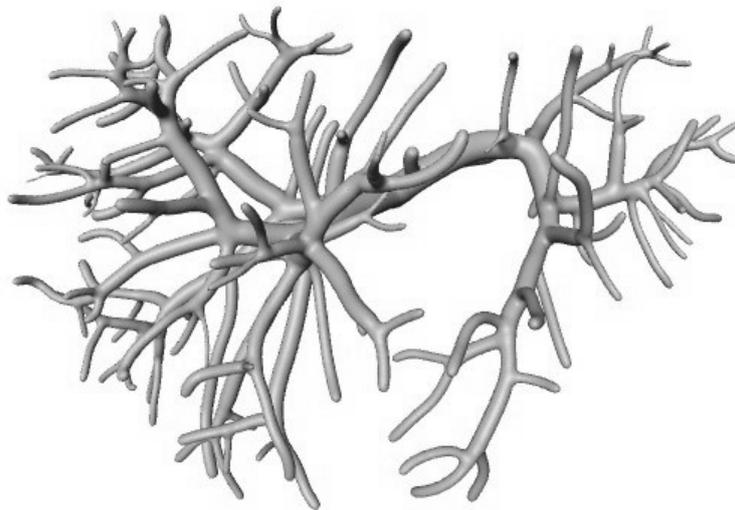


THOMAS LANGE

MODELING PRIOR KNOWLEDGE
FOR IMAGE REGISTRATION
IN LIVER SURGERY





UNIVERSITÄT ZU LÜBECK

From the Institute of Mathematics and Image Computing
of the University of Lübeck

Director: Prof. Dr. Bernd Fischer

Modeling Prior Knowledge for Image Registration in Liver Surgery

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Science/Engineering

Submitted by

Dipl.-Math. techn. Thomas Lange

from Berlin

Lübeck 2011

First referee: Prof. Dr. Bernd Fischer

Second referee: Prof. Dr. Dr. h.c. Peter M. Schlag

Date of oral examination: 22/12/2011

Approved for printing. Lübeck, 22/12/2011

Abstract

The careful consideration of blood vessels and the complete removal of the tumor are essential in oncological liver surgery to preserve healthy liver tissue and to minimize the probability for recurrent tumors. The enormous improvements in medical imaging over the last 20 years enable an accurate computer assisted 3D planning of the surgical intervention. The accurate transfer of the preoperative plan to the patient on the operating table is not trivial as the liver deforms due to intraoperative bedding and mobilization of the organ. Intraoperative 3D ultrasound is a possibility to capture the current shape and position of the liver during a surgical intervention. In the 3D ultrasound volume a navigation system shows the accurate position of the surgical instrument and its spatial relation to the vessels and the tumor.

The key problem for the transfer of the surgical plan is the compensation of the deformations between preoperative images resp. planning models and the intraoperative ultrasound data. Such problems have not yet been solved satisfactory. The image processing technique to compensate this is called non-rigid registration. Non-rigid registration is also needed for the postoperative control based on a comparison between pre- and postoperative images.

The principle difficulty of non-rigid registration is the vast number of theoretically possible non-rigid transformations, of which only a small subset compensates the present anatomical deformations. The fundamental hypothesis, pursued by this thesis, is that the incorporation of a priori knowledge about the image contents or about application-specific transformation properties significantly reduces the number of admissible transformations. We develop a new distance measure which considers the tube-like shapes of vessels by specific local filters, which give high responses, if the preoperative vessel models fit the appearance of vessels at the same position in the intraoperative image. A priori knowledge about anatomical corresponding landmarks is a direct restriction on the transformation. An important property, which sets our method apart from previous work, is that anisotropic tolerances to compensate landmark localization uncertainties are consequently integrated into pure landmark schemes as well as into schemes combining intensity and landmark information. The developed registration methods are validated on clinical image data by a new reference standard.

Zusammenfassung

In der onkologischen Leberchirurgie sind die genaue Beachtung der Blutgefäße und die vollständige Entfernung des Tumors essentiell, um gesundes Lebergewebe zu erhalten und die Wahrscheinlichkeit einer erneuten Tumorbildung zu minimieren. Die enormen Verbesserungen der medizinischen Bildgebung in den letzten 20 Jahren ermöglichen eine genaue computerassistierte 3D Planung chirurgischer Eingriffe. Die genaue Übertragung des präoperativen Plans auf den Patienten im Operationssaal ist nicht trivial, da sich die Leber bei der intraoperativen Lagerung und Mobilisierung deformiert. Intraoperativer 3D Ultraschall ist eine Möglichkeit während des chirurgischen Eingriffs die aktuelle Form und Lage der Leber zu erfassen. Im 3D Ultraschallvolumen zeigt ein Navigationssystem die genaue Lage eines chirurgischen Instruments und dessen räumliche Beziehung zu den Gefäßen und dem Tumor.

Das Hauptproblem bei der Übertragung des chirurgischen Plans ist der Ausgleich der Deformationen zwischen den präoperativen Bilddaten bzw. Planungsmodellen und den intraoperativen Ultraschalldaten, der als nicht-rigide Registrierung bezeichnet wird. Dieses Problem wurde noch nicht zufriedenstellend gelöst. Für den Vergleich von prä- und postoperativen Bilddaten zur postoperativen Kontrolle wird ebenfalls nicht-rigide Registrierung benötigt.

Die Schwierigkeit bei der Registrierung ist die riesige Zahl theoretisch möglicher nicht-rigider Transformationen, von denen nur wenige die anatomischen Deformationen abbilden. Die grundlegende Idee dieser Arbeit ist, dass durch die Einbindung von a priori Wissen über Bildinhalte oder Transformationseigenschaften die Anzahl von zulässigen Transformationen deutlich reduziert wird. Wir entwickeln ein neues Distanzmaß, das die Röhrenform von Gefäßen durch lokale Filter berücksichtigt, die hohe Antworten geben, wenn das präoperative Gefäßmodell zu den abgebildeten Gefäßen im intraoperativen Bild passt. A priori Wissen in Form von anatomisch korrespondierenden Landmarken stellt eine direkte Restriktion der Transformation dar. In unserem Ansatz integrieren wir konsequent anisotrope Toleranzen für die Berücksichtigung von Lokalisierungsungenauigkeiten und zwar sowohl in reinen Landmarkenverfahren, als auch in Verfahren, die Intensitäts- und Landmarkeninformationen kombinieren. Die entwickelten Registrierungsverfahren werden mit einem neuen Referenzstandard auf klinischen Bilddaten validiert.

Acknowledgements

This work has been written in the interdisciplinary field of computer assisted surgery and therefore the knowledge, visions, interdisciplinary understanding, openness and support of many different people have been necessary to enable this work. Three research groups have substantially influenced my academic career during the whole life cycle of my thesis: the surgical oncology group at the Robert-Rössle-Clinic of the Charité, the Zuse Institute Berlin and the Institute of Mathematics and Image Computing at the University of Lübeck.

First of all, I wish to thank Prof. Dr. Dr. h.c. Peter M. Schlag, Director of the Charité Comprehensive Cancer Center and Professor for Surgical Oncology at the Charité – Universitätsmedizin Berlin, for introducing me to the field of tumor and liver surgery. Without his vision of precise computer assisted surgery and his support this work would not have been possible. He gave me the opportunity to learn what surgeons do in the operating room by attending dozens of liver, rectum and pelvic bone resections. In addition, I am grateful to the surgeons Prof. Dr. Michael Hünerbein, Dr. Siegfried Beller and Dr. Per-Ulf Tunn for their enthusiasm, patience, ideas and fruitful discussions how to implement clinically useful navigation systems in the operating room despite their heavy clinical workload at the Robert-Rössle-Clinic. I also thank the surgeons PD Dr. Stefan Gretschel, Susan Burock and Prof. Dr. Beate Rau for their constant support and patience.

Before I started my research in computer assisted surgery, I had to learn how to apply mathematics and computer science to medical problems. I thank Axel Friedrich, my friend from university, who encouraged me to accept a student job in a medical project funded by the DFG Collaborative Research Center 273 "Hyperthermia: Methodology and Clinics" at the Zuse Institute Berlin (ZIB). To Prof. Hans-Christian Hege, head of the Department of Visualization and Data Analysis at the ZIB, I owe the opportunity to learn 3D modeling and visualization of anatomical structures as well as how to develop and implement algorithms for practical medical problems. I also thank Dr. Malte Westerhoff, Dr. Detlev Stalling and Dr. Steffen Prohaska for the development of the Amira software, which has served as the software platform for this work.

The mathematician Prof. Dr. Dr. h.c. Peter Deuffhard, president of ZIB, is Prof. Schlag's equivalent on the methodological side. His vision to use mathe-

matics for solving medical problems and thus enabling a more individual and quantitative medicine was a major inspiration for me. A very important and constant research partner over the years was Dr. Hans Lamecker from ZIB. We wrote our first joint and successful MICCAI paper together with Dr. Martin Seebass in a nightly session. Hans Lamecker taught me how to improve the structure and to emphasize the most important aspects of a paper. I am grateful to Hans Lamecker and Dagmar Kainmüller also from ZIB for our pleasant and successful joint work in liver segmentation and to Dr. Stefan Zachow, head of the Medical Planning group at ZIB, for more general discussions on computer assisted surgery.

It was Hans-Christian Hege, who introduced me to Prof. Dr. Bernd Fischer, an expert of medical image registration, head of the Institute of Mathematics and Image Computing at the University of Lübeck and head of the Project Group Image Registration from Fraunhofer MEVIS. I am grateful to Bernd Fischer and his whole group, in particular Prof. Dr. Jan Modersitzki, Dr. Stefan Heldmann and Janine Olesch for the inspiring, open-minded and uncomplicated collaboration; it has always been a pleasure for me to come to Lübeck. Very special thanks go to Dr. Nils Papenberg for the same reasons and additionally for many explanations to the Lübecker image registration framework and the hospitality of him and his wife Hilke Papenberg.

I am grateful to Prof. Dr. Karl Rohr, head of the Biomedical Computer Vision Group (BMCV) at the DKFZ Heidelberg and Associate Professor at the University of Heidelberg, and Dr. Stefan Wörz also from the BMCV group for inspiring me to consider localization uncertainties in landmark based registration schemes and our corporate work.

Furthermore, I would like to thank Prof. Dr. Heinz Handels, head of the Institute of Medical Informatics at the University of Lübeck for chairing the doctoral committee.

I deeply thank Nadja Wisniewski for very faithful and competent proofreading. My thanks go also to my colleague Sebastian Eulenstein for supporting me in IT questions, fruitful discussions about planning and navigation systems, their joint implementation in a clinical environment and for ensuring an enjoyable everyday work environment.

This work has in no small amount benefited from many discussions with and helpful reviews by too many colleagues from the scientific community, who can be mentioned here and to whom I owe my gratitude for critically considering my work.

Last but not least, I thank my family for their support and enduring patience.

— Thomas Lange —

Contents

1	Introduction	1
1.1	A Whole Greater Than the Sum of Its Parts	1
1.2	The Art of Applying Image Registration Methods to Clinical Problems	3
1.3	Outline and Contributions	4
2	Computer Assisted Liver Surgery	9
2.1	Liver Cancer and Therapies	9
2.2	Liver Surgery	11
2.3	Computer Assisted Surgery Planning	15
2.4	Intraoperative Imaging and Navigation	22
2.5	Conclusions	41
Part I	Variational Image Registration	43
3	The Registration Problem	45
3.1	Introduction	45
3.2	Problem Statement	45
3.3	General Solutions	47
4	Discrete Images and Transformations	49
4.1	Introduction	49
4.2	Cell-Centered Grids	49
4.3	Continuous Images and their Multiscale Representation	52
4.4	Multiresolution Images and Transformations	60
5	Optimization	67
5.1	Basic Definitions	67
5.2	Characterizing Solutions	69
5.3	Algorithms for Unconstrained Optimization	71
5.4	Algorithms for Constrained Optimization	76

6	Nonparametric Image Registration	85
6.1	Introduction	85
6.2	Distance Measures	86
6.3	Regularizers	88
6.4	Nonparametric Image Registration on a Fixed Level	96
6.5	Multilevel Image Registration	98
7	Parametric Image Registration	103
7.1	Introduction	103
7.2	Affine-Linear Transformations	103
7.3	Rigid Transformations	105
7.4	Spline-based Transformations	107
	Part II Modeling Prior Knowledge in Image Registration	109
8	Review of Distance Measures	111
8.1	Introduction	111
8.2	Properties of Distance Measures	112
8.3	General Purpose Distance Measures	112
8.4	Specific Distance Measures	120
9	Distance Measures Incorporating Shape Information	123
9.1	Introduction	123
9.2	Vessels Intensity Models	124
9.3	Correlation Measure Based on the Vessel Intensity Model	128
9.4	Convolution Based Measures	129
10	Possibilities of Constraining Displacement Fields	135
10.1	Introduction	135
10.2	Constraints in Variational Image Registration	135
10.3	Examples for Constraints	136
11	Parametric Landmark Registration	139
11.1	Introduction	139
11.2	Point Landmarks in Clinical Applications	140
11.3	Notations	141
11.4	Principle Landmark Registration Schemes	143
11.5	Landmark Localization Uncertainties	145
11.6	Rigid Landmark Registration	152
11.7	Affine Linear Landmark Registration	157
11.8	Thin-Plate Splines (TPS)	159
11.9	Gaussian Elastic Body Splines (GEBS)	164

12 Combining Landmarks and Intensity-based Registration	169
12.1 Introduction	169
12.2 Landmark Distances Penalizer	169
12.3 Equality Landmark Constraints	170
12.4 Inequality Landmark Constraints	170
12.5 Discretization of Landmark Constraints	172
12.6 Algorithms for Landmark Constraint Registration	174
Part III Applications to Liver Surgery	177
13 Assessment of Registration Algorithms	179
13.1 What Exactly is Accuracy and How Can It Be Determined?	180
13.2 Previous Work	185
13.3 Simulating Rigid Transformations	189
13.4 Reference Standard Based on Vessel Trees	196
13.5 Further Assessment Criteria	201
14 Transfer of Preoperative Models to Intraoperative 3D Ultrasound	
Data	203
14.1 Introduction	203
14.2 Specification of Clinical Image Data	204
14.3 Non-Rigid Registration with Equality Landmark Constraints	205
14.4 Non-Rigid Registration with Inequality Landmark Constraints . . .	211
14.5 Non-Rigid Registration with Shape Distance Measure	215
15 Quantitative Postoperative Control	217
15.1 Introduction	217
15.2 Interactive Vessel Registration	219
15.3 Results	220
15.4 Conclusions	221
Conclusions	225
Bibliography	229

Chapter 1

Introduction

1.1 A Whole Greater Than the Sum of Its Parts

The enormous technical advances in medical imaging in recent years offer improved diagnostics as well as precise and less invasive therapies for many diseases. In particular regional cancer therapies like interventional radiotherapy, radiofrequency thermal ablation, regional thermotherapy, and last but not least surgical strategies benefit strongly from exact 3D medical images of the individual anatomy, physiology and pathology of patients.

Based on those images 3D models of the relevant parts of the anatomy and the tumor might be generated to support the therapy planning and decision. Often simulations computed on those models such as temperature or radiation dose distribution are an essential part of therapy planning. In surgery gentle access paths and the impact resp. risk of different surgical strategies can be analyzed and planned before the intervention using 3D models. In neurosurgery for example a careful consideration of structures traversed to reach a target is required to avoid neurological deficits. In liver surgery the postoperative blood supply and drainage of liver tissue depending on a planned tumor removal can be determined in advance.

Besides this preoperative planning process the use of medical images as a *road-map* for intraoperative localization of tumors and anatomical structures at risk is nowadays an import and sometimes irreplaceable aid for the surgeon or interventional radiologist. Such high-precision intraoperative navigation systems would not have been possible without the increasing resolution, quality and acquisition speed of modern imaging devices.

With the increasing number of available medical images before, during and after therapies the request and need of *comparing and combining different images* rises. The additional information gained by the combination of images from different sources (imaging modalities) or from the quantitative comparison of images from different points in time is often an important benefit for diagnostics, therapy planning and intervention implementation.

Often structures or physiological effects are only recognizable in one of the modalities and vice versa. Prominent examples are the combination of Computed Tomography (CT) and Positron Emission Tomography (PET) or CT and Magnetic Resonance Imaging (MRI). For example, the high contrast of bones in CT and the excellent representation of soft tissue in MRI are combined for neurosurgical intervention planning and navigation to get a more comprehensive understanding of the patient's anatomy.

Next to these combinations of preoperative modalities for better treatment planning the transfer of preoperative onto intraoperative images is also an important issue for the guidance of the surgeon. The acquisition of images during intervention is often indispensable due to tissue deformations like the brain shift effect in neurosurgery. As those intraoperative images are in general of lower quality (e.g., 3D ultrasound) a transfer of high quality preoperative images usually increases the information content for the surgeon significantly. In addition surgery plans prepared on preoperative images can be combined with intraoperative images.

To control the success of an intervention a quantitative comparison of pre- and postoperative image data offers a detailed evaluation method. Which parts of an organ have exactly been removed? Has the preoperative plan been implemented? And on a more general level does the use of a navigation system improve the implementation of a preoperative plan?

For the combination and quantitative comparison of different medical images it is important to know exactly which point in one image corresponds to which point in the other image. Each imaging device has its own coordinate system and the positioning of the patient can not generally be reproduced exactly. Even during an imaging session patients often move or at least parts of the organs and tissue move due to respiration and heartbeat. Soft tissue of organs like the liver usually deforms between different image acquisitions. After surgery even whole parts of the anatomy are missing. To compensate the movements and deformations mathematical algorithms have been developed to compute a transformation which maps each image point of one image to its anatomically corresponding point in the other image. This process is called *image registration*. Using the transformation the information content of both images can be fused into one single image usually leading to an added value as mentioned above. So, since *two images are better than one, but one combined image is even better than two solitary images*, the whole is greater than the sum of its parts.

1.2 The Art of Applying Image Registration Methods to Clinical Problems

Although for rigid structures like bones registration algorithms have been successfully developed and validated the image registration of non-rigid structures like liver tissue is still a challenge and an active field of research. The heart of an image registration algorithm independently of rigid or non-rigid structures is always a suitable distance measure, which measures the differences of two images. The hope is, that the transformation, which minimizes the distance measure is the same transformation, which maps corresponding anatomical points. As the number of theoretically possible non-rigid transformations is huge transformations exist, which might minimize the distance measure, but which do not describe plausible movements or deformations of human organs. In image regions with no or weak image structures the distance measure might not change for very different transformations resp. is driven mainly by image noise. To get plausible transformations either a subset of transformations is explicitly defined via parameterization or a regularizer is added to the distance measure, which penalizes implausible transformations. This second approach is also called non-parametric in contrast to parametric image registration. Both approaches will be used in this work but for different tasks.

In summary the general non-rigid image registration problem can be modeled as the minimization of a functional consisting of the weighted sum of a distance measure, which measures the image differences and a regularizer, which usually penalizes non-smooth transformations. This general variational non-parametric image registration framework has to be adapted to the image classes used in a given medical application. There is no algorithm for all types of medical image registration problems. However the distance measures and assumptions on the transformations are usually quite general. In clinical data sets a wide variety of image quality and image content occurs even for the same modality. The image quality is degraded by noise and artifacts depending on the imaging modality. The imaging characteristics vary significantly between different modalities, i.e. the same anatomy is mapped differently to the acquired images. This means the *appearance* of an anatomical structure differs. In particular the noise, artifacts and imaging characteristics of ultrasound techniques (sonography) are very challenging for medical image processing procedures. Customized distance measures, which consider imaging characteristics, are one way of coping with the challenges of different medical modalities, especially with those of ultrasound imaging.

The consideration of the image content is also an important aspect for the choice of a distance measure and a starting point for the development of new measures. For example, if a characteristic geometry dominates the image content like tubular vessels structures in the liver, a distance measure emphasizing this geometry might lead to better and more robust registration

results. Regarding registration of preoperative to intraoperative images there might even be particular information available about the *individual* geometry of anatomical structures of a specific patient, like the radii of vessels. Although the use of a priori knowledge is a major direction of current research efforts in non-rigid registration there are only very few approaches, which try to incorporate a priori knowledge about the general or specific geometry of image contents into the distance measure.

In addition to adapted distance measures another general possibility to include a priori knowledge is to define additional restrictions for the class of transformations. The regularizer just penalizes non-smooth transformations, but does not rule them out explicitly nor does it guarantee wanted properties like volume preservation or local rigidity. For this purpose the variational registration framework can be extended by adding penalizers or constraints to restrict the possible transformations to a plausible class of transformations for a given medical application. Again general as well as individual a priori knowledge about an image pair might be incorporated into the framework. Volume preservation in the whole image domain is an example for general knowledge and a given set of point correspondences or other geometric features an example for individual knowledge about the image contents. If a set of point correspondences is interactively or automatically provided, it means that at certain points the transformation is already given. However, the correct detection and accurate localization of those features directly influences the accuracy of the whole registration process. On the one hand intuitive tools and semi-automatic support might help to detect and locate features accurately, on the other hand a certain amount of localization uncertainty will always remain. In many cases the uncertainty can be estimated and considered in the registration process. A few approaches incorporating the localization uncertainty of point features have been published, but a systematic overview and assessment of the used methods, as well as some consequent further developments are still missing.

To summarize: the adaptation of a registration process to a specific medical application can encompass the incorporation of geometric features into the distance measure or explicit constraints on the transformation e.g., by given point correspondences.

A major difficulty in the development of non-rigid registration approaches is the determination of the accuracy of the registration results. A couple of methods have been introduced, but up to now there is no generally accepted gold standard.

1.3 Outline and Contributions

The thesis is split into three major parts: part I explains the used image registration framework and its general solution; part II presents the main

theoretical contributions of the thesis: the incorporation of geometrical features into distance measures and explicit constraints on the transformations; part III contains the application of the developed mathematical registration techniques to computer assisted liver surgery and their evaluation.

Chapter 2 starts with a motivating introduction into computer assisted surgery for the removal of tumors from the liver. Those removals are called oncological liver resections. We will explain the anatomical and surgical background, the computer assistance for planning and implementation of surgical interventions, and the necessary fundamentals of the involved medical imaging modalities: computed tomography (CT) and 3D ultrasound (3D US).

Chapters 3 through 7 form Part I – Variational Image Registration. In Chapter 3 the image registration problem is formalized as a variational optimization problem. The defined registration functional is optimized by changing the continuous displacement function, which transforms a template image to a reference image. A general solution for this continuous optimization problem is the transformation to a finite dimensional optimization problem by discretization (discretize-then-optimize approach). The discretization of the images and the displacement field as well as multiscale and multiresolution strategies for fast and robust numerical solutions of the finite-dimensional optimization problem are described in Chapter 4. Some basics and general algorithms for optimization of finite dimensional problems are presented in Chapter 5. An algorithm for constrained optimization is also included to solve registration problems with a priori knowledge formulated as constraints.

There are two general approaches to restrict arbitrary transformations to plausible classes of transformations. The first is presented in Chapter 6; it uses regularization of the registration functional. This nonparametric framework is very flexible and additional constraints can easily be incorporated. The main ideas of the discretization and numerical solution of this framework are described. The second approach to get plausible and smooth transformations is to define the class of transformations explicitly by a parameterization described in Chapter 7. An important class of parametric transformations are rigid and affine transformations. But there are also different spline based schemes with higher degrees of freedom for non-rigid registration.

Chapters 8 through 12 make up Part II – Modeling Prior Knowledge in Image Registration. In Chapter 8 general and specific distance measures are reviewed. New distance measures are developed in Chapter 9 to incorporate a priori knowledge about the shape of anatomical structures into the distance functionals. The measures are based on a geometric modeling of the anatomical structures. The focus is on tube-like structures like vessels as we have the registration of models from preoperative CT data to intraoperative 3D ultrasound images of the liver with its rich vasculature in mind. The main question is how geometrical models can be compared with structures contained in intensity data? The development and discussion of customized

distance measures which consider geometrical features is a novel contribution to the field of non-rigid registration.

Chapter 10 gives an overview how constraints on the deformation field can be incorporated into a variational registration framework. General approaches using penalizers on the registration functional or explicit equality and inequality constraints are distinguished. In the following Chapter 11 a systematic overview on parametric registration approaches based on corresponding point landmarks is given. How can the landmark positions be weighted according to their localization uncertainties, which solutions and new derivations exist? In particular, we develop new schemes in this chapter which incorporate anisotropic localization uncertainties, as they occur in our liver surgery applications.

If only information about corresponding landmarks are considered in the registration process, a large amount of information provided by the intensities of the images is not used. The combination of landmark and intensity information in the nonparametric registration framework is a promising approach and presented in Chapter 12. The consideration of landmark localization uncertainties is a central topic of this chapter. The formulation as an inequality constrained optimization problem is a new contribution to the field of non-rigid registration in particular for anisotropic localization uncertainties.

Most of the contents of Part III – Applications to Liver Surgery (Chapters 13 through 15) have been published by the author and co-workers in different peer-reviewed conference proceedings and journals (see citations below). Also the new registration methods introduced in Part II have been described in these publications.

The application part of the thesis starts in Chapter 13 with the challenging problem of quantitative accuracy assessment of non-rigid registration results. Only few papers address this issue for non-rigid registration. Here we give a systematic overview on registration accuracy assessment in general and accuracy determination methods for non-rigid problems in particular. We introduce a few new methods, which are suitable to our applications. The backbone of the new methods are dense corresponding point sets defined on anatomical structures of real clinical image data in particular vessels [120]. Based on these reference point sets we define accuracy metrics, which partly take into account the inaccuracies involved in the point set determination.

In computer assisted liver surgery we are faced with different kinds of registration problems. It starts with the preoperative planning of liver resections, where different phases of contrast agent enhancement have to be registered to combine liver arteries, portal veins and hepatic veins into one precise 3D liver model. We have investigated the non-rigidity of the transformation between different phases [124, 125] but this is not part of this thesis.

After the resection planning the preoperative image and planning data have

to be transferred to the patient in the OR via intraoperative 3D ultrasound data. Due to deformations of the liver between pre- and intraoperative image acquisition a non-rigid registration scheme is needed. This second registration problem is asymmetric in a sense that preoperatively 3D geometrical models of the liver surface, the vessels and tumors are available in addition to the CT data. The challenge is the quality of the ultrasound data and the narrow time window of approximately 10 minutes (the fewer the better) in the OR. In Chapter 14 the performance of the new distance measures [119, 118] defined in Chapter 9 and the new combined landmark and intensity registration schemes [171, 122, 168, 172, 121, 123] introduced in Chapter 12 are analyzed. The resulting accuracy is determined on clinical data sets using the newly developed accuracy assessment methods explained in Chapter 13.

A third registration problem occurs, if an electromagnetic tracking system is used to realize motions and deformations of the liver during the actual resection. Parametric landmark schemes incorporating localization uncertainties as well as directional information provided by the electromagnetic tracking system can be chosen to solve this problem. While schemes which consider localization uncertainties are part of Chapter 11, the incorporation of directional information is not covered. No explicit section in the application part of the thesis covers registration problems arising from electromagnetic navigation systems.

The qualitative and quantitative comparison of pre- and postoperative image data is an important possibility to validate computer assisted surgical procedures [12]. Due to deformations after surgery caused by the removal of tissue a non-rigid registration scheme is a prerequisite for a precise comparison. This fourth registration task is tackled in Chapter 15 by interactive landmark-based schemes, which incorporate a priori knowledge about the anatomical structures to be registered. In addition to using point landmarks at vessel branchings, we introduce quasi landmarks at vessel segments with anisotropic localization accuracy [126, 127] as described in Chapter `sec:ParametricLandmarkRegistration`. These quasi landmarks help to reduce interaction time and improve accuracy. The accuracy of this challenging task due to missing anatomical structures is investigated on clinical data sets.

Chapter 2

Computer Assisted Liver Surgery

Liver surgery comprises various operations of the liver for different disorders. The most common operation performed on the liver is a resection (removal of a portion of the liver). The reason for a liver resection is usually the removal of a malignant tumor, but also resections are performed on people willing to donate a part of their liver to a person in need of a liver transplant. Although living donor liver transplantations can benefit from computer assistance too we will focus only on oncological resections here.

We will start with some facts about liver cancer, liver anatomy and possible therapies, before turning to medical imaging of the liver as the basis of computer assisted planning and intraoperative navigation systems, which will be covered in the last two sections.

2.1 Liver Cancer and Therapies

Liver tumors either originate in the liver (primary liver cancer) or develop in another organ and spread into the liver (metastatic liver cancer). Hepatocellular carcinomas (HCC) are the most common form of primary liver cancer, but the majority of liver lesions are metastases coming from the lung, breast, colon, pancreas, or stomach, with the colon being the major origin among them. Often multiple metastases are detected. The liver with its capillary vessels is a blood filter and thus is relatively frequently affected by metastasis in particular for primary tumors in stomach, pancreas, colon and rectum, which are connected to the liver via the portal vein system. Up to 50% of the patients with a colorectal carcinoma are developing liver metastasis. Colorectal cancer is the second most common cancer site for men as well as for women. The number of new cases (incidence) per year in Germany is about 37,000 for men and 36,000 for women. Colorectal carcinoma are the second leading cause of cancer-related death for both men and women in Germany [232]. This means about 18,000 patients died due to colorectal cancer in 2007 in Germany. These are 2.2% of all deaths according to the “Statistisches

Bundesamt Deutschland” – Federal Statistical Office Germany.

The treatment of liver cancer depends on the size, location, number and type of the lesions or the primary tumor. Obviously there is a phase in the development of liver related cancer where the location of metastases is limited to the liver and no extrahepatic metastases can be detected. In this phase local therapy principles – surgical as well as non-surgical – are effective. That means an accurate staging based on adequate imaging methods, surgical exploration or biopsies is an important prerequisite for choosing the best therapy option. Surgical resection is the gold standard for the treatment of liver metastases with an achievable five years survival rate of 30% [204]. In the last decades hepatic resection has improved to a safe and efficient procedure due to refinement in diagnostic, anesthetic and operative techniques, like better tissue dissection technologies. As modern cancer therapy is multidisciplinary usually additional chemotherapy and/or radiation therapy is applied to either reduce the number of tumor cells before surgery (adjuvant therapy) or to reduce the risk of recurrent tumors after surgery (neoadjuvant therapy).

For the majority of patients with colorectal liver metastases resectional therapy is not suitable because of multiple metastases in both liver lobes or extrahepatic disease. A potentially curative surgical intervention is possible only in 10 – 15% of the patients with liver metastasis. Other therapy options are available for those patients. Besides the local injection of chemical substances (ethanol injection and chemoembolization), thermal ablation methods are an important therapy possibility. The idea of thermal ablation techniques is to heat (or cold for cryotherapy) a tumor locally in the range between 60°C and 100°C leading to a coagulation necrosis of the tumor tissue. The application of thermoablative methods is limited to a small number of metastases (≤ 5) and the diameters of the lesions are ideally small (different specifications: smaller than 2.5 or 5.0cm). Different technologies exist to induce the heat: laser-induced thermal therapy (LITT), radio-frequency ablation (RFA) and high-intensity focused ultrasound (FUS or HIFUS). For LITT and RFA an applicator needs to be placed under ultrasound, MRI or CT control either through the skin (percutaneously) or directly into an organ during surgery. For bigger lesions multiple applicators can be placed into the tumor. High-intensity focused ultrasound [225, 49] is a promising thermal ablation variant for certain applications, which is applied transcutaneously and hence is completely non-invasive.

Detailed information about regional and minimally invasive tumor therapies can be found in the books of Schlag and Stein [205] and Stroszczyński [223]. We mention these forms of therapies, because they also benefit from similar computer assisted planning and intraoperative navigation techniques as surgical interventions, which we focus on and are explained in more detail in the next section.

2.2 Liver Surgery

2.2.1 Aims of Oncological Liver Surgery

For surgery of malignant liver tumors resection margins are important for long-term outcome [67]. An R0 resection is defined as one in which all margins are histologically free of tumor. On the contrary, an R1 resection is a resection with microscopic residual tumor cells been left behind. An R2 resection is defined as gross residual tumor. Even though the exact width of surgical margins remains controversial [26, 173, 265] patients with a microscopically positive margin (R1) have a worse prognosis than R0-resected. In conclusion the oncological aim of a liver resection is the complete removal (R0 resection) of the tumor with a safety margin of approximately 1 cm around the tumor.

But at the same time as much healthy tissue as possible should be preserved to achieve a sufficient postoperative liver function. For this functional aim the exact knowledge of the liver's vessel system is crucial to assure the blood supply and drainage of the remaining liver tissue. The size, location or distribution of the tumors might be so disadvantageous, that a surgical removal is not possible without losing the liver function. It is the art of surgery to decide, whether a tumor is resectable and to find a compromise between required radicality and optimal preservation of liver tissue function. To ascertain the best compromise the surgeon has to know the principal and individual anatomy of the patient's liver and the 3D location of the tumor in relation to the liver vessels to perform a precise risk analysis and resection planning.

2.2.2 Liver Anatomy

The specific characteristic of the liver vessel system is its dual blood supply consisting of the portal vein and the hepatic arteries. The portal vein coming from stomach and intestine supplies approximately 75% of the liver's blood. This blood has already delivered its oxygen to the organs of the gastrointestinal tract before coming to the liver and therefore cannot supply the liver cells with oxygen. The liver cells receive all of their oxygen from the 25% of blood delivered by the hepatic arteries. The liver is drained by the hepatic veins, which end at the vena cava. A further vessel system in the liver is the bile duct system. Bile is excreted in the liver and required for the digestion of food. Bile ducts transport the bile from the liver either directly into the duodenum or for intermediate storage into the gall bladder.

Usually the portal vein divides immediately before reaching the liver into a left and right branch supplying the left and right liver lobe. The portal vein ramifies further inside the liver into smaller and smaller venous branches. The hepatic arteries and the bile ducts usually run parallel to these portal vein branches inside the liver forming independent functional units with their

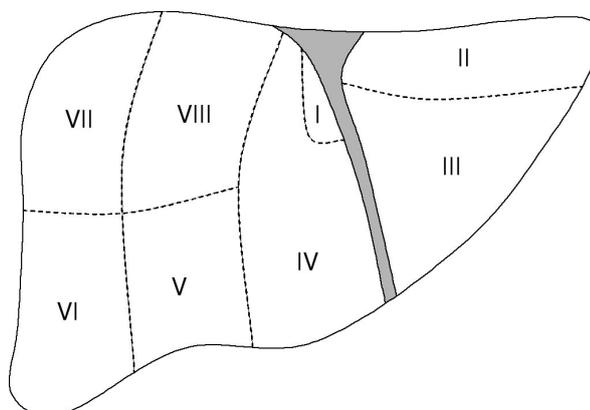


Figure 2.1 Schematic illustration of eight functional liver segments according to Couinaud.

own blood supply and bile drainage. Such a functional unit can therefore be surgically resected without jeopardizing the function of the remaining parts. As the portal vein branches like a tree a hierarchy of self-contained functional units exist. Similarly to the supply system the drainage by the hepatic veins can also be divided into functional units. All parts of the liver tissue, which drain into the same main hepatic vessel branch constitute a functional unit. The supply and drainage units only partially overlap. The art of liver resection planning is to find a good compromise between the damage to the supply and drainage of the liver after resection.

