of the data sets was done, assigning to each slice one of the 6 classes: head, neck, lung, abdomen, upper leg, and lower leg. We perform the dimensionality reduction with Laplacian eigenmaps, Isomap, and PCA, see Figures A.1 and A.2. PCA is used to compare to our previous work [Keil et al., 2006] and illustrate the advantages of the non-linear embedding methods. We further evaluate the influence of image normalization and the integration of the neighboring slices.

Next to graphs of the 2D dimensional embedding of the data, we also present classification results with KNN. To quantify the performance we perform a cross-validation with a leave-one-out strategy. In Tables A.1 to A.5, we show the results of the study of the selected approaches. Each row indicates the percentage of a slice being assigned to one of the 6 classes. Looking at Table A.1, for instance, head slices are assigned with 95% to the head class and with 5% to the neck class.

We obtain the highest overall correct classification rate of 94.0% with Laplacian eigenmaps in combination with the normalization of the images and neighborhood integration. In Table A.1, it can be seen that especially the classification of the head and lower leg slices is very accurate, which is important to determine if the patient moves in head or feet first. Remarkable is that in the case of wrong classifications, labels from neighboring classes are assigned, so that, for example, no head slice wrongly got detected as abdomen or leg. This does not hold for the classification without normalization or neighborhood. For Isomap, the classification accuracy is lower, however, also there the wrong assignments are limited to neighboring classes. The results for PCA show the lowest classification performance.

The low correct classification results for the neck are due to a significantly lower number of slices for training, where even for the medical expert, it is difficult to uniquely assign the transition from head to neck and neck to shoulders.

## A.5 Conclusion

We proposed the application of manifold learning techniques for patient position detection. Our results clearly indicate that the image slices lie on a low-dimensional manifold embedded in the high-dimensional image space. We proposed to apply Laplacian eigenmaps for manifold learning and achieved superior results in comparison to Isomap. Moreover, the proposed adaptation of the method to the specific scenario

Table A.4: Results of classification for Isomap w/ neighborhood and w/ image normalization. The overall correct classification rate is 90.1%.

| | Recognized Class Labels in % | | | | | |
|---|---|---|---|---|---|---|
| | Head | Neck | Lung | Abd. | U. Leg | L. Leg |
| Head | **93.3** | 6.7 | 0.0 | 0.0 | 0.0 | 0.0 |
| Neck | 29.3 | **60.3** | 10.3 | 0.0 | 0.0 | 0.0 |
| Lung | 0.0 | 1.1 | **92.6** | 6.3 | 0.0 | 0.0 |
| Abd. | 0.0 | 0.0 | 7.7 | **89.3** | 2.9 | 0.0 |
| U. Leg | 0.0 | 0.0 | 0.0 | 3.7 | **86.3** | 10.0 |
| L. Leg | 0.0 | 0.0 | 0.0 | 0.0 | 7.4 | **92.6** |

Table A.5: Results of classification for PCA w/ neighborhood and w/ image normalization. The overall correct classification rate is 80.0%.

| | Recognized Class Labels in % | | | | | |
|---|---|---|---|---|---|---|
| | Head | Neck | Lung | Abd. | U. Leg | L. Leg |
| Head | **82.0** | 7.0 | 0.7 | 0.0 | 0.0 | 10.3 |
| Neck | **31.0** | 25.9 | 3.4 | 13.8 | 0.0 | 25.9 |
| Lung | 0.0 | 0.3 | **56.9** | 42.0 | 0.0 | 0.9 |
| Abd. | 0.0 | 0.0 | 16.1 | **82.0** | 1.8 | 0.0 |
| U. Leg | 4.6 | 0.0 | 0.0 | 4.8 | **82.9** | 7.8 |
| L. Leg | 7.9 | 0.1 | 0.5 | 0.0 | 2.4 | **89.1** |

by the consideration of neighboring slices and image normalization led to a further improvement of recognition rates. For the evaluation, we performed a classification with KNN and subsequent cross-validation, leading to very good results.

# Appendix B

# Deformable Mosaicing for Whole-body MRI

*Whole-body magnetic resonance imaging is an emerging application gaining vast clinical interest over the last years. Although recent technological advances shortened the longish acquisition time, this is still the limiting factor, avoiding its wide-spread clinical usage. The acquisition of images with large field of view helps to relieve this drawback, but leads to significantly distorted images. Therefore, we propose a* deformable mosaicing *approach, based on the simultaneous registration to linear weighted averages, to correct for distortions in the overlapping area. This method produces good results on in-vivo data and has the advantage that a seamless integration into the clinical workflow is possible.*

## B.1 Introduction

Whole-body (WB) magnetic resonance imaging is becoming a popular clinical tool due to the recent technological advances in MRI, making faster acquisitions possible. Unlike CT, the acquisition of high-resolution MR images is not feasible during continuous table movement, making a multi-station scanning necessary to cover larger body regions. The compounding of the partially overlapping volumes is straightforward, since the MR scanner keeps track of their exact spatial locations.

The creation of WB images further increases the number of clinical applications for MRI, so far reserved for other modalities, see section B.2. From a current perspective, the major disadvantage using MRI for WB imaging in comparison to CT is the longer scanning time. In this chapter, we use MR acquisitions with a large field of view (FOV), enabling to cover with the same number of scans larger parts of the body. This, however, leads to a degradation of the images by geometrical distortion artifacts towards the boundaries, further described in section B.3. We propose a novel method, originating from groupwise registration (cf. chapter 7), to correct for the geometrical distortion in the overlapping area, see section B.4. Our experiments show the good results on in-vivo data, see section B.5.

## B.2 Clinical Applications of WB-MRI

Recent advances in MRI such as multi-channel receiver, parallel imaging techniques, and automated table movement make high-resolution WB-MRI clinically feasible [Schmidt et al., 2007].

First clinical studies show its value in oncological applications, which focus on the search for metastases of cancer patients in the whole body. [Lauenstein et al., 2004] found out that WB-MRI compares well with standard methods such as CT, PET, and nuclear scintigraphy for the detection of cerebral, pulmonary, and hepatic lesions and is more sensitive for the detection of hepatic and osseous metastases. Additionally, the excellent contrast of MRI provides further information about soft tissue and organs to the physician, supporting his diagnosis [Goyen, 2007].

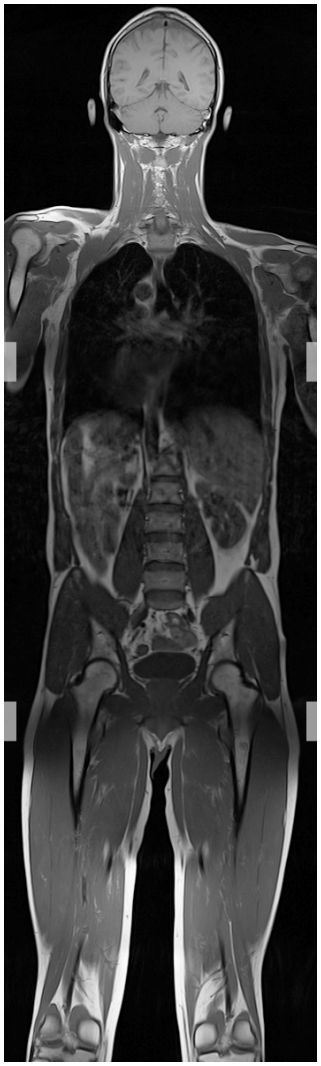


Figure B.1: WB-MRI from 3 stations. Gray bars show overlap.

Non-oncological applications are whole-body fat measurement to evaluate body composition and muscular infections, angiography for the diagnosis of atherosclerosis, and virtual autopsy in forensic science [Goyen, 2007]. WB-MRI is also the method of choice for screening and prevention purposes, which is ethically questionable with CT due to radiation exposure of healthy persons. Moreover, contrast agents used for MRI to highlight specific organs or the vascular tree are relatively safe in comparison to iodine based ones used for CT [Goyen, 2007].

These reasons indicate that MRI is challenging CT as standard WB imaging modality. For head and abdomen, the superior performance of MR was already noted, but imaging the lungs is still delicate [Goyen, 2007]. The major drawback, however, remains the longer acquisition time with MRI. We address this issue by using scans with a larger FOV, allowing for covering an equivalent region with less stations. They have the same resolution as scans with a regular FOV, leading to the same acquisition time, but an increased voxel spacing, leading to a negligible loss of image quality. Typical is a normal FOV of 35 cm and an enlarged one of 50 cm. A disadvantage of the enlarged FOV is that spins are excited, which are farther away from the magnetic iso-center and therefore more sensitive to geometrical distortions, see section B.3.

## B.3 Geometrical Distortion in MRI

Essential for MRI is to know the resonance frequency at each position within the FOV, to relate the frequency spectrum of the received RF impulses to the right spatial location. The Lamor equation relates precession frequency of nuclear spins $\varpi$ to the magnetic field, composed of the static field $B_0$ and the slice and row selection gradient fields $G_z$ and $G_x$, respectively

$$\varpi(x, z) = \gamma(B_0 + x \cdot G_x + z \cdot G_z) \tag{B.1}$$

with $\gamma$ the gyromagnetic ratio. Inhomogeneity of the static field or nonlinearity of gradient fields, more prominent farther away from the magnetic iso-center, lead to uncertainties, causing geometrical distortion artifacts. More specifically, these artifacts are referred to be *system-specific* artifacts in contrast to *patient-induced* artifacts arising from susceptibility effects, chemical shift, and flow [Doran et al., 2005]. Scanner with higher field strength and shorter bore magnets, the current trend in MRI, are more sensitive to distortion effects, putting its correction back into focus [Doran et al., 2005].

In the literature, mainly system-specific artifacts are discussed. [Chang and Fitzpatrick, 1992] correct for $B_0$ distortion by acquiring two almost identical images only differing in the polarity of the read-out gradient. This enables for an exact correction of $B_0$ distortion. In follow-up articles [Kannengiesser et al., 1999] and [Reinsberg et al., 2005] refine this method by using deformable image registration techniques to make it also applicable to real MR images and not only phantom scans. This approach is hardly applicable in the clinical workflow because acquisition time would double.

[Doran et al., 2005] analyze the distortion with phantom scans and apply the deduced correction field to medical data. It is doubtful if corrections based on phantom scans make sense because it is not feasible to build a phantom duplicating a biological system, and distortions calculated at fiducial locations have to be interpolated to create a dense mapping, which limits the accuracy [Chang and Fitzpatrick, 1992]. Our approach has the advantage that no additional images have to be acquired, enabling a seamless integration in the clinical workflow. Additionally, the correction is not limited to system-induced distortions but also covers patient-induced ones.

## B.4 Deformable Mosaicing

In order to introduce our approach of deformable mosaicing, we define the two volumes to be stitched as $I_1 : \Omega_1 \subset \mathbb{R}^3 \to \mathbb{R}$ and $I_2 : \Omega_2 \subset \mathbb{R}^3 \to \mathbb{R}$. The overlapping domain is denoted as $\Omega_o = \Omega_1 \cap \Omega_2$. Since the overlap $\Omega_o$ is the only part where the two images share any information, a naive approach for the mosaicing could be defined as an optimization problem with respect to a certain similarity measure $\rho(\cdot)$

$$\hat{\mathcal{T}}_{1,2} = \arg\min_{\mathcal{T}_{1,2}} \int_{\Omega_o} \rho(I_1(\mathcal{T}_1(\mathbf{x})) - I_2(\mathcal{T}_2(\mathbf{x}))) \, \mathrm{d}\mathbf{x} \tag{B.2}$$

where $\mathbf{x} = (x, y, z)$ denotes a voxel position, and $\mathcal{T}_{1,2}$ are the parameters of the transformations $\mathcal{T}_1$ and $\mathcal{T}_2$ relating the two volumes in the spatial domain. The most common approach in pairwise registration is to assume that one of the two transformations is equal to the identity transformation. In our case, such an approach would lead to several problems: (i) through the selection of a moving and a fixed image, we would introduce a certain bias on the mosaicing result, (ii) since both volumes are distorted due to the inhomogeneous magnetic field in the overlap volume, none of them is actually representing a good reference for the mosaicing, and (iii) a registration performed only within the overlap may result in discontinuities with respect to the rest of the volumes. In order to overcome these problems, we propose an iterative simultaneous registration using a *linear weighted average*. The idea of the weighted average is to account for the underlying physical properties of increasing distortions towards the volume boundaries. Assuming that the boundary information is less reliable, we would like to reduce its influence to the registration.