Different vessel hierarchies have been introduced to get a non-intersecting and complete division (partitioning) of the liver into so called liver segments. The liver segment division of Couinaud[40] is the clinically most relevant. According to Couinaud eight portal venous liver segments are distinguished, which are numbered in clock-wise order starting with the caudate lobe as segment I (see Fig. 2.1). Each segment represents a functional unit, which is supplied by a third-order branch of the portal vein, which leads theoretically to $2^3 = 8$ non-intersecting segments. Current anatomical and radiological investigations have shown, that the regularity of the vessel branchings described by Couinaud is only given in few cases. Actually a multiplicity of portal and hepatic venous branching variations with in some cases very different sizes of supplied parenchyma regions exist [54]. The geometry and topology of the vascular system is highly individual. There are even three different variations of the first main branching [111]. Three very different branching patterns are shown in Fig. 2.2.

Selle [214] tried to find an appropriate hierarchical order on the portal vein tree, for which an automatic algorithm can be used and agrees with the assigned vessel parts (subtrees) to Couinaud's segment partitioning by radio-

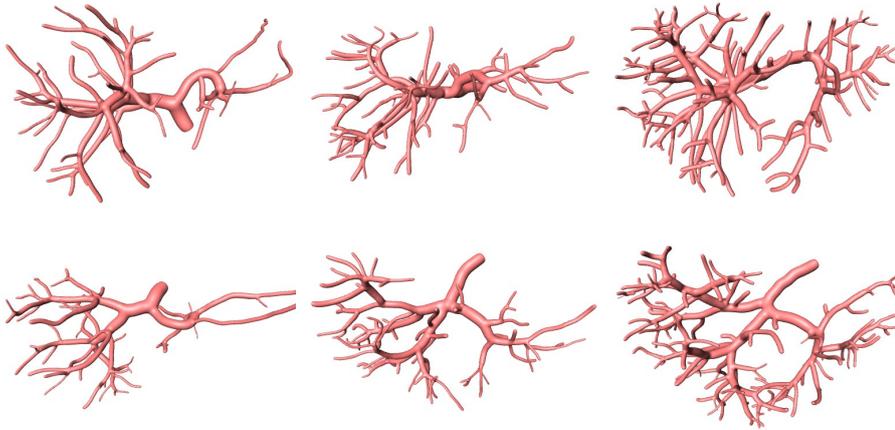


Figure 2.2 Portal veins of three patients (column-wise) with different branching patterns. The first row shows a coronal view while the second row shows a cranio-caudal view.

logical experts. Although he found a hierarchical order and corresponding algorithm which reproduced the manual assignment by an expert quite well, differences still remained. The problem is that in medical practice the Couinaud scheme does not really define a hierarchy, because parallel branches can also be assigned to one liver segment. Radiologists seem to have a regular spatial partitioning of the liver in mind and try to assign vessel subtrees to these spatial regions. From the functional view of the liver it has to be the other way round: which are the biggest functional units, which then define the spatial partitioning of the liver? According to Selle [214] it is a contradiction to find functional independent segments, which are supplied by one main vessel branch on the one hand and on the other hand to use a schematical spatial partitioning in medical practice, where functional self-contained tissue regions are separated and reordered to assign them to different Couinaud segments. In conclusion the spatial partitioning of Couinaud is only a (more or less) standardized coordinate system of the liver to give a coarse orientation in the radiological and surgical practice to describe the approximate position of a lesion inside the liver. We are not aware of any studies which show how reproducibly radiologists are able to partition individual livers exactly into the Couinaud segments. In addition the Couinaud scheme is not sufficient for the exact and reproducible division into functional units of an individual patient. We will see, that with computer assisted modeling of the vessel tree, a general scheme is even not necessary anymore.

2.2.3 Open Liver Surgery

Resections based on the extent of the tumor only (including safety margin) are called atypical. Only in the case of small tumors lying close beneath the liver surface, atypical resections should be performed. In all other cases, the spatial relation of the tumor to the vascular system of the liver has to be taken into account (anatomical resection). As described in the previous section whole functional units of the liver should be resected to ensure a sufficient blood supply and drainage of the remaining liver tissue. This is also called a segment oriented resection. Usually in conventional liver surgery this means Couinaud segments, which ignore the before mentioned imprecisions induced by anatomical variations. A precise comprehension and determination of a segment considering the individual functional units can be achieved by modern medical imaging, image processing and 3D modeling as described in more detail in Sec. 2.3. For conventional as well as for computer assisted liver surgery a risk analysis and determination of the resection volume based on CT images has to be performed.

According to the preoperative imaging and planning, the surgeon aims to resect a defined part of the liver with as little blood loss as possible. The general surgical procedure starts with the opening of the abdomen (the access). Before the actual liver tissue resection can be performed the liver is mobilized and explored. This means the liver is separated from its ligaments, peritoneum and the surrounding septum. Afterwards the liver is explored for additional findings. Nowadays this exploration is often complemented by intraoperative sonography. If the resection of the liver is still indicated, the liver hilum and the vena cava at the junction of the hepatic veins are prepared. The hilum is the central area of the liver, where the bile duct, the portal vein and the hepatic artery enter the liver. The kind and necessity of the preparation of hilum and vena cava depend on the kind and size of the resection. A widespread technique to reduce the bleeding risk is to conduct a Pringle maneuver, which occludes the liver hilus for maximally 30-60 minutes by a vessel clamp. Also the occlusion of hepatic veins at the vena cava is performed in some cases. After this preparation the actual organ tissue (parenchyma) dissection begins. Different dissection technologies are available today to reduce the blood loss: water jet dissectors, ultrasonic dissectors, stapler and diathermy. Small liver vessels are ligated and big vessels are clipped. If the liver resection is finished care of the resection surface and the closing of the abdominal cavity has to be performed. The resected tissue is investigated by a pathologist to classify the tumor and to see if a R0, R1 or R2 resection has been achieved. Instantaneous section methods allow intraoperative pathological investigations. An excellent and detailed German description of modern liver surgery has been published by Lang [111, 112].

2.2.4 Minimally Invasive Surgery

Minimally invasive surgery has become a viable alternative to conventional surgery. Laparoscopic surgery is a minimally invasive technique in which operations in the abdomen are performed through significantly smaller incisions compared to traditional surgical procedures. A camera system, a so-called laparoscope, is inserted through a tube (the trocar) into the abdominal cavity and makes its organs visible on a monitor. Custom-build surgical instruments are inserted to further incisions and the operation is performed under visual control supplied by the camera system. The technical advantages of minimally invasive surgery can be translated into clinical benefits for the patients, i.e. less postoperative pain, better cosmetic results, shorter hospitalization, and earlier convalescence. Laparoscopic operations have replaced a significant proportion of open surgical procedures and are now routinely used. While the role of laparoscopic surgery has been generally accepted for the management of benign disorders, there is an ongoing debate regarding the adequacy of this technique in surgical oncology. There is evidence that minimally invasive surgery can reduce perioperative morbidity in cancer patients. However, definite validation of these procedures for tumor surgery is not yet available due to the lack of prospective randomized trials providing reliable long-term data on disease-free survival and overall survival. It seems likely that minimally invasive procedures will play an important role for the treatment of some kinds of lesions and tumors of limited size.

There are some technical limitations of laparoscopic surgery. The degrees of freedom for the instruments are limited due to the minimal invasive access via the trocars. The absence of direct organ palpation and the lack of the third dimension in the 2D video images are still limits of laparoscopy. The surgeon's orientation and the location of anatomical and pathological structures is therefore more difficult than in open surgery. Modern image-guided surgery systems have the potential to compensate these limitations.

2.3 Computer Assisted Surgery Planning

Although today's imaging methods like multi-detector CT provide excellent visualization of the intrahepatic vascular system, neither the number and distribution nor the extent of functional liver segments can be determined distinctly. Hence areas at risk for devascularization (or drainage loss) can be identified only vaguely and the prediction of the remaining fully vascularized liver parenchyma is inaccurate. With an accurate prediction an optimal surgical planning for an individual patient could be performed. A surgical planning which considers a predicted postoperative function is also called functional planning. On the basis of modern medical imaging the anatomy and function of the liver can be modeled and simulated with computer software. For a gen-

eral introduction to computer assisted planning and in particular functional and model-based planning see also our according (German) book chapter [267].

Before reviewing computer assisted planning for liver surgery in Sec. 2.3.3 preoperative medical imaging of the liver is summarized in the following sections as a basis of the planning process.

2.3.1 Preoperative Medical Imaging

Traditionally the imaging of the liver concentrates on the detection and classification of lesions. The question is whether a lesion is benign or malignant. The aim is to get a high contrast between lesion and surrounding liver tissue. Usually contrast agents are applied in MR or CT imaging to improve the contrast of the lesion, because its blood supply is higher or lower than the surrounding tissue. In addition the contrast of liver vessels itself is increased by the contrast agent such that the location of the lesion in relation to the liver vessels can be displayed. The enormous technical advances in CT imaging has led to high-resolution 3D volume images of the liver. Today CT is the standard diagnostic imaging technique of the liver and will be presented in more detail in the next section.

Contrast-enhanced Multiphase CT Data

We will not explain how computed tomography works in detail. We just outline what is important to know about CT for building detailed 3D digital models of the liver, its vessels and the tumor.

CT is very suitable for the generation of exact geometrical models, because image distortions are very low compared to certain MRI sequences. Streak artifacts caused by metallic objects like dental fillings or implants rarely occur in the liver region. Instead the avoidance of motion artifacts caused by respiration plays an important role in liver imaging. Despite the development of different prospective and retrospective gating techniques, the easiest and most effective way to avoid respiratory motion artifacts is to ask the patient to hold his/her breath during image acquisition. This means the image acquisition has to be performed in a time interval of 30-60 seconds which is ambitious for a scan of the whole liver because it is a relatively big organ measuring approximately 15 cm in cranio-caudal direction.

The introduction of spiral or (mathematically more correct) helical CTs in 1989 allowed to acquire such large volumes of data very quickly while reaching reconstructed slice intervals of 3-5 mm for liver imaging. In contrast to conventional sequential CT helical CTs have a gantry, which rotates continuously while the table with the patient is also moved continuously through the gantry. The resulting trajectory of the X-rays has the shape of a helix. For the reconstruction of planar slices perpendicular to the helix a linear interpolation between adjacent scans is performed [99].

Wang and Vannier [244] analytically investigated the influence of the three most important geometrical imaging parameters (detector collimation width D , table increment per gantry rotation T , reconstructed slice number per collimation n) on the imaging characteristics (effective slice thickness σ , image noise level η , signal-to-noise ratio). Although based on a simplified imaging model for helical CTs their results were consistent with empirical findings on phantoms. The dimensionless ratio of the table feed per rotation of the X-ray source T in mm to the collimated slice thickness D in mm is called the pitch $p = D/T$. The effective slice thickness σ is assumed to be the standard deviation of the slice sensitivity profile (SSP). The SSP is the convolution of the original sensitivity profile induced by the collimator and the table motion function, which is a triangular function in case of linear interpolation [98]. Ideally the collimation profile is a rectangular function.

The most important conclusions Wang and Vannier [244] got from their analytical investigation of helical CT image quality are: 1. The optimal pitch is equal to $\sqrt{2} \approx 1.4$. With a pitch smaller than 1 some regions are not scanned (gaps) and with a pitch close to 2 the image gets blurry. To avoid longitudinal aliasing, at least 2–3 transverse slices should be reconstructed per collimation.

After the introduction of the first four-slice helical CT systems in 1998 an evolution process in CT development started, which W. Kalender called: the slice race. Today there are multi-slice (or multi-detector) CTs, which scan up to 320 slices simultaneously. Multi-slice CTs enable significantly lower scan times for the same scan volume compared to single-slice CTs. Thus it is possible to get high image resolutions (below 0.5 mm) in a short period of time allowing whole liver scans during a respiration rest. Motion artifacts are rare in liver imaging today except for some artifacts in the upper left region of the liver caused by heart motion. Due to the high resolution partial volume effects are decreased and vessels with quite a small diameter can also be detected. The only disadvantage of the high resolution is the increase of image noise. But with image smoothing and slight resampling to 1 or 2 mm slice resolution we get high quality images.

Further and more detailed information about CT principles and technologies with a special focus on modern single- and multi-detector helical CT imaging can be found in the review of Kalender [97] and the books of Terrier [226], Kalender [96] and most recently Buzug [24].

After showing how for a whole liver scan a high resolution in z direction can be obtained during one apnea phase we will now turn towards the issue of reaching a high contrast to depict small lesions and small vessels. As mentioned above usually a contrast agent is applied enhancing the contrast of lesions and vessels to the surrounding liver tissue significantly. Due to the physiology of the liver the timing of the image acquisition phases after mechanical intravenous contrast agent injection is important. Today triphasic liver CT imaging is performed, which means that scans are acquired at three



Figure 2.3 Arterial, portal venous and late venous scans of a multiphase CT scan after 18, 38 and 78 seconds of contrast agent injection.

different time delays after contrast agent injection (see Fig. 2.3). After approximately 18 seconds the contrast agent arrives in the liver arteries resulting in a high contrast between arteries and liver tissue (Fig. 2.3 a). Some tumor entities might be best detectable in this early arterial phase. Delays of 20 and 60 seconds from the beginning of the early arterial scan are used for the portal venous (PV) and late venous phase scans, respectively.

In the portal venous phase the hepatic veins are usually not enhanced (Fig. 2.3 b). However, in the hepatic venous (HV) phase, portal veins are typically also visible, but with lower contrast as in the portal venous phase (Fig. 2.3 c). PV as well as HV images are acquired during respiration hold, which is usually at end-inspiration. Due to patient respiration between the two acquisitions the position and shape of the liver sometimes cannot be reproduced exactly. Thus if portal and hepatic veins from different phases are intended to be integrated into a combined 3D liver model, the phases have to be registered. We evaluated quantitatively 10 image pairs showing that the portal veins move rigidly between 1.3 and 12.3 mm and on average 7.1 (+/- 4.2) mm, while the remaining non-rigid deformations are in the worst cases in the range of 2-3 mm [124, 125]. In conclusion in most cases a rigid registration based on mutual information and on automatically generated liver masks was sufficient. Non-rigid registration was successful, but necessary only in 3 out of 10 cases. Fig. 2.4 shows the portal veins of one case in PV and HV phase in their original position, after masked rigid and non-rigid registration of the HV phase. In summary with modern CT scanners very high resolution preoperative 3D images of the whole liver can be obtained with high contrast of the vessels and with no substantial artifacts.

2.3.2 Geometrical Modeling of Anatomical Structures

An important prerequisite for computer assisted liver resection planning is an accurate geometric 3D model of tumor, vessels and liver surface. These struc-

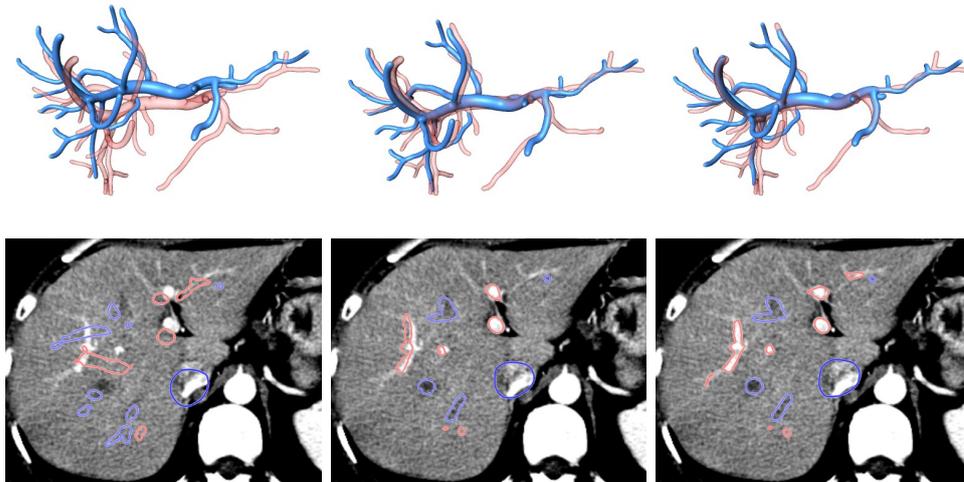


Figure 2.4 The first row shows 3D models of the portal veins in the portal venous phase (transparent) and the hepatic venous phase (opaque). In the second row one slice of the portal venous phase with intersection lines of the liver surface (yellow), the portal veins (pink) and the hepatic veins (blue) from the hepatic venous phase are shown. From left to right the original position of the portal veins, after masked rigid and after non-rigid registration are shown.

tures are not easily extractable with simple image processing techniques even from modern multi-slice CTs and contrast agent application. The 3D modeling process usually consists of two parts. At first each voxel (volume element) of the image data is unambiguously assigned to an anatomical structure: liver parenchyma, tumor tissue, portal or hepatic veins (sometimes also hepatic arteries and bile ducts). The process and the result of this assignment is called segmentation. In a second step a triangle mesh which represents the surface of liver, vessels and tumor is automatically computed by the Marching Cubes algorithm [137] and surface simplification resp. surface smoothing algorithms are applied.

Segmentation is a major research area in medical image processing [110]. Although many different algorithms have been developed the comparison of their performance on clinical relevant image data has hardly been possible until some segmentation challenges started in 2007. The organizers of the challenges provided a data base of clinical image data with reference segmentations where radiological experts manually delineated the contour of the liver in a slice-by-slice fashion. In addition they provided different error metrics, which quantify differences between results obtained from different segmentation methods with the ground truth reference segmentations. The metrics have been summarized into one score to be able to order the performance of the algorithms. The first challenge called “3D Segmentation in the Clinic: A Grand Challenge” covered

the segmentation of the liver surface [84].

Segmenting the liver surface is a challenge because even with contrast agent application the contrast between liver tissue and surrounding tissue is often low and the intensity of the liver varies. Therefore interactive segmentation methods like intelligent scissors/ live wire [203] combined with shape-based interpolation to save interaction time are in wide-spread use [203]. Recently automatic approaches based on statistical shape models were significantly improved [84] and we reached an accuracy (average surface distance) below 1 mm on average [95] with such a method.

In most cases liver tumors are segmented manually due to their low contrast and intensity heterogeneity. But some liver tumors appear quite dark and homogeneous and can be segmented by volume growing based techniques [151]. Similarly to the liver surface segmentation contest, a contest for the segmentation of liver tumors was arranged “3D segmentation in the Clinic, A Grand Challenge II – Liver Tumor Segmentation (2008 MICCAI workshop)”. Most approaches were semi-automatic, but also some automatic approaches were shown.

Liver vessels are mostly segmented by volume growing techniques. Selle et al. [215] presented a preprocessing strategy and a technique to find the threshold for volume growing segmentation of liver vessels automatically. Depending on the image quality, in some cases even an optimal global threshold is not sufficient to segment smaller, but structural important vessels. Therefore we used a volume growing segmentation technique. More sophisticated automatic vessel segmentation algorithms have been presented in the literature, but the application to liver vessels in contrast-enhanced CTs is very rare. A comprehensive review of vessel segmentation methods is given in [131]. An interesting evaluation method for the segmentation of liver vessels has been presented by Lehmann et al. [128]. They acquired CT scans of a porcine liver model and then created in situ corrosion casts by filling the vessels with acrylic resin. Afterwards CT scans of the corrosion cast were acquired. The segmented vessel branches of the living liver and the corrosion cast were verified with the physical corrosion cast.

From the segmented vessels, tumor(s) and liver parenchyma the according surfaces are automatically generated with the marching cubes algorithm and smoothed afterwards. For the generation of the vessel surfaces so-called convolution surfaces [164] were used, which are more suitable for tube-like structures with small radii [164]. Examples of vessel convolution surfaces are shown in Fig. 2.2 and Fig. 2.5.

2.3.3 Risk Analysis and Resection Planning

Several systems have been developed for liver surgery planning in the last couple of years which are based on 3D models generated from CT or MRI

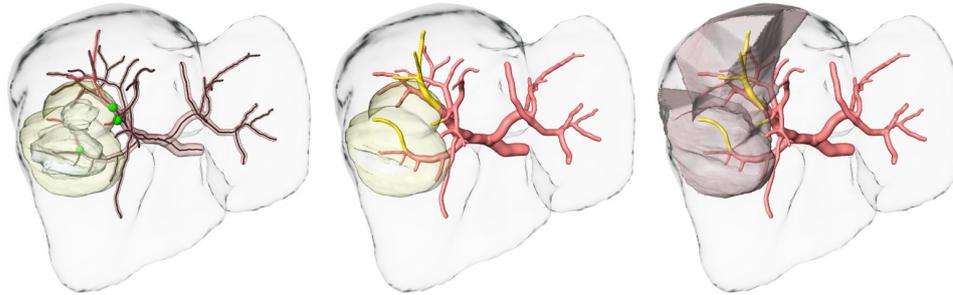


Figure 2.5 a) Vessel center lines computed from the segmented portal veins and the tumor with a safety margin of 1 cm. The vessel center lines inside the safety margin are shown in red. The cut-points of the center lines with the safety margin are shown in green. b) The dependent vessels below the cut points of the vessels with the safety margin. c) The part of the liver tissue (15%) which is supplied by the dependent vessels.

data [221, 215, 144]. There is even a commercial service which offers processing of CT/MRI data for liver surgery planning (MeVis Medical Solutions AG, Distant Services). These systems are applied for planning living donor liver transplantations (LDLT) [69, 81] and oncological liver resections for individual patient anatomies [87, 166, 182, 111, 112]. The visualization of a virtual 3D model of the liver is a valuable tool for the surgeon to get a better imagination of the individual vessel anatomy and in oncological liver surgery the spatial relation of a tumor to these vessels. The distance of the tumor to the surrounding vessels can also be quantified. Anatomical variants such as trifurcation of the portal vein or accessoric hepatic arteries are a common finding and can be appropriately visualized for the surgeon.

For LDLT as well as for oncological resections it is important to know the individual vascular territories as exactly as possible. A vascular territory represents a part of the liver that is supplied or drained by a certain vascular subtree. The root of the vascular subtree might be the cut of a security margin around the tumor with the vascular tree or manually defined by the surgeon. For the determination of the vessel branches which are below (in blood flow direction for portal veins) this cut point, the center lines of the vessels are automatically extracted by a skeletonization algorithm [201] from the segmentation of the vessels (Fig. 2.5a). The center lines are transferred into a directed graph. Now the vessel subtree below the cut point is computed by a breadth-first (or depth-first) search on the directed graph (Fig. 2.5b). The idea of computer-based vascular territory determination is that the closer a liver tissue voxel is to a vascular branch the more likely is it for this voxel to be supplied by the given branch. Although this is only a coarse approximation of the underlying physiology good estimations are obtained [215]. For each

liver tissue voxel the nearest vessel point is determined and for each point on the vessel center lines a list of the liver tissue voxel is saved for which this center line point is the nearest. Now for all points on a vessel subtree the dependent (nearest) liver tissue voxel are collected defining the vascular territory of this subtree (Fig. 2.5c). Then the portal venous supply and also the venous outflow of a vascular territory are computed [81, 182, 113]. With modern multi-detector scanners even hepatic arteries and bile ducts can be incorporated [81].

Based on the vascular territories an automatic risk analysis of blood supply and drainage can be performed according to the location of a tumor or an interactively defined resection plane. By interactively changing the resection plane the impact of different surgical strategies on the arterial devascularization and venous congestion volumes can be predicted. Lang et al. [113] state that computer assisted risk analysis can influence the operation planning of liver resections compared to standard 2D CT. The consistency of the computed virtual vascular territories and real territories has only been evaluated on corrosion casts so far [215]. An evaluation for the liver in vivo is an important task for further research.

2.4 Intraoperative Imaging and Navigation

Although further clinical studies are needed to prove the clinical benefit of computer assisted liver resection planning the main technical difficulties seem to have been solved and information about accurate individual vascular territories have led to changes in the surgical strategy [113]. But to verify the clinical benefit it has to be assured that the planned liver resection is really implemented on the patient. The challenging task is to transfer the plan prepared on preoperative CT data accurately and securely to the intraoperative situation. Intraoperatively the location of the tumor and relevant vessels is hidden underneath the liver surface and the correct location of the resection line can only be estimated. Some tumors close to the surface might be palpable by the surgeon, but deeper tumors where important vessels are usually close by often are not palpable.

Intraoperative navigation systems support the surgeon by visualizing the spatial relation of surgical instruments with respect to invisible anatomical structures. Conventional navigation systems like in maxillofacial-, ENT- and neurosurgery, which are based only on preoperative data, are not suitable for liver surgery. The reason is that the liver significantly deforms between preoperative imaging and the surgical procedure. Even in neurosurgery significant deformations of the brain occur due to gravitation after the opening of the dura: the so called brain-shift [155, 160]. To solve this brain-shift problem intraoperative imaging is used to measure the current shape and position of the brain and its structures. For neurosurgery MRI is an option to acquire

high quality intraoperative images [162, 161]. But it is a challenge to implement an intraoperative MRI and the costs are high. For open visceral surgery intraoperative MRI is not suitable due to the bigger operation field. But there are interventional applications like Laser Induced Thermal Therapy (LITT), which have been used in conventional [241] and open MRI scanners. Even first attempts to perform laparoscopic procedures inside an open MRI scanner have been made [31]. The intraoperative use of CT is limited due to radiation exposure and low soft tissue contrast.

A flexible, relatively cheap imaging modality which can easily be integrated into the OR is ultrasound. Navigation systems based on intraoperative 2D or 3D ultrasound have successfully been developed for neurosurgery and recently for liver surgery [15, 10, 11], too. We developed two navigation systems for liver resections based on intraoperative 3D ultrasound. The first system uses an optical tracking system and the ultrasound volume is updated from time to time. This is also called an iterative system because the intraoperative imaging is not continuously acquired. The problem is that the surgeon does not know when to acquire a new ultrasound volume due to significant movements or deformations. The second navigation system continuously measures the position and orientation of some sensors inside the liver by an electromagnetic tracking system. Before we explain both navigation systems in more detail we will summarize some facts about general ultrasound imaging and explain 3D ultrasound in particular.

2.4.1 Intraoperative Ultrasound Imaging

Ultrasound (also called sonography) is a widespread modality in diagnostics and in particular for abdominal organs like the liver. As ultrasound is a demanding imaging modality for image processing we give some fundamentals about ultrasound image formation, image characteristics and typical artifacts.

Ultrasound Image Formation

For ultrasound imaging, sound waves at a high frequency of 1 to 15 MHz are emitted from a handheld ultrasound probe (transducer) into the respective body part. The sound waves penetrate into soft tissue at a speed of 1450 to 1580 m/s, but are partially reflected at interfaces between different tissue types. This is due to the mismatch between the speed of sound and densities in different tissues. The reflected sound waves are measured by sensors also integrated into the transducer. This means the transducer is generator as well as receiver of sound waves. As the traveled distance of the sound waves is proportional to time, the traveling time is used to determine the position of the acoustic interfaces. The greater the difference between acoustic impedances at the interface, the larger the echo is. The acoustic impedance of a material is the product of the density and the propagation velocity of the ultrasound

waves in this material. At interfaces to air and bone the density differences are so big, that almost all acoustic energy is reflected and no imaging beyond this interface is possible. Thus the physical contact of the probe with the body is essential to avoid air between transducer and body. Therefore coupling gels between probe and body are used.

Not all of the transmitted energy is reflected. Most of the energy is absorbed by the tissue. The higher the frequency is, the greater is the absorption and the lower the penetration depth. Time-gain compensation (TGC) is used to compensate absorption. With increasing time resp. distance the signal is amplified by a pre-determined TGC profile provided by the manufacturer. But the gain can also be changed manually for a set of distances.

The specular reflection is only one part of the returning waves, scattering is another part. Scattering or speckle noise is caused by cellular structures, which are much smaller than the wave length of the ultrasound wave. Speckle results from the constructive and destructive interference of reflections from the sub-resolution scatterers. The image appears grainy. The speckle pattern is characteristic for the kind of tissue, which is used by physicians to differentiate tissue types.

The generation of the ultrasound waves is comparable to usual loudspeakers: an electrical stimulus is transformed into mechanical vibrations in an elastic medium by a transducer. The reverse piezoelectric effect is used in ultrasound probes, by exciting piezoelectric crystals by an electrical voltage. The summation of all waves generated by the piezoelectric crystals forms the ultrasound beam. In pulse-echo imaging short ultrasound pulses with a frequency of 1-15 MHz are emitted through the tissue and after a period of time, sufficient for the previous echoes to return or die out, another burst of ultrasound is emitted and the cycle is repeated. The returning sound wave vibrates the transducer and the transducer turns the mechanical vibrations into electrical pulses (piezoelectrical effect). The ultrasound scanner then processes and transforms the received signals into a digital image by analyzing the travel time.

Four different modes of ultrasound exist: A-mode, B-mode, M-mode and Doppler mode. We will explain B-mode and Doppler mode, because they are the only relevant modes for the imaging of the liver. In the B-mode or brightness mode, the resulting pixel represents the strength of the echoes at the respective body parts. An array of transducers simultaneously scans a plane through the body resulting in a 2D image.

Three general types of transducers exist: linear, curvilinear and phased array transducers. Linear array transducers produce parallel beams imaging a rectangular area, curvilinear arrays generate radial beams imaging a sectional (fan-like) area. Phased array transducers offer a high flexibility to form different wave fronts by certain phase delay patterns. This is also called beam

forming and can be used to steer the wave front into a certain direction which allows ultrasound probes with a small contact area or to focus the ultrasonic pulses to one or several distances.

Doppler ultrasound measures blood flow using the Doppler effect. This means Doppler imaging measures *movements* of scatters. A series of pulses (also called a wave train) is transmitted into the same direction. Echoes from stationary tissue remain the same from pulse to pulse. For moving scatterers the frequency differs from pulse to pulse due to the Doppler effect. It can be measured whether the blood is moving towards or away from the ultrasound probe, and its relative velocity. The direction (towards or away from the probe) of the flow is colored with blue and red (Color Doppler). But there is no standard which color represents which direction. Usually the frequency shift is determined via the phase difference. The phase differences between consecutive pulses are detected by autocorrelation. Power Doppler differs from conventional Color Doppler in the way the Doppler signals are processed. Instead of estimating the mean frequency and variance, the integral of the power spectrum is estimated. The colours in the power Doppler image indicate only that blood flow is present. No information on flow direction and velocity is given.

For liver surgery it is important to know where a vessel is located. The flow velocity and direction in the vessels is not relevant. Thus power Doppler ultrasound is suitable for intraoperative imaging of the liver vessels anatomy. In Fig. 2.6 intraoperative images of the liver are shown. In B-mode imaging the vessels appear dark and the reflections on the backside (diaphragm) of the liver (lower part of the image) appear bright. The interface between liver and diaphragm is large and smooth reflecting almost in normal direction. If the angle between interface and beam direction is low, the ultrasound waves will not be reflected back to the transducer and the bright curve will be interrupted. Due to (periportal) fat tissue the region around the portal veins appear bright. The tumor (upper left) also appears dark. The power Doppler image is shown as a yellow overlay onto the B-mode image (Fig. 2.6a) and as an original intensity image (Fig. 2.6b). If no significant flash-artifacts are present the vessels can be directly visualized by 3D volume rendering of the power Doppler image (Fig. 2.6c). In comparison the portal (pink) and hepatic (blue) veins extracted from a preoperative CT of the same patient are shown from a similar view direction (Fig. 2.6d). In particular two main branches of the hepatic veins shown in the CT model can be identified well in the 3D volume rendering of the intraoperative power Doppler image.

More details about the technical basics of ultrasound imaging can be found for example in the books of Suri et al. [224] or Gibbs et al. [70] and the thesis of Wein [248].

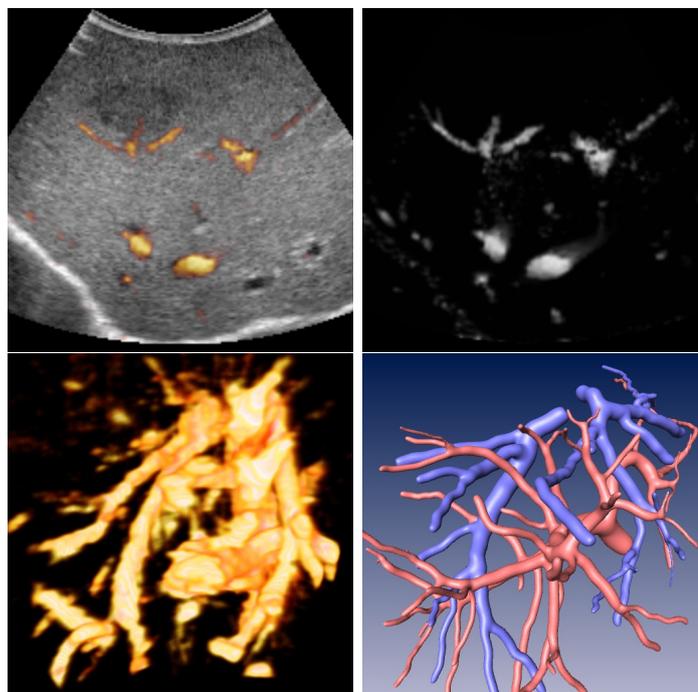


Figure 2.6 Upper left: An intraoperative B-mode ultrasound image of the liver with overlaid power Doppler image. Upper right: Intraoperative power Doppler image without B-Mode. Lower left: 3D volume rendering of the power Doppler image. Lower right: 3D model of the vessels extracted from preoperative CT data of the same patient.

Ultrasound Imaging Characteristics

The axial resolution (in beam direction) of an ultrasound image is determined by the frequency resp. wave length. The higher the frequency the shorter the wavelength and the higher the resolution. The wavelength of a 5 MHz ultrasound beam is approximately 0.3 mm. It would not be possible to resolve objects which are less than 0.3 mm apart. As already mentioned above the drawback of higher frequencies is the decrease of depth penetration due to stronger attenuation. The liver is imaged at lower frequencies of about 3-6 MHz to get a sufficient depth penetration. The lateral resolution (perpendicular to the beam) depends on the beam width, which again is inversely related to the frequency, but depends also on the beam forming and the aperture of the probe (probe diameter). The width of the beam varies with the depth. The smallest width and thus the highest resolution is reached in the depth of the focal zone. It is therefore clinically important to focus the target structure in the focal zone. The ultrasound beam has also a width perpendicular to the image plane which is in particular important in 3D ultrasound for the

elevational resolution between image slices. The beam width perpendicular to the image plane is the same as the beam width inside the image plane for circular apertures, but may differ for transducers with rectangular apertures like linear or phased array transducers.