### B.4.1 Simultaneous Registration to Linear Weighted Average

Let us define another volume $S : \Omega_s$ on the union of the two volume domains $\Omega_s = \Omega_1 \cup \Omega_2$. The intensities of $S$ are set using our average model, or

$$S(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega_o \\ I_1(\mathcal{T}_1(\mathbf{x})), & \text{if } \mathbf{x} \in \Omega_1 \setminus \Omega_2 \\ I_2(\mathcal{T}_2(\mathbf{x})), & \text{if } \mathbf{x} \in \Omega_2 \setminus \Omega_1 \end{cases} \tag{B.3}$$

where $f(\cdot)$ is a function computing the linear weighting in the overlap volume, or

$$f(\mathbf{x}) = (1 - h(\mathbf{x})) \cdot I_1(\mathcal{T}_1(\mathbf{x})) + h(\mathbf{x}) \cdot I_2(\mathcal{T}_2(\mathbf{x})). \tag{B.4}$$

The linear function $h(\cdot)$ has a range of $(0, 1)$ and is defined for the overlap domain $\Omega_o$ with respect to the stitching direction. In our application, this direction is usually along the head-feet axis which corresponds to the y-axis of our common 3D coordinate system for all the MRI volumes.

The setup for the deformable stitching and the initialization of the linear weighted average is illustrated in Fig. B.2. We can reformulate the naive registration in Eq. (B.2) in order to pose a simultaneous registration based on the linear weighted average $S$. In terms of an energy function (which is to be minimized), we define

$$E_{\text{data}}(\mathcal{T}_{1,2}) = \sum_{i=1}^{2} \int_{\Omega_o} \rho(S(\mathbf{x}) - I_i(\mathcal{T}_i(\mathbf{x}))) \, \mathrm{d}\mathbf{x}. \tag{B.5}$$

In order to reduce the dimensionality of the problem, we consider free-form deformations [Rueckert et al., 1999] as the transformation model for the two images. A deformation grid $G : [1, K] \times [1, L] \times [1, M]$ is superimposed onto the volume domain $\Omega_s$. By deforming the grid (with a 3D displacement vector $\mathbf{d_p}$ for each

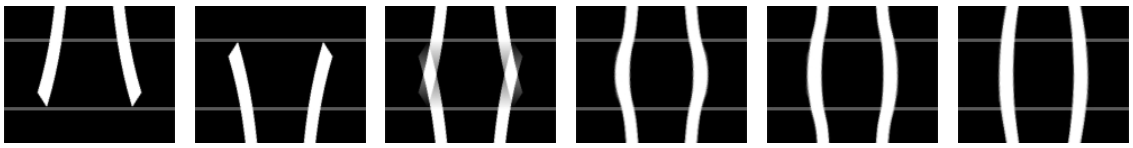


Figure B.2: Synthetic example of a deformable stitching. The first and second image are to be stitched where both are significantly distorted. The initialization of our linear weighted average is shown in the third image. The horizontal gray lines indicate the borders of the overlap area. Fourth to sixth image is an illustration of the registration progress and the iterative improvement of the linear weighted average.

control point) the underlying structures are aligned. The transformation of a voxel $\mathbf{x}$ can be expressed using a combination of basis functions, or

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathcal{D}(\mathbf{x}) \quad \text{with} \quad \mathcal{D}(\mathbf{x}) = \sum_{\mathbf{p} \in G} \eta(|\mathbf{x} - \mathbf{p}|)\, \mathbf{d}_\mathbf{p} \tag{B.6}$$

where $\eta(\cdot)$ is the weighting function (based on cubic B-Splines) measuring the contribution of the control point $\mathbf{p}$ to the displacement field $\mathcal{D}$.

Now, we can rewrite the objective function defined in Eq. (B.5) based on the two deformation grids $G_1$ and $G_2$, or

$$E_{\text{data}}(\mathcal{T}_{1,2}) = \sum_{i=1}^{2} \frac{1}{|G_i|} \sum_{\mathbf{p} \in G_i} \int_{\Omega_o} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \cdot \rho(S(\mathbf{x}) - I_i(\mathcal{T}_i(\mathbf{x})))\, \mathrm{d}\mathbf{x}. \tag{B.7}$$

where $\hat{\eta}(\cdot)$ computes the influence of a voxel $\mathbf{x}$ to a control point $\mathbf{p}$. Such a function acts as a *projection* of the distance/similarity measure computed from the volume domain back to the coarser level of control points. Different definitions of the $\hat{\eta}(\cdot)$ have to be considered with respect to the used similarity measure. We use NCC, which is robust to intensity variations common in MRI. We define

$$\hat{\eta}(|\mathbf{x} - \mathbf{p}|) = \begin{cases} 1, & \text{if } \eta(|\mathbf{x} - \mathbf{p}|) > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{B.8}$$

Basically, this function masks voxels influenced by a control point $\mathbf{p}$ resulting in a local image patch centered at the control point. From this patch, a *local* similarity measure can then be computed.

The simultaneous registration to an average should overcome the problems for the reference selection, mentioned before. This is very similar to atlas construction approaches where the average is used as the reference image in order to achieve an unbiased coordinate frame (e.g. for shape models) [Joshi et al., 2004]. In addition, we try to account for the increasing distortions using a linear weighted average.

### B.4.2 Optimization through Discrete Labeling

We propose to define the simultaneous registration as a discrete optimization problem. Discrete optimization has been recently shown to provide very good results in the case of standard pairwise registration [Glocker et al., 2007]. Based on the previous assumptions, we define a set of discrete variables $G_{\text{mrf}} = G_1 \cup G_2$. Thus, each variable corresponds to a control point of one of the two deformation grids. Similar to [Glocker et al., 2007], we consider a discrete set of labels $L = \{l^1, ..., l^i\}$ corresponding to a quantized version of the deformation space $\Theta = \{\mathbf{d}^1, ..., \mathbf{d}^i\}$. A label assignment $l_\mathbf{p}$ to a grid node $\mathbf{p}$ is associated with displacing the node by the corresponding vector $\mathbf{d}^{l_\mathbf{p}}$. If a label is assigned to every node we get a *discrete labeling* $\mathbf{l}$. A popular model for representing discrete labeling problems are Markov Random Fields (MRFs) [Li, 2001]. The general form of a first-order MRF is

$$E_{\text{mrf}}(\mathbf{l}) = \sum_{\mathbf{p} \in G_{\text{mrf}}} V_\mathbf{p}(l_\mathbf{p}) + \sum_{(\mathbf{p},\mathbf{q}) \in \mathcal{E}_{\text{mrf}}} V_{\mathbf{pq}}(l_\mathbf{p}, l_\mathbf{q}) \tag{B.9}$$

where $V_\mathbf{p}(\cdot)$ are the unary potentials representing the data term, $V_{\mathbf{pq}}(\cdot,\cdot)$ are the pairwise potentials representing the smoothness term, and $\mathcal{E}_{\text{mrf}}$ represents the neighborhood system represented by edges between nodes. We define the unary potentials (in iteration $t$) according to our data term $\forall \mathbf{p} \in G_i$ as

$$V_\mathbf{p}(l_\mathbf{p}) = \int_{\Omega_o} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \cdot \rho(S(\mathbf{x}) - I_i(\mathcal{T}_i^{t-1}(\mathbf{x}) + \mathbf{d}^{l_\mathbf{p}}))d\mathbf{x}. \tag{B.10}$$

The pairwise potentials encode a penalty term for assigning different labels to neighboring nodes. The FFD transformation model already inherits some implicit smoothness properties. Additionally, one can consider explicit regularization constraints on the grid domain using the pairwise potentials. These are defined $\forall(\mathbf{p},\mathbf{q}) \in \mathcal{E}_{\text{mrf}} \wedge \mathbf{p},\mathbf{q} \in G_i$ as

$$V_{\mathbf{pq}}(l_\mathbf{p}, l_\mathbf{q}) = \lambda \left| (\mathcal{R}_i(\mathbf{p}) + \mathbf{d}^{l_\mathbf{p}}) - (\mathcal{R}_i(\mathbf{q}) + \mathbf{d}^{l_\mathbf{q}}) \right| \tag{B.11}$$

where $\lambda$ denotes a weighting factor for the smoothness term and $\mathcal{R}_i(\cdot)$ back-projects the accumulated displacement field (of iteration $t-1$) on the control point level

$$\mathcal{R}_i(\mathbf{p}) = \int_{\Omega_s} \hat{\eta}(|\mathbf{x} - \mathbf{p}|)\, \mathcal{D}_i^{t-1}(\mathbf{x})\, d\mathbf{x}. \tag{B.12}$$

In contrast to the data term energy, the smoothness energy affects the whole stitching domain $\Omega_s$. Such an approach together with the use of smooth FFD transformations guarantees continuous and seamless transitions between the overlapping and non-overlapping areas of the stitched volume while the actual stitch is (softly) constrained to the volume of overlap.

Many optimization algorithms exist for efficiently solving discrete labeling problems in forms of an MRF [Li, 2001]. We use a recently proposed method called Fast-PD [Komodakis et al., 2007].

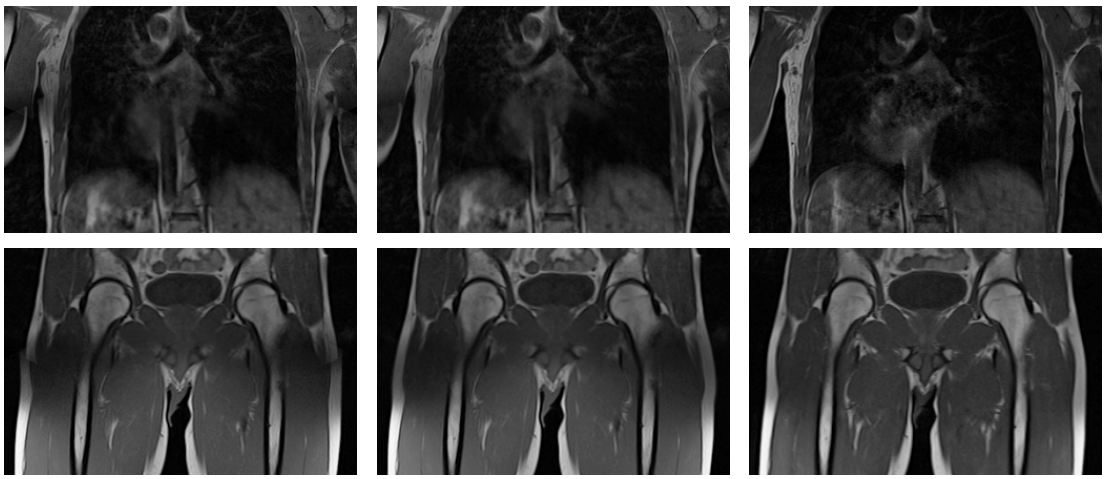


Figure B.3: The two rows are magnifications of the stitching areas of the WB-MRI shown in Fig. B.1. Left: Initial average. Middle: Final stitching result after 3 optimization cycles. Right: Reference scan where the overlap volume is centered within the MR scanner. Our method is able to reproduce similar smooth and continuous transitions as present in the reference images.

### B.4.3 Iterative Multi-Scale Approach

We propose to embed the simultaneous registration into a common iterative multi-scale approach. The simultaneous registration of the two volumes $I_1$ and $I_2$ to the linear weighted average $S$ is performed in a pyramidal setup where different levels of resolution for the volume as well as for the deformation grids are considered. On each level, several discrete labelings are computed where the set of displacements $\Theta$ is successively refined each time and the displacement fields are incrementally updated. After the registration converges, the linear weighted average $S$ is recomputed and a new registration cycle is started. In Fig. B.2 three of such cycles are illustrated for the case of synthetic data. Usually, only a few cycles are needed until the average shows no dramatic changes anymore.

## B.5 Experimental Validation

We evaluate our method on 8 whole-body T1- and T2-weighted data sets from three different Siemens MR scanners: Avanto 1.5T, Trio 3T, and Espree 1.5T. The overlaps vary between 5 and 27 cm. For all stitching results, we obtained very positive feedback from our clinical experts in the radiology department who inspected the images visually. An example mosaic is shown in Fig. B.1, consisting of three volumes having a FOV of $50 \times 50 \times 28$ cm$^3$, a resolution of $448 \times 448 \times 35$ voxel, and an overlap of 5 cm. In Fig. B.3, one can clearly see the influence of the distortion by regarding the initial average, and the improvement after the deformation by comparing the result to the reference scan. The final resolution for the 3 station stitch shown in Fig. B.1 is $448 \times 1256 \times 35$ where the two stitches take together approximately 25

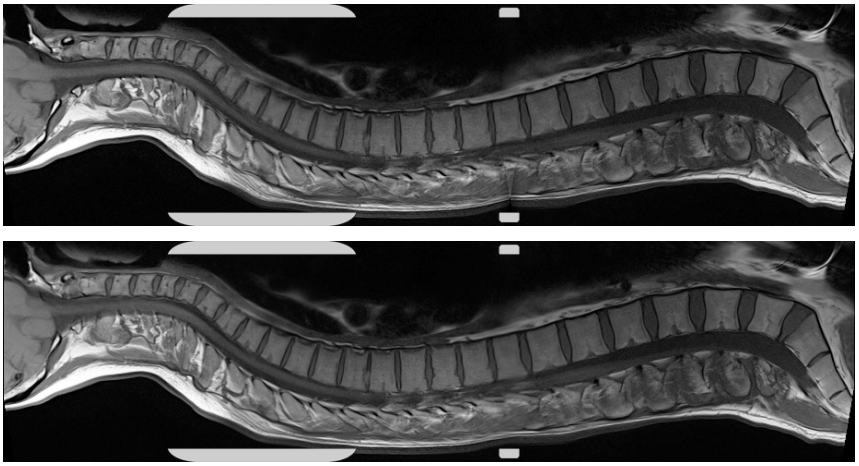


Figure B.4: Stitching of 3 spine volumes. Top: initial average. Bottom: result. Gray bars indicate overlap.

min of computational time.

To illustrate that the proposed method also works for varying overlaps, we show the stitching of 3 volumes for whole-spine MR, see Fig. B.4. The first overlap is with 15.2 cm very large and our method arrives at producing a sharper average. The second one, with only 1.4 cm, shows discontinuities in the initial average, which are removed after deformable mosaicing.

## B.6 Conclusion

Speeding up the acquisition for WB-MRI with large FOV images leads to significant distortions towards the boundaries. Methods for distortion correction proposed in the literature are not applicable to the WB imaging setup because they either elongate the workflow or only correct for specific system-induced distortions. So far, the overlap in WB-MRI has not been used to correct for distortion. We propose the usage of simultaneous deformable registration in a mosaicing scenario. Key for the simultaneous registration is the creation of a linearly weighted average, each of the two images is registered to. Our experiments on synthetic and in-vivo data show the ability of the method to correct for distortions. The unaltered clinical workflow makes our approach very interesting for being integrated into further MR scanner generations.

# List of Publications and Patents

## Books

- C. Wachinger. Insights Into Three-Dimensional Ultrasound Mosaicing: Strategies for the alignment of multiple ultrasound volumes. In *VDM Verlag Dr. Müller*, Stuttgart, ISBN: 3836488264, 2008.

## Journal Publications

- C. Wachinger, N. Navab. Entropy and Laplacian Images: Structural Representations for Multi-Modal Registration. In *Medical Image Analysis*, accepted for publication.

- C. Wachinger, W. Wein, N. Navab. Registration Strategies and Similarity Measures for Three-Dimensional Ultrasound Mosaicing. In *Journal of Academic Radiology*, 2008.

## Conference Publications

- S. Demirci, A. Bigdelou, L. Wang, C. Wachinger, M. Baust, R. Tibrewal, R. Ghotbi, H. Eckstein, N. Navab. 3D Stent Recovery from One X-ray Projection. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2011.

- C. Wachinger, T. Klein, N. Navab. The 2D Analytic Signal on RF and B-mode Ultrasound Images. In *Information Processing in Medical Imaging (IPMI)*, 2011.

- M. Yigitsoy[1], C. Wachinger[1], N. Navab. Temporal Groupwise Registration for Motion Modeling. In *Information Processing in Medical Imaging (IPMI)*, 2011.

- M. Yigitsoy, C. Wachinger, N. Navab. Manifold Learning for Image-Based Breathing Gating in MRI. In *SPIE Medical Imaging*, 2011.

[1] Joint first authors.

- C. Wachinger, M. Yigitsoy, N. Navab. Manifold Learning for Image-Based Breathing Gating with Application to 4D Ultrasound. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 2, pp. 26-33, 2010.

- C. Wachinger, N. Navab. Manifold Learning for Multi-Modal Image Registration. In *21st British Machine Vision Conference (BMVC)*, pp. 82.1-82.12, 2010.

- C. Wachinger, N. Navab. Structural Image Representation for Image Registration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pp. 23-30, 2010.

- C. Wachinger, D. Mateus, A. Keil, N. Navab. Manifold Learning for Patient Position Detection in MRI. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1353 - 1356, 2010.

- C. Wachinger, N. Navab. Alignment of Viewing-Angle Dependent Ultrasound Images. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 1, pp. 667-674, 2009.

- C. Wachinger, N. Navab. Similarity Metrics and Efficient Optimization for Simultaneous Registration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-786, 2009.

- C. Wachinger, S. Baumann, J. Zeltner, B. Glocker, N. Navab. Sphere Extraction in MR Images with Application to Whole-Body MRI. In *SPIE Medical Imaging*, pp. 72594J.1-72594J.8, 2009.

- C. Wachinger, R. Shams, N. Navab. Towards an Estimation of Acoustic Impedance from Multiple Ultrasound Images. In *Bildverarbeitung für die Medizin*, pp. 405-409, 2009.

- B. Glocker, C. Wachinger, J. Zeltner, N. Paragios, N. Komodakis, M. Hansen, N. Navab. MRI Composing for Whole Body Imaging. In *Bildverarbeitung für die Medizin*, pp. 420-424, 2009.

- C. Wachinger, B. Glocker, J. Zeltner, N. Paragios, N. Komodakis, M. S. Hansen, N. Navab. Deformable Mosaicing for Whole-Body MRI. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 2, pp. 113-121, 2008.

- C. Wachinger, R. Shams, N. Navab. Estimation of Acoustic Impedance from Multiple Ultrasound Images with Application to Spatial Compounding. In *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pp. 1-8, 2008.

- C. Wachinger, N. Navab. Ultrasound Specific Similarity Measures for Three-Dimensional Mosaicing. In *SPIE Medical Imaging*, pp. 69140F.1-69140F.9, 2008.
- C. Wachinger, W. Wein, N. Navab. Three-Dimensional Ultrasound Mosaicing. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 2, pp.327-335, 2007.
- A. Keil, C. Wachinger, G. Brinker, S. Thesen, N. Navab. Patient Position Detection for SAR Optimization in Magnetic Resonance Imaging. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 2, pp.49-57, 2006.

# Patents

- C. Wachinger, N. Navab. Anwendung eines zweidimensionalen analytischen Signals in der Sonographie. Patent Pending.
- C. Wachinger, B. Glocker, J. Zeltner, N. Navab. Method for combining images and magnetic resonance scanner, Patent application number: 12/461993, Date: 03/18/2010.

# Bibliography

[Abd-Elmoniem et al., 2002] Abd-Elmoniem, K., Youssef, A., and Kadah, Y. (2002). Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion. *Biomedical Engineering, IEEE Transactions on*, 49(9):997–1014.

[Aja-Fernandez et al., 2007] Aja-Fernandez, S., Martin-Fernandez, M., and Alberola-Lopez, C. (2007). Tissue identification in ultrasound images using rayleigh local parameter estimation. *Bioinformatics and Bioengineering. Proceedings of the 7th IEEE International Conference on*, pages 1129–1133.

[Ali et al., 2008] Ali, R., Gooding, M., Christlieb, M., and Brady, M. (2008). Advanced phase-based segmentation of multiple cells from brightfield microscopy images. In *ISBI*, pages 181–184.

[Anderson et al., 1997] Anderson, M., Soo, M., Bentley, R., and Trahey, G. (1997). The detection of breast microcalcifications with medical ultrasound. *The Journal of the Acoustical Society of America*, 101:29.

[Andronache et al., 2008] Andronache, A., von Siebenthal, M., Székely, G., and Cattin, P. (2008). Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical Image Analysis*, 12(1):3–15.

[Arandjelovic and Cipolla, 2007] Arandjelovic, O. and Cipolla, R. (2007). A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. *ICCV*.

[Arsigny et al., 2006] Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A Log-Euclidean framework for statistics on diffeomorphisms. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 924–931.

[Ashburner and Friston, 2005] Ashburner, J. and Friston, K. (2005). Unified segmentation. *Neuroimage*, 26(3):839–851.

[Averbuch et al., 2006] Averbuch, A., Coifman, R., Donoho, D., Elad, M., and Israeli, M. (2006). Fast and accurate polar Fourier transform. *Applied and Computational Harmonic Analysis*, 21(2):145–167.

[Azar et al., 2006] Azar, A., Xu, C., Pennec, X., and Ayache, N. (2006). An interactive hybrid non-rigid registration framework for 3d medical images. In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, pages 824–827.

[Bai and Brady, 2009] Bai, W. and Brady, S. (2009). Spatio-temporal image registration for respiratory motion correction in pet imaging. In *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro*, pages 426–429. IEEE Press.

[Baker and Matthews, 2004] Baker, S. and Matthews, I. (2004). Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, 56(3):221–255.

[Balci et al., 2007] Balci, S., Golland, P., and Wells, W. (2007). Non-rigid groupwise registration using b-spline deformation model. In *MICCAI Workshop on Open-Source and Open-Data*, pages 106–122. Citeseer.

[Bardera et al., 2006] Bardera, A., Feixas, M., Boada, I., and Sbert, M. (2006). High-dimensional normalized mutual information for image registration using random lines. *Lecture Notes in Computer Science*, 4057:264–271.

[Basarab et al., 2007] Basarab, A., Aoudi, W., Liebgott, H., Vray, D., and Delachartre, P. (2007). Parametric deformable block matching for ultrasound imaging. In *Image Processing (ICIP). IEEE International Conference on*, volume 2, pages II –429 –II –432.

[Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359.

[Behar et al., 2003] Behar, V., Adam, D., and Friedman, Z. (2003). A new method of spatial compounding imaging. *Ultrasonics*, 41:377–384.

[Belkin and Niyogi, 2003] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6).

[Belongie et al., 2002] Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 509–522.

[Benhimane, 2006] Benhimane, S. (2006). *Vers une approche unifiee pour le suivi temps-reel et l'asservissement visuel*. Docteur en sciences - specialite: Informatique temps-reel, automatique et robotique, Ecole Nationale Superieure des Mines de Paris.

[Benhimane and Malis, 2004] Benhimane, S. and Malis, E. (2004). Real-time image-based tracking of planes using efficient second-order minimization. In *IEEE/RSJ*, pages 943–948.

[Biesdorf et al., 2009] Biesdorf, A., Wörz, S., Kaiser, H.-J., Stippich, C., and Rohr, K. (2009). Hybrid spline-based multimodal registration using local measures for joint entropy and mutual information. In Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., and Taylor, C., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5761 of *Lecture Notes in Computer Science*, pages 607–615. Springer.

[Bishop, 2008] Bishop, C. (2008). A New Framework for Machine Learning. *Lecture Notes in Computer Science*, 5050:1.

[Bishop and Lasserre, 2007] Bishop, C. and Lasserre, J. (2007). Generative or discriminative? getting the best of both worlds. *Bayesian Statistics*, 8:3–24.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag.

[Bock and Krischer, 1998] Bock, R. K. and Krischer, W. (1998). *The Data Analysis Briefbook*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition.

[Bossa et al., 2007] Bossa, M., Hernandez, M., and Olmos, S. (2007). Contributions to 3d diffeomorphic atlas estimation: Application to brain images. In *Medical Image Computing and Computer-Assisted Intervention*, volume 4791 of *Lecture Notes in Computer Science*, pages 667–674. Springer Berlin / Heidelberg.

[Bossa et al., 2008] Bossa, M., Zacur, E., and Olmos, S. (2008). Algorithms for computing the group exponential of diffeomorphisms: Performance evaluation. In *Computer Vision and Pattern Recognition Workshops. IEEE Computer Society Conference on*, pages 1 –8.