Besides the image resolution contrast is the second most important image characteristic. If we can not differentiate relevant structures by different intensities even high resolution images are useless. The contrast resolution depends on the electrical and acoustical properties of the probe and the amplifier system as well as on the properties of the imaged object. If the acoustic impedance difference between neighbored tissues is low also the contrast will be low. It is important to know that the original signal intensities are compressed to the intensities on the display. The dynamic range between the smallest and biggest signal intensity is 30-40 dB. This original range is reduced by logarithmic compression to match the smaller dynamic range (256 intensities) of the display and to emphasize objects with weak backscatter.

Even a high contrast can be substantially corrupted by noise. Usually the signal- or contrast-to-noise ratio is used to quantify how much the image is corrupted by noise. In addition to the usual noise of the electronic of the system (e.g., the amplifier) we already mentioned speckle noise. Standard smoothing filters (e.g., Gaussian smoothing) are not suitable to reduce speckle noise. But anisotropic diffusion filters have been developed which consider the special structure and distribution of speckle noise [266, 107]. Speckle noise can be modeled by a Rayleigh distribution [242] but the accurate description of the speckle statistics is still an active field of research [107]. In addition the logarithmic compression of the displayed ultrasound images has to be taken into account leading to a log compressed Rayleigh distribution.

Ultrasound Artifacts

Many different types of artifacts exist for B-mode and power Doppler sonography. We will focus on the most important artifacts with implications for computer assisted liver surgery.

B-mode imaging is based on the following assumptions: the speed of sound and the attenuation are constant in the imaged tissue, the axis of the ultrasound beam is straight, the ultrasound pulse travels only to objects that are on the beam axis and directly back to the transducer. If one of these assumptions is significantly disturbed visible artifacts are likely to be present in the image.

The geometry of the ultrasound image may be degraded by speed of sound errors. Usually it is assumed that the ultrasound beam propagates at an average speed of 1540 m/s but for fat the speed is 1450 m/s and for liver tissue 1550 m/s. In addition tissue is a non-homogeneous and imperfect transmission medium. If the assumed speed and the actual speed differ the computed

depth of a returning echo is incorrect leading to image distortions. A slightly incorrect image geometry is usually no problem for diagnostic ultrasound, but if ultrasound imaging is used as a basis of a navigation system this fact is relevant for the accuracy of the whole system [135]. Refraction is another cause for image distortions and loss of resolution. Starting with the skin surface at each tissue interface refraction occurs and the effects are cumulated such that echoes from tissue parts, which do not lie in the direction of the ultrasound beam, will appear in the image. This violates the assumption that a detected echo comes from the shortest and straight sound path between reflector and transducer. A similar problem arises from the fact that the ultrasound beam is not an ideal line but has a width depending on the depth as described in the above section about the lateral resolution. In addition the ultrasound beam has weaker side lobes such that highly reflective object parts, which do not lie in the main beam direction, can be present in the image.

Strongly reflecting interfaces (bone, calcifications) or attenuating structures (solid tissue, malignant masses) lead to shadowing. The ultrasound beam can not reach the tissue behind the interface or structure such that no echoes will be received from this region and the region will be represented dark in the image. Enhancement is the opposite effect to shadowing. Cysts or other liquid-filled structures are usually less attenuating as assumed for the normal tissue. This means tissue behind such structures is enhanced by the usual attenuation correction because the ultrasound beam has not been attenuated inside the cyst as expected. But shadowing and enhancements are also diagnostically useful. The physician may differentiate tissues or pathologies by the occurrence of shadowing or enhancement behind the structure in question. Detailed physical explanations of reflection and refraction artifacts are given by Robinson et al. [187].

Strong reflectors might cause a second artifact: reverberations or multiple reflections. The sound waves (at least a part of them) are bouncing back and forth between the tissue interface and the transducer. Bright bands of decreasing intensity will appear in the image. Reverberations can also occur between two strong parallel reflectors.

Now we come to artifacts of power Doppler ultrasound imaging. In general power Doppler ultrasound is less sensitive to the angle between the ultrasound beam and the vessel direction than color Doppler but is still direction dependent. If the blood flow direction is approximately the same as the beam direction there will be no Doppler signal.

As power Doppler ultrasound measures blood motions with regard to a static background additional motion of the tissue causes so called “flash artifacts”. The suppression of the vessel wall motion by high-pass filters is standard. Those filters suppress low frequency tissue motion but may also suppress low-velocity blood flow. Depending on the location of a tumor inside the liver other tissue motion due to heart motion or respiration can cause flash

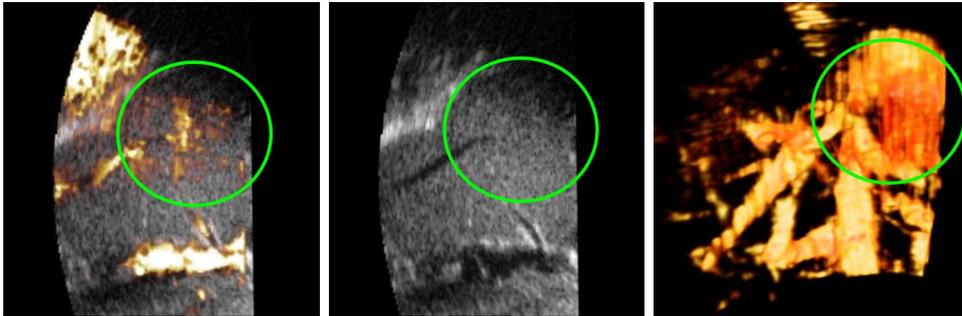


Figure 2.7 Power Doppler flash artifacts in the liver caused by heart motion. Left: B-mode image with overlaid power Doppler and severe flash artifacts in the upper left corner and spurious artifacts inside the green circle. Middle: B-mode image alone showing that there are no vessels inside the green circle. Right: 3D volume rendering of the power Doppler volume with a severe flash artifact inside the green circle.

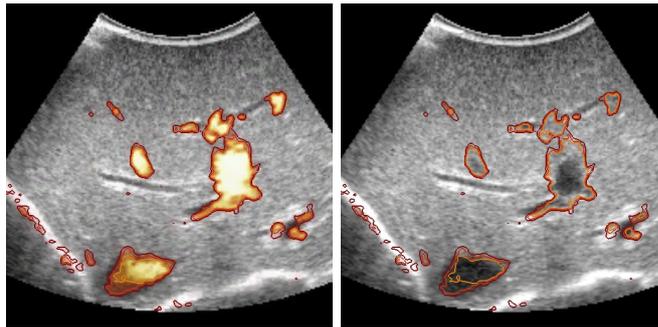


Figure 2.8 Blooming artifacts: the vessels appear brighter in the power Doppler image than they are. a) B-Mode with overlaid power Doppler signal and b) the same image with power Doppler isolines compared to the appearance of vessels in the corresponding B-mode image.

artifacts (see Fig. 2.7). A quite common power Doppler artifact is “blooming”; the Doppler signal spreads out from inside the vessel and appears outside the vessel such that the vessel appears broader than it is (color bleeding, see Fig.2.8 a, b). The reason may be multiple reflections back and forth between blood cells similar to reverberation artifacts in B-mode imaging. In particular if ultrasound contrast agents are applied very strong blooming effects occur soon after bolus injection and vanish after a while. Blooming can be reduced by reducing the gain at the ultrasound device, but lower signals in smaller vessels might vanish, too.

We observe a further power Doppler artifact which occurs with the motor-driven 3D ultrasound probe of the Voluson system of General Electric (GE)

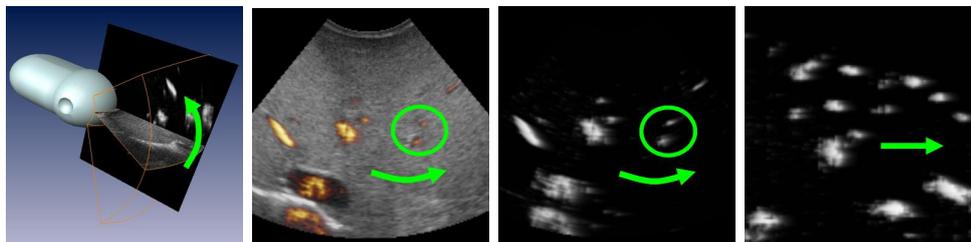


Figure 2.9 a) 3D power Doppler artifacts: the signal is smeared out in the direction of the transducer motion, b) power Doppler as color overlay on B-mode, c) power Doppler intensities alone d) power Doppler signal if the acquisition geometry is rectified. A clear direction dependence of the smoothing effect can be observed.

Healthcare and which to our knowledge has not been described in the literature so far. The principles of 3D ultrasound will be explained in the next section. In Fig. 2.9 the 3D power Doppler artifact is illustrated. A smoothing of the power Doppler signal in the direction of the transducer motion is shown. The reason might be, that not all signals have been returned to the transducer when the next image slice is acquired. A possibility to restore the true signal from the vessels is a deconvolution with an anisotropic point spread function in the direction of the transducer motion.

More detailed descriptions of artifacts in B-mode and power Doppler imaging can be found in the literature [83, 227, 197, 157, 159].

3D Ultrasound

In diagnostics usually conventional 2D ultrasound seems to satisfy the clinical requirements. But with the increasing availability of 3D ultrasound machines the number of publications increases, which investigate the potential of 3D and even 4D (time series of 3D) ultrasound. In particular in obstetrics the number of 3D ultrasound machines is dramatically increasing, but this seems to be at least partially for marketing reasons as parents like to see their babies. One obvious advantage of 3D ultrasound compared with 2D ultrasound is the possibility to better quantify lengths and volumes in 3D. In the context of computer assisted surgery 3D ultrasound is a basis for modeling anatomical structures and even their dynamical behavior in case of time series. In particular 3D ultrasound can easily be integrated into interventional and surgical procedures allowing intraoperative imaging as a prerequisite or extension of navigation systems.

Four different principle 3D ultrasound technologies exist:

1. tracked 2D probes, also called freehand 3D ultrasound,

2. 3D probes with mechanically (or electronically) steered 1D transducer lines (like in 2D probes), also called wobbler,
3. 3D ultrasound probes consisting of a 2D transducer array,
4. sensorless techniques based on decorrelation of speckle noise.

Sensorless tracking is done by analyzing the speckle in the US images using decorrelation [229] or linear regression [180]. Encouraging results are presented, but practical performance of in vivo imaging has to be further evaluated. In freehand 3D ultrasound [193, 181] a position sensor of a localizer system is clamped to a conventional 2D US-transducer and the transducer is manually swept over the volume of interest while the position and orientation of the imaging planes are recorded by the localizer system. After scanning, the 2D image planes are composed to a 3D image volume. In contrast to freehand 3D ultrasound mechanically steered 3D probes do not rely on tracking sensors. Instead a 2D probe with a 1D transducer array is swept by a motor contained in a specific 3D ultrasound probe. One example of such a system using a motor is the Voluson system developed by Kretztechnik and now distributed by General Electric (GE). A just recently commercially available alternative are volumetric 3D probes containing a 2D array of transducer elements, such that 3D volumes can be directly measured. This is a very promising technology and will be the future of 3D ultrasound, but until now the image quality is not as high as with swept 1D transducer arrays.

Generally all 3D ultrasound technologies are suitable as a basis for intraoperative navigation systems. We prefer mechanically steered 3D probes, because they are very easily manageable in the OR. In open liver surgery the 3D probe is held directly onto the liver for only a few seconds (see Fig. 2.10). The scan time depends on the image resolution and scan angles. These 3D probes allow for high quality volume acquisitions because a regular sampling of the volume is guaranteed in contrast to freehand 3D ultrasound. In addition the export of the volumetric ultrasound data is possible in original quality. On the contrary freehand 3D ultrasound systems usually use the low resolution video output to export the 2D scans. The abdominal 3D probe (RAB 4-8) of the Voluson 730 system contains a steered curvilinear 1D transducer array. The transducer array lies in lateral direction and the ultrasound beams are emitted in axial direction (see Fig. 2.10 for the nomenclature of the directions). A conventional 2D sector scan lying in the axial-lateral plane is acquired. By moving the scan plane in elevational direction a 3D volume is swept. The shape of this volume is a part of a torus, because the center of the sector in the scan plane and the center of the movement of the image plane differ [105]. Hence, the original image geometry is given in torus coordinates.

Usually the original data sampled on torus coordinates are reformatted to isotropic cuboid-shaped voxel allowing the use of standard image processing

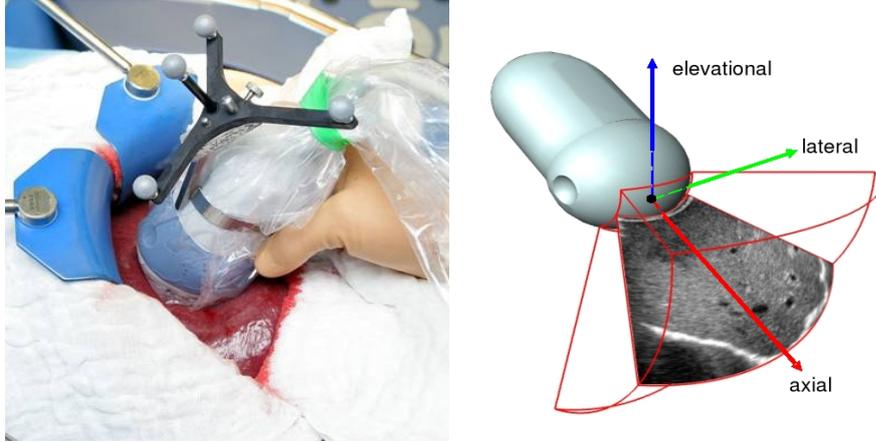


Figure 2.10 3D ultrasound probe. Left: Intraoperative application of a 3D ultrasound probe (RAB 4-8) from a Voluson 730 system directly onto the liver. The probe is covered by a sterile drape and equipped with an optical tracker to determine the spatial position and orientation of the probe. Right: Nomenclature of imaging directions. The axial and lateral directions define the usual 2D scan plane. The elevational direction is perpendicular to the scan plane and gives the sweep direction to acquire a 3D ultrasound volume.

and visualization techniques. But it is an interesting field of research to process the originally sampled data (see for example [152]) to avoid a loss of quality as sampling always smooths the data. For the resampling of the data one has to know how torus coordinates can be transformed to Cartesian coordinates and vice versa.

Torus Coordinates. We start with a parametric representation of a torus, which directly leads to the definition of torus coordinates. Let the outer radius from the center of the hole to the center line of the torus tube be R and the inner radius of the tube be r . For a standard ring torus the outer radius R is bigger than the inner radius r . The torus surface is generated by rotating a circle γ_M with radius r around the z axis. This circle γ_M is drawn in a (ρ, z) half plane M , which is perpendicular to the (x, y) plane (see Fig. 2.11 left). The position of this meridian plane M is parameterized by the angle ϕ between the x -axis and the plane. The circular center line of the torus ring (the outer circle) is parameterized by the radius $R \in \mathbb{R}^+$ and the angle $\phi \in [-\pi, \pi)$:

$$\begin{aligned} x &= R \cos(\phi), \\ y &= R \sin(\phi). \end{aligned} \tag{2.1}$$

The circle γ_M is parameterized inside the meridian plane M by the angle

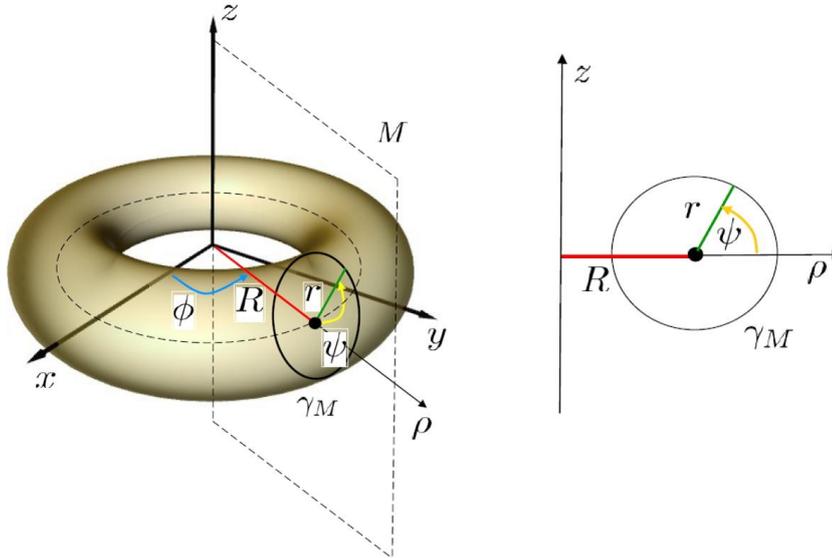


Figure 2.11 Parameterization of a torus. The torus ring is parameterized by the radius R (red) and the angle ϕ (blue). The circular cross section γ_M of the torus tube in the half plane M is parameterized by the radius r (green) and the angle ψ (yellow). The half plane M is spanned by the z - and ρ -axis, which is shown in detail on the right.

$\psi \in [-\pi, \pi)$ and the inner radius $r \in \mathbb{R}^+$ (see also Fig. 2.11 right):

$$\begin{aligned} \rho &= R + r \cos(\psi), \\ z &= r \sin(\psi). \end{aligned} \tag{2.2}$$

Now the two parameterizations of the outer (2.1) and inner circle (2.2) can be composed to the parameterization of the torus. A vector pointing onto the surface of the torus can be divided into a vector v_1 pointing onto the center line of the torus ring (see red line in Fig. 2.11) and a second vector v_2 from the center line to the surface (see green line in Fig. 2.11). The first vector v_1 has a length of R and the z coordinate is always 0. The x and y coordinates are simply given by the parameterization 2.1 of the outer circle. The second vector v_2 has a length of r . The z coordinate is given by the parameterization 2.2 of the inner circle and depends only on the angle ψ and not on ϕ . The x and y coordinates of v_2 in the (x, y) plane are $(r \cos(\phi), r \sin(\phi))$. In the z direction with increasing angle ψ these coordinates are decreasing with a factor of $\cos(\psi)$.

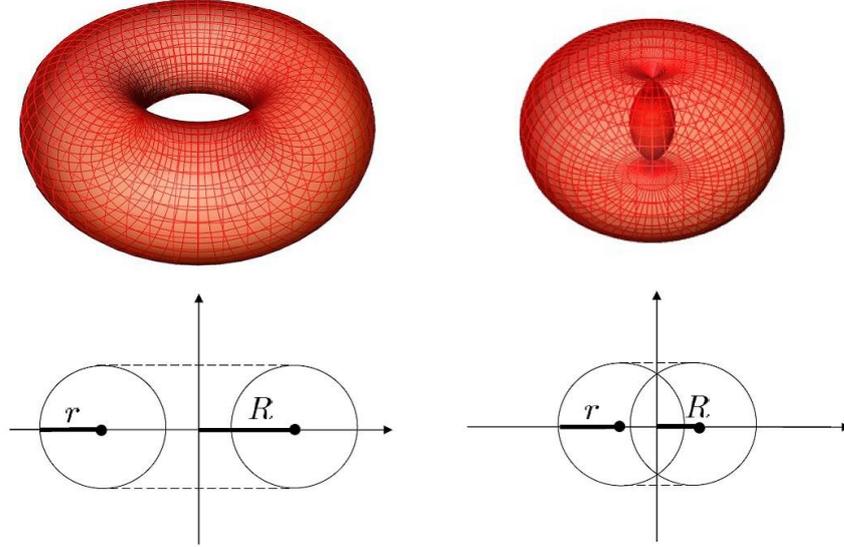


Figure 2.12 Different Kinds of Tori. On the left a common ring and on the right a self intersecting spindle torus is shown.

The resulting parametric representation of a torus by adding v_1 and v_2 is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{pmatrix} + r \begin{pmatrix} \cos(\psi) \cos(\phi) \\ \cos(\psi) \sin(\phi) \\ r \sin(\psi) \end{pmatrix} = \begin{pmatrix} (R + r \cos(\psi)) \cos(\phi) \\ (R + r \cos(\psi)) \sin(\phi) \\ r \sin(\psi) \end{pmatrix} \quad (2.3)$$

with the parameters $r \in \mathbb{R}^+$, $\phi \in [-\pi, \pi)$, $\psi \in [-\pi, \pi)$ and the constant outer radius $R \in \mathbb{R}^+$.

For a point represented in torus coordinates (r, ϕ, ψ) given by the ultrasound acquisition geometry the corresponding Cartesian coordinates (x, y, z) are determined by formula 2.3.

Ultrasound Acquisition Geometry in Torus Coordinates. The torus-shaped acquisition geometry results from the fact that the center of the scan plane sector and the rotation center of the sweeping motor are not identical. In addition the radius of the torus ring R is smaller than even the smallest radius coordinate r_{min} of the torus tube ($R < r_{min}$). This results in a self-intersecting spindle torus instead of a standard ring torus (see Fig. 2.12)

We now describe the location of the 3D ultrasound volume (see Fig. 2.10) in the torus coordinate system. The initial position of the ultrasound image plane (2D scan) is associated to the (x, y) plane of the torus (see Fig. 2.11). A point on a scan line is defined by the inner radius r of a torus and the direction of a scan line in the image plane is determined by the angle ϕ . The image plane is then swept by the angle ψ in elevational direction. The image

volume is delimited by minimal and maximal torus coordinates $r_{min}, r_{max}, \phi_{min}, \phi_{max}, \psi_{min}, \psi_{max}$.

The discrete 3D ultrasound image is then given on a torus grid inside this bounded volume. Let $m = (m_r, m_\phi, m_\psi) \in \mathbb{N}^3$ be the number of grid points in the coordinate directions r, ϕ and ψ . The grid is uniformly sampled in each coordinate direction leading to the following grid point distances in torus coordinates:

$$h_r = \frac{1}{m_r - 1}(r_{max} - r_{min}), h_\phi = \frac{1}{m_\phi - 1}(\phi_{max} - \phi_{min}), h_\psi = \frac{1}{m_\psi - 1}(\psi_{max} - \psi_{min}). \quad (2.4)$$

In contrast to Cartesian grids the size and volume of the grid cells is not uniform but is increasing with the radius r because the radial lines are running apart with increasing r . The volume of a grid element or the whole scan volume can be determined by the integral substitution rule for multiple variables. The determinant of the Jacobian matrix of torus coordinates is

$$\det J(r, \phi, \psi) = \left| \frac{\partial(x, y, z)}{\partial(r, \phi, \psi)} \right| = r(R + r \cos(\psi)) \quad (2.5)$$

giving the volume of a torus part

$$V^{\text{torus part}} = \int_{\psi_{min}}^{\psi_{max}} \int_{\phi_{min}}^{\phi_{max}} \int_{r_{min}}^{r_{max}} r(R + \cos(\psi)) dr d\phi d\psi. \quad (2.6)$$

To get the volume of a single grid cell only the limits of the definite integral have to be changed.

Conversion from Torus to Cartesian Grid. For reformatting an ultrasound volume originally given in torus coordinates onto a Cartesian grid, we will use the inverse coordinate transformation, because for each Cartesian grid point (x, y, z) its representation in torus coordinates (r, ϕ, ψ) has to be determined:

$$\begin{pmatrix} r \\ \phi \\ \psi \end{pmatrix} = \begin{pmatrix} \sqrt{(x^2 + y^2 - R)^2 + z^2} \\ \arctan\left(\frac{y}{x}\right) \\ \arctan\left(\frac{z}{\sqrt{x^2 + y^2 - R}}\right) \end{pmatrix}. \quad (2.7)$$

Three steps are necessary to interpolate the ultrasound intensity value on a Cartesian grid point (x_i, y_j, z_k) from the given ultrasound intensity values on the neighboured torus grid points:

1. Compute for each Cartesian grid point (x_i, y_j, z_k) the corresponding torus coordinates (r, ϕ, ψ) via equation 2.7.
2. Determine the cell of the torus grid which contains the point (r, ϕ, ψ) .
3. Interpolate (linearly) the ultrasound intensity value at the torus coordinate position (r, ϕ, ψ) from the intensity values on the eight torus grid points of the found torus grid cell.

2.4.2 Iterative 3D Ultrasound-Based Navigation

In section 2.3 we explained how computer assisted modeling and planning systems offer surgeons the possibility to perform detailed risk analysis and to define optimal individual resection plans for liver surgery. The challenge is to transfer the preoperative plan precisely and securely to the intraoperative situation. The location of the tumor and relevant vessels is hidden underneath the liver surface and the correct location of the resection line can only be estimated. Intraoperative navigation systems support the surgeon by visualizing the spatial relation of surgical instruments with respect to invisible anatomical structures [154]. Conventional navigation systems based only on preoperative data are not suitable for liver surgery, because the liver significantly deforms between preoperative imaging and the surgical procedure. Thus, dedicated navigation systems are based on either intraoperative liver surface information acquired by range scanners [27] or intraoperative 2D [15, 167, 176] or 3D ultrasound [10, 11].

Range scanners only capture the frontal part of the liver surface, yet significant features are rare on the surface. Furthermore, it is not clear how well deformations deep inside the liver can be detected. This impedes the use of range scanners for registration purposes. By contrast, 3D ultrasound directly recovers volumetric regions, including tumors and important vessels.

First navigation systems based on 2D [19] and freehand 3D ultrasound [75, 235, 228] have been developed for neurosurgery to compensate brain-shift. Commercial systems for 2D ultrasound (IGSonic from Brainlab and SonoNav from Medtronic) and freehand 3D ultrasound (SonoWand from SONOWAND AS, BrainNavigator from Localite) are available. The main difference of our novel system developed at the Charité [10, 11] to the SonoWand system is the used 3D ultrasound technology. We use a mechanically steered 3D probe instead of freehand 3D ultrasound. This 3D ultrasound technology and its advantages are explained in section 2.4.1. The general principle of a 3D ultrasound based navigation system is similar to a conventional neuronavigation system regarding instrument tracking and visualization. The difference is the use of intraoperative instead of preoperative image data and the kind of registration to relate image space and physical patient space. No explicit registration is necessary. A position sensor is attached to the ultrasound probe and the location of the probe during acquisition is measured. The location is represented by a rigid transformation T_{S2} , which describes the position and the orientation of the sensor with respect to the global coordinate system of the tracking camera (see Fig.2.13). If in addition the position and orientation of the ultrasound image coordinate system is known in relation to the location of the position sensor on the probe (T_{C2}), also the spatial relationship between image space and physical patient space is determined (see Fig. 2.13). The process of computing the transformation (T_{C2}) that converts the ultrasound image space into the coordinate system of the position sensor attached

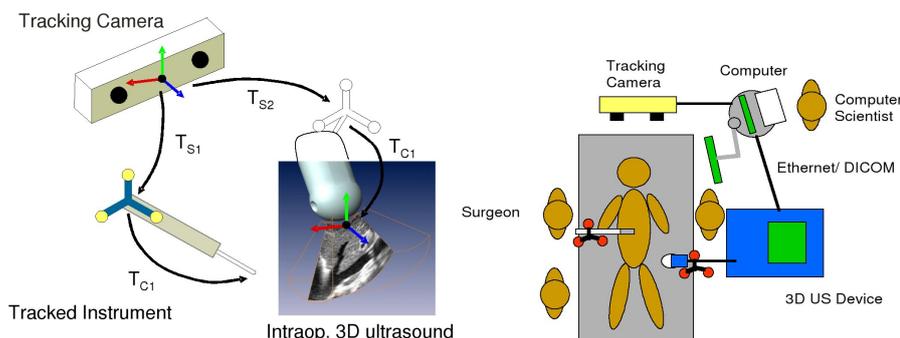


Figure 2.13 Principle of a navigation system based on intraoperative 3D ultrasound using a tracked 3D probe. On the left the different rigid transformations used in the system and on the right the principle assembly in the OR are outlined.

to the probe is called calibration. As calibration is a time-consuming process that has to be performed very precisely it is usually performed once and a suitable mounting ensures a reproducible attachment of the position sensor. A comprehensive overview of calibration techniques is given by Mercier et al. [145]. Lindseth et al. [135] found that improper probe calibration was the main contributor to the overall navigation inaccuracy of the SonoWand system.

The location of the surgical instrument is measured (T_{S1}) by a position sensor attached to it. Again a rigid calibration transformation (T_{C1}) is needed to determine the tip of the instrument (see Fig. 2.13). Now the current location of the instrument tip in the 3D ultrasound volume can be visualized to the surgeon (see Fig. 2.14). Display techniques may be conventional orthogonal slices oriented to the patient (axial, sagittal, coronal), from the surgeon's view, or only defined by the position and orientation of the surgical tool. In any plane slicing, only one slice defined by the position and orientation of the surgical tool is displayed from each 3D volume. In our system we usually use two slices (see Fig. 2.14). The first slice is approximately sagittally oriented and moves with the instrument tip from left to right. The second slice is approximately coronally oriented and usually manually aligned to the location of the tumor. It can also be tilted to adjust it to the direction of an important vessel.

It is important to keep in mind that an intraoperatively acquired ultrasound volume is only a snapshot of the anatomy at this point in time. If the liver tissue is significantly deformed during tissue resection another 3-D update (ultrasound acquisition) has to be acquired. Such types of navigation systems are also called iterative systems [134]. Until now we just visualize the preoperative 3D model separately to the intraoperative ultrasound slices. Oldhafer et al. [167] rigidly registered the preoperative CT data and intraoperative 2D



Figure 2.14 A photo of our 3D ultrasound-based navigation system in the OR (left) and the intraoperative visualization (right) are shown.

ultrasound data via vascular landmark features. No information about the resulting accuracy is given. Peterhans et al. [176] used external landmarks on the liver surface for a rigid registration. They report a median root mean square distance of 6.3 mm between corresponding landmarks. This is also called the fiducial registration error (FRE), which does not allow a statement about the more important target registration error (TRE). For a detailed discussion on accuracy aspects concerning the FRE and TRE see chapter 13. There might be cases where the tumor location is close to possible point landmarks for example in the left liver lobe, but according to our experiences the registration error after rigid registration based on external landmarks is higher than 1 cm in many situations. Even if vessel center lines are used for rigid registration significant deformations remain [115]. Due to the mentioned deformations of the liver a reliable non-rigid registration algorithm is needed to transfer the preoperative model and plan to the intraoperative 3D ultrasound data.

2.4.3 Continuous Navigation Based on Electromagnetic Tracking

The disadvantage of a navigation concept with an iterative intraoperative imaging modality as described in the previous section is the lack of information about organ movements and deformations in the time period between two intraoperative image acquisitions. The surgeon does not know when it is necessary to acquire a new image volume due to movements or deformations. The ultimate solution would be a continuous 3D imaging, which is theoretically possible with MRI, 3D C-Arms or 3D ultrasound, but only few attempts have been undertaken to use it in clinical navigation [31]. A compromise is to directly measure motion at some points (or even surface regions) on or inside the liver. These measurements can then also be used to indicate to the surgeon that significant deformations have occurred. The movements and deformations also might be compensated based on the measurements and a deformation model.

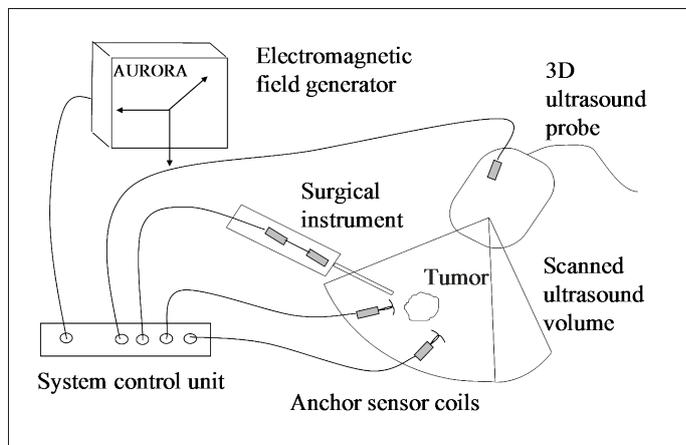


Figure 2.15 Principle of the electromagnetic navigation system. The gray rectangles are the sensor coils; their positions and orientations are measured by the electromagnetic tracking system Aurora (NDI inc.).

Possibilities for such a continuous monitoring are optical fiducial needles introduced by Meier-Hein et al. [139] for percutaneous radiofrequency ablation of liver lesions or needles with color-coded heads which can be tracked with an endoscopic video camera for laparoscopic interventions [9]. The disadvantage of such optical trackers is that they have to be placed outside the liver due to line-of-sight restrictions of optical systems.

A promising direction to measure motion and deformations directly *inside* an organ are electromagnetic tracking technologies. The standard tracking technology for intraoperative navigation is optical tracking. This is due to the high accuracy and reliability of such systems. The main disadvantage of optical tracking systems is again the free line-of-sight issue. In contrast to that electromagnetic systems also can measure locations inside the body, because the used electromagnetic fields permeate tissue. Electromagnetic systems consist of a control unit where a field generator emitting the electromagnetic fields and sensor coils are connected via cables. The principle is that electromagnetic fields with three different spatial directions are emitted which induce currents in the sensor coils. By measuring the currents the 3D position and orientation of the sensor coils can be determined.

The main limitations of electromagnetic systems are their smaller measurement volume, lower accuracy and most importantly their higher interference liability due to metallic instruments and other electromagnetic fields. Although compensation strategies and redundancy control have been developed this is still an active field of research. Further and more detailed information about electromagnetic tracking systems can be found in [14]. Recently sensor coils with ever smaller diameters below one mm have been developed enabling

their insertion directly into the liver [269]. We developed an electromagnetic navigation system for continuous movement and deformation detection [116]. A drawing of the developed electromagnetic navigation system is shown in Fig. 2.15. The basis of the navigation system is the electromagnetic tracking system Aurora from Northern Digital Inc.(NDI). The spatial position and orientation of 5 and 6 degrees of freedom (DOF) sensor coils are measured by the system. This means a 5 DOF sensor coil measures translations in three directions and rotations about two axes. The rotation around the sensor axis can not be determined, in contrast to a 6 DOF sensor coil. Like for the optical navigation system in the previous section a 3D ultrasound machine (GE Voluson 730 Expert) is again used for intraoperative imaging. The position and orientation of the ultrasound probe is measured by a 6 DOF sensor coil while the position and orientation of a surgical instrument are measured by two 5 DOF sensor coils. We developed anchor sensor coils, which are inserted and fixed in the liver tissue, can be sterilized and are biocompatible. A 5D sensor coil with a diameter of 0.8 mm and a length of 11.0 mm has been integrated into a shrinkable tubing together with a marker wire usually used for the preoperative marking of breast lesions (Fig.2.16 left). This marker wire has two wire tips which are anchored in the tissue (Fig.2.16 right). The anchor sensor coil is inserted by a cannula and can also easily be pulled back into the cannula to reposition or remove the anchor. At least two of the 5 DOF anchor sensor coils have to be inserted into the liver to measure a translation (three directions) and rotation (around three axes) of the tissue. We assume that in clinical practice not more than four sensor coils will be inserted into the liver. The anchor sensor coils are placed close to relevant anatomical structures like important vessels or a tumor.

Each 5 DOF anchor sensor coil provides the position of a point inside the liver and a direction at this point (rotation about two axes). When the intraoperative ultrasound image is acquired the reference positions p_1, \dots, p_n and directions d_1, \dots, d_n of the anchor sensor coils are also measured. After image acquisition during the actual intervention the positions and directions of the sensors are continuously measured. If the liver tissue in the surgical region of interest moves without deforming significantly and the sensors are close to this region the movement of the region can be measured by the sensors. For the compensation of the movement a rigid transformation has to be determined which transforms the ultrasound volume to the current position of the surgical region. This transformation can be computed by a landmark-based rigid registration of the positions and directions of the sensors during image acquisition onto their current positions and directions. We have published a method how to detect deformations from the sensor information [116]. Deformations might be compensated by non-rigid registration schemes on landmarks with directions, e.q., Rohr et al. [194] presented a thin-plate spline approach which considers directional information.



Figure 2.16 The size of a small anchor sensor coil compared to a one cent coin (left) and inserted into the liver during a resection (right).

2.5 Conclusions

In summary rigid and non-rigid registration is needed for different tasks in computer assisted liver surgery: in intraoperative navigation systems based on 3D ultrasound to transfer the preoperative plan onto the patient, also during the actual resection to track liver motions and detect deformations, and last but not least for the quantitative postoperative control to validate the navigation system and individual success of a liver resection.

Part I

Variational Image Registration

Chapter 3

The Registration Problem

3.1 Introduction

Although there are a large variety of registration methods, most of them are based on similar principles. In the majority of cases a functional depending on a) the intensities or features of the involved images and b) on the wanted transformation function is formulated, such that the minimum of the functional is achieved for a transformation, which maps each point in the one image to the corresponding anatomical point in the other image. The topic of this chapter will be: How the registration problem is formalized to an optimization problem, how the continuous problem can be discretized and numerically solved and how multiscale and multiresolution strategies can help to avoid local minima and accelerate the registration process. For the ease of notation and understanding everything is formulated for 3D images, but 2D and higher dimensions are also possible in most cases.

3.2 Problem Statement

We start with a formulation of the general registration problem of two images.

General Registration Problem. Let be given two images; a reference image \mathcal{R} and a template image \mathcal{T} . Find a *plausible* mapping (transformation) y which transforms \mathcal{T} onto \mathcal{R} , such that $\mathcal{T} \circ y$ gets more *similar* to \mathcal{R} . Optionally some *constraints* on y might have to be fulfilled.

Although digital medical images are discrete data, we first define the used registration approach in a continuous setting and than present suitable discretizations. We formalize \mathcal{R}, \mathcal{T} as compactly supported images on a usually cube-shaped domain $\Omega \subset \mathbb{R}^3$:

$$\mathcal{R}, \mathcal{T} : \Omega \rightarrow \mathbb{R} \tag{3.1}$$

and let \mathcal{T} be distorted by a transformation

$$y : \mathbb{R}^3 \rightarrow \mathbb{R}^3. \tag{3.2}$$

3. THE REGISTRATION PROBLEM

The main idea is to formalize registration as a variational problem, that means an optimization over a functional. The modeling of this functional has to answer the following questions:

- What does *plausible* transformation mean?
- What does *similar* mean?
- How can *constraints* be incorporated?

We first give some general and intelligible answers, before a detailed mathematically modeling is performed.

The first question concerns the comparison of images. Two images are said to be similar, if spatial positions of corresponding image contents are close to each other. To quantify this similarity we define distance (or similarity) measures \mathcal{D} which depend on the reference image \mathcal{R} and the transformed template image $\mathcal{T} \circ y$. Those measures are functionals which return a single value for the two image functions and a given transformation. This means the distance measure depends on the transformation: $\mathcal{D} = \mathcal{D}[y]$. The distance measures are either based on the original intensity data of the images or on processed data emphasizing or extracting special image features. Distance measures are presented in detail in Part II.

The second question concerns the transformation. The definition of a distance measure alone is not sufficient to ascertain a registration task, as the class of transformations is huge and some really “ugly” ones are among them, which might also give low distance measure values. A class of admissible transformations $y \in \mathcal{M}$ has to be defined, which includes only *plausible* transformations for a given application from a function space \mathcal{M} . For medical images those transformations are usually smooth and abrupt changes might only occur, if organs slip along other organs or surgical interventions have been performed between both images. It is often assumed that the tissue behaves approximately like an elastic material. The set of admissible transformations might be explicitly defined by parameterization or implicitly by a so-called regularizer \mathcal{S} or additional constraints of desired properties of the transformation. The second approach is called the nonparametric registration approach. Standard regularizers are defined in Chapter 6. Additional constraints on the transformation y can be introduced as a penalizer \mathcal{P} to the registration functional \mathcal{J} or as hard equality or inequality constraints on y . Constraints and penalizers are explained in more detail in Chapter 10.

Now we assemble the distance measure \mathcal{D} , the regularizer \mathcal{S} and an optional penalizer \mathcal{P} to a general framework for registration. Let $y_0 \in \mathcal{M}$ be a start or pre-registration transformation which is part of the current overall transformation y but should not be considered by the regularizer. The scalar factors $\alpha, \beta \in \mathbb{R}^+$ weight the distance measure \mathcal{D} , the regularizer \mathcal{S} and the optional

penalizer \mathcal{P} . Additional equality or inequality constraints on y are described by y belonging to some function subspace \mathcal{M} fulfilling the constraints. Now the joint registration functional \mathcal{J} is defined as:

$$\mathcal{J}[y] = \mathcal{D}[\mathcal{T}(y), \mathcal{R}] + \alpha\mathcal{S}[y - y_0] + \beta\mathcal{P}[y] \xrightarrow{y} \min \quad (3.3)$$

$$s.t. \quad y \in \mathcal{M} \quad (3.4)$$

Depending on the imaging properties and additional knowledge about the image contents particular choices for the distance measure, the regularizer, the penalizer and the function space are performed. Usually only a subset of these building blocks is used and specified for a given application. For example, in the standard nonparametric approach only the distance measure and the regularizer are defined:

$$\mathcal{J}[y] = \mathcal{D}[\mathcal{T}(y), \mathcal{R}] + \alpha\mathcal{S}[y - y_0] \xrightarrow{y} \min. \quad (3.5)$$

Further and other specifications of the building blocks of the registration functional will be presented in this work.

In some cases it may be advantageous to parameterize the transformation y . This means y can be represented as a linear combination of p basis functions $q_k(x) : \mathbb{R}^3 \rightarrow \mathbb{R}^3, k = 1, \dots, p$ and the linear coefficients $\gamma_k \in \mathbb{R}$ are the parameters which are optimized. If the basis functions are collected into a matrix $Q(x) = [q_1(x) \dots q_p(x)]$ we get the compact representation of the parameterized transformation y :

$$y(\gamma; x) = Q(x)\gamma \quad \text{with} \quad Q \in \mathbb{R}^{3 \times p}, \gamma \in \mathbb{R}^p. \quad (3.6)$$

Typical parametric transformations are rigid, affine-linear and spline transformations (see Sec. 7). Usually no regularizer is applied for parameterized transformations. A simple parametric registration functional looks like:

$$\mathcal{J}(\gamma) = \mathcal{D}[\mathcal{T}(y(\gamma)), \mathcal{R}] + \alpha\mathcal{S}(\gamma) \xrightarrow{\gamma} \min. \quad (3.7)$$

Now the optimization is performed with respect to the parameters γ and not directly with respect to y .

3.3 General Solutions

Only few registration problems like rigid and some non-rigid landmark-based schemes are analytically solvable. In most cases the continuous optimization problems have to be solved numerically. Different general numerical approaches exist [36]. We will highlight two approaches to show their principal differences and will then concentrate on one of the approaches, which is used in our work.

3. THE REGISTRATION PROBLEM

Important questions are when and how to discretize the problem to get a numerically solvable formulation. One possibility is to transfer the continuous optimization problem into partial differential equations (PDEs). These equations are also called Euler-Lagrange equations and their solution fulfills the necessary optimality condition of the continuous problem. Then the PDEs are discretized and the resulting linear systems solved [147]. This approach is also called Optimize-Discretize approach. Fast solution methods for the resulting highly structured linear systems have been developed for different regularizers [147].

In this work we focus on an alternative approach which is called Discretize-Optimize approach. First the continuous optimization problem is discretized to a finite-dimensional optimization problem. Then the whole framework of optimization algorithms for finite-dimensional problems can be applied. The advantage of this approach is that by discretization the registration problems can be transferred to quite general optimization problems, for which a lot of theoretical and practical knowhow exists.

In the following, we will define discrete images and how they can be transformed by discrete transformations, we will explain some basics and algorithms for general unconstrained and constrained finite-dimensional optimization problems. Then we will define some regularizers and how they are discretized and discuss a distance measure and its discretization exemplarily. Distance measures are discussed in more detail in chapter 8. We then have all ingredients to apply the general nonparametric registration framework developed by the registration group of Prof. Dr. Bernd Fischer at the University of Lübeck. More details and examples about the framework can be found in the PhD thesis of Dr. Nils Papenberg [170] and the book of Prof. Dr. Jan Modersitzki [150].

Chapter 4

Discrete Images and Transformations

4.1 Introduction

For the development of reliable and efficient registration algorithms based on numerical optimization schemes appropriate discretizations of the images and the transformation are essential. Although the original images coming from an imaging device are already discrete we will generate a continuous imaging model and discretize the images again. This is because we have to be able to transform the template image and to use resolutions which differ from the original resolution e.g., in a multiresolution framework or just to match different resolutions of reference and template image. The discrete images and discrete transformations are defined on regular grids where each grid point is assigned a scalar-valued intensity value (for images) or a vector valued displacement vector (for transformations). The regular grids are the basis of the finite differences schemes used for discretization of the registration problems. We start with the definition of regular cell-centered grids.

4.2 Cell-Centered Grids

4.2.1 Introduction

A grid is a partitioning of a part of the space (for example the image domain) with a set of grid cells. The grid cells are defined by a set of grid points, which are connected by a set of grid lines. Grids with different topology and geometry exist. An example of a curvilinear grid results from the torus imaging geometry of the 3D ultrasound probe (Sec. 2.4.1). Here we only discuss regular grids. The grid cells of regular grids are identical as opposed to the mentioned curvilinear grids where the size of the cells differs. In our case the grid cells are (anisotropic) cuboids and not cubes, because in tomographic images the slice thickness is often bigger than the resolution in the slice. The grid cells are called voxels (volume elements) in image processing. We identify each image voxel with a cell-centered grid point.

We will focus on cell-centered grids \mathbf{x}^{cc} because in addition to the discrete images the distance functionals and most of the regularizers will also be discretized on cell-centered grids. For the discretization of the elastic regularization functional it is important to use short differences for the discretization of the derivatives of the displacement vector field [78]. Short differences are able to capture high frequent oscillations and lead to numerically more stable discretization schemes. These short differences are defined on points in the middle between the cell-centered grid points. These intermediate points can be interpreted as an own type of grid. This grid type is called staggered grid (\mathbf{x}^{stg}) and will be defined in more detail in Chapter 6 when the elastic registration functional is discretized.

4.2.2 Notations and Terms

Some notations and terms are introduced which are helpful for the definition of three dimensional grids.

1. Let $A \in \mathbb{R}^{m \times n}$ be a matrix with m rows and n columns. Following the MATLAB notation $A_{k,:}$ denotes the k -th row and $A_{:,l}$ the l -th column of the matrix A .
2. Let $\mathbf{x} \in \mathbb{R}^n$ be a n -dimensional column vector. Again in MATLAB notation a subvector $\mathbf{x}_{m:l}$ represents the elements m to l of the vector \mathbf{x} (with $1 \leq m \leq l \leq n$).
3. Let $\mathbf{1}_n \in \mathbb{R}^n$ denote a column vector where all entries are one: $\mathbf{1}_n = (1, 1, \dots, 1)^T$.
4. For a matrix $A \in \mathbb{R}^{m \times n}$ and a matrix $B \in \mathbb{R}^{p \times r}$ their **Kronecker product** $C = A \otimes B \in \mathbb{R}^{mp \times nr}$ is defined as:

$$C = (a_{ij} \cdot B)_{i=1,\dots,m;j=1,\dots,n} = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}.$$

The Kronecker product is not commutative: $A \otimes B \neq B \otimes A$.

4.2.3 One-Dimensional Grids

At first we will define one-dimensional cell-centered grids which can be used to define three-dimensional cell-centered grids. If an interval $(\omega_1, \omega_2) \subset \mathbb{R}$ is divided into $m \in \mathbb{N}$ equally sized cells of size $h = (\omega_2 - \omega_1)/m \in \mathbb{R}$ the midpoints of the cells define a one-dimensional **cell-centered grid**. The coordinates of the grid points will be stored in vectors. The according vector for a cell-centered grid looks like

$$\mathbf{x}^{cc} \in \mathbb{R}^m \quad \text{with} \quad (\mathbf{x}^{cc})_k = \omega_1 + \left(k - \frac{1}{2}\right)h, \quad k = 1, \dots, m. \quad (4.1)$$

4.2.4 Three-Dimensional Grids

Let a rectangular cuboid $\Omega = (\omega_1, \omega_2) \times (\omega_3, \omega_4) \times (\omega_5, \omega_6)$ be the 3D image domain and $m = (m_1, m_2, m_3)$ the number of cells/voxel in each coordinate direction. Let the number of all grid cells be $n = m_1 \cdot m_2 \cdot m_3$ and the size of the cells in each coordinate direction be $h_l = \frac{1}{m_l}(\omega_{2l} - \omega_{2l-1})$, $l = 1, 2, 3$. For the cell (i, j, k) with $i = 1, \dots, m_1, j = 1, \dots, m_2, k = 1, \dots, m_3$ the midpoint $x_{i,j,k} \in \mathbb{R}^3$ is:

$$x_{i,j,k} = \left(\omega_1 + \left(i - \frac{1}{2}\right)h_1, \omega_3 + \left(j - \frac{1}{2}\right)h_2, \omega_5 + \left(k - \frac{1}{2}\right)h_3 \right)^T. \quad (4.2)$$

If we traverse all cells (i, j, k) in lexicographical ordering: firstly increase the index i in x_1 direction, secondly the index j in x_2 direction and lastly index k in x_3 direction, then we can represent the grid as a matrix $\bar{\mathbf{x}}^{cc} \in \mathbb{R}^{m_1 m_2 m_3 \times 3}$ containing all cell midpoint coordinates. One row of the matrix looks like:

$$(\bar{\mathbf{x}}^{cc})_{i+m_1 j+m_1 m_2 k,:} = x_{i,j,k}^T. \quad (4.3)$$

For the optimization framework we need a representation of the grid as one long vector. Therefore we write the matrix $\bar{\mathbf{x}}^{cc}$ column-wise in a vector $\mathbf{x}^{cc} \in \mathbb{R}^{3m_1 m_2 m_3}$ by means of the vec-operator:

$$\mathbf{x}^{cc} = \text{vec}(\bar{\mathbf{x}}^{cc}) = \begin{pmatrix} \bar{x}_{:,1}^{cc} \\ \bar{x}_{:,2}^{cc} \\ \bar{x}_{:,3}^{cc} \end{pmatrix}. \quad (4.4)$$

The long vector notation can be transferred back to the matrix (coordinate vector) representation by:

$$\bar{\mathbf{x}}^{cc} = (\mathbf{x}_{1:n}^{cc} \ \mathbf{x}_{(n+1):2n}^{cc} \ \mathbf{x}_{(2n+1):3n}^{cc}) \quad (4.5)$$

with $n = m_1 m_2 m_3$.

We illustrate this notation by a small example. Let a 2D grid with dimensions $m = (3, 2)$ be defined on the domain $\Omega = (0, 3) \times (5, 7)$. In matrix representation the cell-centered grid is given by:

$$\bar{\mathbf{x}}^{cc} = ((0.5 \ 5.5); (1.5 \ 5.5); (2.5 \ 5.5); (0.5 \ 6.5); (1.5 \ 6.5); (2.5 \ 6.5)) \quad (4.6)$$

and the corresponding long vector notation looks like:

$$\mathbf{x}^{cc} = (0.5 \ 1.5 \ 2.5 \ 0.5 \ 1.5 \ 2.5 \ 5.5 \ 5.5 \ 5.5 \ 6.5 \ 6.5 \ 6.5)^T. \quad (4.7)$$

Three-dimensional grids can be constructed from one-dimensional grids using the Kronecker product defined in the notation section 4.2.2. Let $\mathbf{x}_1^{cc} \in \mathbb{R}^{m_1}$ be a 1D cell-centered grid in x_1 -direction, $\mathbf{x}_2^{cc} \in \mathbb{R}^{m_2}$ in x_2 -direction and

$\mathbf{x}_3^{cc} \in \mathbb{R}^{m_3}$ in x_3 -direction. Then a 3D cell-centered grid $\mathbf{x}^{cc} \in \mathbb{R}^{3m_1m_2m_3}$ represented by one long vector can be constructed from these 1D grids by

$$\mathbf{x}^{cc} = \begin{pmatrix} \mathbf{1}_{m_3} & \otimes & \mathbf{1}_{m_2} & \otimes & \mathbf{x}_1^{cc} \\ \mathbf{1}_{m_3} & \otimes & \mathbf{x}_2^{cc} & \otimes & \mathbf{1}_{m_1} \\ \mathbf{x}_3^{cc} & \otimes & \mathbf{1}_{m_2} & \otimes & \mathbf{1}_{m_1} \end{pmatrix}. \quad (4.8)$$

We explain this definition by the above 2D example. Let the one-dimensional grids be $\mathbf{x}_1^{cc} = (0.5 \ 1.5 \ 2.5)^T$ and $\mathbf{x}_2^{cc} = (5.5 \ 6.5)^T$. Then the two-dimensional grid is constructed by:

$$\mathbf{x}^{cc} = \begin{pmatrix} \mathbf{1}_{m_2} \otimes \mathbf{x}_1^{cc} \\ \mathbf{x}_2^{cc} \otimes \mathbf{1}_{m_1} \end{pmatrix} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0.5 \\ 1.5 \\ 2.5 \end{pmatrix}; \begin{pmatrix} 5.5 \\ 6.5 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \quad (4.9)$$

$$= (0.5 \ 1.5 \ 2.5 \ 0.5 \ 1.5 \ 2.5 \ 5.5 \ 5.5 \ 5.5 \ 6.5 \ 6.5 \ 6.5) \quad (4.10)$$

Be aware of the effect if the $\mathbf{1}_n$ vector is Kronecker multiplied before or after a 1D coordinate vector. In the first case the coordinate vector is repeated as a whole while in the second case the single coordinates are repeated.

We define the notation $\mathbf{x}^{cc}(\Omega, m)$ for a cell-centered grid with dimensions m on the domain Ω in vector notation and $\bar{\mathbf{x}}^{cc}(\Omega, m)$ in matrix (coordinate vectors) notation.

4.3 Continuous Images and their Multiscale Representation

In the introduction of this chapter we already mentioned that we need to model continuous images from the original discrete images acquired by imaging devices. A *continuous* image \mathcal{T} is just a function which maps each point in 3D space (resp. in the image domain Ω) to a scalar intensity value: $\mathcal{T} : \mathbb{R}^3 \rightarrow \mathbb{R}$. A *discrete* image $T \in \mathbb{R}^n$ is a vector which contains intensity values of the continuous image model at the grid points of a cell-centered grid $\mathbf{x} = \mathbf{x}^{cc}(\Omega, m)$:

$$T := \mathcal{T}(\mathbf{x}) = \mathcal{T}(\bar{\mathbf{x}}) = [\mathcal{T}(\bar{\mathbf{x}}_{j,:})]_{j=1}^n. \quad (4.11)$$

The dimension $n = m_1m_2m_3$ of the vector matches the number of grid points and the intensity values are sorted in the same ordering as the grid points.

The resolution of the cell-centered grid can theoretically be chosen arbitrarily such that the resolution of an original discrete image T_{orig} can be changed by generating a continuous image model \mathcal{T} based on the original image and sample the continuous image on a grid with the new resolution. This makes it possible to match the resolution of reference and template image. In addition multi-resolution approaches are possible as will be discussed later.

In addition the intensity values of the transformed template image can be computed. Therefore the grid of the reference image is transformed by a discrete transformation $\mathbf{y} \in \mathbb{R}^{3n}$ and the intensity values of the continuous template image \mathcal{T} at the transformed grid positions are evaluated. This means for the reference and the template image the same number of grid points are evaluated in the discrete distance measure as will be shown later. Before we formalize the transformation of the template image we determine a discrete transformation from the continuous transformation $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. The discrete transformation $\bar{\mathbf{y}} \in \mathbb{R}^{n \times 3}$ defined on the grid \mathbf{x} (see above) contains the transformed position of each grid point by row:

$$\bar{\mathbf{y}} = y(\bar{\mathbf{x}}) = [y(\bar{\mathbf{x}}_{j,:})^T]_{j=1}^n. \quad (4.12)$$

Like for the grids the $(n \times 3)$ matrix $\bar{\mathbf{x}}$ is transferred into a long vector $\mathbf{y} \in \mathbb{R}^{3n}$ to fit to the general optimization framework by stacking the matrix column-wise by the vec-operator:

$$\mathbf{y} = \text{vec}(\bar{\mathbf{y}}) = \begin{pmatrix} \bar{\mathbf{y}}_{:,1} \\ \bar{\mathbf{y}}_{:,2} \\ \bar{\mathbf{y}}_{:,3} \end{pmatrix}. \quad (4.13)$$

Now we can express the transformed template image with the discrete transformation \mathbf{y} . The continuous model of the template image \mathcal{T} is evaluated at the transformed grid positions:

$$T_{\mathbf{y}} := \mathcal{T}(\mathbf{y}) = \begin{pmatrix} \mathcal{T}(\mathbf{y}_1, \mathbf{y}_{1+n}, \mathbf{y}_{1+2n}) \\ \mathcal{T}(\mathbf{y}_2, \mathbf{y}_{2+n}, \mathbf{y}_{2+2n}) \\ \vdots \\ \mathcal{T}(\mathbf{y}_n, \mathbf{y}_{2n}, \mathbf{y}_{3n}) \end{pmatrix} \quad (4.14)$$

with the abbreviation $T_{\mathbf{y}} \in \mathbb{R}^n$ for the transformed discrete template image.

Because the registration functional has to be optimized with respect to the transformation \mathbf{y} the derivative (Jacobian) of the transformed template image is important to enable efficient optimization algorithms. Since the j th component of $T_{\mathbf{y}}$ depends only on $\mathbf{y}_j, \mathbf{y}_{j+n}, \mathbf{y}_{j+2n}$ the $(n \times 3n)$ Jacobian matrix of $T_{\mathbf{y}}$ is a block matrix with diagonal blocks:

$$dT_{\mathbf{y}} = \left(\frac{\partial T_j(\mathbf{y})}{\partial \mathbf{y}_k} \right)_{j=1, \dots, n, k=1, \dots, 3n} \quad (4.15)$$

$$= (d_{\mathbf{y}_{1:n}} T_{\mathbf{y}}, d_{\mathbf{y}_{n+1:2n}} T_{\mathbf{y}}, d_{\mathbf{y}_{2n+1:3n}} T_{\mathbf{y}}) \quad (4.16)$$

$$= (\text{diag}(\partial_1 \mathcal{T}(\mathbf{y})), \text{diag}(\partial_2 \mathcal{T}(\mathbf{y})), \text{diag}(\partial_3 \mathcal{T}(\mathbf{y}))). \quad (4.17)$$

Until now we can discretize continuous images and transformations onto arbitrary regular grids, determine a discrete transformed template image from the

continuous template image and a transformation and compute the derivative of the transformed template image with respect to the transformation.

The open question is how we can generate a good continuous image model \mathcal{T} from a given original discrete image T^{orig} ? Two ingredients are important for a good image model: an interpolation (resp. approximation) method and a scale-space approach. The interpolation method enables the computation of image intensities at arbitrary points that do not lie on the grid of a discrete image. In a scale-space an image is considered on different scales by representing the image as a one-parameter (the scale) family of smoothed images. The higher the scale parameter the more details in the image are smoothed away. A scale-space or multi-scale approach is important to find an adequate scale for an aimed resolution defined by the grid onto which the image is sampled. The adequate scale prevents undersampling and resulting aliasing artifacts. In image registration different scales and resolutions are used to make the optimization process more robust and faster. We will discuss multi-scale and multi-resolution approaches in more detail.

In summary for the computation of a continuous image we first represent an image on an adequate scale and then interpolate image intensities based on this scale. We will start with the presentation of interpolation methods for 3D images. Let the original discrete image T^{orig} be defined on the original grid $\mathbf{x}^{\text{orig}} = \mathbf{x}^{cc}(\Omega, m^{\text{orig}})$ with the original number of voxels $n^{\text{orig}} = m_1^{\text{orig}} \cdot m_2^{\text{orig}} \cdot m_3^{\text{orig}}$ and voxel sizes h^{orig} . To build a continuous image model \mathcal{T} , we usually look for a continuous interpolation function $I : \mathbb{R}^3 \rightarrow \mathbb{R}$ with:

$$I(\bar{\mathbf{x}}_{j,:}) = T_j^{\text{orig}} \quad \text{for } j = 1, \dots, n. \quad (4.18)$$

This means the image model agrees at the grid points with the intensity values of the original discrete image.

Many different interpolation functions for 3D images can be found in literature [143]. Linear interpolation is a easy to implement and efficient scheme but the resulting interpolation function is not differentiable at the grid points, which is an important prerequisite for many efficient optimization schemes. Interpolation with cubic splines (piecewise polynomial functions) is a good compromise between smoothness and efficiency. Splines possess many outstanding theoretical and practical properties. For a nice overview see the work of Michael Unser [233]. Due to the noise in medical images an exact fit by interpolating splines may not be desirable. In addition splines tend to produce overshoots between grid points.

Therefore we relax the interpolation constraint to an approximation constraint:

$$I(\bar{\mathbf{x}}_{j,:}) \approx T_j^{\text{orig}} \quad \text{for } j = 1, \dots, n. \quad (4.19)$$

Smoothness conditions on the approximating function are added to get a well-posed problem. Interestingly there are natural smoothness conditions, which

lead again to splines as solutions of the approximation problem. These splines are also called smoothing splines and will be presented in the next section.

4.3.1 One-dimensional Splines

Splines are smooth piecewise polynomial functions. The paper of Schoenberg [207] from 1946 is probably the first mathematical reference where the word spline is used for such functions. The term spline stems from ship building. A spline is a thin wooden strip which is fixed at some points and is bending like a cubic spline with natural boundary conditions. The spline attempts to minimize the inner strain by the bending.

The bending energy is approximated by integration over the square of the second derivate of the interpolation function $I : \mathbb{R} \rightarrow \mathbb{R}$:

$$\mathcal{S}^{\text{spline}}[I] = \int_{\Omega} (I''(x))^2 dx. \quad (4.20)$$

Given some data $T^{\text{orig}} \in \mathbb{R}^n$ on a one-dimensional grid $\mathbf{x} \in \mathbb{R}^n$ the solution of the optimization problem constrained by the interpolation conditions

$$\mathcal{S}^{\text{spline}}[I] \xrightarrow{I} \min \quad \text{subject to} \quad I(\mathbf{x}_j) = T_j^{\text{orig}} \quad \text{for } j = 1, \dots, n \quad (4.21)$$

is a cubic spline (see for example [147]) and can be represented by a linear combination of some basis functions:

$$I^{\text{spline}}(x) = \sum_{j=1}^n c_j b^j(x). \quad (4.22)$$

In order to get an easy indexing of the basis functions we linearly map the domain $\Omega = (\omega_1, \omega_2)$ onto the interval $(0.5, n + 0.5)$ where $x' \rightarrow x = (x' - \omega_1)/h + 0.5$. Thus the grid cell centers $x_j = \omega_1 + (j - 0.5)h$ are mapped onto j . Now the basis functions $b^j(x)$ are determined by shifting ($b^j(x) = b(x - j)$) only one single basis function:

$$b(x) = \begin{cases} (x + 2)^3, & -2 \leq x < -1, \\ -x^3 - 2(x + 1)^3 + 6(x + 1), & -1 \leq x < 0, \\ x^3 + 2(x - 1)^3 - 6(x - 1), & 0 \leq x < 1, \\ (2 - x)^3, & 1 \leq x < 2, \\ 0, & \text{else.} \end{cases} \quad (4.23)$$

The coefficients $c = [c_1; \dots; c_n]$ can easily be computed by solving the linear system $\sum_{j=1}^n c_j b^j(x_k) = T_j^{\text{orig}}$ or with the tridiagonal matrix $B = [b^j(x_k)] \in \mathbb{R}^{n \times n}$ in matrix notation: $Bc = T^{\text{orig}}$.

The coefficients are computed once and then for each point x in the domain the spline interpolation function can be efficiently evaluated. At first the grid

cell, which contains the point x , is determined. It is assumed that the domain has been linearly mapped onto $(0.5, n + 0.5)$ (see above). The point x is split into an integer part $p = \lfloor x \rfloor$ and a remainder $\xi = x - p$ leading to: $x = p + \xi$. At most four basis functions are nonzero for x because $b(x) = 0$ for $x \notin (-2, 2)$. Thus the evaluation of the spline interpolation function reduces to:

$$I^{\text{spline}}(x) = c_{p-1}b(\xi + 1) + c_p b(\xi) + c_{p+1}b(\xi - 1) + c_{p+2}b(\xi - 2). \quad (4.24)$$

Approximative Smoothing Splines

As we mentioned already in the previous section relaxing the interpolation conditions to an approximative scheme enables the consideration of noise and prevents overshoots. Therefore a data-fitting term is defined which measures how well the continuous model function $I : \mathbb{R} \rightarrow \mathbb{R}$ fits the given data $T^{\text{orig}} \in \mathbb{R}^n$ on the grid points $\mathbf{x} \in \mathbb{R}^n$:

$$\mathcal{D}[I] = \left\| I(\mathbf{x}_j) - T_j^{\text{orig}} \right\|^2. \quad (4.25)$$

Now a model function is needed, which is smooth on the one hand and fits the data as well as possible on the other hand:

$$\mathcal{D}[I] + \theta \mathcal{S}^{\text{spline}}[I] \xrightarrow{I} \min \quad (4.26)$$

where $\mathcal{S}^{\text{spline}}$ is again an approximation to the bending energy of I . A smooth function I has a low value of $\mathcal{S}^{\text{spline}}$. With the weighting factor $\theta > 0$ the smoothing is weighted against the data fitting term. For big values of θ the resulting model function will be smooth but the approximation of the data points will be low. For small values of θ the approximation of the data will be better but the model function will be less smooth. This is illustrated in Fig. 4.1 compared to the interpolating cubic spline. The overshooting (oscillations of polynomials) of the interpolating spline between data points is recognizable. For the smoothing splines the overshooting is decreased with increasing θ . For $\theta \rightarrow \infty$ the smoothing splines degrade to a straight line the smoothest possible model function. Interestingly the solution of the approximation problem is again a cubic spline which is again parameterizable like above. Those approximating splines are also called smoothing splines. The solution can be found by rewriting the approximation problem such that it depends on the coefficients $c \in \mathbb{R}^n$. Then the bending energy $\mathcal{S}^{\text{spline}}$ can be written as:

$$\mathcal{S}^{\text{spline}}(I) = \int_{\Omega} (I''(x))^2 dx = \int_{\Omega} \left(\sum_{j=1}^n c_j b_j''(x) \right)^2 dx \quad (4.27)$$

$$= \sum_{i=1}^n \sum_{j=1}^n c_i c_j \int_{\Omega} b_i'' b_j''(x) dx = c^T M c = \|c\|_M \quad (4.28)$$

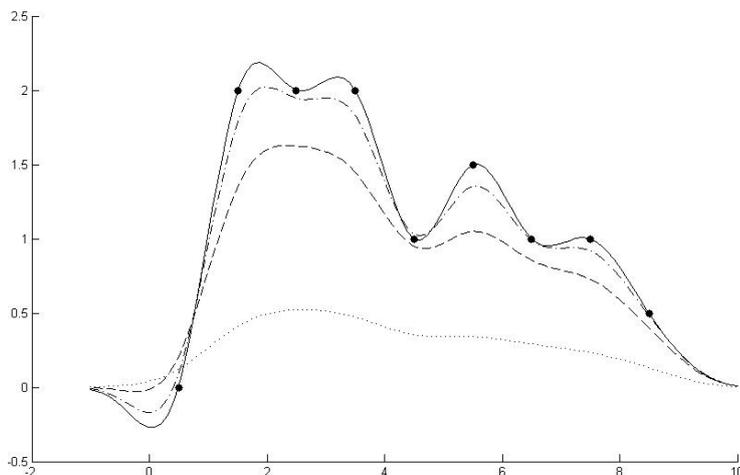


Figure 4.1 Smoothing splines approximating the black dots with smoothing weights $\theta = 2$ (dashed-dotted line), $\theta = 10$ (dashed line) and $\theta = 100$ (dotted line). The interpolating cubic spline is plotted as a solid line.

with $I(x) = \sum_{j=1}^n c_j b_j(x)$ and $M_{ij} = \int_{\Omega} b_i'' b_j''(x) dx$.

Now the approximation problem 4.26 can be written depending on the spline coefficients using the model matrix B defined for the interpolating splines:

$$\|Bc - T^{\text{orig}}\|^2 + \theta c^T M c \xrightarrow{c} \min. \quad (4.29)$$

The solution of this minimization problem can be found via the following normal equations:

$$(B^T B + \theta M)c = B^T T^{\text{orig}}. \quad (4.30)$$

We see that the matrix M acts as a regularizer compared to the pure interpolation problem which we get for $\theta = 0$.