[Bouhlel and Sevestre-Ghalila, 2009] Bouhlel, N. and Sevestre-Ghalila, S. (2009). Nakagami markov random field as texture model for ultrasound rf envelope image. *Computers in Biology and Medicine*, 39(6):535–544.

[Boukerroui et al., 2003] Boukerroui, D., Noble, J., and Brady, M. (2003). Velocity estimation in ultrasound images: a block matching approach. *Inf Process Med Imaging*, 18:586–98.

[Boukerroui et al., 2004] Boukerroui, D., Noble, J., and Brady, M. (2004). On the Choice of Band-Pass Quadrature Filters. *Journal of Mathematical Imaging and Vision*, 21(1):53–80.

[Brant and Helms, 2007] Brant, W. and Helms, C. (2007). *Fundamentals of diagnostic radiology*. Lippincott Williams & Wilkins.

[Bronstein et al., 2010] Bronstein, M., Bronstein, A., Michel, F., and Paragios, N. (2010). Data fusion through cross-modality metric learning using similarity-sensitive hashing. In *Computer Vision and Pattern Recognition (CVPR). IEEE Conference on*, pages 3594–3601. IEEE.

[Brown, 1992] Brown, L. G. (1992). A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376.

[Brown, 1959] Brown, T. (1959). Direct contact ultrasonic scanning techniques for the visualization of abdominal masses. In *Proceedings of the second International Conference on Medical Electronics*, page 358.

[Brox et al., 2009] Brox, T., Bregler, C., and Malik, J. (2009). Large Displacement Optical Flow . In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Burckhardt, 1978] Burckhardt, C. (1978). Speckle in ultrasound B-mode scans. *IEEE Transations on Sonics and Ultrasonics*, 25(1):1–6.

[Buzug et al., 1997] Buzug, T., Weese, J., Fassnacht, C., and Lorenz, C. (1997). Image registration: Convex weighting functions for histogram-based similarity measures. *Lecture Notes in Computer Science*, pages 203–212.

[Bystrov et al., 2009] Bystrov, D., Vik, T., Schulz, H., Klinder, T., and Schmidt, S. (2009). Local motion analysis in 4D lung CT using fast groupwise registration. In *Proceedings of the 16th IEEE international conference on Image processing*, pages 1729–1732. IEEE Press.

[Cahill et al., 2008] Cahill, N., Schnabel, J., Noble, J., and Hawkes, D. (2008). Revisiting overlap invariance in medical image alignment. In *Computer Vision and Pattern Recognition Workshops*.

[Carneiro and Jepson, 2002] Carneiro, G. and Jepson, A. D. (2002). Phase-based local features. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 282–296, London, UK. Springer-Verlag.

[Caruana and Niculescu-Mizil, 2006] Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *ICML*, pages 161–168.

[Castillo et al., 2010] Castillo, E., Castillo, R., Martinez, J., Shenoy, M., and Guerrero, T. (2010). Four-dimensional deformable image registration using trajectory modeling. *Physics in Medicine and Biology*, 55:305.

[Chan et al., 2006] Chan, T., Esedoglu, S., and Nikolova, M. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM journal on applied mathematics*, 66(5):1632–1648.

[Chang and Fitzpatrick, 1992] Chang, H. and Fitzpatrick, J. (1992). A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE TMI*, 11(3):319–329.

[Chefd'hotel et al., 2002] Chefd'hotel, C., Hermosillo, G., and Faugeras, O. (2002). Flows of diffeomorphisms for multimodal image registration. In *IEEE International Symposium on Biomedical Imaging*, pages 753–756.

[Chen and Varshney, 2003] Chen, H. and Varshney, P. (2003). Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. *IEEE Transactions on medical imaging*, 22(9):1111–1119.

[Choi and Lee, 2000] Choi, Y. and Lee, S. (2000). Injectivity conditions of 2d and 3d uniform cubic b-spline functions. *Graphical Models*, 62(6):411 – 427.

[Chung et al., 2002] Chung, A. C. S., III, W. M. W., Norbash, A., and Grimson, W. E. L. (2002). Multi-modal image registration by minimising kullback-leibler distance. In *Medical Image Computing and Computer-Assisted Intervention*.

[Cincotti et al., 2001] Cincotti, G., Loi, G., and Pappalardo, M. (2001). Frequency decomposition and compounding of ultrasound medical images with wavelet packets. *IEEE transactions on medical imaging*, 20(8):764–771.

[Cobbold, 2007] Cobbold, R. (2007). *Foundations of biomedical ultrasound*. Oxford University Press, USA.

[Cohen and Dinstein, 2002] Cohen, B. and Dinstein, I. (2002). New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition*, 35(2):455–463.

[Cole-Rhodes et al., 2003] Cole-Rhodes, A., Johnson, K., LeMoigne, J., and Zavorin, I. (2003). Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *Image Processing, IEEE Transactions on*, 12(12):1495–1511.

[Colgan et al., 2008] Colgan, R., McClelland, J., McQuaid, D., Evans, P., Hawkes, D., Brock, J., Landau, D., and Webb, S. (2008). Planning lung radiotherapy using 4D CT data and a motion model. *Physics in Medicine and Biology*, 53:5815.

[Collignon et al., 1995] Collignon, A., Vandermeulen, D., Suetens, P., and Marchal, G. (1995). 3D multi-modality medical image registration using feature space clustering. In *Computer Vision, Virtual Reality and Robotics in Medicine*.

[Cootes et al., 2004] Cootes, T., Marsland, S., Twining, C., Smith, K., and Taylor, C. (2004). Groupwise Diffeomorphic Non-rigid Registration for Automatic Model Building. In *European Conference on Computer Vision*.

[Cox et al., 2008] Cox, M., Lucey, S., Sridharan, S., and Cohn, J. (2008). Least squares congealing for unsupervised alignment of images. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Cox et al., 2009] Cox, M., Sridharan, S., Lucey, S., and Cohn, J. (2009). Least-squares congealing for large numbers of images. In *IEEE International Conference on Computer Vision*, pages 1949–1956.

[Crum et al., 2004] Crum, W., Hartkens, T., and Hill, D. (2004). Non-rigid image registration: theory and practice. *British journal of radiology*, 77(Special Issue 2):S140.

[Crum et al., 2005] Crum, W., Tanner, C., and Hawkes, D. (2005). Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. *Physics in Medicine and Biology*, 50:5153.

[D'Agostino et al., 2006] D'Agostino, E., Maes, F., Vandermeulen, D., and Suetens, P. (2006). A unified framework for atlas based brain image segmentation and registration. In *WBIR*, pages 136–143.

[D'Agostino and Stephens, 1986] D'Agostino, R. and Stephens, M. (1986). Goodness-of-fit techniques. *Marcel Dekker, Inc. New York, NY, USA*, page 560.

[de Winter et al., 2003] de Winter, S., Hamers, R., Degertekin, M., Tanabe, K., Lemos, P., Serruys, P., Roelandt, J., and Bruining, N. (2003). A novel retrospective gating method for intracoronary ultrasound images based on image properties. In *Computers in Cardiology*.

[Destrempes and Cloutier, 2010] Destrempes, F. and Cloutier, G. (2010). A Critical Review and Uniformized Representation of Statistical Distributions Modeling the Ultrasound Echo Envelope. *Ultrasound in medicine & biology*, 36(7):1037–1051.

[Destrempes et al., 2009] Destrempes, F., Meunier, J., Giroux, M.-F., Soulez, G., and Cloutier, G. (2009). Segmentation in ultrasonic b-mode images of healthy carotid arteries using mixtures of nakagami distributions and stochastic optimization. *IEEE Trans Med Imaging*, 28(2):215–229.

[Dietrich et al., 2002] Dietrich, C., Ignee, A., Gebel, M., Braden, B., and Schuessler, G. (2002). Imaging of the abdomen. *Z Gastroenterol*, 40:965–970.

[Doran et al., 2005] Doran, S., Charles-Edwards, L., Reinsberg, S., and Leach, M. (2005). A complete distortion correction for MR images: I. Gradient warp correction. *Physics in Medicine and Biology*, 50(7):1343–1361.

[Dowson et al., 2008] Dowson, N., Kadir, T., and Bowden, R. (2008). Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1841–1857.

[Du et al., 2010] Du, S., Zheng, N., Ying, S., and Liu, J. (2010). Affine iterative closest point algorithm for point set registration. *Pattern Recogn. Lett.*, 31(9):791–799.

[Dussik, 1942] Dussik, K. (1942). Über die Möglichkeit, hochfrequente mechanische Schwingungen als diagnostisches Hilfsmittel zu verwerten. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 174(1):153–168.

[Dutt, 1995] Dutt, V. (1995). *Statistical analysis of ultrasound echo envelope*. PhD Thesis, Mayo Graduate School.

[Dutt and Greenleaf, 1996] Dutt, V. and Greenleaf, J. (1996). Statistics of the log-compressed echo envelope. *The Journal of the Acoustical Society of America*, 99:3817.

[Dutt and Greenleaf, 1994] Dutt, V. and Greenleaf, J. F. (1994). Ultrasound echo envelope analysis using a homodyned k-distribution signal model. *Ultrason. Imag.*, 16(4-5):265–287.

[Ehrhardt et al., 2006] Ehrhardt, J., Werner, R., Frenzel, T., Säring, D., Lu, W., Low, D., and Handels, H. (2006). Reconstruction of 4D-CT data sets acquired during free breathing for the analysis of respiratory motion. In *Proc. of SPIE Vol*, volume 6144, pages 614414–1. Citeseer.

[El-Baz and Gimel'farb, 2008] El-Baz, A. and Gimel'farb, G. (2008). Global image registration based on learning the prior appearance model. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 1–7.

[Elen et al., 2008] Elen, A., Choi, H. F., Loeckx, D., Gao, H., Claus, P., Suetens, P., Maes, F., and D'hooge, J. (2008). Three-dimensional cardiac strain estimation using spatio temporal elastic registration of ultrasound images: A feasibility study. *Medical Imaging, IEEE Transactions on*, 27(11):1580 –1591.

[Eltoft, 2003] Eltoft, T. (2003). Speckle: Modeling and filtering. In *Norwegian Signal Process. Symp.*

[Estepar et al., 2006] Estepar, R., Washko, G., Silverman, E., Reilly, J., Kikinis, R., and Westin, C. (2006). Accurate airway wall estimation using phase congruency. *MICCAI*.

[Esther Leung et al., 2008] Esther Leung, K., van Stralen, M., Nemes, A., Voormolen, M., van Burken, G., Geleijnse, M., ten Cate, F., Reiber, J., de Jong, N., van der Steen, A., and Bosch, J. (2008). Sparse registration for three-dimensional stress echocardiography. *Medical Imaging, IEEE Transactions on*, 27(11):1568–1579.

[Feldmann et al., 2006] Feldmann, T., Bouattour, S., Paulus, D., and Deinzer, F. (2006). Kombination verschiedener Ähnlichkeitsmaße für die 2D/3D-Registrierung von Röntgenbildern mittels Demokratischer Integration. In *Bildverarbeitung für die Medizin 2006*, pages 226–230. Bildverarbeitung für die Medizin, Springer Verlag.

[Felsberg and Sommer, 2001] Felsberg, M. and Sommer, G. (2001). The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144.

[Felsberg and Sommer, 2004] Felsberg, M. and Sommer, G. (2004). The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and vision*, 21(1):5–26.

[Flampouri et al., 2006] Flampouri, S., Jiang, S., Sharp, G., Wolfgang, J., Patel, A., and Choi, N. (2006). Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4D-CT data and Monte Carlo simulations. *Physics in Medicine and Biology*, 51:2763.

[Fleet et al., 1991] Fleet, D., Jepson, A., and Jenkin, M. (1991). Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210.

[Gauß, 1808] Gauß, K. F. (1808). Letter to bolyai.

[Gay-bellile et al., 2007] Gay-bellile, V., Bartoli, A., and Sayd, P. (2007). Feature-driven direct non-rigid image registration. In *British Machine Vision Conference (BMVC)*.

[Gee et al., 2003] Gee, A. H., Treece, G. M., Prager, R. W., Cash, C. J. C., and Berman, L. H. (2003). Rapid registration for wide field-of-view freehand 3d ultrasound. *IEEE Trans. Med. Imaging*, 22(11):1344–1357.

[Gemmeke and Ruiter, 2007] Gemmeke, H. and Ruiter, N. (2007). 3d ultrasound computer tomography for medical imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 580(2):1057 – 1065.

[Georg et al., 2008] Georg, M., Souvenir, R., Hope, A., and Pless, R. (2008). Manifold learning for 4d ct reconstruction of the lung. In *Computer Vision and Pattern Recognition Workshops*.

[Georgel et al., 2008] Georgel, P., Benhimane, S., and Navab, N. (2008). A Unified Approach Combining Photometric and Geometric Information for Pose Estimation . In *19th British Machine Vision Conference (BMVC)*.

[Gerber et al., 2009] Gerber, S., Tasdizen, T., Joshi, S., and Whitaker, R. (2009). On the manifold structure of the space of brain images. In *MICCAI*, volume 5761.

[Glocker et al., 2007] Glocker, B., Komodakis, N., Paragios, N., Tziritas, G., and Navab, N. (2007). Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *IPMI*, Kerkrade, Netherlands.

[Glocker et al., 2008] Glocker, B., Komodakis, N., Tziritas, G., Navab, N., and Paragios, N. (2008). Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741.

[Gonzalez and Woods, 2002] Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing (2nd Edition)*. Prentice Hall.

[Goodman, 2006] Goodman, J. (2006). *Speckle phenomena in optics: theory and applications*. Roberts & Co.

[Goyen, 2007] Goyen, M. (2007). *Real Whole Body MRI: Requirements, Indications, Perspectives*. McGraw-Hill.

[Grady, 2006] Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1768–1783.

[Granlund and Knutsson, 1995] Granlund, G. H. and Knutsson, H. (1995). *Signal Processing for Computer Vision*. Kluwer Academic Publishers.

[Grau et al., 2007] Grau, V., Becher, H., and Noble, J. (Sept. 2007). Registration of multiview real-time 3-d echocardiographic sequences. *Medical Imaging, IEEE Transactions on*, 26(9):1154–1165.

[Grau et al., 2006] Grau, V., Becher, H., and Noble, J. A. (2006). Phase-based registration of multi-view real-time three-dimensional echocardiographic sequences. In *MICCAI*, pages 612–619.

[Grau and Noble, 2005] Grau, V. and Noble, J. A. (2005). Adaptive multiscale ultrasound compounding using phase information. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 589–596.

[Greenleaf and Bahn, 1981] Greenleaf, J. and Bahn, R. (1981). Clinical imaging with transmissive ultrasonic computerized tomography. *Biomedical Engineering, IEEE Transactions on*, BME-28(2):177–185.

[Haber and Modersitzki, 2007] Haber, E. and Modersitzki, J. (2007). Intensity gradient based registration and fusion of multi-modal images. *Methods of information in medicine*, 46(3):292–299.

[Hacihaliloglu et al., 2008] Hacihaliloglu, I., Abugharbieh, R., Hodgson, A., and Rohling, R. (2008). Bone segmentation and fracture detection in ultrasound using 3d local phase features. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 287–295.

[Hagenlocker and Fujimura, 1998] Hagenlocker, M. and Fujimura, K. (1998). CFFD: a tool for designing flexible shapes. *The Visual Computer*, 14(5):271–287.

[Hajnal et al., 2001] Hajnal, J., Hawkes, D., and Hill, D., editors (2001). *Medical Image Registration*. CRC Press, Baton Rouge, Florida.

[Hamm et al., 2009] Hamm, J., Davatzikos, C., and Verma, R. (2009). Efficient large deformation registration via geodesics on a learned manifold of images. In *MICCAI*, volume 5761, pages 680–687.

[Hartkens et al., 2002] Hartkens, T., Hill, D., Castellano-Smith, A., Hawkes, D., Maurer, C., Martin, A., Hall, W., Liu, H., and Truwit, C. (2002). Using points and surfaces to improve voxel-based non-rigid registration. *Lecture Notes in Computer Science*, pages 565–572.

[Hedrick et al., 2004] Hedrick, W. R., Hykes, D. L., and Starchman, D. E. (2004). *Ultrasound Physics and Instrumentation*. Mosby, 4 edition.

[Henrich et al., 2003] Henrich, W., Schmider, A., Kjos, S., Tutschek, B., and Dudenhausen, J. W. (2003). Advantages of and applications for extended field-of-view ultrasound in obstetrics. *Archives of Gynecology and Obstetrics*, V268:121–127.

[Hermosillo et al., 2002] Hermosillo, G., Chefd'Hotel, C., and Faugeras, O. (2002). Variational Methods for Multimodal Image Matching. *International Journal of Computer Vision*, 50(3):329–343.

[Hill et al., 2001] Hill, D., Batchelor, P., Holden, M., and Hawkes, D. (2001). Medical image registration. *Physics in medicine and biology*, 46(3):1–1.

[Hoaglin et al., 1983] Hoaglin, D., Mosteller, F., and Tukey, J. (1983). *Understanding robust and exploratory data analysis*, volume 3. Wiley New York.

[Holden, 2008] Holden, M. (2008). A review of geometric transformations for nonrigid body registration. *Medical Imaging, IEEE Transactions on*, 27(1):111–128.

[Holmes et al., 1954] Holmes, J., Howry, D., Posakony, G., and Cushman, C. (1954). The ultrasonic visualization of soft tissue structures in the human body. *Transactions of the American Clinical and Climatological Association*, 66:208.

[Howry and Gordon, 1964] Howry, D. and Gordon, D. (1964). Ultrasonic tomography. *Ultrasound as a diagnostic and surgical tool. Ed. D. GORDON (Livingstone, Edinburgh & London.*

[Huang et al., 2007] Huang, G., Jain, V., and Learned-Miller, E. (2007). Unsupervised Joint Alignment of Complex Images. In *IEEE International Conference on Computer Vision*, pages 1–8.

[Huber et al., 2002] Huber, S., Wagner, M., Medl, M., and Czembirek, H. (2002). Real-time spatial compound imaging in breast ultrasound. *Ultrasound Med Biol*, 28(2):155–63.

[Huijssen et al., 2003] Huijssen, J., Bouakaz, A., Verweij, M., and de Jong, N. (2003). Simulations of the nonlinear acoustic pressure field without using the parabolic approximation. In *Ultrasonics, IEEE Symposium on*, volume 2, pages 1851 – 1854.

[Hyvärinen et al., 2001] Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. John Wiley and Sons.

[Ionasec et al., 2010] Ionasec, R., Wang, Y., Georgescu, B., Voigt, I., Navab, N., and Comaniciu, D. (2010). Robust motion estimation using trajectory spectrum learning: Application to aortic and mitral valve modeling from 4D TEE. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1601–1608. IEEE.

[Jakeman and Pusey, 1976] Jakeman, E. and Pusey, P. N. (1976). A model for non-rayleigh sea echo. *IEEE Trans. Antennas Propag.*, 24(4-5):806–814.

[Jespersen et al., 1998] Jespersen, S., Wilhjelm, J., and Sillesen, H. (1998). Multi-angle compound imaging. *Ultrasonic imaging*, 20(2):81–102.

[Jespersen et al., 2000] Jespersen, S., Wilhjelm, J., and Sillesen, H. (2000). In vitro Spatial Compound Scanning for Improved Visualization of Atherosclerosis. *Ultrasound in Medicin and Biology*, 26(8):1357–1362.

[Johnson and Christensen, 2002] Johnson, H. and Christensen, G. (2002). Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461.

[Joshi et al., 2004] Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:151–160.

[Jurie and Dhome, 2002] Jurie, F. and Dhome, M. (2002). Hyperplane approximation for template matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):996–1000.

[Kadir and Brady, 2001] Kadir, T. and Brady, M. (2001). Saliency, scale and image description. *Int. J. Comput. Vision*, 45(2):83–105.

[Kadir and Brady, 2005] Kadir, T. and Brady, M. (2005). Estimating statistics in arbitrary regions of interest. In *British Machine Vision Conference*, volume 2.

[Kannengiesser et al., 1999] Kannengiesser, S., Wang, Y., and Haacke, E. (1999). Geometric distortion correction in gradient-echo imaging by use of dynamic time warping. *Magnetic Resonance in Medicine*, 42(3):585–590.

[Kaplan and Ma, 1994] Kaplan, D. and Ma, Q. (1994). On the statistical characteristics of log-compressed Rayleigh signals: Theoretical formulation and experimental results. *The Journal of the Acoustical Society of America*, 95:1396.

[Karadayi et al., 2006] Karadayi, K., Hayashi, T., and Kim, Y. (2006). Automatic image-based gating for 4d ultrasound. In *Engineering in Medicine and Biology Society*.

[Karamalis et al., 2010] Karamalis, A., Wein, W., and Navab, N. (2010). Fast ultrasound image simulation using the westervelt equation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Beijing, China.

[Keil et al., 2006] Keil, A., Wachinger, C., Brinker, G., Thesen, S., and Navab, N. (2006). Patient position detection for SAR optimization in magnetic resonance imaging. In *MICCAI*, volume 4191, pages 49–57.

[Kennedy et al., 2003] Kennedy, J., Ter Haar, G., and Cranston, D. (2003). High intensity focused ultrasound: surgery of the future? *British Journal of Radiology*, 76(909):590.

[Kim et al., 2003] Kim, S. H., Choi, B. I., Kim, K. W., Lee, K. H., and Han, J. K. (2003). Extended Field-of-View Sonography: Advantages in Abdominal Applications. *Journal of Ultrasound in Medicine*, 22(4):385–394.

[King et al., 2009] King, A., Boubertakh, R., Rhode, K., Ma, Y., Chinchapatnam, P., Gao, G., Tangcharoen, T., Ginks, M., Cooklin, M., Gill, J., et al. (2009). A subject-specific technique for respiratory motion correction in image-guided cardiac catheterisation procedures. *Medical Image Analysis*, 13(3):419–431.

[King et al., 2010] King, A., Buerger, C., and Schaeffter, T. (2010). Cardiac Respiratory Motion Modelling by Simultaneous Registration and Modelling from Dynamic MRI Images. *Biomedical Image Registration*, pages 222–233.

[Klein and Huesman, 2002] Klein, G. and Huesman, R. (2002). Four-dimensional processing of deformable cardiac PET data. *Medical Image Analysis*, 6(1):29–46.

[Klein et al., 2011] Klein, T., Hansson, M., and Navab, N. (2011). Spatial statistics based feature descriptor for rf ultrasound data. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*.

[Koenderink, 1984] Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, V50(5):363–370.

[Kokiopoulou et al., 2011a] Kokiopoulou, E., Chen, J., and Saad, Y. (2011a). Trace optimization and eigenproblems in dimension reduction methods. *Numerical Linear Algebra with Applications*, 18(3):565–602.

[Kokiopoulou and Frossard, 2009] Kokiopoulou, E. and Frossard, P. (2009). Minimum distance between pattern transformation manifolds: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1225–1238.

[Kokiopoulou et al., 2011b] Kokiopoulou, E., Zervos, M., Kressner, D., and Paragios, N. (2011b). Optimal Similarity Registration of Volumetric Images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Kolev et al., 2009] Kolev, K., Klodt, M., Brox, T., and Cremers, D. (2009). Continuous global optimization in multiview 3d reconstruction. *International Journal of Computer Vision*, 84:80–96.

[Komodakis et al., 2007] Komodakis, N., Tziritas, G., and Paragios, N. (2007). Fast, approximately optimal solutions for single and dynamic mrfs. In *CVPR*.

[Kovesi, 1999] Kovesi, P. (1999). Image features from phase congruency. *Videre: Journal of Computer Vision Research*, 1(3).

[Krücker et al., 2002] Krücker, J., LeCarpentier, G., Fowlkes, J., and Carson, P. (2002). Rapid elastic image registration for 3-d ultrasound. *IEEE Transactions on Medical Imaging*, 21(11):1384–1394.

[Kybic and Unser, 2003] Kybic, J. and Unser, M. (2003). Fast parametric elastic image registration. *IEEE Transactions on Image Processing*, 12(11):1427–1442.

[Laplace, 1812] Laplace, P.-S. (1812). *Théorie Analytique des Probabilitiés*. Courcier, Paris.

[Larose, 2001] Larose, D. A. (2001). *Iterative X-Ray/CT registration using accelerated volume rendering*. PhD thesis, Carnegie Mellon University.

[Larrue and Noble, 2011] Larrue, A. and Noble, J. A. (2011). Nakagami imaging with small windows. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 887–890.