4.3.2 Three-dimensional Splines

One-dimensional splines can be extended quite easily to multivariate splines. Here we only consider the case of three-dimensional splines. Three-dimensional spline basis functions $b_{ijk} : \mathbb{R}^3 \rightarrow \mathbb{R}$ can be written as a product of one-dimensional spline basis functions $b_i, b_j, b_k : \mathbb{R} \rightarrow \mathbb{R}$:

$$b_{ijk}(x) = b_i(x_1) \cdot b_j(x_2) \cdot b_k(x_3) \quad (4.31)$$

with $x = (x_1, x_2, x_3)$. For each coordinate direction a one-dimensional basis function is applied.

Now the three-dimensional spline function $I^{\text{spline}} : \mathbb{R}^3 \rightarrow \mathbb{R}$ can be written as a linear combination:

$$I^{\text{spline}}(x) = \sum_{k=1}^{m_1} \sum_{j=1}^{m_2} \sum_{i=1}^{m_3} c_{ijk} b_{ijk} \quad (4.32)$$

where m_1, m_2, m_3 are the number of grid points in the respective coordinate directions.

The $n = m_1 \cdot m_2 \cdot m_3$ coefficients $c \in \mathbb{R}^n$ for an interpolating three-dimensional spline are again computed via a linear system of equations

$$Bc = T^{\text{orig}}. \quad (4.33)$$

The system matrix $B \in \mathbb{R}^{n \times n}$ is constructed from the system matrices $B_1 \in \mathbb{R}^{m_1 \times m_1}, B_2 \in \mathbb{R}^{m_2 \times m_2}, B_3 \in \mathbb{R}^{m_3 \times m_3}$ of the one-dimensional splines. We explain the basic idea for the 2D case. First the coefficients for the interpolation in x_1 -direction are determined by the system matrix B_1 for each of the m_2 grid lines. This can be expressed by the Kronecker product: $(I_{m_2} \otimes B_1)$. Then the resulting coefficients are interpolated in x_2 -direction for each of the m_1 grid lines, which again can be expressed by the Kronecker product: $(B_2 \otimes I_{m_1})$. At the end we get for the 2D case:

$$(B_2 \otimes I_{m_1})(I_{m_2} \otimes B_1)c = T^{\text{orig}}. \quad (4.34)$$

The matrix product of the two Kronecker products can be simplified using $(A \otimes B)(C \otimes D) = AC \otimes BD$ leading to the 2D system matrix

$$B^{2D} = B_2 \otimes B_1. \quad (4.35)$$

The derivation of the system matrix B^{3D} for the 3D case is straight forward and leads to: $B^{3D} = (B_3 \otimes I_{m_2} \otimes I_{m_1})(I_{m_3} \otimes B_2 \otimes I_{m_1})(I_{m_3} \otimes I_{m_2} \otimes B_1)$ (see also in Papenberg [170]). Again the term can be simplified (see Modersitzki [150]) and we get:

$$B^{3D} = B_3 \otimes B_2 \otimes B_1. \quad (4.36)$$

The determination of the coefficients for *approximating* 3D splines is also carried out by means of Kronecker products and can be found in Papenberg [170] and Modersitzki [150].

Derivatives of 3D Splines

The partial derivatives of a 3D spline function can be computed quite easily because only one of the three one-dimensional splines depends on the given coordinate direction:

$$\partial_1 I^{\text{spline}}(x) = \sum_{k=1}^{m_1} \sum_{j=1}^{m_2} \sum_{i=1}^{m_3} c_{ijk} b'_1(x_1) b_2(x_2) b_3(x_3). \quad (4.37)$$

The derivative $b'_1(x_1)$ can be computed from the definition of the one-dimensional spline basis function (see Eq. 4.23). The derivatives with respect to the coordinate directions x_2 and x_3 are computed in the same fashion.

4.3.3 Multiscale Representation of Images

Lindeberg expressed the importance of the scale concept in the introduction of a review paper [132]:

We perceive objects in the world as meaningful entities only over certain ranges of scale. A simple example is the concept of a branch of a tree, which makes sense only at a scale from, say, a few centimeters to at most a few meters. It is meaningless to discuss the tree concept at the nanometer or the kilometer level. At those scales it is more relevant to talk about the molecules that form the leaves of the tree, or the forest in which the tree grows. [...]

This fact, that objects in the world appear in different ways depending on the scale of observation, has important implications if one aims at describing them. It shows that the scale concept and the notion of multi-scale representation are of crucial importance.

A nice example for the need of multiscale representations are maps. In cartography maps are produced at different levels of abstraction. A map of the whole world only contains countries, big islands and some of the major cities. In a regional map towns, small villages, streets and small rivers are shown. In a city guide even buildings might be mapped. On Google maps the scale slider is an important feature to adjust the needed level of detail.

In image processing multiscale representations of images are a fundamental concept to extract information from the images on a reasonable level of detail or to find an appropriate model of the image as a basis of image processing algorithms for segmentation or registration. The theory of multiscale representations is called scale-space theory [255, 102, 133]. The term scale-space has been introduced by Witkin [255]. A multiscale representation of an image is a family of continuous models $\mathcal{L}(x, t)$ of the image with the continuous spatial scale parameter $t \in \mathbb{R}_0^+$. The bigger the parameter t the smoother the image and the less details are recognizable. For $t = 0$ one gets the original image $\mathcal{L}(x, 0) = \mathcal{T}^{\text{orig}}$.

The most prominent scale-space is the linear (diffusion) scale-space [255]. The linear scale-space is generated by convolution of an image with Gaussian kernels of increasing width (variation σ^2). The resulting scale-space family is equivalent to the solution of the linear diffusion equation. We will use the Gaussian scale-space to model vessel trees and define a new distance measure in Chapter 9.

The intuitive properties a scale-space should possess have been formalized into different scale-space axiom systems [132]. Koenderink [102] introduced the concept of causality, which means that new level sets must not appear with increasing scale parameter. This formalizes the crucial requirement that structures at coarse scales should be simplifications of corresponding structures at finer scales. No new structures should accidentally appear on coarser scales by the smoothing operation. If causality is combined with isotropy and homogeneity Koenderink [102] showed that such a scale-space representation must satisfy the diffusion equation and thus leads uniquely to a Gaussian scale-space. Isotropy means that all spatial positions and homogeneity that all scale levels must be treated in a similar manner. Other properties (axioms) like decreasing number of local extrema, semi-group structure and scale invariance are described by Lindeberg [132]. Combinations of these alternative axioms lead also to a Gaussian scale-space.

An alternative scale-space to the Gaussian scale-space can be constructed from approximating splines as defined in the previous section. In this case the weighting factor θ is the scale parameter and the continuous image is modeled by a smoothing spline: $\mathcal{L}(x, t) = \mathcal{T}^{spline}(x, \theta)$. For a detailed discussion of a scale-space derived from splines see also the paper of Wang and Lee [246]. Splines are good approximations of the Gaussian kernel. They converge to the Gaussian function if the order of the spline tends to infinity. This follows from the central limit theorem [234]. Thus many good properties of the Gaussian kernel are inherited. In particular the important causality property is shared by the spline scale-space at least in a discrete sense [246]. Even advantages of the spline scale-space over the Gaussian scale-space like a higher computational efficiency are listed by Wang and Lee [246].

In the next section a multiscale representation based on approximative splines in comparison to a multiresolution representation of a 2D slice of a liver CT is shown in Fig. 4.2. A combination of a multiscale and multiresolution representation will be used for our nonparametric image registration approach as described in Chapter 6.

4.4 Multiresolution Images and Transformations

In the previous section we showed how images can be represented on different *scales* enabling a focus on coarse and important features if necessary. Starting with reference and template images on coarser scales and switching to ever finer scales leads to more robust registration processes because the number of local minima is reduced at the beginning and on finer scales the current transformation is already close to the 'correct' local minimum.

However, in a discrete multiscale representation of an image the *resolution* of the underlying grid is not changed. If in addition multiresolutions are also used the registration process can be accelerated significantly. On a coarse scale

also a coarse resolution is sufficient. The idea is to start on a coarse resolution where each optimization iteration is efficiently computable. Then on finer levels only few iterations have to be performed, because only corrections to the solutions on the coarser resolutions have to be found. For a visual impression of multiscale and multiresolution images see Fig. 4.2.

A multilevel registration algorithm based on multiscale and multiresolution representations of the images and the transformation will be described in section 6.5. In principle the algorithm uses a small number of finite levels where on each level a different discretization of the registration functional is optimized. The resulting transformation of a coarser level serves as an initial transformation on the next finer level. For the implementation of such a strategy two questions have to be answered:

1. how to *restrict* the fine original *images* to coarser images, and
2. how to *prolongate* a coarse *transformation* to a finer one?

An image pyramid with images of different resolutions is computed before the actual optimization iterations start. The prolongation of the transformation is performed at the end of each level. The images and the transformation are based on the same grid resolution on each level. We will only consider resolution changes by the factor 2. If h is the resolution on a fine level then the resolution on the next coarser level is $H = 2h$. We formalize the restriction by an operator P_H^h which acts on a fine image: $T^H = P_H^h T^h$ and the prolongation by an operator P_h^H which acts on a coarse transformation: $\mathbf{y}^h = P_h^H \mathbf{y}^H$.

4.4.1 Restriction of Fine to Coarse Image

The restriction of a fine to a coarse 3D image just means averaging 8 neighboured voxel to one coarse voxel. We will explain and define the averaging operator P_H^h first for 1D and 2D images and show how the 3D and 2D operators are constructed from the 1D operator. For a compact description the MATLAB notation $T_{1:k:n}$ is used to extract certain elements from the image vector T . The notation means starting from index 1 take every k -th element until n is reached: $1, 1+k, 1+2k, \dots, n-k, n$, if n can be divided by k without rest. Now the averaging of two neighboured voxels of a 1D image $T^h \in \mathbb{R}^m$ to a coarser image $T^H \in \mathbb{R}^{m/2}$ can be described with this notation as:

$$T^H = \frac{1}{2}(T_{1:2:m-1}^h + T_{2:2:m}^h) \quad (4.38)$$

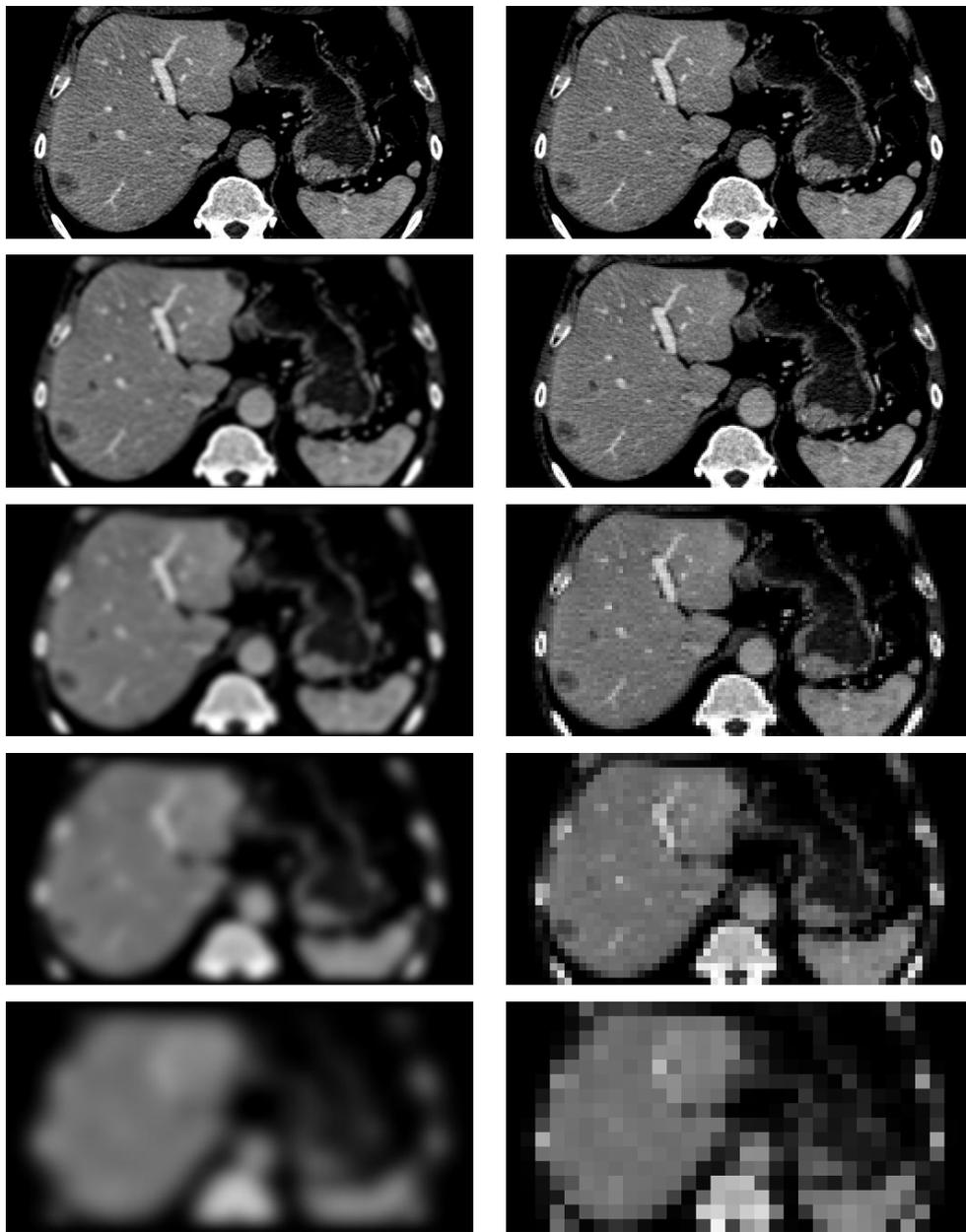


Figure 4.2 In the first column one 2D slice of the liver on different *scales* (multiscale) with weighting factors $\theta = 0, 10, 100, 1000, 10000$, in the second column the same slice with different *resolutions* (multiresolution) on the levels $l = 8, 7, 6, 5, 4$ with $2 * 2^l * 2^l$ voxel is shown. The first row contains in both columns the original image for $\theta = 0$ and the original resolution $512 * 256$.

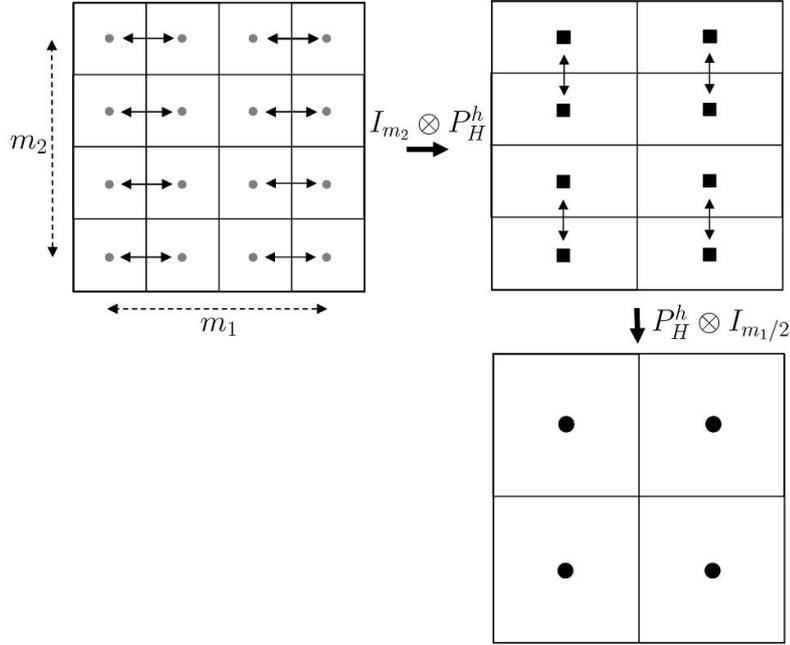


Figure 4.3 Restriction of a fine image $T^h \in \mathbb{R}^{m_1 m_2}$ (upper left) to a coarser image $T^H \in \mathbb{R}^{m_1 m_2/4}$ (lower right). At first voxel value pairs are averaged in x_1 -direction (horizontal arrows) and then the resulting values (upper right) are averaged in x_2 -direction (vertical arrows).

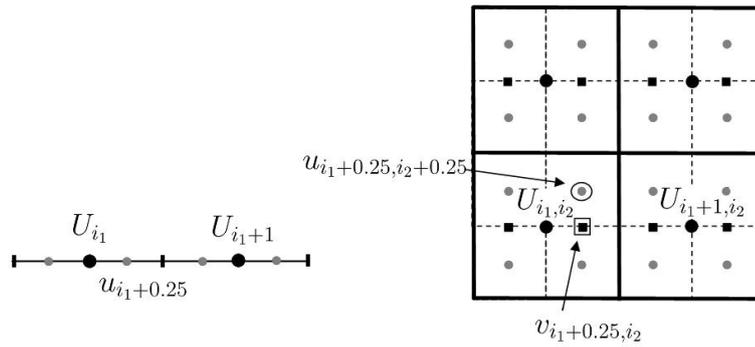


Figure 4.4 A transformation U given on a 1D (left) or a 2D (right) cell-centered grid is linearly prolonged to a finer transformation u . The big black dots are the points of the coarse grid and the gray small dots of the finer grid. In the 2D case intermediate values v are determined on points illustrated as black squares.

Chapter 5

Optimization

The key for fast registration are efficient optimization algorithms. Here, we will present some basic facts and the optimization algorithms which have been used for this thesis. We will discuss the advantages of the used algorithms and will distinguish differences to other well known approaches. The optimization algorithms are used as a tool, but not as a research topic of their own. The focus is on the choice of suitable existing optimization techniques for image registration problems. The optimization algorithms are generally formulated. An important work of the previous and following chapters is to reformulate the registration problems such that they fit to the generally formulated optimization techniques. The optimization algorithms and notations are primarily taken from the standard book on Numerical Optimization of Nocedal and Wright [163]. This is also a good book for more details and further reading. All optimization problems will be formulated as *minimization* problems.

5.1 Basic Definitions

Two types of optimization problems have to be solved in the registration framework: unconstrained and constrained problems.

Unconstrained Optimization

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a smooth objective function. We are looking for a vector $x_* \in \mathbb{R}^n$ is searched, which solves the following minimization problem:

$$f(x) \xrightarrow{x} \min. \quad (5.1)$$

For image registration it is of particular importance to differentiate between local and global minimizers. Generally we are interested in the global minimizer. A point x_* is a global minimizer if $f(x_*) \leq f(x)$ for all $x \in \mathbb{R}^n$ (or at least in the domain $\Omega \subset \mathbb{R}^n$ of interest). Such a global minimizer is usually difficult to find, because we do not have a good picture of the overall shape

of f . Since optimization algorithms only sample f , there may be global minimizers in regions, which have not been sampled. Most algorithms are able to find only *local* minimizers. A point x_* is a *local minimizer* if there exists a neighborhood \mathcal{N} of x_* such that $f(x_*) \leq f(x)$ for all $x \in \mathcal{N}$. A point x_* is a *strict local minimizer* if there exists a neighborhood \mathcal{N} of x_* such that $f(x_*) < f(x)$ for all $x \in \mathcal{N}$ with $x \neq x_*$.

In general the optimization algorithms presented here are only guaranteed to find one of the local minimizer. However when f is convex or the starting point of the algorithm is close enough to the global minimizer the local minimizer found by the algorithm is also a global minimizer. In image registration optimization algorithms often get trapped in local minima and do not reach the global minimum because the registration functional is not smooth and convex or the starting transformation is too far from the globally optimal transformation. Therefore it will be important to use multiscale approaches and regularizers to make the registration functional convex or at least more convex.

Constrained Optimization

The optimization problem might be extended by constraints on the variables. These constraints might be equalities or inequalities in the variables. Although in the most general case both kinds of constraints might occur in the same optimization problem, we here define two separate problems, because only those are needed for the registration problems in this thesis.

Let $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be the constraints function and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ the objective function. The equalities constrained problem is defined as:

$$f(x) \xrightarrow{x} \min \quad \text{s.t.} \quad c(x) = 0 \quad (5.2)$$

and the inequalities constrained problem as:

$$f(x) \xrightarrow{x} \min \quad \text{s.t.} \quad c(x) \geq 0. \quad (5.3)$$

For optimization problems which are constrained by equality as well as inequality constraints we introduce the index set $\mathcal{I} \subseteq 1, \dots, m$ for the inequality constraints $c_i(x) \geq 0, i \in \mathcal{I}$ and the index set $\mathcal{E} = 1, \dots, m \setminus \mathcal{I}$ for the equality constraints $c_i(x) = 0, i \in \mathcal{E}$.

Important for constrained optimization are the terms: feasible and active set (or region). The *feasible set* (or region) F is the set of points which satisfies the constraints: $F = \{x | c(x) = 0\}$ or $F = \{x | c(x) \geq 0\}$. The *active set* $A(x)$ at any feasible point x are the indices of the inequality constraints $c_i(x) \geq 0, i \in \mathcal{I}$ for which equality holds: $A(x) = \{i \in \mathcal{I} | c_i(x) = 0\}$. For equality constraints $c_i(x) = 0, i \in \mathcal{E}$ the active set is the whole index set $A(x) = \mathcal{E}$ because x is feasible by definition. If equality and inequality constraints exist the active set is the union of both active sets.

5.2 Characterizing Solutions

It is helpful to know the analytical characteristics of a smooth objective function f at a local minimum because some algorithms depend on this knowledge.

Unconstrained Optimization

Necessary and sufficient conditions for local minimizers of smooth functions are known from basic analysis. Their proof is based on Taylor's theorem, which will also be central for some of the optimization algorithms.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuously differentiable function and $x, p \in \mathbb{R}^n$. Then f can be approximated at a point $x + p$ by the function value and the gradient at the point x by:

$$f(x + p) = f(x) + \nabla f(x)^T p + \mathcal{O}(\|p\|^2). \quad (5.4)$$

The approximation by a second order model is even better. Therefore let f be a twice continuously differentiable function. Then f can be approximated by the function value, the gradient and the Hessian at the point x :

$$f(x + p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p + \mathcal{O}(\|p\|^3). \quad (5.5)$$

A (first-order) necessary condition for a local minimizer is: If f is continuously differentiable and x_* is a local minimizer of f , then the gradient vanishes at x_* : $\nabla f(x_*) = 0$. This means the only candidates for local minimizers are the stationary points with $\nabla f(x_*) = 0$. To guarantee that x_* is a local minimizer the following (second-order) sufficient condition can be used. Let f be a twice continuously differentiable function. If $\nabla f(x_*) = 0$ and $\nabla^2 f(x_*)$ is positive definite, then x_* is a *strict* local minimizer of f .

Constrained Optimization

For constrained optimization the gradient of the constraints plays a central role. If we consider an optimization problem with equality constraints and follow the contour line (level set) of the constraints function for $c(x) = 0$, then the value of the objective function f will usually vary along the constraint contour line. This means the contour line of c usually intersects the contour lines of f . Only if the contour line $c(x) = 0$ meets the contour lines of f *tangentially*, the value of f does not change any longer locally. Since the gradient of a function is perpendicular to the contour lines, saying that the contour lines meet tangentially is equivalent to saying that the gradients are parallel. This means that at a local minimum x_* there are scalars λ such that $\nabla f(x_*) = \lambda \nabla c(x_*)$ holds.

To formalize the necessary conditions for local minimizers the Lagrangian function is used. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the objective function and $c : \mathbb{R}^n \rightarrow$

\mathbb{R}^m the equality constraints. The Lagrangian function $L : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ with Lagrange multipliers $\lambda \in \mathbb{R}^m$ is defined as:

$$L(x, \lambda) = f(x) - \lambda^T c(x). \quad (5.6)$$

The gradient of the Lagrangian function is $\nabla_x L(x, \lambda) = \nabla f(x) - \lambda^T \nabla c(x)$, representing the parallelism condition of the gradients of objective function and constraints if set to 0. Thus at a local minimum x_* , there are scalars λ_* , such that $\nabla_x L(x_*, \lambda_*) = 0$.

By means of the Lagrangian function we can formulate the *Karush-Kuhn-Tucker (KKT)* first order necessary conditions for constrained optimization problems. At first the KKT conditions are presented for *equality* constraints. Let f and c be continuously differentiable. If x_* is a local minimizer then there exist constants $\lambda_* \in \mathbb{R}^m$ such that

$$\nabla_x L(x_*, \lambda_*) = 0, \quad (5.7)$$

$$c(x_*) = 0. \quad (5.8)$$

We explained the first (parallelism) condition already above. Since the gradients might also be parallel at points not located on the zero contour line of the constraints, the second condition ensures that the constraints hold (feasibility condition).

For constrained optimization problems with *inequality* constraints the KKT conditions look like:

$$\nabla_x L(x_*, \lambda_*) = 0, \quad (5.9)$$

$$c(x_*) \geq 0, \quad (5.10)$$

$$\lambda_* \geq 0, \quad (5.11)$$

$$\lambda_*^i c_i(x_*) = 0, \text{ for } i = 1, \dots, m. \quad (5.12)$$

Again the parallelism of the gradients and the feasibility is required (Eq. 5.9 and Eq. 5.10) like for equality constraints. In addition the Lagrange multipliers have to be positive to ensure that the singular point lies on the correct side of the constraints or in other words the gradients point in the same direction (Eq. 5.11). The last condition (Eq. 5.12) implies that a Lagrange multiplier is 0, if the corresponding inequality is *inactive* ($c_i(x_*) > 0$). In this case the inactive inequalities theoretically can be omitted. This equation is also called complementary (slackness) condition. For the complex proof and a more detailed description see Nocedal and Wright [163]. For mixed constraints both sets of KKT conditions are combined.

For a minimum x_* to satisfy the KKT conditions, some regularity conditions (constraint qualification) have to hold to exclude degenerated cases. For example the linear independence constraint qualification (LICQ) ensures

that the gradient of the active inequalities (resp. the gradient of the equality constraints) are linearly independent at x_* .

Sufficient conditions for a local minimum of a constrained optimization problem consist of a) the KKT conditions and b) the second order derivative of the Lagrangian function $\nabla_{xx}L$ has to be positive definite on a certain set of directions (see Nocedal and Wright [163] for details).

5.3 Algorithms for Unconstrained Optimization

All presented optimization algorithms are iterative, i.e., starting from a point $x_0 \in \mathbb{R}^n$ a series of points $x_0, x_1, \dots, x_k \in \mathbb{R}^n$ is generated which converge to a local minimizer x_* . The values of the objective function are monotonically decreasing ($f(x_k) \geq f(x_{k+1})$). We will call the point x_k in iteration step k the ‘iterate’ of this step.

There are two fundamental strategies to find the next iterate x_{k+1} from the last one x_k : line search and trust-region strategies. In the line search strategy, in each iteration step k a search direction $p_k \in \mathbb{R}^n$ and a step length $\alpha \in \mathbb{R}^+$ is chosen, such that the new iterate is determined by

$$x_{k+1} = x_k + \alpha p_k. \quad (5.13)$$

Different methods exist to determine the search direction. If and how they use derivatives of f and the last search directions makes them different. We will present the important Newton resp. Gauß-Newton search direction in Sec. 5.3.1. The objective of the step length computation is to solve the one-dimensional minimization problem along the given search direction p_k :

$$f(x_k + \alpha p_k) \xrightarrow{\alpha} \min. \quad (5.14)$$

Practically this minimization is solved only approximately (inexact line search) because each evaluation of f (or ∇f) is computationally expensive, in particular, in image registration of large 3D images. Therefore only few candidate values for α are tried out to reduce the computational costs. How the candidates can be chosen to get an adequate reduction in f is shown in Sec. 5.3.2.

In the trust-region optimization strategy, at the current point x_k the objective function f is approximated by a model function m_k and an optimizer of this model function is computed. Usually, the model function behaves like f only near x_k . Thus the search for a step p_k , such that $x_k + p_k$ minimizes m_k is restricted to some region around x_k . Often the region is a sphere ($\|p\| \leq \Delta \in \mathbb{R}^+$) but may also be elliptical or box-shaped. Δ is also called the trust region radius. For a spherical region, the following optimization problem on m_k is solved in each iteration step instead of the original problem:

$$m_k(x_k + p) \xrightarrow{p} \min, \text{ where } \|p\| \leq \Delta. \quad (5.15)$$

If the determined solution does not decrease the value of f sufficiently, the model function is a bad approximation for the whole trust region of the given size. In this case the trust region is diminished and the optimization problem on the model m_k is computed again.

The model function m_k is usually quadratic. A natural choice is the Taylor approximation $m_k(x_k+p) = f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2 f_k p$, but often approximations to the Hessian $\nabla^2 f_k$ are used instead of the full Hessian to save computation time.

The line search and trust-region strategies differ in the order they use to compute the direction and the distance of the next iteration point is computed. Line search first identifies a search direction and then determines the distance in this direction. In the trust-region approach first a maximal distance (the trust-region radius) is computed and then a direction and step size inside the trust region.

As we will use only line search strategies for the optimization of the registration functional, we will explain line search optimization in more detail in the following sections. But trust-region optimization has also been used for registration purposes, e.g., by Henn [85].

5.3.1 Search Directions for Line Search Optimization

Steepest (or gradient) descent optimization uses the most natural search direction for a line search algorithm: $p_k = -\nabla f(x_k)$. Steepest descent is quite popular in image registration because it is easy to implement and requires only gradient computation and no second order derivatives. In gradient direction, the variation of the objective function is strongest. But this greedy strategy is often not efficient because of the high number of iterations resulting from zig-zag in the search directions. Nocedal and Wright [163] show that the steepest descent line search algorithm converges to a local optimum, but the convergence rate even for exact step length choices may be unacceptable low.

Newton Search Direction

The aim of Newton search direction is to use more information (second order derivatives) about the objective function to find search directions leading to more efficient optimization schemes (better convergence rate).

Like for the trust-region approach we consider the second-order Taylor series approximation of the objective function:

$$f(x_k + p) \approx m_k(x_k, p) = f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T \nabla^2 f(x_k) p. \quad (5.16)$$

If we assume that $\nabla^2 f_k$ is positive definite, we can find the vector p that minimizes $m_k(p)$ by setting the derivative of $m_k(p)$ with respect to p to 0.

The Newton direction follows from $D_p m_k(p) = \nabla f(x_k) + \nabla^2 f(x_k)p = 0$ as a minimum direction of the model $m_k(p)$:

$$p_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k). \quad (5.17)$$

The search direction is computed by solving the symmetric linear system: $\nabla^2 f(x_k)p_k = -\nabla f(x_k)$.

If $\nabla^2 f_k$ is positive definite, the Newton direction is a descent direction. A line search optimization algorithm based on the Newton search direction has a fast local quadratic convergence rate. The disadvantage of the Newton direction is that the Hessian $\nabla^2 f_k$ may not always be positive definite, or is close to being singular. In this case the Newton direction may not be defined or is not necessarily a descent direction. In nonparametric image registration, the Hessian contains information from the noisy image data and the distance measures are not convex for all possible image pairs. Thus the Hessian of the registration functional can not be assumed to be positive definite even with a regularizer. In addition, the computation of the full Hessian is usually expensive in particular for large 3D image data.

There are two general strategies to ensure optimization steps of good quality and to reduce the computational effort. In both strategies an approximative Newton direction is used. The first strategy is to use the exact full Hessian, but to solve the linear system only approximately. This is called the *inexact* Newton approach. If an iterative linear system solver is applied, the iterations can be terminated before the exact solution is reached. This is also called the *truncated* Newton method [156, 44]. The iterations can also be stopped if negative curvature is encountered (Newton-CG method [163]). The second general strategy for the computation of an approximative Newton direction is to approximate or modify the Hessian matrix. For a simple modification of the Hessian we add a positive diagonal matrix or a full matrix to the Hessian to get a sufficiently positive definite matrix (modified Newton method [163]). To avoid the computation of the full Hessian an approximation of the Hessian can be used. Quasi-Newton methods [163] are typical members of this class of optimization techniques. They are based only on gradient information from the current and previous iterations to approximate the *inverse* of the Hessian. Another approximation strategy for the Hessian is to exploit a special structure of the objective function and hence of the Hessian. Least squares problems are an example for such special objective functions which lead to the Gauß-Newton search direction presented in the next section.

Gauß-Newton Search Direction

The structure of non-linear least squares problems with the objective function $f(x) = \|r(x)\|^2$, $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ allows an effective approximation of the Hessian. Modersitzki [150] and Papenberg [170] extended the formulation of the least

squares objective function such that it fits to the general form of the distance measures defined in sec. 6.2. He considers objective functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form:

$$f(x) = \phi(r(x)) \quad \text{with } \phi : \mathbb{R}^m \rightarrow \mathbb{R}, \quad r : \mathbb{R}^n \rightarrow \mathbb{R}^m. \quad (5.18)$$

The function r is called the residual function and ϕ the outer function. The least squares function is a special case of this more general formulation where the outer function is the Euclidean norm.

The gradient and the Hessian of the special objective function f are

$$\nabla f(x) = d\phi(r(x)) \cdot dr(x), \quad (5.19)$$

$$\nabla^2 f(x) = dr(x)^T d^2\phi(r(x)) \cdot dr(x) + d\phi(r(x)) \cdot d^2r(x). \quad (5.20)$$

If ϕ is a simple and smooth function whose second derivative $d^2\phi$ is easy to compute we only need to know the first derivative dr of the residual function to compute the first part of the Hessian $\nabla^2 f(x)$. The first part is often more important than the second summation term if the residuals are small or the model is close to linearity near the solution, i.e., d^2r is small. In image registration, the residuals also contain the noise of the images, such that the second derivative of the residuals does not provide much information. Therefore we get a good approximation of the Hessian by neglecting the second summation term:

$$\nabla^2 f(x) \approx H^{GN}(x) = dr(x)^T d^2\phi(r(x)) \cdot dr(x). \quad (5.21)$$

Instead of solving the standard Newton equations $\nabla^2 p = -\nabla f$ now the Gauß-Newton equations with the Hessian approximation $H^{GN} p = -\nabla f$ are solved to get the Gauß-Newton search direction:

$$p_k = -H_{GN}^{-1}(x_k) \nabla f(x_k). \quad (5.22)$$

5.3.2 Step Length Determination for Line Search Optimization

If we have determined a search direction, we need to choose a step length which actually reduces the objective function value. An ideal choice would be the global minimizer along the search direction which means to minimize the univariate function $\psi(\alpha) = f(x_k + \alpha p_k)$ with respect to the step length $\alpha > 0$. Usually an inexact line search is performed to determine a step length that provides a good reduction in f at low computational costs. Different candidates on the search direction (different α) are checked for a sufficient decrease of the objective function value. For each candidate the objective function value and often the gradient value have to be computed which is computationally expensive for large optimization problems like 3D image registration. Often, the simple condition on α_k to provide a reduction in f : $f(x_k + \alpha_k p_k) < f(x_k)$

does not lead to a sufficient decrease. The Armijo condition is a standard criterion to define a sufficient decrease:

$$f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^T p_k, \quad c \in (0, 1). \quad (5.23)$$

The standard choice of the tolerance is $c = 10^{-4}$. We can interpret the Armijo condition to mean the reduction to be proportional to the step length α_k and the directional derivative $\nabla f_k^T p_k$ in the search direction p_k .