[Lauenstein et al., 2004] Lauenstein, T. C., Goehde, S. C., Herborn, C. U., Goyen, M., Oberhoff, C., Debatin, J. F., Ruehm, S. G., and Barkhausen, J. (2004). Whole-Body MR Imaging: Evaluation of Patients for Metastases. *Radiology*, 233(1):139–148.

[Learned-Miller, 2006] Learned-Miller, E. G. (2006). Data driven image models through continuous joint alignment. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 28(2):236–250.

[Ledesma-Carbayo et al., 2005] Ledesma-Carbayo, M., Kybic, J., Desco, M., Santos, A., Sühling, M., Hunziker, P., and Unser, M. (2005). Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *IEEE Transactions on Medical Imaging*, 24(9):1113.

[Lee and Elgammal, 2007] Lee, C. and Elgammal, A. (2007). Modeling view and posture manifolds for tracking. In *ICCV*.

[Lee et al., 2009] Lee, D., Hofmann, M., Steinke, F., Altun, Y., Cahill, N., and Scholkopf, B. (2009). Learning similarity measure for multi-modal 3d image registration. In *Computer Vision and Pattern Recognition. IEEE Conference on*, pages 186 –193.

[Lee and Moore, 2005] Lee, P. and Moore, J. (2005). Gauss-Newton-on-manifold for Pose Estimation. *Journal of Industrial and Management Optimization*, 1(4):565.

[Legg et al., 2009] Legg, P. A., Rosin, P. L., Marshall, D., and Morgan, J. E. (2009). A robust solution to multi-modal image registration by combining mutual information with multi-scale derivatives. In Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., and Taylor, C., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5761 of *Lecture Notes in Computer Science*, pages 616–623. Springer.

[Lempitsky et al., 2008] Lempitsky, V., Roth, S., and Rother, C. (2008). FusionFlow: Discrete-continuous optimization for optical flow estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8.

[Leotta and Martin, 1999] Leotta, D. and Martin, R. (1999). Three-dimensional spatial compounding of ultrasound scans with incidence angle weighting. *Ultrasonics Symposium*, 2:1605–1608.

[Lester and Arridge, 1999] Lester, H. and Arridge, S. R. (1999). A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149.

[Leung et al., 2005] Leung, Y., Roshier, A., Johnson, S., Kerslake, R., and McNally, D. (2005). Demonstration of the appearance of the paraspinal musculoligamentous structures of the cervical spine using ultrasound. *Clin Anat*, 18(2):96–103.

[Leventon et al., 1998] Leventon, M. E., Eric, W., and Grimson, L. (1998). Multimodal volume registration using joint intensity distributions. In *Medical Image Computing and Computer-Assisted Interventation*, pages 1057–1066. Springer.

[Li et al., 2008] Li, G., Citrin, D., Camphausen, K., Mueller, B., Burman, C., Mychalczak, B., Miller, R., and Song, Y. (2008). Advances in 4D medical imaging and 4D radiation therapy. *Technology in cancer research & treatment*, 7(1):67.

[Li, 2009] Li, S. (2009). *Markov random field modeling in image analysis*. Springer-Verlag New York Inc.

[Li et al., 1998] Li, S., Wang, H., and Soh, W. (1998). Robust estimation of rotation angles from image sequences using the annealing M-estimator. *Journal of Mathematical Imaging and Vision*, 8(2):181–192.

[Li, 2001] Li, S. Z. (2001). *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc.

[Loeckx et al., 2010] Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., and Suetens, P. (2010). Nonrigid image registration using conditional mutual information. *Medical Imaging, IEEE Transactions on*, 29(1):19 –29.

[Lord, 1954] Lord, R. D. (1954). The use of the hankel transform in statistics. i. *Pattern Recogn. Lett.*, 41(4-5).

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.

[Madsen et al., 2004] Madsen, K., Nielsen, H., and Tingleff, O. (2004). Methods for Non-Linear Least Squares Problems. *Technical University of Denmark*.

[Maes et al., 1997] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual-information. *IEEE transactions on Medical Imaging*, 16(2):187–198.

[Mahony and Manton, 2002] Mahony, R. and Manton, J. (2002). The Geometry of the Newton Method on Non-Compact Lie Groups. *Journal of Global Optimization*, 23(3):309–327.

[Maintz et al., 1996] Maintz, J., van den Elsen, P., and Viergever, M. (1996). Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical image analysis*, 1(2):151–161.

[Maintz et al., 1997] Maintz, J., van der Elsen, P., and Viergever, M. (1997). Registration of 3D Medical Images Using Simple Morphological Tools. In *International Conference on Information Processing in Medical Imaging*, pages 204–217.

[Maintz and Viergever, 1998] Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36.

[Mansoury and Pasha, 2008] Mansoury, S. and Pasha, E. (2008). Determination of Maximum Entropy Probability Distribution via Burg's Measure of Entropy. *Applied Mathematical Sciences*, 2(57):2851–2858.

[Martinez-Möller et al., 2007] Martinez-Möller, A., Bundschuh, R., Riedel, M., Navab, N., Ziegler, S., Schwaiger, M., and Nekolla, S. (2007). Comparison of respiratory sensors and its compliance for respiratory gating in emission tomography. In *Journal of Nuclear Medicine*, volume 48, page 426.

[McClelland et al., 2006] McClelland, J., Blackall, J., Tarte, S., Chandler, A., Hughes, S., Ahmad, S., Landau, D., and Hawkes, D. (2006). A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. *Medical Physics*, 33:3348.

[Mellor and Brady, 2005] Mellor, M. and Brady, M. (2005). Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4):330–343.

[Mikolajczyk and Schmid, 2004] Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *Int. J. Comput. Vision*, 60(1):63–86.

[Mikolajczyk and Schmid, 2005] Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630.

[Modersitzki, 2009] Modersitzki, J. (2009). *FAIR: flexible algorithms for image registration*. Society for Industrial and Applied Mathematics (SIAM).

[Moradi et al., 2007] Moradi, M., Mousavi, P., and Abolmaesumi, P. (2007). Tissue characterization using fractal dimension of high frequency ultrasound rf time series. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 10, pages 900–908.

[Mulet-Parada and Noble, 2000] Mulet-Parada, M. and Noble, J. (2000). 2D+ T acoustic boundary detection in echocardiography. *Medical Image Analysis*, 4(1):21–30.

[Mumford, 1994] Mumford, D. (1994). The bayesian rationale for energy functionals. In *Geometry-driven diffusion in Computer Vision*, pages 141–153. Kluwer Academic.

[Murray et al., 1994] Murray, R. M., Li, Z., and Sastry, S. S. (1994). *A Mathematical Introduction to Robotic Manipulation*. CRC Press.

[Myronenko, 2010] Myronenko, A. (2010). *Non-rigid Image Registration: Regularization, Algorithms and Applications*. PhD thesis, Oregon Health and Science University.

[Myronenko and Song, 2009a] Myronenko, A. and Song, X. (2009a). Image registration by minimization of mapping complexity. In *CVPR Workshops, IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*.

[Myronenko and Song, 2009b] Myronenko, A. and Song, X. (2009b). Image Registration by Minimization of Residual Complexity. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Myronenko et al., 2009] Myronenko, A., Song, X., and Sahn, D. (2009). Maximum Likelihood Motion Estimation in 3D Echocardiography through Non-rigid Registration in Spherical Coordinates. *Functional Imaging and Modeling of the Heart*, pages 427–436.

[Nakagami, 1960] Nakagami, N. (1960). The m-distribution, a general formula for intensity distribution of rapid fadings. In Hoffman, W. G., editor, *Statistical Methods in Radio Wave Propagation*. Oxford, England: Pergamon.

[Nguyen et al., 2009] Nguyen, T., Moseley, J., Dawson, L., Jaffray, D., and Brock, K. (2009). Adapting liver motion models using a navigator channel technique. *Medical physics*, 36:1061.

[Ni et al., 2008] Ni, D., Qu, Y., Yang, X., Chui, Y.-P., Wong, T.-T., Ho, S. S. M., and Heng, P.-A. (2008). Volumetric ultrasound panorama based on 3D SIFT. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

[Padfield, 2010] Padfield, D. (2010). Masked FFT registration. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Papoulis and Pillai, 1991] Papoulis, A. and Pillai, U. S. (1991). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill; 3rd edition.

[Pearson, 1901] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Mag.*

[Peetrons, 2002] Peetrons, P. (2002). Ultrasound of muscles. *European Radiology*, 12(1):35–43.

[Peng et al., 2010] Peng, Y., Ganesh, A., Wright, J., Xu, W., and Ma, Y. (2010). RASL: Robust Alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Penney et al., 2008] Penney, G., Griffin, L., King, A., and Hawkes, D. (2008). A novel framework for multi-modal intensity-based similarity measures based on internal similarity. *SPIE*, 6914.

[Penney et al., 1998] Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., and Hawkes, D. (1998). A comparison of similarity measures for use in 2-d-3-d medical image registration. *Medical Imaging, IEEE Transactions on*, 17(4):586–595.

[Perperidis et al., 2005] Perperidis, D., Mohiaddin, R., and Rueckert, D. (2005). Spatio-temporal free-form registration of cardiac MR image sequences. *Medical Image Analysis*, 9(5):441–456.

[Peyrat et al., 2010] Peyrat, J., Delingette, H., Sermesant, M., Xu, C., and Ayache, N. (2010). Registration of 4D Cardiac CT Sequences Under Trajectory Constraints With Multichannel Diffeomorphic Demons. *IEEE transactions on medical imaging*.

[Pinton et al., 2009] Pinton, G., Dahl, J., Rosenzweig, S., and Trahey, G. (2009). A heterogeneous nonlinear attenuating full-wave model of ultrasound. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 56(3):474–488.

[Pizarro and Bartoli, 2007] Pizarro, D. and Bartoli, A. (2007). Shadow resistant direct image registration. In *SCIA*, pages 928–937.

[Pluim et al., 2000a] Pluim, J., Maintz, J., and Viergever, M. (2000a). Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8):809–814.

[Pluim et al., 2000b] Pluim, J., Maintz, J., and Viergever, M. (2000b). Interpolation artefacts in mutual information-based image registration. *Computer vision and image understanding*, 77(2):211–232.

[Pluim et al., 2003] Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2003). Mutual information based registration of medical images: A survey. *IEEE Trans. Med. Imaging*, 22(8):986–1004.

[Pohl et al., 2006] Pohl, K., Fisher, J., Grimson, W., Kikinis, R., and Wells, W. (2006). A Bayesian model for joint segmentation and registration. *Neuroimage*, 31(1):228–239.

[Poon and Rohling, 2005] Poon, T. and Rohling, R. (2005). Three-dimensional extended field-of-view ultrasound. *Ultrasound in Medicine and Biology*, 32(3):357–369.

[Prager et al., 2003] Prager, R. W., Gee, A. H., Treece, G. M., and Berman, L. H. (2003). Decompression and speckle detection for ultrasound images using the homodyned k-distribution. *Pattern Recognition Letters*, 24(4-5):705 – 713.

[Rayleigh, 1877] Rayleigh, J. (1877). *The theory of sound*. Number Vols. I and II in The Theory of Sound. MacMillan.

[Reddy and Chatterji, 1996] Reddy, B. and Chatterji, B. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, 5(8):1266 –1271.

[Reinsberg et al., 2005] Reinsberg, S., Doran, S., Charles-Edwards, E., and Leach, M. (2005). A complete distortion correction for MR images: II. Rectification of static-field inhomogeneities by similarity-based profile mapping. *Phys Med Biol*, 50(11):2651–2661.

[Remmert et al., 2007] Remmert, G., Biederer, J., Lohberger, F., Fabel, M., and Hartmann, G. (2007). Four-dimensional magnetic resonance imaging for the determination of tumour movement and its evaluation using a dynamic porcine lung phantom. *Physics in medicine and biology*, 52:N401.

[Revell et al., 2004] Revell, J., Mirmehdi, M., and McNally, D. (2004). Combined ultrasound speckle pattern similarity measures. In *Medical Image Understanding and Analysis*, pages 149–153. BMVA Press.

[Rice, 1945] Rice, S. (1945). *Mathematical analysis of random noise*. American Telephone and Telegraph Co.