Unfortunately the Armijo condition is always satisfied for sufficiently small step lengths α . To avoid an unacceptable small step length the Wolfe or Goldstein conditions (see [163]) or a backtracking line search approach is chosen together with the sufficient decrease Armijo condition. The backtracking algorithm works as follows: Start with an initial step length α . The natural choice for a Newton (or Gauß-Newton) scheme is $\alpha = 1$. Now the Armijo condition is tested for the α . If the reduction is sufficient (the condition is satisfied), we are done and let $\alpha_k = \alpha$. If not, the step length α is iteratively shortened by a factor $\rho \in (0, 1)$: $\alpha = \rho\alpha$ until the Armijo condition is satisfied. We use a shortening factor $\rho = \frac{1}{2}$. As a safeguard, the line search (and the whole optimization iteration) is terminated, if the step length gets too small.

5.3.3 Gauß-Newton Scheme

Now we have all ingredients for a line search optimization algorithm based on the Gauß-Newton search direction and the Armijo backtracking step length computation. If an implementation of the objective function f , its gradient ∇f and the approximation to the Hessian H is provided the Gauß-Newton algorithm works as follows:

Algorithm 5.1. (Gauß-Newton Algorithm)

1. Set $k = 0$, let $x_k = x_0$ be the starting value
2. Compute $f(x_k), \nabla f(x_k), H(x_k)$
3. Determine search direction p_k by solving $H(x_k)p_k = -\nabla f(x_k)$
4. Find the step length α_k by Armijo line-search on x_k in direction p_k
5. Break if line search fails
6. Update current value to: $x_{k+1} = x_k + \alpha_k p_k$
7. Set $k = k + 1$
8. If stopping rules (see below) are not satisfied continue with step 2

We use the following established stopping criteria introduced by Gill, Murray and Wright [71]:

1. $\|f(x_{k-1}) - f(x_k)\| < \tau(1 + \|f(x_k)\|)$, (relative change in objective function value)
2. $\|x_{k-1} - x_k\| < \sqrt{\tau}(1 + \|x_k\|)$, (relative change in parameters)
3. $\|\nabla f(x_k)\| < \sqrt[3]{\tau}(1 + \|f(x_k)\|)$, (length of the gradient)
4. $\|\nabla f(x_k)\| < \epsilon$, (gradient below machine precision)
5. $k > k_{max}$, (maximal number of iterations)

The factor $\tau \in \mathbb{R}^+$ is defined by the user. If the objective function value at the minimum is close to zero ($f(x_*) \approx 0$), the criterion $\|f(x_k) - f(x_{k+1})\| \leq \tau \|f(x_k)\|$ can not be satisfied (see [71]). Therefore an additional summand τ is introduced to the first criterion. A similar argument holds for the second and third criterion. The parameter $\epsilon \in \mathbb{R}^+$ represents the machine precision and $k_{max} \in \mathbb{N}$ the maximal number of iterations.

The iterations are stopped if all of the first three criteria or if one of the two safeguard criteria (4 or 5) are met. For more details see [71].

5.4 Algorithms for Constrained Optimization

There are two main categories of algorithms for constrained optimization. Either the original problem is replaced by a sequence of unconstrained problems or a sequence of simpler (quadratic) subproblems. One possible choice for the first category is to transfer the constraints to a quadratic penalty term which is added to the objective function. The quadratic term penalizes violations of the constraints. This approach is called *quadratic penalty* method [163], p.490. The iterates may violate the constraints (are not feasible) *during* the optimization process, but with a severe weighting of the penalty term it can be ensured that the optimizer fulfills the constraints at the *end* of the process.

An alternative are *log-barrier* methods [163], p.498. In contrast to the penalty approach each iterate is feasible during the whole optimization process. The algorithm starts in the feasible region and logarithmic penalty terms make sure that the iteration points stay away from the boundary of the feasible region.

The third main group of algorithms which replace the original constrained problem to a sequence of unconstrained problems are augmented Lagrangian methods [163], p.511. Augmented Lagrangian methods are similar to quadratic penalty methods, but the subproblems are in general better conditioned. In contrast to log-barrier methods the iterates do not have to be feasible during the optimization process but are feasible at the end.

Besides penalty, barrier, and augmented Lagrangian methods, another category of important approaches for constrained optimization problems is sequential quadratic programming (SQP). The idea of SQP methods is to model the

constrained problem in each iteration by a quadratic (constrained) subproblem which is easier to solve. The search direction p_k is the solution of the subproblem. The step length is determined by a so-called merit function, which measures the progress of the optimization. The merit function has to find a balance between a reduction of the objective function and satisfying the constraints.

We will present a SQP method, which uses a Gauß-Newton approximation to the Hessian and is also called generalized Gauß-Newton method. In addition, the generalized Gauß-Newton approach is combined with an Augmented Lagrangian method.

5.4.1 Sequential Quadratic Programming

At first we will focus only on equality constrained problems. Sequential quadratic programming can be motivated by applying Newton's method onto the KKT conditions for equality constrained problems (Eq. 5.8). We formulate the KKT conditions as a nonlinear function F which is set to 0:

$$F(x, \lambda) = \begin{pmatrix} F_1(x, \lambda) \\ F_2(x, \lambda) \end{pmatrix} = \begin{pmatrix} \nabla_x L(x, \lambda) \\ c(x) \end{pmatrix} = 0 \quad (5.24)$$

where $\nabla_x L(x, \lambda) = \nabla f(x) - \lambda^T \nabla c(x)$ is the gradient of the Lagrangian. Newton's method for solving these non-linear equations is defined by the iterations $x_{k+1} = x_k + p_k$, $\lambda_{k+1} = \lambda_k + q_k$ where the next Newton step (p_k, q_k) is determined by solving the linear Newton equations:

$$J_F(x_k, \lambda_k)(p_k, q_k) = -F(x_k, \lambda_k), \quad (5.25)$$

where the Jacobian J_F of F is:

$$J_F = \begin{pmatrix} \nabla_x F_1 & \nabla_\lambda F_1 \\ \nabla_x F_2 & \nabla_\lambda F_2 \end{pmatrix} = \begin{pmatrix} \nabla_{xx} L & -\nabla_x c^T \\ \nabla_x c & 0 \end{pmatrix}. \quad (5.26)$$

Hence, we get the Newton equations:

$$\begin{pmatrix} \nabla_{xx} L(x_k, \lambda_k) & -\nabla_x c^T(x_k) \\ \nabla_x c(x_k) & 0 \end{pmatrix} \begin{pmatrix} p_k \\ q_k \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) + \nabla c^T(x_k) \lambda_k \\ -c(x_k) \end{pmatrix}. \quad (5.27)$$

If we subtract $\nabla c^T(x_k) \lambda_k$ from the first equation we obtain the equivalent system:

$$\begin{pmatrix} \nabla_{xx} L(x_k, \lambda_k) & -\nabla_x c^T(x_k) \\ \nabla_x c(x_k) & 0 \end{pmatrix} \begin{pmatrix} p_k \\ \lambda_k + q_k \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) \\ -c(x_k) \end{pmatrix}. \quad (5.28)$$

With $x_{k+1} = x_k + p_k$, $\lambda_{k+1} = \lambda_k + q_k$ we get the next iterate which defines the new search direction (x_{k+1}, λ_{k+1}) of the original constrained optimization problem.

We will now show that the solution of this Newton system can be identified with the solution of a quadratic approximation of the constrained optimization problem leading to the SQP framework. Therefore we define the following sequence of quadratic optimization problems with linear constraints to find a new search direction and new Lagrangian multipliers at an iterate (x_k, λ_k) :

$$\begin{aligned} \frac{1}{2}p^T H_k p + \nabla f_k^T p &\xrightarrow{p} \min, \\ \text{s.t. } C_k p + c_k &= 0. \end{aligned} \tag{5.29}$$

The quadratic optimization problem has a unique solution (p_k, z_k) (under certain assumptions on C_k and H_k):

$$\begin{aligned} H_k p_k + \nabla f_k - C_k^T z_k &= 0, \\ C_k p_k + c_k &= 0. \end{aligned} \tag{5.30}$$

Let H_k be $H_k = \nabla_{xx}L(x_k)$ and $C_k = \nabla c(x_k)$, then we get exactly the Newton equations 5.28 from above where $z_k = \lambda_{k+1} = \lambda_k + q_k$. This linear system is also called a KKT system because it is derived from the KKT conditions.

Now we can formulate the (local) SQP algorithm:

Algorithm 5.2. (Sequential Linear Programming (SQP))

1. Set $k = 0$
2. Determine $f_k, \nabla f_k, H_k, c_k, C_k$
3. Solve the KKT system (Eq. 5.30) to get p_k and z_k
4. Set $x_{k+1} = x_k + p_k, \lambda_{k+1} = z_k$
5. Set $k = k+1$
6. If convergence test is not satisfied, continue with step 2

The algorithm is called local, because it converges only if the iterates are close to the solution. As in unconstrained problems the algorithm also converges from remote starting points and for nonconvex problems if a line-search or trust-region strategy is added. Again we will focus only on a line-search variant. Two ingredients are important for a practical line-search SQP approach: a) A modification of the Hessian H_k to get a positive definite approximation leading to a convex problem and b) a merit function which guides the algorithm to the solution. Therefore we combine the SQP framework with an augmented Lagrangian and a generalized Gauß-Newton approach in the next sections.

5.4.2 Augmented Lagrangian and SQP

The idea of the augmented Lagrangian method is to convert a constrained into an unconstrained optimization problem for which algorithms are already available by adding a quadratic penalizer term $\frac{1}{2\mu}c(x)^T c(x)$ to the Lagrangian function instead to the objective function like in the quadratic penalty method. We get the following augmented Lagrangian function L_A which is parameterized by the penalty weight μ :

$$L_A(x, \lambda; \mu) = f(x) - \lambda c(x) + \frac{1}{2\mu}c(x)^T c(x). \quad (5.31)$$

The violation of the constraints is measured by the penalty term, which is just the norm of the constraints. The penalty weight $\mu \in \mathbb{R}^+$ is decreased during the iterations to ensure feasibility at the end of the optimization process. If the violation of the constraints cannot sufficiently be reduced during an iteration, the penalty weight is decreased in the next iteration ($\mu_{k+1} = \frac{1}{2}\mu_k$). Because the KKT conditions require that $\nabla_x L(x_*, \lambda_*) = 0$ and $c(x_*) = 0$, the term $c(x)^T c(x)$ vanishes at the optimum (x_*, λ_*) and the augmented Lagrangian L_A coincides with the Lagrangian L . Then the penalty weight μ no longer needs to be small as in the quadratic penalty method leading to fewer ill conditioned subproblems.

The necessary condition $(\nabla_x L_A(x_*, \lambda_*), \nabla_\lambda L_A(x_*, \lambda_*)) = 0$ for the unconstrained augmented Lagrangian function at the optimum (x_*, λ_*) leads to a very similar linear system compared to the KKT system of the SQP framework (Eq. 5.28). With $\nabla_x L_A = \nabla_x L + \frac{1}{\mu}(\nabla_x c)^T c$ and $\nabla_\lambda L_A = \nabla_\lambda L = -c$ we get:

$$\begin{pmatrix} \nabla_{xx} L_A(x_k, \lambda_k) & -\nabla_x c^T(x_k) \\ \nabla_x c(x_k) & 0 \end{pmatrix} \begin{pmatrix} p_k \\ \lambda_k + q_k \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) - \frac{1}{\mu}(\nabla_x c)^T c \\ -c(x_k) \end{pmatrix}. \quad (5.32)$$

There are only two differences to the usual SQP system 5.28: a) the Hessian of the augmented Lagrangian $\nabla_{xx} L_A = \nabla_{xx} L + \frac{1}{\mu}(\nabla c)^T \nabla c$ and b) the additional summand $\frac{1}{\mu}(\nabla_x c)^T c$ on the right hand side of the first equation. The additional term $\frac{1}{\mu}(\nabla c)^T \nabla c$ in the Hessian yields a stabilization by adding positive curvature (see also [163], p. 540). The theoretical justification for the augmented Lagrangian approach can be found in the book of Nocedal and Wright [163], p. 517. They proof that when the exact Lagrange multipliers λ_* are known, the solution x_* of the original constrained problem is a minimizer of the augmented Lagrangian $L_A(x, \lambda_*; \mu)$ for all μ sufficiently small. Although we do know only a reasonable estimate of the exact λ_* in practice they show in a second theorem that we still obtain a good estimate of x_* .

As in the previous section an alternative derivation leads to the same system by using the SQP framework. The quadratic approximation of the equality

constrained optimization problem was already defined in Eq. 5.29:

$$\begin{aligned} \frac{1}{2}p^T H_k p + \nabla f_k^T p &\xrightarrow{p} \min, \\ \text{s.t. } C_k p + c_k &= 0. \end{aligned} \quad (5.33)$$

If we formulate the augmented Lagrangian \tilde{L}_A for this quadratic problem, we get:

$$\tilde{L}_A(p, \lambda; \mu) = f_k + \nabla f_k^T p + \frac{1}{2}p^T H_k p - \lambda^T (c_k + C_k p) + \frac{1}{2\mu} (c_k + C_k p)^T (c_k + C_k p). \quad (5.34)$$

For ease of notation we ignore the index k for a moment. With $H = \nabla^2 f$, $C = \nabla c$ the necessary conditions for a minimum (p_*, λ_*) of this augmented Lagrangian

$$\nabla_p \tilde{L}_A(p, \lambda; \mu) = \nabla f^T + H p - \lambda^T C + \frac{1}{\mu} C^T (c + C p) = 0, \quad (5.35)$$

$$\nabla_\lambda \tilde{L}_A(p, \lambda; \mu) = c + C p = 0 \quad (5.36)$$

lead to the same KKT system as above:

$$\begin{pmatrix} H + \frac{1}{\mu} C^T C & -C^T \\ C & 0 \end{pmatrix} \begin{pmatrix} p \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f - \frac{1}{\mu} (C^T c) \\ -c \end{pmatrix}. \quad (5.37)$$

If we add $\mu\lambda$ to the second equation and rearrange the whole system, we derive the following iteration step for the determination of the next search direction:

$$(H_k + \frac{1}{\mu} C_k^T C_k) p_{k+1} = C_k^T \lambda_{k+1} - \nabla f_k^T - \frac{1}{\mu} C_k^T c_k, \quad (5.38)$$

$$\lambda_{k+1} = \lambda_k - \frac{1}{\mu} (c_k + C_k p_k) \quad (5.39)$$

Now the equations are independent of each other. We can update λ_{k+1} by the second equation and then solve the first equation for p_{k+1} using the update λ_{k+1} . On this new search direction (p_{k+1}, λ_{k+1}) a strong Wolfe line search strategy (see [163]) using a merit function is applied to find the step length α_k leading to the next iterate $x_{k+1} = x_k + \alpha_k p_{k+1}$.

Now we motivate the use of a merit function. For *unconstrained* optimization problems the optimization progress can directly be measured by the objective function ($f(x_{k+1}) < f(x_k)$). In *constrained* optimization there is a conflict in each iteration between a reduction in the objective function and satisfying the constraints. Therefore we need a measure to find the best compromise between these two objectives. Merit functions quantify the compromise and

control the algorithm: a new search direction p_{k+1} is only accepted if a sufficient decrease in the merit function can be reached. As a merit function we use the augmented Lagrangian function itself which is also called ‘Fletcher’s augmented Lagrangian merit function’ in the literature [163].

Until now we have described the determination of the search direction and a step length determination strategy by means of a merit function for a fixed penalty weight factor μ . At the end, we will sketch when and how the penalty weight factor μ_k is changed in an iteration step k . The violation of the constraints in step $k + 1$ is compared to the previous step k :

$$\kappa \|c_k\| < \|c_{k+1}\| \quad (5.40)$$

with $\kappa \in (0, 1)$, e.g., $\kappa = 0.5$. If no sufficient decrease is reached the penalty weight μ is decreased by $\mu_{k+1} = \frac{1}{2}\mu_k$.

Extension to Inequality Constraints

Until now, we only considered equality constrained optimization problems. Now, we extend the augmented Lagrangian approach to inequality constraints. For simplicity suppose that only inequality and no equality constraints are given. Again two derivations are possible: either we take the augmented Lagrangian of the original inequality constrained problem or we formulate a sequential quadratic program for inequality constraints and setup an augmented Lagrangian for this quadratic program. In both cases so-called slack variables $s \in \mathbb{R}^m$ are introduced to convert the inequality constraints into equality constraints plus bound constraints on the slack variables. The bound constraints are later eliminated resulting in a transformed augmented Lagrangian which can be treated like the augmented Lagrangian of the equality constrained problem. The first derivation can be found in Nocedal and Wright [163], p.514. We will sketch the second derivation here.

At first the inequality constrained optimization problem is approximated by a sequential quadratic model:

$$\begin{aligned} \frac{1}{2}p^T H_k p + \nabla f_k^T p &\stackrel{p}{\rightarrow} \min, \\ \text{s.t. } C_k p + c_k &\geq 0. \end{aligned} \quad (5.41)$$

This problem can be converted into an *equality* constrained problem by introducing slack variables $s = s_1, \dots, s_m \in \mathbb{R}^m$ and replacing the inequality constraints

$$C_k p + c_k \geq 0 \quad \text{by} \quad C_k p + c_k - s = 0, \quad s \geq 0. \quad (5.42)$$

5. OPTIMIZATION

We now have an equality constrained problem with additional bound constraints ($s \geq 0$):

$$\begin{aligned} & \frac{1}{2}p^T H_k p + \nabla f_k^T p \xrightarrow{p} \min, \\ \text{s.t. } & C_k p + c_k - s = 0, \quad s \geq 0. \end{aligned} \quad (5.43)$$

The bound constraints are either treated explicitly by special algorithms, e.g., the LANCELOT package, or are eliminated to get a pure unconstrained problem. We use the second variant and formulate an augmented Lagrangian function L_A^I for the equality and bound constraint problem 5.43. The index I indicates that the augmented Lagrangian is used to optimize the originally inequality constrained problem. For a simpler notation we leave out the index k . Let the model of the objective function be $m_f(p) = (\nabla f_k)^T p + \frac{1}{2}p^T H_k p$ and let the new equality constraints be $\tilde{c}(p, s) = c + Cp - s$, then we get the following subproblem:

$$\begin{aligned} L_A^I(p, \lambda; \mu, s) = m_f(p) - \lambda^T \tilde{c}(p, s) + \frac{1}{2\mu} \tilde{c}(p, s)^T \tilde{c}(p, s) \xrightarrow{p} \min, \\ \text{s.t. } s \geq 0. \end{aligned} \quad (5.44)$$

The slack variables s_i appear in just two summands of the augmented Lagrangian L_A^I . Since the augmented Lagrangian is convex with respect to the slack variables a solution with respect to s can be computed explicitly. By setting $\nabla_s L_A^I = 0$ one gets the minimum of the augmented Lagrangian with respect to s : $s = c + Cp - \lambda\mu$. If a component s_i of this unconstrained minimizer is smaller than 0, the optimal value of this component for the whole bound constraint problem 5.44 is 0 because of the bound constraint $s_i \geq 0, i = 1, \dots, m$ and the convexity of L_A^I with respect to s . This means the solution s^* of Eq. 5.44 with respect to the slack variables s is given by:

$$s_i^* = \max((c + Cp)_i - \mu\lambda_i, 0). \quad (5.45)$$

Now the slack variables s can be substituted if we differentiate for each component s_i two cases. If $(c + Cp)_i - \mu\lambda_i > 0$ we substitute s_i by $(c + Cp)_i - \mu\lambda_i$ in the augmented Lagrangian L_A^I else s_i is substituted by 0. This leads to an augmented Lagrangian which no longer depends explicitly on the slack variables and bound constraints (for further details see [163]):

$$L_A^I(p, \lambda; \mu) = m_f(p) + \mathbf{1}_m^T \psi(p, \lambda; \mu) \quad (5.46)$$

with $\mathbf{1}_m = (1, \dots, 1)$ and

$$\psi_i(p, \lambda; \mu) = \begin{cases} -\lambda_i(c + Cp)_i + \frac{1}{2\mu}(c + Cp)_i^2, & \text{if } (c + Cp)_i - \mu\lambda_i \leq 0, \\ -\frac{\mu}{2}\lambda_i^2, & \text{else} \end{cases} \quad (5.47)$$

As in the equality constrained problem, a new search direction is now defined by the KKT system of the augmented Lagrangian $L_A^I(p, \lambda; \mu)$ using its gradients $\nabla_p L_A^I$ and $\nabla_\lambda L_A^I$. The update strategy for the Lagrangian multipliers differs slightly from the equality constraints case (see [163], p.514). We update by setting:

$$\lambda_{k+1} = \max \left(\lambda_k - \frac{1}{\mu}(c + Cp), 0 \right). \quad (5.48)$$

Afterwards the new search direction p_{k+1} is computed based on λ_{k+1} and a line search strategy is applied similarly to the equality constrained case in the previous section.

5.4.3 Generalized Gauss-Newton

As in the unconstrained case, calculating the Hessian of the original objective function may lead to high computational costs. Therefore we again use a Gauss-Newton type approximation of the Hessian $\nabla^2 f \approx H^{GN}$ as described in equation 5.21 of section 5.3.1. For equality constrained least squares problems the use of the Gauss-Newton approximation was described by Bock et al. [18] and Olesch [169]. They call it the *generalized* Gauss-Newton approach because the Gauss-Newton iteration rule determines a search direction (x_k, λ_k) for the Lagrangian function $L(x, \lambda)$ instead of the original objective function $f(x)$. Here we use the Gauss-Newton approximation H^{GN} of the Hessian in the augmented Lagrangian framework for equality and inequality constraints.

Chapter 6

Nonparametric Image Registration

6.1 Introduction

The general nonparametric registration problem has been already modeled as a minimization problem on functionals of continuous functions in Chapter 3. We omit additional penalizers for soft constraints and also hard constraints on the deformation here. They will be discussed in Chapter 10. This means we concentrate on nonparametric image registration defined by a functional J consisting of a distance measure \mathcal{D} on the images \mathcal{T}, \mathcal{R} and a regularizer \mathcal{S} on the deformation $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$:

$$\mathcal{J}[y] = \mathcal{D}[\mathcal{T}(y), \mathcal{R}] + \alpha \mathcal{S}[y - y^{\text{ref}}] \xrightarrow{y} \min \quad (6.1)$$

where y^{ref} is a reference transformation for the regularizer, which is explained more detail in Chapter 6.3. To solve this problem numerically, it has to be discretized such that the Gauß-Newton optimization scheme as described in Sec. 5 can be applied. The discretization of the images $R = \mathcal{R}(\mathbf{x}^{cc}), T = \mathcal{T}(\mathbf{x}^{cc})$ and the transformation $\mathbf{y} = y(\mathbf{x}^{cc})$ on a cell-centered grid \mathbf{x}^{cc} have been shown in Chapter 4. Now the distance measures and the regularizers have to be defined and discretized to get a discrete version of the whole registration functional:

$$J(\mathbf{y})^h = D^h(T_{\mathbf{y}}, R) + \alpha S^h(\mathbf{y} - \mathbf{y}^{\text{ref}}) \quad (6.2)$$

where the spatial discretization h corresponds to the grid resolution. For ease of presentation, we will often omit h in the notation. Based on the discrete functional and the Gauß-Newton optimization scheme, we define a non-rigid nonparametric image registration algorithm on a fixed discretization level. Since the distance functional is usually non-convex there might be several local minima. Thus multi-level strategies are necessary which also substantially improve the efficiency of the registration. The main idea of a multi-level strategy is to solve the minimization problem first on a coarse discretization where only the most important features are considered. The probability to

get trapped in a local minimum is lower on a coarse level. If no further improvements can be achieved the coarse solution is prolonged to a finer discretization and used as a good starting guess for the finer level. On the finer level, usually fewer iterations have to be executed because the coarse solution was already close to the optimum and only corrections have to be computed.

6.2 Distance Measures

Distance measures are a key part of registration methods. In each iteration step they determine how similar the transformed template image looks to the reference image. A review of general and special-purpose as well as some new distance measures are presented in detail in the chapters 8 and 9. Here we define only the simple ‘sum of squared differences’ distance measure to show how a distance measure is discretized in general and what the derivative of a discrete distance measure looks like.

The sum of squared differences distance measure \mathcal{D}^{SSD} is defined by:

$$\mathcal{D}^{SSD}[\mathcal{T}, \mathcal{R}, y] = \frac{1}{2} \int_{\Omega} (\mathcal{T}(y(x)) - \mathcal{R}(x))^2 dx. \quad (6.3)$$

Chapter 8 gives an overview for which types of image pairs this distance measure is suitable, what its properties are and compares it to other measures.

The sum of squared differences measure is an example of a distance measure where for each image point in the reference image a kind of difference (in this case the Euclidean distance) to the transformed template image is computed (by the function ψ) and then all the differences are integrated. All of the distance measures defined in this thesis will be of this integrated differences type. We formalize them by

$$\mathcal{D}[\mathcal{T}, \mathcal{R}] = \int_{\Omega} \psi(\mathcal{T}(x), \mathcal{R}(x)) dx. \quad (6.4)$$

6.2.1 Discretization of Distance Measures

Distance measures are discretized by a numerical integration method also called quadrature for one-dimensional functions. Many different numerical integration schemes exist. For our purposes the simple and robust midpoint quadrature is adequate. Let $f : \Omega \rightarrow \mathbb{R}$ be an integrable function and x_j the cell centers of a one-dimensional cell-centered grid $\mathbf{x}^{cc}([\omega_1, \omega_2], m)$. The cell size is $h = (\omega_2 - \omega_1)/m$. Then the integral of f can be approximated by:

$$\int_{\Omega} f(x) dx = h \sum_{j=1}^m f(x_j) + \mathcal{O}(h^2). \quad (6.5)$$

The approximation error is of the order h^2 . With h small enough the integration error will be below any given error tolerance resp. with $h \rightarrow 0$ the approximation equals the continuous formulation. For image registration a natural choice for h is the original image resolution, but we will see that also other choices are possible and reasonable.

By means of the midpoint quadrature rule we are now able to discretize the SSD distance measure (and all other measures):

$$\mathcal{D}^{\text{SSD}}[\mathcal{T}, \mathcal{R}, y] = \frac{1}{2} \int_{\Omega} (\mathcal{T}(y(x)) - \mathcal{R}(x))^2 dx \quad (6.6)$$

$$\approx D^{\text{SSD}}(T_{\mathbf{y}}, R) = \frac{1}{2} \cdot \bar{h} \cdot \|T_{\mathbf{y}} - R\|^2 \quad (6.7)$$

where $\bar{h} = h_1 \cdot h_2 \cdot h_3$ is the product of the voxel sizes in each coordinate direction. Due to the already mentioned special form (integrated differences) of all distance measures used in this thesis a discrete distance measure can be decomposed into an inner (residual) and an outer function. For a discretized distance measure $D : \mathbb{R}^{3n} \rightarrow \mathbb{R}$ this can be phrased as:

$$D(\mathbf{y}) = \phi(r(\mathbf{y})) \quad (6.8)$$

with $\phi : \mathbb{R}^{\tilde{n}} \rightarrow \mathbb{R}$ being the outer function and $r : \mathbb{R}^{3n} \rightarrow \mathbb{R}^{\tilde{n}}$ being the residual or inner function. For the sum of squared differences measure we get: $\phi(r(\mathbf{y})) = \frac{\bar{h}}{2} \|r(\mathbf{y})\|^2$ and $r(\mathbf{y}) = T_{\mathbf{y}} - R$. In this case $\tilde{n} = 1$ holds, but \tilde{n} might be bigger than one.

6.2.2 Derivatives of Discrete Distance Measures

To enable fast optimization schemes, we determined the analytical first order derivative and an analytical approximation of the second order derivative of a discrete distance measure $D(\mathbf{y}) = D(T_{\mathbf{y}}, R) = \phi(r(\mathbf{y}))$ with respect to the transformation \mathbf{y} . The derivatives of the decomposed distance measure can be computed based on the chain and product rule:

$$dD(\mathbf{y}) = d\phi(r(\mathbf{y})) \cdot dr(\mathbf{y}), \quad (6.9)$$

$$d^2D(\mathbf{y}) = dr(\mathbf{y})^T d^2\phi(r(\mathbf{y})) \cdot dr(\mathbf{y}) + d\phi(r(\mathbf{y})) \cdot d^2r(\mathbf{y}). \quad (6.10)$$

For the Gauß-Newton optimization scheme we only use an *approximation* of the Hessian d^2D . The second summand including the second order derivatives of the residual function r is neglected because the residual depends on the noisy image data while the outer function ϕ is usually smooth:

$$d^2D(\mathbf{y}) \approx dr(\mathbf{y})^T d^2\phi(r(\mathbf{y})) \cdot dr(\mathbf{y}). \quad (6.11)$$

This means for each distance measure we only need to determine the outer and inner functions ϕ and r as well as their first derivatives $d\phi, dr$ and the

second derivative $d^2\phi$ of the generally simple and smooth outer function ϕ . For example, the inner and outer functions of the sum of squared differences measure are

$$r(\mathbf{y}) = T_{\mathbf{y}} - R, \quad \phi(r) = \frac{\bar{h}}{2} r^T r \quad (6.12)$$

and the derivatives of the inner and outer function

$$dr(\mathbf{y}) = dT_{\mathbf{y}}, \quad d\phi(r) = \bar{h}r, \quad d^2\phi(r) = \bar{h} \quad (6.13)$$

lead to the derivatives of the distance measure

$$dD(\mathbf{y}) = \bar{h} \cdot (T_{\mathbf{y}} - R)^T \cdot dT_{\mathbf{y}}, \quad d^2D(\mathbf{y}) \approx dT_{\mathbf{y}}^T \cdot \bar{h} \cdot dT_{\mathbf{y}}. \quad (6.14)$$

The transformed template image $T_{\mathbf{y}}$ and its derivative $dT_{\mathbf{y}}$ are given in sec. 4.3. The derivative $dT_{\mathbf{y}}$ uses the derivative of the 3D spline model function determined in sec. 4.3.2.

6.3 Regularizers

The task of a regularizer is to alter the registration functional to get a well-posed problem with a unique solution. Adding a regularizer on the transformation y to the registration functional tries to get a convex objective function for the optimization process. Most regularizers are L_2 -norms on derivatives of the transformation y or more precisely of the displacements $u = y - y^{\text{ref}}$. The reference transformation y^{ref} might be simply the identity $y^{\text{ref}}(x) = x$ or an initial transformation computed by a landmark- or parametric intensity-based registration.

6.3.1 Continuous Formulation

First we will define the regularizers in a continuous setting before their discretization is discussed. Three different regularizers are presented for three-dimensional transformations y .

Diffusion Regularizer

The diffusion regularizer has been introduced by Horn and Schunck [91] to image processing (optical flow) and by Fischer and Modersitzki [56] in particular for image registration. The simple idea of the regularizer is to penalize high gradients of y to get a smooth transformation:

$$\mathcal{S}^{\text{diff}}[y] = \frac{1}{2} \int_{\Omega} \sum_{j=1}^3 \left\| \nabla(y_j - y_j^{\text{ref}}) \right\|^2 dx. \quad (6.15)$$

The coordinate directions are independently regularized. A deformation in one coordinate direction does not imply a deformation in other directions. This is an important difference to the elastic regularizer which we will consider in the next section.

Elastic Regularizer

The term elastic registration is sometimes used for non-rigid registration in general. We use the term ‘elastic registration’ for an approach based on the regularizer introduced by Broit [22]. The regularizer is physically motivated and represents the linear elastic potential measuring the energy resulting from deforming an elastic material:

$$\mathcal{S}^{\text{elas}}[y] = \frac{1}{2} \int_{\Omega} \sum_{j=1}^3 \mu \left\| \nabla(y_j - y_j^{\text{ref}}) \right\|^2 + (\mu + \lambda) \text{div}^2(y - y^{\text{ref}}) dx \quad (6.16)$$

where the divergence of a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is defined by $\text{div} = \partial_1 f_1 + \partial_2 f_2 + \partial_3 f_3$. The divergence operator indicates a change of volume. Materials with different elastic properties are modeled by the material constants $\lambda, \mu \geq 0$ which are also called Lamé constants.

The elastic potential is also used for landmark registration based on Gaussian Elastic Body Splines (see sec. 11.9).

Curvature Regularizer

The curvature regularizer introduced by Fischer and Modersitzki [59] is based on the second order derivatives:

$$\mathcal{S}^{\text{curv}}[y] = \frac{1}{2} \int_{\Omega} \sum_{j=1}^3 \left\| \Delta(y_j - y_j^{\text{ref}}) \right\|^2 dx \quad (6.17)$$

with the Laplace operator $\Delta f = \partial_{11} f + \partial_{22} f + \partial_{33} f$ for a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$. The name of the regularizer is due to the fact that each of the summands Δy_j is an approximation to the curvature of the function component y_j . The curvature regularizer leads to smoother transformations than the other two regularizers. In addition, the curvature regularizer has the property that linear parts of the deformation do not influence the value of the regularizer: $\mathcal{S}^{\text{curv}}[y] = 0$ for $y = Ax + b$ resp. $\Delta y_j = 0$.

6.3.2 Discretization

The two most important ingredients for the discretization of the regularizers are the discretizations for the first and second order partial derivatives and the numerical computation of the integral. For numerical integration again the

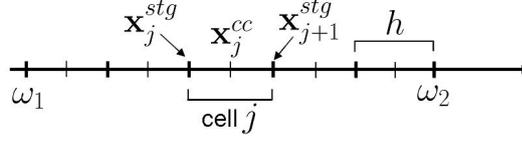


Figure 6.1 One-dimensional cell-centered and staggered grid on (ω_1, ω_2) .

midpoint quadrature rule is applied on a cell-centered grid like for the distance measures (see Sec. 6.2.1). For the divergence term of the elastic regularizer, central short finite differences are recommended as a discretization of the first order derivatives [150, 170]. The derivatives have to be evaluated on a cell-centered grid to match the discretization of the distance measure. But short finite differences use function values at points on the cell *boundaries*. We will define so-called staggered grids which are defined on the cell boundaries. The function values on the staggered grids can be interpolated from the values on the cell-centered grids and vice versa.

The discretization of the diffusion regularizer is treated as a special case of the elastic regularizer. Since the curvature regularizer is based only on second order derivatives no short finite differences and therefore no staggered grids are needed.