[Rijkhorst et al., 2010] Rijkhorst, E.-J., Heanes, D., Odille, F., Hawkes, D., and Barratt, D. (2010). Simulating dynamic ultrasound using mr-derived motion models to assess respiratory synchronisation for image-guided liver interventions. In Navab, N. and Jannin, P., editors, *Information Processing in Computer-Assisted*

*Interventions*, volume 6135 of *Lecture Notes in Computer Science*, pages 113–123. Springer Berlin / Heidelberg.

[Rényi, 1961] Rényi, A. (1961). On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Stat. and Probability*, pages 547–561.

[Roche et al., 2000] Roche, A., Malandain, G., and Ayache, N. (2000). Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80.

[Roche et al., 1998] Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The correlation ratio as a new similarity measure for multimodal image registration. In *Proc. of First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of *LNCS*, pages 1115–1124, Cambridge, USA. Springer Verlag.

[Rohde et al., 2009] Rohde, G., Aldroubi, A., and Healy, D. (2009). Interpolation artifacts in sub-pixel image registration. *IEEE Transactions on Image Processing*, 18(2):333–345.

[Rohde et al., 2008] Rohde, G. K., Wang, W., Peng, T., and Murphy, R. F. (2008). Deformation-based nonlinear dimension reduction: Applications to nuclear morphometry. In *ISBI*.

[Rohlfing et al., 2001] Rohlfing, T., Maurer, Jr., C. R., O'Dell, W. G., and Zhong, J. (2001). Modeling liver motion and deformation during the respiratory cycle using intensity-based free-form registration of gated MR images. In Mun, S. K., editor, *Medical Imaging: Visualization, Display, and Image-Guided Procedures*, volume 4319 of *Proceedings of SPIE*, pages 337–348.

[Rohlfing et al., 2004] Rohlfing, T., Maurer, Jr., C. R., O'Dell, W. G., and Zhong, J. (2004). Modeling liver motion and deformation during the respiratory cycle using intensity-based free-form registration of gated MR images. *Medical Physics*, 31(3):427–432.

[Rohr et al., 2004] Rohr, K., Cathier, P., and Wörz, S. (2004). Elastic registration of electrophoresis images using intensity information and point landmarks. *Pattern recognition*, 37(5):1035–1048.

[Roweis and Saul, 2000] Roweis, S. T. and Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326.

[Rueckert et al., 2006] Rueckert, D., Aljabar, P., Heckemann, R., Hajnal, J., and Hammers, A. (2006). Diffeomorphic registration using B-splines. *Medical Image Computing and Computer-Assisted Intervention*, pages 702–709.

[Rueckert et al., 2000] Rueckert, D., Clarkson, M. J., Hill, D. L. G., and Hawkes, D. J. (2000). Non-rigid registration using higher-order mutual information. In *SPIE*, volume 3979, pages 438–447.

[Rueckert et al., 1999] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Trans Med Imaging*, 18(8):712–721.

[Russakoff et al., 2004] Russakoff, D. B., Tomasi, C., Rohlfing, T., Maurer, C. R., and Jr. (2004). Image similarity using mutual information of regions. In *8th European Conference on Computer Vision (ECCV*, pages 596–607. Springer.

[Sarrut et al., 2006] Sarrut, D., Boldea, V., Miguet, S., and Ginestet, C. (2006). Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. *Medical physics*, 33:605.

[Scherzer, 2010] Scherzer, O. (2010). *Handbook of mathematical methods in imaging*. Springer Verlag.

[Schmidt et al., 2007] Schmidt, G. P., Reiser, M. F., and Baur-Melny, A. (2007). Whole-body imaging of the musculoskeletal system: the value of MR imaging. *Skel Rad*, 36(12):1109–1119.

[Schreibmann et al., 2006] Schreibmann, E., Chen, G., and Xing, L. (2006). Image interpolation in 4D CT using a BSpline deformable registration model. *International Journal of Radiation Oncology*Biology*Physics*, 64(5):1537–1550.

[Seshamani et al., 2009] Seshamani, S., Rajan, P., Kumar, R., Girgis, H., Dassopoulos, T., Mullin, G., and Hager, G. (2009). A meta registration framework for lesion matching. In Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., and Taylor, C., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5761 of *Lecture Notes in Computer Science*, pages 582–589. Springer.

[Shams et al., 2008] Shams, R., Hartley, R., and Navab, N. (2008). Real-time simulation of medical ultrasound from ct images. In *Medical Image Computing and Computer Assisted Intervention*.

[Shams et al., 2007] Shams, R., Kennedy, R. A., Sadeghi, P., and Hartley, R. (2007). Gradient intensity-based registration of multi-modal images of the brain. In *ICCV*.

[Shankar, 2000] Shankar, M. (2000). A general statistical model for ultrasonic backscattering from tissues. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 47(3):727 –736.

[Shankar et al., 2001] Shankar, P., Dumane, V., Reid, J., Genis, V., Forsberg, F., Piccoli, C., and Goldberg, B. (2001). Classification of ultrasonic B-mode images of breast masses using Nakagami distribution. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 48(2):569–580.

[Shankar et al., 2002] Shankar, P., Dumane, V., Reid, J., Genis, V., Forsberg, F., Piccoli, C., and Goldberg, B. (2002). Classification of ultrasonic b-mode images of breast masses using nakagami distribution. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 48(2):569–580.

[Shankar et al., 1993] Shankar, P. M., Reid, J. M., Ortega, H., Piccoli, C. W., and Goldberg, B. B. (1993). Use of non-rayleigh statistics for the identification of tumors in ultrasonic b-scans of the breast. *IEEE Trans. Med. Imag.*, 12(4-5):687–692.

[Shechtman and Irani, 2007] Shechtman, E. and Irani, M. (2007). Matching Local Self-Similarities across Images and Videos. *IEEE Conference on Computer Vision and Pattern Recognition*.

[Shekhovtsov et al., 2008] Shekhovtsov, A., Kovtun, I., and Hlavac, V. (2008). Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 112:91–99.

[Sidorov et al., 2009] Sidorov, K., Richmond, S., and Marshall, D. (2009). An Efficient Stochastic Approach to Groupwise Non-rigid Image Registration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Sizintsev et al., 2008] Sizintsev, M., Derpanis, K., and Hogue, A. (2008). Histogram-based search: a comparative study. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[Souvenir and Pless, 2007] Souvenir, R. and Pless, R. (2007). Image distance functions for manifold learning. *Image Vision Comput.*, 25(3):365–373.

[Stefanescu et al., 2004] Stefanescu, R., Pennec, X., and Ayache, N. (2004). Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis*, 8(3):325–342.

[Strintzis and Kokkinidis, 1997] Strintzis, M. and Kokkinidis, I. (1997). Maximum likelihood motion estimation in ultrasound image sequences. *Signal Processing Letters, IEEE*, 4(6):156–157.

[Studholme and Cardenas, 2004] Studholme, C. and Cardenas, V. (2004). A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recogn. Lett.*, 25(10):1191–1202.

[Studholme et al., 1999] Studholme, C., Hill, D., and Hawkes, D. (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86.

[Sundar et al., 2009a] Sundar, H., Khamene, A., Yatziv, L., and Xu, C. (2009a). Automatic image-based cardiac and respiratory cycle synchronization and gating of image sequences. In *MICCAI*, pages 381–388.

[Sundar et al., 2009b] Sundar, H., Litt, H., and Shen, D. (2009b). Estimating myocardial motion by 4D image warping. *Pattern recognition*, 42(11):2514–2526.

[Szilágyi and Brady, 2009] Szilágyi, T. and Brady, S. M. (2009). Feature extraction from cancer images using local phase congruency: a reliable source of image descriptors. In *ISBI'09: Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging*, pages 1219–1222, Piscataway, NJ, USA. IEEE Press.

[Tempany et al., 2003] Tempany, C., Stewart, E., McDannold, N., Quade, B., Jolesz, F., and Hynynen, K. (2003). Mr imaging–guided focused ultrasound surgery of uterine leiomyomas: A feasibility study1. *Radiology*, 226(3):897.

[Tenenbaum et al., 2000] Tenenbaum, J., Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319.

[Teng et al., 2006] Teng, C.-C., Shapiro, L. G., and Kalet, I. (2006). Head and neck lymph node region delineation using a hybrid image registration method. In *ISBI*, pages 462–465.

[Timinger et al., 2005] Timinger, H., Krueger, S., Dietmayer, K., and Borgert, J. (2005). Motion compensated coronary interventional navigation by means of diaphragm tracking and elastic motion models. *Physics in Medicine and Biology*, 50:491.

[Toews et al., 2005] Toews, M., Collins, D., and Arbel, T. (2005). Maximum a posteriori local histogram estimation for image registration. In *MICCAI*, volume 3750, page 163.

[Tola et al., 2008] Tola, E., Lepetit, V., and Fua, P. (2008). A fast local descriptor for dense matching. In *Conference on Computer Vision and Pattern Recognition*, Alaska, USA.

[Trahey et al., 1987] Trahey, G., Allison, J., and von Ramm, O. (1987). Angle independent ultrasonic detection of blood flow. *Biomedical Engineering, IEEE Transactions on*, BME-34(12):965–967.

[Tran et al., 2008] Tran, D., Kamani, A., Lessoway, V., and Rohling, R. N. (2008). Adaptive spatial compounding for improving ultrasound images of the epidural space. In Emelianov, S. Y. and McAleavey, S. A., editors, *Medical Imaging 2008: Ultrasonic Imaging and Signal Processing*, volume 6513. SPIE.

[Treece et al., 2002] Treece, G., Prager, R., Gee, A., Cash, C., and Berman, L. (2002). Grey-scale gating for freehand 3D ultrasound. In *IEEE International Symposium on Biomedical Imaging*, pages 993–996.

[Trobin et al., 2008] Trobin, W., Pock, T., Cremers, D., and Bischof, H. (2008). Continuous energy minimization via repeated binary fusion. *ECCV*, pages 677–690.

[Twining et al., 2005] Twining, C., Cootes, T., Marsland, S., Schestowitz, R., Petrovic, V., and Taylor, C. (2005). A unified information-theoretic approach to groupwise non-rigid registration and model building. In Christensen, G. and Sonka, M., editors, *19th International Conference on Information Processing in Medical Images (IPMI'05)*, volume 3565 of *Lecture Notes in Computer Science*, pages 1–14. Springer.

[Twining et al., 2004] Twining, C., Marsland, S., and Taylor, C. (2004). Groupwise non-rigid registration: The minimum description length approach. In *British Machine Vision Conference*.

[Van den Elsen et al., 1995] Van den Elsen, P., Maintz, J., Pol, E.-J., and Viergever, M. (1995). Automatic registration of ct and mr brain images using correlation of geometrical features. *Medical Imaging, IEEE Transactions on*, 14(2):384 –396.

[Vandemeulebroucke et al., 2007] Vandemeulebroucke, J., Sarrut, D., and Clarysse, P. (2007). The popi-model, a point-validated pixel-based breathing thorax model. In *XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, Toronto, Canada.

[Vercauteren, 2008] Vercauteren, T. (2008). *Image Registration and Mosaicing for Dynamic In Vivo Fibered Confocal Microscopy*. Phd thesis, École Nationale Supérieure des Mines de Paris.

[Vercauteren et al., 2007] Vercauteren, T., Pennec, X., Malis, E., Perchant, A., and Ayache, N. (2007). Insight into efficient image registration techniques and the demons algorithm. *IPMI*.

[Vercauteren et al., 2008a] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008a). Diffeomorphic demons using ITK's finite difference solver hierarchy. *Insight Journal*.

[Vercauteren et al., 2008b] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008b). Symmetric log-domain diffeomorphic registration: A demons-based approach. *Medical Image Computing and Computer-Assisted Intervention*, pages 754–761.

[Vercauteren et al., 2009] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.

[Vercauteren et al., 2006] Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., and Ayache, N. (2006). Robust mosaicing with correction of motion distortions

and tissue deformation for in vivo fibered microscopy. *Medical Image Analysis*, 10(5):673–692.

[Viola and Wells, 1997] Viola, P. and Wells, W. (1997). Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154.

[Viola, 1995] Viola, P. A. (1995). *Alignment by Maximization of Mutual Information*. Ph.d. thesis, Massachusetts Institute of Technology.

[von Siebenthal, 2008] von Siebenthal, M. (2008). *Analysis and Modelling of Respiratory Liver Motion using 4DMRI*. PhD thesis, Eidgenössische Technische Hochschule ETH Zürich.