Discrete First Derivatives in the One-Dimensional Case

Let a cell-centered grid $\mathbf{x}^{cc}(\Omega, m)$ on $\Omega = (\omega_1, \omega_2)$ with m cells of size $h = (\omega_2 - \omega_1)/m$ be given. For a one-dimensional function $f : \mathbb{R} \rightarrow \mathbb{R}$ the first derivative at a grid point \mathbf{x}_j^{cc} can be approximated by the short central finite difference

$$\partial f(\mathbf{x}_j^{cc}) = \frac{f(\mathbf{x}_j^{cc} + 0.5h) - f(\mathbf{x}_j^{cc} - 0.5h)}{h} + \mathcal{O}(h^2). \quad (6.18)$$

The approximation order $\mathcal{O}(h^2)$ follows from Taylor's theorem for a twice continuously differentiable function f .

The problem is that we have to evaluate the function f at points on the *boundaries* of the cells $(\mathbf{x}_j^{cc} + 0.5h, \mathbf{x}_j^{cc} - 0.5h)$ and not at the cell centers of the given grid. The points on the cell boundaries can be interpreted as a grid type of their own. Modersitzki [150] and Papenberg [170] call them nodal grids, but they can also be seen as one-dimensional staggered-grids which are defined by Modersitzki and Papenberg only for two and three dimensions. The grid points of a one-dimensional staggered-grid $\mathbf{x}^{stg}(\Omega, m) \in \mathbb{R}^{m+1}$ are defined as

$$(\mathbf{x}^{stg})_j = \omega_1 + jh, j = 0, \dots, m. \quad (6.19)$$

The relation of a staggered to a cell-centered grid is illustrated in Fig. 6.1. Note that there is one more cell boundary point than cell center points such that there are m cell-centered grid points and $m + 1$ staggered grid points.

The short finite differences can be written in terms of the staggered grid:

$$\partial f(\mathbf{x}_j^{cc}) = \frac{f(\mathbf{x}_j^{stg}) - f(\mathbf{x}_{j-1}^{stg})}{h} + \mathcal{O}(h^2) \quad (6.20)$$

because $\mathbf{x}_j^{cc} + 0.5h = \mathbf{x}_j^{stg}$ and $\mathbf{x}_j^{cc} - 0.5h = \mathbf{x}_{j-1}^{stg}$.

Now we introduce discrete derivation operators in matrix form. They allow compact descriptions and an extension to three dimensions by combining them with matching identity matrices via Kronecker products (\otimes). But the matrices are not explicitly generated in program code because they would consume too much memory even if coded as sparse matrices.

Let $\mathbf{x}^{stg} = \mathbf{x}^{stg}(\Omega, m)$ be the staggered grid and $\mathbf{x}^{cc} = \mathbf{x}^{cc}(\Omega, m)$ be the corresponding cell-centered grid. The first derivative of f at the positions \mathbf{x}^{cc} of the one-dimensional cell-centered grid can be rephrased as

$$\partial f(\mathbf{x}^{cc}) \approx D \cdot f(\mathbf{x}^{stg}) \quad (6.21)$$

with

$$D(h, m) = \frac{1}{h} \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{m \times (m+1)}. \quad (6.22)$$

The operator is a $(m \times (m + 1))$ matrix because the staggered grid has $m + 1$ and the cell-centered grid m grid points.

Discrete First Derivatives in the Three-Dimensional Case

The 3D discrete derivatives are again based on staggered grids. Thus, at first, we construct 3D staggered grids from 1D grids. In 3D, there is one staggered grid for each of the three coordinate directions. We first explain and illustrate staggered grids in 2D for an easier understanding.

The two 2D staggered grids \mathbf{x}^{stg1} and \mathbf{x}^{stg2} for the x_1 - and x_2 -direction are defined on the boundaries of the rectangular cells. The grid points of \mathbf{x}^{stg1} lie on the vertical rectangle sides and the grid points of \mathbf{x}^{stg2} on the horizontal rectangle sides (see Fig. 6.2). Both staggered grids can easily be generated from the one-dimensional staggered grids. If we consider \mathbf{x}^{stg1} , the x_1 coordinates of the grid points are exactly the grid points $\mathbf{x}_1^{stg} \in \mathbb{R}^{m_1+1}$ of a 1D staggered grid and the x_2 coordinates are the grid points $\mathbf{x}_2^{cc} \in \mathbb{R}^{m_2}$ of a cell-centered grid. As the values \mathbf{x}_1^{stg} stay the same for each of the m_2 horizontal cell boundary lines they have to be repeated m_2 times in the description of the grid. The same holds for the values \mathbf{x}_2^{cc} which stay the same

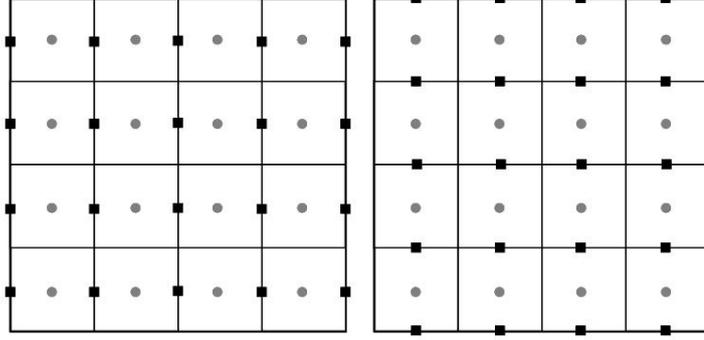


Figure 6.2 Two-dimensional cell-centered (gray dots) and staggered grids (black squares) in x_1 -direction on the left and in x_2 -direction on the right.

for each of the $m_1 + 1$ vertical cell boundary lines (see Fig. 6.2). For the formal description of the whole grid as one vector the repetitions can be expressed by the Kronecker product:

$$\mathbf{x}^{stg_1} = \begin{pmatrix} \mathbf{1}_{m_2} \otimes \mathbf{x}_1^{stg} \\ \mathbf{x}_2^{cc} \otimes \mathbf{1}_{m_1+1} \end{pmatrix} \in \mathbb{R}^{2(m_1+1)m_2}. \quad (6.23)$$

$\mathbf{1}_n$ the column vector where all n elements are equal to 1. For the x_2 coordinates the order of the factors has to be changed to get an analog vector notation like for the cell-centered grids (see Sec. 4.2.4).

The second staggered grid \mathbf{x}^{stg_2} is similarly derived:

$$\mathbf{x}^{stg_2} = \begin{pmatrix} \mathbf{1}_{m_2+1} \otimes \mathbf{x}_1^{cc} \\ \mathbf{x}_2^{stg} \otimes \mathbf{1}_{m_1} \end{pmatrix} \in \mathbb{R}^{2m_1(m_2+1)}. \quad (6.24)$$

For the discrete derivatives of the transformation y we need a mix of both staggered grids: the x_1 components of \mathbf{x}^{stg_1} and the x_2 components of \mathbf{x}^{stg_2} . We define the mixed components as their own staggered grid:

$$\mathbf{x}^{stg} = \begin{pmatrix} \mathbf{1}_{m_2} \otimes \mathbf{x}_1^{stg} \\ \mathbf{x}_2^{stg} \otimes \mathbf{1}_{m_1} \end{pmatrix} \in \mathbb{R}^{(m_1+1)m_2+m_1(m_2+1)}. \quad (6.25)$$

Like in the 1D and 2D case three-dimensional staggered grids live on the cell boundaries. In 3D, the grid cells are cuboids and the staggered grid points lie on the centers of the cuboid's faces. For each spatial direction a staggered grid exists and for our purpose we combine components of them to a special staggered grid like in the 2D case:

$$\mathbf{x}^{stg} = \begin{pmatrix} \mathbf{1}_{m_3} \otimes \mathbf{1}_{m_2} \otimes \mathbf{x}_1^{stg} \\ \mathbf{1}_{m_3} \otimes \mathbf{x}_2^{stg} \otimes \mathbf{1}_{m_1} \\ \mathbf{x}_3^{stg} \otimes \mathbf{1}_{m_2} \otimes \mathbf{1}_{m_1} \end{pmatrix}. \quad (6.26)$$

Since the one-dimensional staggered grids have one grid point more than the cell-centered grids and the number of grid points m_1, m_2, m_3 might be different in the three coordinate directions, we get different sizes for the three components of the 3D staggered grid. The sizes n_1, n_2, n_3 for the three components are:

$$n_1 = (m_1 + 1)m_2m_3, \quad n_2 = m_1(m_2 + 1)m_3, \quad n_3 = m_1m_2(m_3 + 1). \quad (6.27)$$

The overall length of the staggered grid vector \mathbf{x}^{stg} is then $n_1 + n_2 + n_3 = 3m_1m_2m_3 + m_1m_2 + m_1m_3 + m_2m_3$.

As in the one-dimensional case, we now define matrices as discrete differential operators on the transformation $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ with $y = (y_1, y_2, y_3)^T$. Let D_{ij} be the discrete operator for the first derivative of the i -th component function y_i with respect to the spatial direction x_j . We give two examples:

$$\partial_2 y_1(\mathbf{x}^{cc}) \approx D_{12} \mathbf{y}_{1:n_1}^{stg} \quad \text{and} \quad \partial_1 y_2(\mathbf{x}^{cc}) \approx D_{21} \mathbf{y}_{(n_1+1):n_1+n_2}^{stg}. \quad (6.28)$$

The discrete transformation \mathbf{y}^{stg} has to be given on a staggered grid but the derivatives are evaluated at grid points of the according cell-centered grid. The transformation on the staggered grid is defined as:

$$\mathbf{y}^{stg} = (\mathbf{y}_{1:n_1}^{stg}, \mathbf{y}_{n_1+1:n_1+n_2}^{stg}, \mathbf{y}_{n_1+n_2+1:n_1+n_2+n_3}^{stg})^T \quad (6.29)$$

$$= (y_1(\mathbf{x}^{stg_1}), y_2(\mathbf{x}^{stg_2}), y_3(\mathbf{x}^{stg_3}))^T. \quad (6.30)$$

As for the grids, the discrete differential operators in 3D are constructed from the 1D operators (see eq. 6.22) by means of Kronecker products. Below we list all staggered grid-based discrete derivative operators in three dimensions:

$$\begin{aligned} D_{11} &= I_{m_3} \otimes I_{m_2} \otimes D(h_1, m_1) \\ D_{12} &= I_{m_3} \otimes D(h_2, m_2 - 1) \otimes I_{m_1+1} \\ D_{13} &= D(h_3, m_3 - 1) \otimes I_{m_2} \otimes I_{m_1+1} \\ \\ D_{21} &= I_{m_3} \otimes I_{m_2+1} \otimes D(h_1, m_1 - 1) \\ D_{22} &= I_{m_3} \otimes D(h_2, m_2) \otimes I_{m_1} \\ D_{23} &= D(h_3, m_3 - 1) \otimes I_{m_2+1} \otimes I_{m_1} \\ \\ D_{31} &= I_{m_3+1} \otimes I_{m_2} \otimes D(h_1, m_1 - 1) \\ D_{32} &= I_{m_3+1} \otimes D(h_2, m_2 - 1) \otimes I_{m_1} \\ D_{33} &= D(h_3, m_3) \otimes I_{m_2} \otimes I_{m_1} \end{aligned} \quad (6.31)$$

Discretization of Elastic Regularizer

The continuous elastic regularizer can be rewritten in operator form as:

$$\mathcal{S}^{\text{elas}}[y] = \frac{1}{2} \int_{\Omega} \sum_{j=1}^3 \mu \left\| \nabla(y_j - y_j^{\text{ref}}) \right\|^2 + (\mu + \lambda) \text{div}^2(y - y^{\text{ref}}) dx \quad (6.32)$$

$$= \frac{1}{2} \int_{\Omega} |\mathcal{B}[y]|^2 dx \quad (6.33)$$

where the differential operator \mathcal{B} is defined as

$$\mathcal{B}[y] = \begin{pmatrix} \sqrt{\mu} \nabla & & & \\ & \sqrt{\mu} \nabla & & \\ & & \sqrt{\mu} \nabla & \\ \sqrt{\mu + \lambda} \partial_1 & \sqrt{\mu + \lambda} \partial_2 & \sqrt{\mu + \lambda} \partial_3 & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}. \quad (6.34)$$

At first we discretize the differential operator \mathcal{B} . The partial derivatives of the gradient and the summands of the divergence term are approximated by the 3D differential operators defined in equation 6.31:

$$B^{\text{elas}} = \begin{pmatrix} \sqrt{\mu} D_{11} & & & & & & & & \\ \sqrt{\mu} D_{12} & & & & & & & & \\ \sqrt{\mu} D_{13} & & & & & & & & \\ & \sqrt{\mu} D_{21} & & & & & & & \\ & \sqrt{\mu} D_{22} & & & & & & & \\ & \sqrt{\mu} D_{23} & & & & & & & \\ & & \sqrt{\mu} D_{31} & & & & & & \\ & & \sqrt{\mu} D_{32} & & & & & & \\ & & \sqrt{\mu} D_{33} & & & & & & \\ \sqrt{\mu + \lambda} D_{11} & \sqrt{\mu + \lambda} D_{22} & \sqrt{\mu + \lambda} D_{33} & & & & & & \end{pmatrix}. \quad (6.35)$$

The discrete elastic regularizer can now be approximated by the midpoint integration rule:

$$\mathcal{S}^{\text{elas}}[y] = \frac{1}{2} \int_{\Omega} |\mathcal{B}[y]|^2 dx \quad (6.36)$$

$$\approx \frac{\bar{h}}{2} \left\| B^{\text{elas}}(\mathbf{y}_{stg} - \mathbf{y}_{stg}^{\text{ref}}) \right\|^2 = S^{\text{elas}}(\mathbf{y}_{stg} - \mathbf{y}_{stg}^{\text{ref}}) \quad (6.37)$$

Discretization of Diffusive Regularizer

The diffusive regularizer is just a special case of the elastic one with the physically meaningless material constants $\lambda = 1, \mu = -1$. Thus the divergence term vanishes and for the discrete representation of the diffusive regularizer we get:

$$S^{\text{diff}}(\mathbf{y}) = \frac{\bar{h}}{2} \left\| B^{\text{diff}}(\mathbf{y} - \mathbf{y}^{\text{ref}}) \right\|^2 \quad (6.38)$$

with B being B^{elas} , B^{diff} or B^{curv} . The gradient and Hessian of such L_2 -norm based regularizers are simply:

$$d_{\mathbf{y}}S(\mathbf{y}) = \bar{h}B^T B(\mathbf{y} - \mathbf{y}^{\text{ref}}) \quad (6.44)$$

$$d_{\mathbf{y}}^2S(\mathbf{y}) = \bar{h}B^T B \quad (6.45)$$

6.4 Nonparametric Image Registration on a Fixed Level

Now all the described ingredients can be composed into a numerical algorithm for nonparametric image registration (NPIR) on a *fixed* level. We use a discretization based on a grid with a fixed resolution. This algorithm will be the main building block in the next section for a multilevel approach. Theoretically, the fixed level approach can also be applied without a multilevel strategy, but in practice this is not recommendable as we will explain in the multilevel section.

The core of the NPIR algorithm consists of a Gauß-Newton-type optimization algorithm (see Sec. 5.3) applied to an objective function defined by the discretized registration functional $J^h(T^h, R^h, \mathbf{y}^h)$.

We define a function object `JobjFctn` with an initialization and an evaluation method which is provided to the Gauß-Newton algorithm. The initialization method is called only once, before the actual Gauß-Newton routine is applied, and stores some static objects which do not change during the whole iteration process as they are members of the function object. The objects to be stored are the input of the routine:

$$\text{JobjFctn.init}(\mathbf{x}^{cc}, R_0, T^{\text{coeff}}, \mathbf{y}^{\text{ref}}). \quad (6.46)$$

$\mathbf{x}^{cc}(\Omega, m)$ is the cell-centered grid on which the discretization of the registration functional is based. This means the transformation and the images have to be sampled on this grid. If the reference image is not already given on \mathbf{x}^{cc} , it has to be resampled once at the beginning leading to the image R_0 . For given interpolation coefficients R^{coeff} the sampled image R_0 is determined by the interpolation routine $R_0 = \text{interpolate}(R^{\text{coeff}})$. The template image has to be resampled in each iteration of the registration process according to the current transformation. But the interpolation coefficients can also only be determined once at the beginning. For the computation of the interpolation coefficients T^{coeff} (or R^{coeff}) of the original template image T^{orig} (or R^{orig}) we define the routine: $T^{\text{coeff}} = \text{computeInterCoeff}(T^{\text{orig}})$. Remember that the regularizer does not smooth the whole transformation \mathbf{y} , but the displacements $\mathbf{y} - \mathbf{y}^{\text{ref}}$. Usually the reference image is set to the identity $\mathbf{y}^{\text{ref}} = \mathbf{x}^{cc}$, but also a special reference transformation \mathbf{y}^{ref} might be given, for example coming from a manual or parametric pre-registration.

The evaluation method

$$[J_k, dJ_k, H_k] = \text{JobjFctn.evaluate}(\mathbf{y}_k) \quad (6.47)$$

returns the objective function value $J_k = J(\mathbf{y}_k) \in \mathbb{R}$, its derivative $dJ_k = d_{\mathbf{y}}J(\mathbf{y}_k) \in \mathbb{R}^n$ and an approximation $H_k^J = H^J(\mathbf{y}_k)$ to its Hessian for the current transformation \mathbf{y}_k in iteration k . This routine is used by the Gauß-Newton optimization scheme.

The NPIR algorithm computes an optimal transformation \mathbf{y}^{opt} with respect to the registration functional J , the given images $T^{\text{orig}}, R^{\text{orig}}$ on the image domain Ω and a resolution $m = (m_1, m_2, m_3)$ on which the discretization is based on.

Algorithm 6.1. (Nonparametric Image Registration (NPIR))

1. Determine $\mathbf{x}^{cc}(\Omega, m)$ for given Ω and m
2. $T^{\text{coeff}} = \text{computeInterCoeff}(T^{\text{orig}})$, $R^{\text{coeff}} = \text{computeInterCoeff}(R^{\text{orig}})$
3. $R_0 = \text{interpolate}(\mathbf{x}^{cc}, R^{\text{coeff}})$
4. Set $\mathbf{y}_0 = \mathbf{y}^{\text{ref}} = \mathbf{x}^{cc}$ (or to given start and reference transformation)
5. $\text{JobjFctn.init}(\mathbf{x}^{cc}, R_0, T^{\text{coeff}}, \mathbf{y}^{\text{ref}})$
6. $\mathbf{y}^{\text{opt}} = \text{GaußNewton}(\text{JobjFctn}, \mathbf{y}_0)$

Now we shortly explain what the evaluation routine JobjFctn.evaluate does internally. The value of the functional and its derivatives are computed based on the corresponding values of the distance measure (D_k, dD_k, H_k^D) and the regularizer (S_k, dS_k, d^2S_k):

$$J_k = D_k + \alpha S_k, \quad dJ_k = dD_k + \alpha dS_k \quad H_k^J = H_k^D + \alpha d^2S_k. \quad (6.48)$$

First, we consider the determination of the value and derivatives of the distance measure. Remember that we only cover discrete distance measures of the form $D(\mathbf{y}) = \phi(r(\mathbf{y}))$. General properties of distance measures and a simple example are given in section 6.2. Further distance measures are described in the Chapters 8 and 9. The distance measure needs the transformed image $T_k = T_{\mathbf{y}} = \mathcal{T}(\mathbf{y}_k)$ and the reference image R_0 . Therefore at first the transformed template image T_k is computed by interpolation on the current transformation using the pre-determined interpolation coefficients: $T_k = \text{interpolate}(\mathbf{y}_k, T^{\text{coeff}})$. Be aware that the current transformation is represented by the transformed grid $\mathbf{y}_k = y_k(\mathbf{x}^{cc})$. For the computation of dD_k we also need the derivative $dT_{\mathbf{y}}$ of the transformed template image (see sec. 4.3) as part of the derivative of the residual function $dr(\mathbf{y})$. Now the value and the derivatives of the distance measure are determined as:

$$\begin{aligned} D_k &= D(T_k, R_0), \\ dD_k &= d\phi(r(\mathbf{y}_k)) \cdot dr(\mathbf{y}_k), \quad d^2D_k \approx H_k^D = dr(\mathbf{y}_k)^T d^2\phi(r(\mathbf{y}_k)) \cdot dr(\mathbf{y}_k) \end{aligned} \quad (6.49)$$

One possibility to accelerate the computations further is to take the diagonal of the distance measure Hessian matrix: $H^D \approx \text{diag}(dr^T d^2\psi \cdot dr)$.

The determination of the current value $S_k = S(\mathbf{y}_k)$ of the discrete regularizer is just based on the current transformation \mathbf{y}_k and the reference \mathbf{y}^{ref} . The images are not involved in the regularization. The analytic derivative $d_{\mathbf{y}}S$ and Hessian $d_{\mathbf{y}}^2S$ for the regularizers can be found in Section 6.3:

$$S(\mathbf{y}) = \frac{\bar{h}}{2} \left\| B(\mathbf{y} - \mathbf{y}^{\text{ref}}) \right\|^2, \quad d_{\mathbf{y}}S(\mathbf{y}) = \bar{h}B^T B(\mathbf{y} - \mathbf{y}^{\text{ref}}), \quad d_{\mathbf{y}}^2S(\mathbf{y}) = \bar{h}B^T B \quad (6.50)$$

where B is the discrete differential operator of the regularizer. No approximative Hessian is used here but the full analytic second derivative of the regularizer.

If the elastic regularizer is used, the transformation has to be discretized on a staggered grid, but the images and thus the distance measure are discretized on the respective cell-centered grid. To get an overall discretization on a cell-centered grid the staggered grid has to be interpolated on the cell-centered grid by $\mathbf{y}^{\text{cc}} = P\mathbf{y}^{\text{stg}}$ where the matrix P is an averaging operator which averages each two neighboured staggered grid points to one cell-centered grid point:

$$P_m = \frac{1}{2} \begin{pmatrix} 1 & 1 & & & \\ & \ddots & \ddots & & \\ & & & 1 & 1 \end{pmatrix} \in \mathbb{R}^{m \times (m+1)}. \quad (6.51)$$

The three-dimensional averaging operator is constructed by means of the Kronecker product:

$$P = \begin{pmatrix} I_{m_3} \otimes I_{m_2} \otimes P_{m_1} & & & \\ & I_{m_3} \otimes P_{m_2} \otimes I_{m_1} & & \\ & & & P_{m_3} \otimes I_{m_2} \otimes I_{m_1} \end{pmatrix}. \quad (6.52)$$

For a concrete practical implementation of the presented NPIR algorithm matrix-free computations are performed (see [170, 150]) because storing the matrices is a waste of memory, even if sparse matrices are used. For testing purposes the presented matrix version of the algorithm can be used on small problems because it can be straight forwardly implemented, for example in MATLAB.

6.5 Multilevel Image Registration

The basic idea of a multilevel image registration approach was already sketched in the introduction of this chapter: instead of using one fixed discretization of the registration functional as in the previous section the multilevel algorithm

runs on a sequence of discretizations J^{h_1}, J^{h_2}, \dots running from coarse to fine (h_i are decreasing) of the continuous registration functional \mathcal{J} . The resulting transformation \mathbf{y}^h of one level is a good starting point $\mathbf{y}_0^{h/2}$ for the next finer level $h/2$.

There are four different possibilities to incorporate multiscale (Sec. 4.3.3) and multiresolution (Sec.4.4) approaches into image registration. The first possibility is to use multiscale representations of the images alone [170, 150]. There are no theoretical disadvantages for this strategy but the computational costs are impractically high for 3D medical images as the computations are performed on all scales with the same (fine) resolution. The second possibility is to apply a multiresolution (also multilevel) strategy for the images and the transformation alone [170, 150]. This leads to much faster computations compared to the first multiscale strategy but features like vessels are not smeared over a broader range on coarser levels (Fig. 6.3, left column) which would lead to a smoother distance functional.

The remaining two possibilities are a combination of multiscale *and* multiresolution strategies potentially using the advantages of both methods. The first alternative for a combination is to start applying a multiresolution approach and then to compute different scales of the images for each resolution [170]. But this is not a good strategy because information might already be destroyed by the resolution reduction as in the pure multi-resolution approach from above. A better strategy is to determine a multiscale representation of the original images *first* and *then* to compute different resolutions on each scale [170]. This offers the possibility to choose a resolution which fits the current scale. We do not use the whole space spanned by different scales and different resolutions but only one appropriate resolution per scale (see Fig. 6.3, right column). For this combination of multi-scale and multiresolution strategies we need three additional subroutines (compared to the NPIR algorithm): one for the multiscale image interpolation, one for the image restriction of the multiresolution approach, and one for the prolongation of the transformation.

The multiscale interpolation by approximating splines is explained in section 4.3.3. The interpolation is separated into two steps like in the single level NPIR registration method in the previous section: the computation of the interpolation coefficients and the actual interpolation based on these coefficients. Thus we replace the routine $T^{\text{coeff}} = \text{computeInterCoeff}(T^{\text{orig}})$ by $T^{\text{coeff}} = \text{computeInterCoeffMultiScale}(T^{\text{orig}}, \theta)$ where the smoothing factor θ is added as an input. It is not necessary to replace the actual interpolation routine $T = \text{Interpolate}(\mathbf{x}^{cc}, T^{\text{coeff}})$ because the approximation is already contained in the computed coefficients T^{coeff} .

Now on each level l , the reference and template image on a scale with weighting factor θ_l are determined and the resolution of the images is reduced to the resolution of this level (see Sec. 4.4). We assume that the number of

voxels in each dimension are a power of 2: $m^l = (2^l, 2^l, 2^l)$ where the exponent defines the level number l . The original image is given on the finest level l_{\max} with $T^{l_{\max}} = T^{\text{orig}}, R^{l_{\max}} = R^{\text{orig}}$ defined on the grid $\mathbf{x}^{cc}(\Omega, m^l)$. The original image might also be interpolated to this finest level. The routine $T^l = \text{Restrict}(T_{\text{coeff}}^l, l)$ interpolates an image onto the finest grid based on the interpolation coefficients T_{coeff}^l and restricts the resulting discrete image to the resolution m^l of the current level l . The restriction operator P_{avg}^{3D} (see sec. 4.4.1) is applied several times until the desired coarser resolution m^l is reached.

The prolongation of a transformation $\mathbf{y}_{\text{opt}}^{l-1}$ computed on level $l-1$ to an initial transformation \mathbf{y}_0^l in the next finer level l is performed by the subroutine Prolongate. As explained in section 4.4.2 only the displacements $\mathbf{y}^{l-1} - \mathbf{x}^{l-1}$ are prolonged leading to the following update formula: $\mathbf{y}_0^l = \mathbf{x}^l + \text{Prolongate}(\mathbf{y}_{\text{opt}}^{l-1} - \mathbf{x}^{l-1})$.

By means of the three defined subroutines computeInterCoeffMultiScale, Restrict and Prolongate now the multilevel image registration (MLIR) algorithm using different image scales and resolutions can be formulated. The algorithm is similar to the NPIR algorithm on a fixed level, but now we have a loop over the different scales resp. resolutions. The initialization of the registration and an initialization step in each iteration. The initialization of the objective function and the Gauß-Newton optimization algorithm remain the same as in the fixed level algorithm (see step 9 and 10). The inputs of the algorithm are the original template and reference images $T^{\text{orig}}, R^{\text{orig}}$, a minimal (coarsest) and maximal (finest) level l_{\min}, l_{\max} and for each level a smoothing factor for the approximating spline representation of the images θ^l .

Algorithm 6.2. (Multilevel Image Registration (MLIR))

1. Set $l = l_{\min}$ (coarsest resolution)
2. Set $m^l = (2^l, 2^l, 2^l); x^l = x^{cc}(\Omega, m^l)$
3. $(R_{\text{coeff}}^l, T_{\text{coeff}}^l) = \text{computeInterCoeffMultiScale}(R^{\text{orig}}, T^{\text{orig}}, \theta^l)$
4. $(R_0^l, T_0^l) = \text{Restrict}(R_{\text{coeff}}^l, T_{\text{coeff}}^l, l)$ (contains interpolation)
5. $\tilde{T}_{\text{coeff}}^l = \text{computeInterCoeff}(T_0^l)$
6. If $l = l_{\min}$ then $\mathbf{y}_0^l = y(\mathbf{x}^l)$
7. If $l \neq l_{\min}$ then $\mathbf{y}_0^l = \mathbf{x}^l + \text{Prolongate}(\mathbf{y}_{\text{opt}}^{l-1} - \mathbf{x}^{l-1})$
8. $\mathbf{y}_{\text{ref}}^l = \mathbf{x}^l$ (or determined for given parametric ref. transformation).
9. $\text{JobjFctn.init}(\mathbf{x}^l, R_0^l, \tilde{T}_{\text{coeff}}^l, \mathbf{y}_{\text{ref}}^l)$
10. $\mathbf{y}_{\text{opt}}^l = \text{Gauß-Newton}(\text{JobjFctn}, \mathbf{y}_0^l)$

11. If $l = l_{\max}$ stop, else increment $l = l + 1$ and go back to step 2.

If a staggered grid for the regularizer is needed in step 2, a staggered grid is generated $\mathbf{x}^l = \mathbf{x}^{stg}(\Omega, m^l)$. Step 7 is changed to an appropriate prolongation step for staggered grids (Sec. 4.4.2). The staggered grid and the staggered grid based transformation are interpolated onto a corresponding cell-centered grid resp. cell-centered grid based transformation (see operator P defined at the end of Sec. 6.4) for the evaluation of the distance measure. For the regularizer the staggered grid based transformation is used.

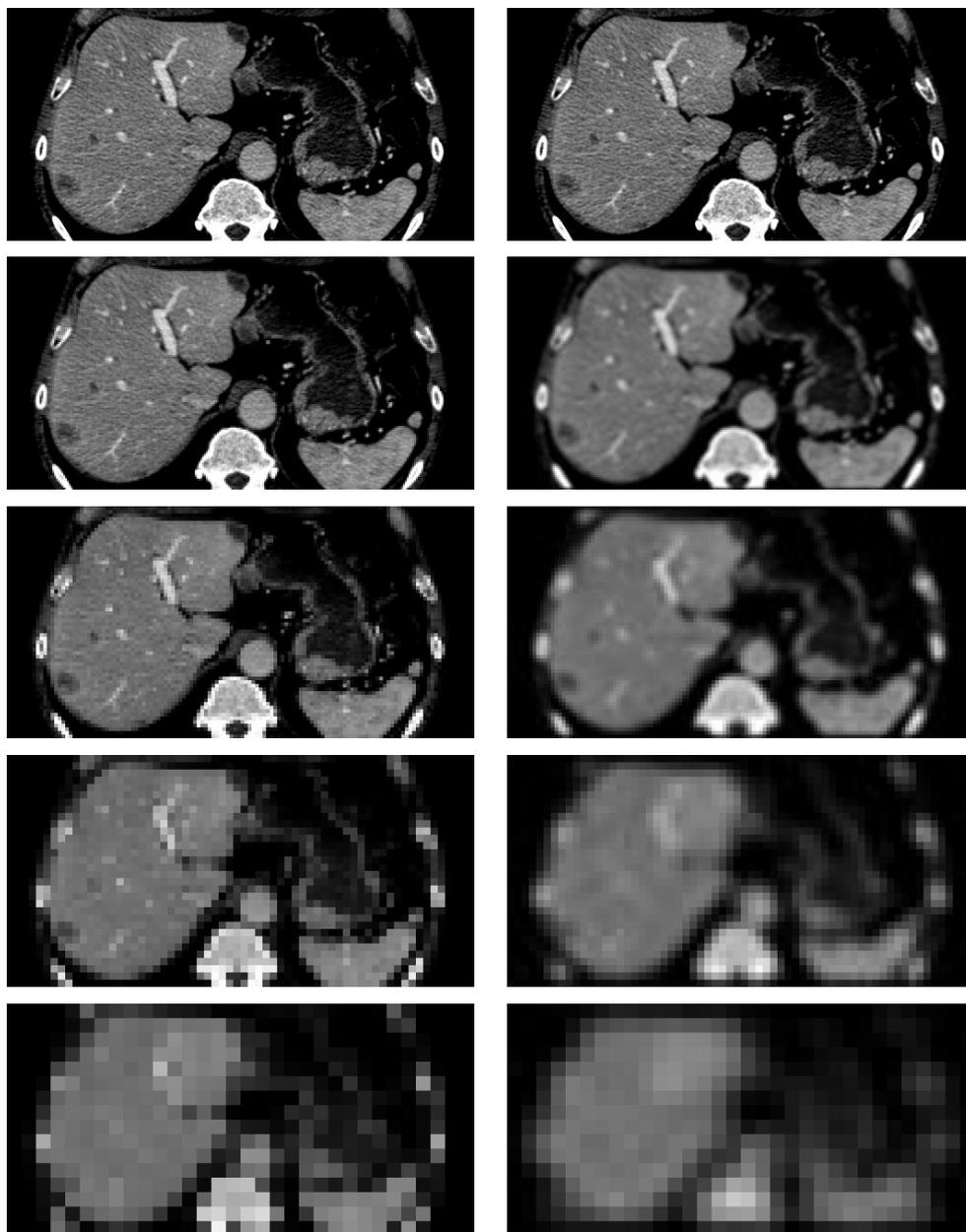


Figure 6.3 The first column shows one 2D slice of the liver with different resolutions based on the original image. In the second column first images on different scales with $\theta = 0; 10; 100; 1000; 10,000$ are generated and then the resolution is reduced to the same levels as in the first column. The levels $l = 8, 7, 6, 5, 4$ with $(2 \cdot 2^l) \times 2^l$ voxel are shown. The first row contains in both columns the original image for $\theta = 0$ and the original resolution 512×256 .

Chapter 7

Parametric Image Registration

7.1 Introduction

Instead of characterizing the registration transformation y by a smoothing functional the space of admissible transformations can be defined directly. The registration functional is then optimized over this space. Usually the function spaces used for image registration are parameterizable via a linear combination of some basis functions. We already defined the parameterized transformation y (see Sec. 3.2):

$$y(\gamma, x) = Q(x)\gamma \quad \text{with} \quad Q \in \mathbb{R}^{3 \times p}, \gamma \in \mathbb{R}^p \quad (7.1)$$

where $p \in \mathbb{N}$ is the number of basis functions. Now the registration functional for the parametric registration problem depends on the parameters $\gamma \in \mathbb{R}^p$ and not directly on the transformation y (see Section 3.2). Image-based parametric registration will not be a part of this thesis but can be found in [170, 150]. Nevertheless parametric transformations will be presented because they are needed for landmark-based registration schemes discussed in Chapter 11. The parametric transformations are defined on a collection of points which might be a regular grid but also a list of landmark points.