[von Siebenthal et al., 2007] von Siebenthal, M., Székely, G., Gamper, U., Boesiger, P., Lomax, A., and Cattin, P. (2007). 4D MR imaging of respiratory organ motion and its variability. *Physics in medicine and biology*, 52:1547.

[Wachinger, 2007] Wachinger, C. (2007). Three-dimensional ultrasound mosaicing. Master's thesis, Technische Universität München.

[Wachinger et al., 2009] Wachinger, C., Baumann, S., Zeltner, J., Glocker, B., , and Navab, N. (2009). Sphere extraction in mr images with application to whole-body mri. In *SPIE Medical Imaging*, Orlando, Florida, USA.

[Wachinger et al., 2008a] Wachinger, C., Glocker, B., Zeltner, J., Paragios, N., Komodakis, N., Hansen, M. S., and Navab, N. (2008a). Deformable Mosaicing for Whole-body MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, New York, USA.

[Wachinger et al., 2011] Wachinger, C., Klein, T., and Navab, N. (2011). The 2D Analytic Signal on RF and B-mode Ultrasound Images. In *Information Processing in Medical Imaging (IPMI)*.

[Wachinger et al., 2010a] Wachinger, C., Mateus, D., Keil, A., and Navab, N. (2010a). Manifold Learning for Patient Position Detection in MRI. In *IEEE International Symposium on Biomedical Imaging*, Rotterdam, The Netherlands.

[Wachinger and Navab, 2008] Wachinger, C. and Navab, N. (2008). Ultrasound specific similarity measures for three-dimensional mosaicing. In *SPIE Medical Imaging*, San Diego, California, USA.

[Wachinger and Navab, 2009a] Wachinger, C. and Navab, N. (2009a). Alignment of viewing-angle dependent ultrasound images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

[Wachinger and Navab, 2009b] Wachinger, C. and Navab, N. (2009b). Similarity Metrics and Efficient Optimization for Simultaneous Registration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Wachinger and Navab, 2010a] Wachinger, C. and Navab, N. (2010a). Manifold learning for multi-modal image registration. In *11st British Machine Vision Conference (BMVC)*.

[Wachinger and Navab, 2010b] Wachinger, C. and Navab, N. (2010b). Structural image representation for image registration. In *CVPR Workshops, IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, San Francisco, California.

[Wachinger and Navab, 2011] Wachinger, C. and Navab, N. (2011). Entropy and Laplacian images: Structural representations for multi-modal registration. to appear.

[Wachinger et al., 2008b] Wachinger, C., Shams, R., and Navab, N. (2008b). Estimation of acoustic impedance from multiple ultrasound images with application to spatial compounding. In *CVPR Workshops, IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, Anchorage, Alaska.

[Wachinger et al., 2007] Wachinger, C., Wein, W., and Navab, N. (2007). Three-dimensional ultrasound mosaicing. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Brisbane, Australia.

[Wachinger et al., 2008c] Wachinger, C., Wein, W., and Navab, N. (2008c). Registration strategies and similarity measures for three-dimensional ultrasound mosaicing. *Academic Radiology*, 15:1404–1415.

[Wachinger et al., 2010b] Wachinger, C., Yigitsoy, M., and Navab, N. (2010b). Manifold Learning for Image-Based Breathing Gating with Application to 4D Ultrasound. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Beijing, China.

[Wacker and Deinzer, 2009] Wacker, M. and Deinzer, F. (2009). Automatic robust medical image registration using a new democratic vector optimization approach with multiple measures. In Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., and Taylor, C., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5761 of *Lecture Notes in Computer Science*, pages 590–597. Springer.

[Wagner et al., 1983] Wagner, R. F., Smith, S. W., Sandrik, J. M., and Lopez, H. (1983). Statistics of Speckle in Ultrasound B-Scans. *IEEE Trans. Sonics and Ultrasonics*, 30(3):156–163.

[Wang et al., 2009a] Wang, P., Kelly, C., and Brady, M. (2009a). Application of 3d local phase theory in vessel segmentation. In *ISBI'09: Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging*, pages 1174–1177, Piscataway, NJ, USA. IEEE Press.

[Wang et al., 2009b] Wang, Q., Yap, P.-T., Wu, G., and Shen, D. (2009b). Attribute vector guided groupwise registration. In Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., and Taylor, C., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5761 of *Lecture Notes in Computer Science*, pages 656–663. Springer.

[Wang and Feng, 2004] Wang, X. and Feng, D. D. (2004). Automatic hybrid registration for 2-dimensional ct abdominal images. In *Proceedings of the Third International Conference on Image and Graphics*, pages 208–211, Washington, DC, USA. IEEE Computer Society.

[Wei and Tao, 2010] Wei, Y. and Tao, L. (2010). Efficient Histogram-Based Sliding Window. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[Wein et al., 2008] Wein, W., Brunke, S., Khamene, A., Callstrom, M., and Navab, N. (2008). Automatic CT-Ultrasound Registration for Diagnostic Imaging and Image-guided Intervention. *Medical Image Analysis*, 12:577–585.

[Wells et al., 1996] Wells, W., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*.

[West et al., 1997] West, J., Fitzpatrick, J., Wang, M., Dawant, B., Maurer Jr, C., Kessler, R., Maciunas, R., Barillot, C., Lemoine, D., Collignon, A., et al. (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554.

[Wietzke et al., 2009] Wietzke, L., Sommer, G., and Fleischmann, O. (2009). The geometry of 2d image signals. In *CVPR*, pages 1690–1697.

[Wild and Reid, 1952] Wild, J. and Reid, J. (1952). Application of echo-ranging techniques to the determination of structure of biological tissues. *Science*, 115(2983):226.

[Wilhjelm et al., 2004] Wilhjelm, J., Jensen, M., Jespersen, S., Sahl, B., and Falk, E. (Feb. 2004). Visual and quantitative evaluation of selected image combination schemes in ultrasound spatial compound scanning. *Medical Imaging, IEEE Transactions on*, 23(2):181–190.

[Woods et al., 2006] Woods, N., Galatsanos, N., and Katsaggelos, A. (2006). Stochastic Methods for Joint Registration, Restoration, and Interpolation of Multiple Undersampled Images. *IEEE Transactions on Image Processing*, 15(1):201.

[Wörz and Rohr, 2008] Wörz, S. and Rohr, K. (2008). Hybrid Physics-Based Elastic Image Registration Using Approximating Splines. *SPIE Medical Imaging*.

[Wu et al., 2005] Wu, F., Wang, Z., Zhu, H., Chen, W., Zou, J., Bai, J., Li, K., Jin, C., Xie, F., and Su, H. (2005). Feasibility of us-guided high-intensity focused ultrasound treatment in patients with advanced pancreatic cancer: Initial experience1. *Radiology*, 236(3):1034.

[Wu and Fan, 2009] Wu, Y. and Fan, J. (2009). Contextual Flow. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Wyatt and Noble, 2003] Wyatt, P. and Noble, J. (2003). MAP MRF joint segmentation and registration of medical images. *Medical Image Analysis*, 7(4):539–552.

[Xiao et al., 2002] Xiao, G., Brady, J., Noble, J., Burcher, M., and English, R. (2002). Nonrigid registration of 3-D free-hand ultrasound images of the breast. *Medical Imaging, IEEE Transactions on*, 21(4):405–412.

[Xiaohua et al., 2004] Xiaohua, C., Brady, M., and Rueckert, D. (2004). Simultaneous segmentation and registration for medical image. In *MICCAI*, pages 663–670. Springer.

[Xiaoxun and Yunde, 2006] Xiaoxun, Z. and Yunde, J. (2006). Local Steerable Phase (LSP) Feature for Face Representation and Recognition. In *Conference on Computer Vision and Pattern Recognition*, volume 2.

[Yang and Fox, 2004] Yang, Z. and Fox, M. D. (2004). Speckle reduction and structure enhancement by multichannel median boosted anisotropic diffusion. *EURASIP J. Appl. Signal Process.*, 2004(1):2492–2502.

[Yi and Soatto, 2009] Yi, Z. and Soatto, S. (2009). Nonrigid Registration Combining Global and Local Statistics. In *CVPR*, pages 2200–2207.

[Yigitsoy et al., 2011a] Yigitsoy, M., Wachinger, C., and Navab, N. (2011a). Manifold learning for image-based breathing gating in mri. In *SPIE Medical Imaging*, Orlando, Florida, USA.

[Yigitsoy et al., 2011b] Yigitsoy, M., Wachinger, C., and Navab, N. (2011b). Temporal Groupwise Registration for Motion Modeling. In *Information Processing in Medical Imaging (IPMI)*.

[Ying and Sin, 2005] Ying, M. and Sin, M.-H. (2005). Comparison of extended field of view and dual image ultrasound techniques: Accuracy and reliability of distance measurements in phantom study. *Ultrasound in Medicine & Biology*, 31:79–83.

[Zagzebski, 1996] Zagzebski, J. (1996). *Essentials Of Ultrasound Physics*. Mosby, 1 edition.

[Zang et al., 2007] Zang, D., Wietzke, L., Schmaltz, C., and Sommer, G. (2007). Dense optical flow estimation from the monogenic curvature tensor. *Scale Space and Variational Methods in Computer Vision*, pages 239–250.

[Zass and Shashua, 2008] Zass, R. and Shashua, A. (2008). Probabilistic graph and hypergraph matching. In *Computer Vision and Pattern Recognition, IEEE Conference on*.

[Zefran et al., 1998] Zefran, M., Kumar, V., and Croke, C. (1998). On the generation of smooth three-dimensional rigid body motions. *Robotics and Automation, IEEE Transactions on*, 14(4):576–589.

[Zetsche and Barth, 1990] Zetsche, C. and Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two dimensional signals. *Vision Research*, 30.

[Zhang et al., 2006] Zhang, Q., Souvenir, R., and Pless, R. (2006). On Manifold Structure of Cardiac MRI Data: Application to Segmentation. *CVPR*, 1:1092–1098.

[Zhang et al., 2007] Zhang, W., Noble, J. A., and Brady, J. M. (2007). Spatio-temporal registration of real time 3d ultrasound to cardiovascular mr sequences. In *MICCAI*, pages 343–350.

[Zheng, 2008] Zheng, G. (2008). Effective incorporation of spatial information in a mutual information based 3d-2d registration of a ct volume to x-ray images. In *MICCAI*.

[Zheng and Zhang, 2006] Zheng, G. and Zhang, X. (2006). A unifying map-mrf framework for deriving new point similarity measures for intensity-based 2d-3d registration. In *International Conference on Pattern Recognition*, volume 2, pages 1181–1185.

[Zhu et al., 2003] Zhu, H., Oakeson, K., and Friedman, M. (2003). Retrieval of cardiac phase from IVUS sequences. In *Proceedings of SPIE*, volume 5035, page 135.

[Zhuang et al., 2009] Zhuang, X., Hawkes, D., and Ourselin, S. (2009). Unifying encoding of spatial information in mutual information for nonrigid registration. In *Information Processing in Medical Imaging*, volume 5636 of *Lecture Notes in Computer Science*, pages 491–502. Springer Berlin / Heidelberg.

[Zikic et al., 2010] Zikic, D., Glocker, B., Kutter, O., Groher, M., Komodakis, N., Kamen, A., Paragios, N., and Navab, N. (2010). Linear intensity-based image registration by markov random fields and discrete optimization. *Medical Image Analysis*, 14(4):550–562.

[Zikic et al., 2006] Zikic, D., Wein, W., Khamene, A., Clevert, D.-A., and Navab, N. (2006). Fast deformable registration of 3d-ultrasound using a variational approach. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 4190:915–923.

[Zitova and Flusser, 2003] Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.

[Zöllei, 2006] Zöllei, L. (2006). *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*. Ph.d. thesis, MIT; MIT-CSAIL.

[Zöllei et al., 2003] Zöllei, L., Fisher III, J., and Wells III, W. (2003). A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration. In *IPMI*.

[Zöllei et al., 2007] Zöllei, L., Jenkinson, M., Timoner, S., and Wells, W. (2007). A Marginalized MAP Approach and EM Optimization for Pair-Wise Registration. *Information Processing in Medical Imaging (IPMI)*, 4584:662.

[Zöllei et al., 2005] Zöllei, L., Learned-Miller, E., Grimson, E., and Wells, W. (2005). Efficient Population Registration of 3D Data. In *Computer Vision for Biomedical Image Applications, ICCV*.