Typical low-dimensional function spaces are the spaces of rigid and affine transformations. High-dimensional spaces allowing local deformations are usually spline-based transformations where the basis functions are defined according to irregular points (Thin-plate Splines) or on regular grids (B-Splines). Usually intensity-based schemes are defined on regular grids of control points and feature-based schemes on irregular points located at the features. In this thesis, we will only use spline-based transformations on irregular points.

7.2 Affine-Linear Transformations

We start with affine-linear transformations because they are quite easy to define and to treat numerically. No constraints on the linear mapping like for

rigid transformations are needed. Compared to rigid transformations, affine-linear transformations can also have different scalings in different spatial directions and shearing. An affine-linear transformation is defined as

$$\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{v} \quad (7.2)$$

with an arbitrary matrix $\mathbf{A} \in \mathbb{R}^3 \times \mathbb{R}^3$ and $\mathbf{v}, \mathbf{x} \in \mathbb{R}^3$. We assume \mathbf{A} to be full-rank to exclude projections. The 9 degrees of freedom (number of parameters) of \mathbf{A} describe three rotations, three scalings and three shearings one for each coordinate direction. The vector \mathbf{v} describes a translation.

For the minimization of a given registration functional $J[\mathbf{y}]$ the optimal parameters \mathbf{A} and \mathbf{v} have to be computed. To be able to use the generic optimization framework defined in Chapter 5 we need one common parameter vector $\boldsymbol{\gamma}$ for the matrix \mathbf{A} and the vector \mathbf{v} . Therefore we show how an affine-linear transformation can be rewritten to fulfill this requirement. Each matrix and translation vector element is assigned an element of the 12-dimensional parameter vector $\boldsymbol{\gamma} \in \mathbb{R}^{12}$.

The discrete transformation $\mathbf{y} \in \mathbb{R}^{3n}$ is defined on the same grid as the discrete reference image. Thus let $\bar{\mathbf{x}} = \bar{\mathbf{x}}^{cc}(\Omega, m)$ be a cell-centered grid in matrix representation and the number of grid points be $n = m_1 \cdot m_2 \cdot m_3$. As mentioned in the introduction all transformations can also be defined on an arbitrary collection of points instead of the cell-centered grid, e.g., a list of landmark points. For one affine linearly transformed grid point (or landmark) $x_j = \bar{\mathbf{x}}_{j,:}$: (one row of matrix) holds:

$$y(x_j) = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_5 & \gamma_6 & \gamma_7 \\ \gamma_9 & \gamma_{10} & \gamma_{11} \end{pmatrix} x_j^T + \begin{pmatrix} \gamma_4 \\ \gamma_8 \\ \gamma_{12} \end{pmatrix} = \begin{pmatrix} x_j, 1 & & \\ & x_j, 1 & \\ & & x_j, 1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{12} \end{pmatrix}. \quad (7.3)$$

The last term allows to write all transformed vectors $y(x_j)$ in one big vector in the special ordering (first all coordinates in x -, than in y - and z -direction). Therefore let the matrix $\mathbf{Q} \in \mathbb{R}^{3n \times 12}$ be defined as:

$$\mathbf{Q}(\bar{\mathbf{x}}) = \mathbf{I}_3 \otimes (\bar{\mathbf{x}}, \mathbf{1}_n) = \begin{pmatrix} \bar{\mathbf{x}}, \mathbf{1}_n & & \\ & \bar{\mathbf{x}}, \mathbf{1}_n & \\ & & \bar{\mathbf{x}}, \mathbf{1}_n \end{pmatrix} \quad (7.4)$$

with $\mathbf{1}_n$ being a n -dimensional vector where all entries are 1. For ease of notation, we will replace $\mathbf{Q}(\bar{\mathbf{x}})$ by $\mathbf{Q}(\mathbf{x})$ in the following formulas. The matrix $(\bar{\mathbf{x}})$ and long vector (\mathbf{x}) notations of the cell-centered grid can easily be converted into each other (see Sec. 4.2.4).

Now we can express an arbitrary discrete affine-linear transformation $\mathbf{y}^{\text{aff}} \in \mathbb{R}^{3n}$ on a discrete grid $\mathbf{x} = \mathbf{x}^{cc}(\Omega, m)$ by a parameter vector $\boldsymbol{\gamma}^{\text{aff}} \in \mathbb{R}^{12}$ using the matrix $\mathbf{Q}(\mathbf{x})$:

$$\mathbf{y}^{\text{aff}}(\boldsymbol{\gamma}^{\text{aff}}, \mathbf{x}) = \mathbf{Q}(\mathbf{x})\boldsymbol{\gamma}^{\text{aff}}. \quad (7.5)$$

For efficient optimization schemes the derivative of the discrete affine-linear transformation \mathbf{y}^{aff} regarding the parameters γ^{aff} is needed:

$$d_{\gamma}\mathbf{y} : \mathbb{R}^{12} \rightarrow \mathbb{R}^{n \times 12} \quad (7.6)$$

$$\gamma \mapsto d_{\gamma}\mathbf{y} = Q \quad (7.7)$$

7.3 Rigid Transformations

Rigid transformations play an important role in computer assisted surgery. For example, the mechanical properties of bones can be well approximated as rigid bodies. This means they are translated and rotated, but their shape and volume do not change. The rigidity property of the rotation can be described by an orthogonal matrix $Q^{\text{rot}} \in \mathbb{R}^{3 \times 3}$ with $\det(Q^{\text{rot}}) = 1$. Together with the translation vector $v \in \mathbb{R}^3$ a rigid transformation $y^{\text{rig}} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ on a point $x \in \mathbb{R}^3$ is defined by:

$$y^{\text{rig}}(x) = Q^{\text{rot}}x + v. \quad (7.8)$$

The main difficulty with rigid registrations is the treatment of the rotation. Many ways exist to describe rotations mathematically: rotation matrices, one axis and one angle, unit quaternions and Euler angles. These representations have different characterizations concerning ease of use, numerical stability and compactness. The already described matrix representation has the disadvantage that the orthogonality constraint has always to be considered. The matrix has 9 elements, but there are only 3 degrees of freedom. Leonhard Euler investigated the properties of rigid transformations and found two important representations. Every rigid transformation can be represented by an axis and an angle around this axis. Euler proved this in his work in 1776: *Formulae generales pro translatione quacunque corporum rigidorum* (General formulas for the translation of arbitrary rigid bodies). The original citation in Latin is:

Quomodocunque sphaera circa centrum suum conuertatur, semper assignari potest diameter, cuius directio in situ translato conueniat cum situ initiali.

This can be translated as: ‘In whatever way a sphere is turned about its center, it is always possible to assign a diameter, whose direction in the translated state agrees with that of the initial state.’ This representation possesses 4 parameters: 3 for the axis and 1 for the angle. Quaternions are a closely related representation which are usually used today due to the reduction of necessary calculations by the quaternion algebra and numerical precision. Nevertheless we use Euler angles because they are a very compact parameterization (3 instead of 4 parameters), which leads to a smaller search space for optimization. Leonhard Euler introduced what we now call Euler angles in his book:

rum una in ipfo axe ID capiatur. Dari igitur affumo relationem inter coordinatas $IX = x$, $XY = y$ et $YZ = z$, quarum prima in ipfa recta IR, fecunda in plano ad vim normali, ac tertia ipfi vi Rr parallela capiatur. Ex Y primo ad IA perpendicularis YX' ducatur, in plano autem ad tabulam normali AID perpendicularis $X'y$ ipfi YZ et yZ ipfi $X'Y$ parallela, erit ut ante vidimus:

$IX' = x \cos \eta - y \sin \eta$; $X'y = YZ = z$; $X'Y = yZ = x \sin \eta + y \cos \eta$.
Tum in plano normali ex y ad ID ducatur perpendicularis yx , et habebuntur novae coordinatae, quales desideramus, quae sint $Ix = X$; $xy = Y$ et $yZ = Z$, atque ita per praecedentes determinantur.

$X = x \cos \eta \cos \theta - y \sin \eta \cos \theta + z \sin \theta$; $Y = z \cos \theta - x \cos \eta \sin \theta + y \sin \theta$; $Z = x \sin \eta + y \cos \eta$.

Figure 7.1 Text part of the original publication of Euler's: 'Theory of the motion of solid or rigid bodies'. The text part describes the Euler angles.

Theroria motus corporum solidorum seu rigidorum [53], Chapter IX, De prima motus generatione in corporibus rigidis, problema 59, Theory of the motion of solid or rigid bodies. Part of the original publication is shown in Fig. 7.1. According to Euler's Rotation theorem, any rotation can be represented by three angles, which are now called Euler angles. These Euler angles describe three subsequent rotations around three axes. Since rotations do not commute, the order in which they are applied is important. There are at least 24 standard Euler angles conventions [217] depending on the chosen axes, the order of the applied rotations and the extrinsic or intrinsic interpretation of the angles (frame of reference). We use a convention where the axes are in right-hand rule order, the rotations are first around the z-axis, then around the y-axis and then around the x-axis and the reference coordinate system is fixed and not rotating (extrinsic). The range of the angle around the y-axis has to be restricted to $[0, \pi]$ or $[-\pi/2, \pi/2]$. The rotations around the x-,y-,z-axis can be described by matrices depending on the corresponding Euler angle $\gamma_1, \gamma_2, \gamma_3$:

$$R_1(\gamma_1) = \begin{pmatrix} 1 & & \\ & c_1 & -s_1 \\ & s_1 & c_1 \end{pmatrix}, R_2(\gamma_2) = \begin{pmatrix} c_2 & & s_2 \\ & 1 & \\ -s_2 & & c_2 \end{pmatrix}, R_3(\gamma_3) = \begin{pmatrix} c_3 & -s_3 & \\ s_3 & c_3 & \\ & & 1 \end{pmatrix}$$

with the abbreviations: $c_j = \cos \gamma_j, s_j = \sin \gamma_j, j = 1, 2, 3$. Be aware that for the rotation about the y-axis the signs of the sines have been changed such that the right-hand rule holds. Applying the rotations about the axes subsequently is the same as multiplying the corresponding matrices leading to the rotation matrix:

$$\begin{aligned} Q^{\text{rot}}(\gamma_1, \gamma_2, \gamma_3) &= R_1(\gamma_1)R_2(\gamma_2)R_3(\gamma_3) \\ &= \begin{pmatrix} c_2c_3 & -c_2s_3 & s_2 \\ s_1s_2c_3 + c_1s_3 & -s_1s_2s_3 + c_1c_3 & -s_1c_2 \\ -c_1s_2c_3 + s_1s_3 & c_1s_2s_3 + s_1c_3 & c_1c_2 \end{pmatrix} \end{aligned}$$

We are searching for a parameterization of the rigid transformation based on one parameter vector $\gamma^{\text{rig}} = (\gamma_1, \dots, \gamma_6) \in \mathbb{R}^6$ and not a rotation matrix and a translation vector. The first three parameters $\gamma_1, \gamma_2, \gamma_3$ represent the Euler angles and the second three $\gamma_4, \gamma_5, \gamma_6$ the translations in x-,y-,z-direction.

We define the non-linear function $q : \mathbb{R}^6 \rightarrow \mathbb{R}^{12}$ which maps the parameter vector γ^{rig} to the (row-wise) entries of the rotation matrix Q^{rot} and keeps the translation entries:

$$q(\gamma^{\text{rig}}) = \begin{pmatrix} c_2 c_3 \\ -c_2 s_3 \\ -s_2 \\ \gamma_4 \\ -s_1 s_2 c_3 + c_1 s_3 \\ s_1 s_2 s_3 + c_1 c_3 \\ -s_1 c_2 \\ \gamma_5 \\ s_1 s_3 + c_1 s_2 c_3 \\ s_1 c_3 - c_1 s_2 s_3 \\ c_1 c_2 \\ \gamma_6 \end{pmatrix}. \quad (7.9)$$

Now the rigid transformation \mathbf{y}^{rig} on a cell-centered grid \mathbf{x} can be written as:

$$\mathbf{y}^{\text{rig}}(\gamma^{\text{rig}}, \mathbf{x}) = Q(\mathbf{x})q(\gamma^{\text{rig}}) \quad (7.10)$$

where $Q(\mathbf{x}) = I_3 \otimes (\bar{x}, \mathbf{1}_n)$ is the same matrix as for the affine transformation. The significant difference to the affine transformation is the non-linearity introduced by the function q .

The derivative of the rigid transformation \mathbf{y}^{rig} with respect to the parameters γ^{rig} is:

$$d_\gamma \mathbf{y}^{\text{rig}}(\gamma, \mathbf{x}) = Q(\mathbf{x})d_\gamma q(\gamma) = Q(\mathbf{x}) \begin{pmatrix} \frac{\partial q_1}{\partial \gamma_1} & \dots & \frac{\partial q_1}{\partial \gamma_6} \\ \vdots & \ddots & \vdots \\ \frac{\partial q_{12}}{\partial \gamma_1} & \dots & \frac{\partial q_{12}}{\partial \gamma_6} \end{pmatrix}. \quad (7.11)$$

The explicit computation of the partial derivatives $\frac{\partial q_i}{\partial \gamma_j}$ is straight forward.

7.4 Spline-based Transformations

The degrees of freedom of a rigid or an affine transformation are very low: 6 for the rigid and 12 for the affine transformation in 3D space. Thus only global transformation effects can be modeled. For the consideration of local deformations a larger number of parameters is necessary. Splines are an efficient and flexible possibility to get a theoretically arbitrarily high number of

degrees of freedom. The spline coefficients are the parameters of a continuous parametric transformation. For each coordinate direction one spline-based transformation function is defined. The spline basis functions are either given on a regular grid (one basis function for each grid point) or on irregularly distributed points. The first variant using spline functions as defined in section 4.3.2 is quite popular in image based parametric non-rigid registration. Details can be found in [199, 109, 192] but will not be described further in this thesis. The second variant is usually used for landmark-based non-rigid registration schemes. Most popular is the thin-plate spline approach, which will be explained in more detail in Section 11.8. An interesting alternative to thin-plate splines are Gaussian elastic body splines which will also be discussed in section 11.9.

Part II

Modeling Prior Knowledge in Image Registration

Chapter 8

Review of Distance Measures

8.1 Introduction

Distance measures are a very important building block of rigid as well as non-rigid intensity-based registration methods. There are many different distance measures and their number is still increasing. We will explain the main principles of the most important distance measures. Most of the other measures are more or less variations of these basic measures. They differ mainly in the kind of dependency they assume between the intensities of the reference and template image: identity, linear, functional or stochastic. Another important aspect, in particular for non-rigid registration, is whether a measure evaluates the intensity values globally or compares a local neighborhood for each image point.

What is considered a reasonable choice for a distance measure depends highly on the given application. Although some distance measures can be used in most applications, e.g., mutual information, they all have some advantages and disadvantages for particular applications, which will be discussed. Instead of choosing suitable general distance measures, specifically designed distance measures for given applications might lead to better registration results. Specific distance measures can take a priori information about the imaging modalities or even on the image content like geometrical properties of imaged structures into account.

Existing general and specific distance measures are reviewed and new measures incorporating shape information on image contents are introduced. We will only consider distance measures on scalar-valued images, no vector-valued (like color images) or tensor-valued images (like diffusion tensor images). A general introduction into distance measures, their discretization and derivatives is given in section 6.2.

8.2 Properties of Distance Measures

The performance of a distance measure depends on the image classes of the reference and template image. Thus an essential classification of distance measures is their suitability for images from the same (*monomodal*) or different (*multimodal*), e.g., CT and MRI, imaging modalities. Another important property of a distance measure is the *kind of dependency* assumed between template and reference image: identity, linear, functional, statistical. A distance measure can be based directly on the intensity values of the images or on their *morphology*, for example based on gradient information or other differential operators. A further discrimination of distance measures is their *locality*. Often for each image point in the reference image some difference to the template image is computed and then we integrate over all differences. There are distance measures which use instead of single points a whole local neighborhood. There will be an own section about such local distance measures.

Besides the properties of a distance measure concerning the *image contents* also the *efficiency* of a measure is important because in each iteration step of the optimizer the distance measure and its derivative have to be computed. This means the distance measure is part of the objective function and hence should be also *differentiable*. As we use the Gauß-Newton resp. the generalized Gauß-Newton optimization algorithm for all registration problems in this thesis we only need first order derivatives of the residual function and second order derivatives of the outer function. The special form of the (discrete) distances measures considered here is: $D(\mathbf{y}) = \phi(r(\mathbf{y}))$ (see Sec. 6.2). We will define the outer and inner functions ϕ and r for all distance measures as well as their first derivatives $d\phi, dr$ and the second derivative $d^2\phi$ of the generally simple and smooth outer function ϕ .

8.3 General Purpose Distance Measures

Roche et al. [188] classified the most important general purpose measures except morphological measures by the implicit assumptions on which the measures rely on. They enumerate different specific functional relationships between the intensities of template and reference image: identity, affine-linear, functional and statistical. The functional dependency $f : \mathbb{R} \rightarrow \mathbb{R}$ of the image intensities of the reference \mathcal{R} and template \mathcal{T} for the ideal transformation y^{ideal} is formalized as

$$\mathcal{R}(x) = f(\mathcal{T}(y^{\text{ideal}}(x))) + \epsilon(x) \tag{8.1}$$

where ϵ is some additive noise.

In addition Roche et al. [188] casted the search for an optimal distance measure for a class of images into a maximum likelihood estimation problem.

The image intensities are treated as random variables. They derive two popular measures (correlation ratio and mutual information) from the modeling assumptions on the images in the maximum likelihood framework. Wein [248] showed that also the normalized cross correlation is consistent with the maximum likelihood framework. This important approach offers the possibility to make the implicit assumptions explicit and helps to understand which distance measure is the most suitable for a class of images.

In the following we will present distance measures for the most important functional dependencies f . The formal derivation of the according distance measures from the functional dependencies is left to the paper of Roche et al. [188].

8.3.1 Identity Relationship

If the reference and template image differ only by Gaussian noise, the optimal measure in the likelihood sense is the sum of squared differences (SSD) measure already defined in Section 6.2. Also the discretization and the residual and outer function are given there. It is a simple measure which can be computed very efficiently.

The assumed functional dependency of the reference and template image intensities is the identity ($f = \text{id}$) leading to:

$$\mathcal{R} \approx f(\mathcal{T}(y^{\text{ideal}})) = \text{id}(\mathcal{T}(y^{\text{ideal}})) = \mathcal{T} \quad (8.2)$$

for the ideal transformation y^{ideal} . This assumption is very restrictive in the sense that the SSD measure is only suitable for monomodal registration problems. Even for images from the same modality the intensity variations of different acquisitions can be too high.

One advantage of the SSD measure is the local error contribution of every single voxel pair compared to the mutual information measure which will be explained in Section 8.3.4. Thus the SSD measure is particularly suitable for non-rigid registration problems.

One of the disadvantages of the SSD measure is its sensitivity to outliers. To reduce this sensitivity outer functions from robust statistics such as $\phi(x) = \frac{x^2}{\sigma^2 + x^2}$ with a sensitivity parameter σ can be used (see for example [190]). In robust statistics, estimators are investigated which are not unduly affected by small deviations from model assumptions like the assumption of Gaussian distributed image noise for the SSD measure.

8.3.2 Affine-Linear Relationship

If the contrast α and/or the brightness β between the reference and template image is different there is no identity relationship between the intensity values

anymore, but the intensities might still be affine-linearly related:

$$\mathcal{R} \approx f(\mathcal{T}(y^{\text{ideal}})) = \alpha\mathcal{T}(y^{\text{ideal}}) + \beta. \quad (8.3)$$

The according optimal distance measure in the maximum likelihood sense [248] is based on the normalized cross-correlation (NCC) of reference and template image, also denoted as correlation coefficient:

$$\text{NCC}[\mathcal{T}, \mathcal{R}] = \frac{\text{Cov}(\mathcal{T}, \mathcal{R})}{\sqrt{\text{Var}(\mathcal{T})}\sqrt{\text{Var}(\mathcal{R})}} \quad (8.4)$$

with the expectation value $E(\mathcal{T}) = |\Omega|^{-1} \int_{\Omega} \mathcal{T}(x) dx$, the variance $\text{Var}(\mathcal{T}) = E((\mathcal{T} - E(\mathcal{T}))^2)$ and the covariance $\text{Cov}(\mathcal{T}, \mathcal{R}) = E((\mathcal{T} - E(\mathcal{T}))^T(\mathcal{R} - E(\mathcal{R})))$.

The value of the normalized cross-correlation is between -1 and 1. Its absolute value is high if the images are well registered and low if not. If the sign is negative the intensities in one image are inverted compared to the other images. As we like to get a minimum for well registered images we define the (squared) normalized cross-correlation distance measure as:

$$\mathcal{D}^{\text{NCC}}[\mathcal{T}, \mathcal{R}, y] = 1 - \text{NCC}^2[\mathcal{T}(y), \mathcal{R}]. \quad (8.5)$$

The NCC distance measure is very useful for monomodal images but hardly for multi-modal images. But if it is computed locally for each image point it is also applicable to multi-modal image pairs (see Sec. 8.3.6).

Discretization and Derivatives

The discretization of \mathcal{D}^{NCC} is straightforward based on the numerical integration scheme presented in Section 6.2. Let R and T_y be the discretized reference and transformed template image on the same grid with n grid points. Then the expectation value $E(\mathcal{R})$ can simply be approximated by the mean of the intensities on the grid points:

$$E(\mathcal{R}) \approx \frac{1}{n} \sum_{i=1}^n R_i =: \bar{R}. \quad (8.6)$$

Let be $\hat{R} := (R - \bar{R}) / \|R - \bar{R}\|$ and $\hat{T}_y := (T_y - \bar{T}) / \|T_y - \bar{T}\|$ the normalized reference and template image, then the NCC can be discretized by:

$$\text{NCC}(\mathcal{T}_y, \mathcal{R}) \approx \hat{R}^T \hat{T}_y := \text{NCC}(T_y, R). \quad (8.7)$$

It follows the discrete version of the NCC distance measure:

$$\mathcal{D}^{\text{NCC}}(T_y, R) = 1 - \text{NCC}^2(T_y, R). \quad (8.8)$$

Now we define the residual function r and the outer function ϕ for the use of the distance measure in a Gauß-Newton optimization scheme:

$$r(\mathbf{y}) = T_{\mathbf{y}} - \bar{T}, \quad \phi(r) = 1 - \frac{(r^T \bar{R})^2}{r^T r} \quad (8.9)$$

where the normalized reference image \hat{R} has to be computed only once at the beginning of the optimization process.

The derivatives r and ϕ needed for the Gauß-Newton optimization scheme are:

$$dr(\mathbf{y}) = dT_{\mathbf{y}} - \mathbf{1}_n, \quad d\phi = -2 \frac{\hat{R}^T r}{r^T r} \hat{R}^T + 2 \left(\frac{\hat{R}^T r}{r^T r} \right)^2 r^T, \quad d^2\phi = \frac{2}{r^T r}. \quad (8.10)$$

8.3.3 General Functional Relationship

In the previous section we assumed a linear relation between the intensity values of the reference and template image. Roche et al. [189] introduced a distance measure which is capable of handling also a general (but unknown) functional dependency f between the intensity values of reference and template image:

$$\mathcal{R} \approx f(\mathcal{T}(y^{\text{ideal}})). \quad (8.11)$$

The idea of the distance measure is that in well registered images all image points with a certain intensity in the template image are mapped to image points with a corresponding (possibly very different) intensity in the reference image. But this holds only if the images are correctly registered. Thus, the degree of functional dependence is a possible distance measure for registration.

The functional dependence can be measured by the correlation ratio distance measure [189]:

$$\mathcal{D}^{\text{CR}}[\mathcal{T}(y), \mathcal{R}] = 1 - \eta(\mathcal{R} | \mathcal{T}(y)) = 1 - \frac{\text{Var}(E(\mathcal{R} | \mathcal{T}(y)))}{\text{Var}(\mathcal{R})}. \quad (8.12)$$

If \mathcal{R} is completely independent of \mathcal{T} , then the expectation $E(\mathcal{R} | \mathcal{T}(y))$ is constant and its variance is zero. Thus η is zero and \mathcal{D}^{CR} is one. If every value of \mathcal{R} can be predicted from $\mathcal{T}(y)$ (complete functional dependency), the expectation is $E(\mathcal{R} | \mathcal{T}(y)) = \mathcal{R}$ resulting in $\eta = 1$ and \mathcal{D}^{CR} is zero. Note that the Correlation Ratio is not symmetric because $\eta(\mathcal{R} | \mathcal{T}(y)) \neq \eta(\mathcal{T}(y) | \mathcal{R})$. The correlation ratio distance measure is not very popular and we do not use it in this thesis. Therefore we refer to Roche et al. [189] for its discretization and implementation.

8.3.4 Statistical Relationship

The assumption of an explicit intensity mapping f as in the previous section is often too restrictive for multi-modal image registration. A more general approach is to assume only statistical dependencies between the intensity values of the reference and template image. A very popular distance measure suitable for this general assumption is based on mutual information MI. This measure from information theory was introduced into the field of image registration both by Collignon et al. [37] and Viola and Wells [239, 240, 251]. Plum et al. [178] give a comprehensive survey on mutual information used for medical image registration methods.

The main idea of mutual information is that accurately registered images minimize the amount of information in a shared representation of the images. If the intensity values of the reference image can be well explained by the intensity values of the registered template image, the information content of a shared representation is as low as for the reference image alone.

A very common measure for the information content of an image $\mathcal{T} : \Omega \rightarrow \mathbb{R}$ on $\Omega \subset \mathbb{R}^3$ is the Shannon entropy. Therefore let the intensity value distribution of the image be described by the continuous probability density $\rho_{\mathcal{T}} : \mathbb{R} \rightarrow \mathbb{R}$ with $\rho_{\mathcal{T}} \geq 0$ and $\int_{\mathbb{R}} \rho_{\mathcal{T}}(t)dt = 1$. Then the Shannon entropy H of the density is defined by:

$$H[\rho_{\mathcal{T}}] = - \int_{\mathbb{R}} \rho_{\mathcal{T}}(t) \log(\rho_{\mathcal{T}}(t)) dt. \quad (8.13)$$

An image with almost a single intensity value has a very low entropy value ($H = 0$ for a constant image). It contains very little information. An image with many different intensities, which occur more or less equally in the image yields a high entropy value. The image contains a high amount of information. Hence, the entropy is also a measure for the dispersion of a probability distribution. If the distribution has a single sharp peak, the entropy is low. A dispersed distribution yields a high entropy value.

The same properties hold for the entropy of the joint density $\rho_{\mathcal{T},\mathcal{R}} : \mathbb{R}^2 \rightarrow \mathbb{R}$ of the reference and template image intensity pairs $(t, r) \in \mathbb{R}^2$:

$$H[\rho_{\mathcal{T},\mathcal{R}}] = \int_{\mathbb{R}^2} \rho_{\mathcal{T},\mathcal{R}}(t, r) \log(\rho_{\mathcal{T},\mathcal{R}}(t, r)) d(t, r). \quad (8.14)$$

If we find a transformation that minimizes the joint entropy, the two images should be well aligned. But the joint entropy also decreases, if the entropy of one of the images decreases. Therefore also the information content of the single images should be considered in a distance measure. The mutual information relates the information content of the single images to the joint information content of both images:

$$\text{MI}[\rho_{\mathcal{R}}, \rho_{\mathcal{T}}] = H[\rho_{\mathcal{R}}] + H[\rho_{\mathcal{T}}] - H[\rho_{\mathcal{T},\mathcal{R}}]. \quad (8.15)$$

The mutual information is zero, if the images are stochastically independent. Therefore we use the negative mutual information for the definition of the distance measure to get a minimum for well aligned images:

$$\mathcal{D}^{\text{MI}}[\mathcal{R}, \mathcal{T}, y] = -\text{MI}[\rho_{\mathcal{R}}, \rho_{\mathcal{T}(y)}]. \quad (8.16)$$

Implementation

The main challenge for the implementation of mutual information is the determination of the joint probability distribution $\rho_{\mathcal{R}, \mathcal{T}}$. There are two main methods for estimating the joint (and also the marginal) probability densities: histogram-based [191] and Parzen-window-based [240] estimation. The use of histograms is widespread because they can easily be implemented [191, 178] by sorting intensity values into a number of bins. The problem with histogram estimators is that they depend on several parameters, e.g., number of bins, bin width, and lead to a non-differentiable distance measure. The idea of the alternative Parzen-window-based estimator is to approximate the intensity value distribution by the sum of shifted copies of a continuous and compactly supported kernel function, e.g., a cubic spline. In contrast to the histogram-based estimator the Parzen-window estimator [240] is smooth and differentiable, but it depends also on different parameters. The width of the kernel is comparable to the bin width of a histogram. In addition the number of kernel copies and the number of image samples is crucial and lead to different estimations.

Based on the estimation of $\rho_{\mathcal{R}, \mathcal{T}}$ the integral in Eq. 8.14 is approximated by a midpoint quadrature rule (see Sec. 6.2.1). A detailed derivation of the discretization, the reformulation of the mutual information measure into a residual and an outer function and the derivatives can be found in [170].

Besides the mentioned implementational difficulties a main drawback of the mutual information distance measure is that the dependence of intensity values of neighboring voxels is ignored. Only the intensity *distributions* of reference and template image are considered, not the intensity *differences* for each single image point as the SSD measure does. Pluim et al. [177] try to improve this weakness of mutual information by combining mutual information with gradient information. This leads us to morphological distance measures.

8.3.5 Morphological Measures

The aim of morphological (shape and structure) distance measures is to get independent of the actual intensities which might vary across the image domain or between different modalities for the same anatomical structures. This means the focus is on image features and structures (image morphology) and not on intensities. The idea is that image contents are mainly represented by intensity changes. Thus morphological distance measures are suitable for multimodal image registration.

The image gradient is a simple morphological feature measuring the magnitude and direction of intensity changes representing edges of objects. As the magnitude depends on the modality only the direction information should be used to define a morphological distance measure. Therefore Haber and Modersitzki [80] introduce the normalized gradient field (NGF) $n[\mathcal{T}]$ of an image \mathcal{T} :

$$n[\mathcal{T}, \eta] = \frac{\nabla \mathcal{T}}{\sqrt{|\nabla \mathcal{T}|^2 + \xi^2}} \quad (8.17)$$

with an edge sensitivity parameter $\xi \in \mathbb{R}^+$. Ignoring ξ for a moment the normalization leads to gradient vectors of length one. The strength of the intensity change does not matter anymore. The problem is that due to noise even in regions with constant intensity small gradient vectors occur which are scaled up to length one by the normalization. The parameter ξ determines which gradients are considered as an edge ($|\mathcal{T}| > \xi$) and which gradients are considered resulting from noise ($|\mathcal{T}| \leq \xi$).

For well registered images the intensity changes approximately point in the same direction and so do the normalized gradients: $n[\mathcal{T}](x) \approx \pm n[\mathcal{R}](x)$. In this case the gradients are maximally linear dependent (zero angle). The scalar product quantifies the linear dependency of the gradients. If the gradients are linear dependent, the scalar product is one, if the gradients are perpendicular to each other the scalar product is zero. The scalar product is also zero, if one of the normalized gradients is zero. In a registration process the aim is to maximize the scalar product or to minimize its negative value. To get positive values we add additionally a one:

$$d(n[\mathcal{T}](x), n[\mathcal{R}](x)) = 1 - ((n[\mathcal{T}](x))^T n[\mathcal{R}](x))^2. \quad (8.18)$$

If we integrate this measure over all image points we get the NGF distance measure:

$$\mathcal{D}^{\text{NGF}}[\mathcal{T}, \mathcal{R}, y] = \int_{\Omega} d(n[\mathcal{T}(y)](x), n[\mathcal{R}](x)) \, dx. \quad (8.19)$$

A detailed derivation of the residual r and outer function ϕ , the discretization and the derivatives of the NGF measure can be found in [80, 150, 170].

Also distance measures based on more complex differential operators than the gradient operator have been published. Maintz et al. [140] generate feature images from the original reference resp. template image and determine the cross correlation of the feature images. They determine the feature images by fuzzy edgeness and rigdeness operators. Holden et al. [88] use multi-channel mutual information incorporating derivatives of linear scale space into the different channels. Droske et al. [48] introduce a distance measure based on very general morphological features considering the image level sets $\mathcal{M}[\mathcal{T}, c] = \{x \in \Omega | \mathcal{T}(x) = c\}$.

8.3.6 Local Distance Measures

Local distance measures are an alternative approach to morphological measures concerning intensity inhomogeneities across the image domain. They are based on the assumption that the intensity variation is low inside a local window region.

For example, one localized version of the mutual information distance measure has been published (see [178]). One problem with the local determination of mutual information is that the results can suffer from the small number of samples. It is a statistical measure and if the number of samples are too small, the statistics will be bad. Therefore relatively large subimages have to be used. A detailed discussion can be found in [178].

Here we will focus on the local correlation coefficient (LCC) distance measure, which will be used in Sec. 9.3 to incorporate shape information into the registration process. The LCC distance measure is even suitable for multimodality registration as explained by Weese et al. [247]. First, we define the local mean, local variance and local covariance at an image point $x \in \mathbb{R}^3$. The locality is reached by a convolution with a kernel function (also called a window or weighting function) $\mathcal{K} : \mathbb{R}^3 \rightarrow \mathbb{R}$. For example the local mean at $x \in \mathbb{R}^3$ for an image \mathcal{T} can be determined by:

$$E_{\mathcal{K}}(\mathcal{T})(x) = |\Omega|^{-1} \int_{\Omega} \mathcal{K}(s-x) \mathcal{T}(s) ds. \quad (8.20)$$

The kernel function can be a simple spherical (or cubic) binary function with $\mathcal{K}_{S_r}(x) = 1$ if $x \in S_r = \{x \mid \|x\| \leq r\}$ and 0 elsewhere, as in [247] or a Gaussian function, as in [25, 174]. The local version of the covariance of a reference \mathcal{R} and a template image \mathcal{T} at a point $x \in \mathbb{R}^3$ looks like:

$$\text{Cov}_{\mathcal{K}}(\mathcal{T}, \mathcal{R})(x) = |\Omega|^{-1} \int_{\Omega} \mathcal{K}(s-x) (\mathcal{T}(s) - E_{\mathcal{K}}(\mathcal{T})(x))^T (\mathcal{R}(s) - E_{\mathcal{K}}(\mathcal{R})(x)) ds. \quad (8.21)$$

Now with $\text{Var}_{\mathcal{K}}(\mathcal{R}) := \text{Cov}_{\mathcal{K}}(\mathcal{R}, \mathcal{R})$ the local correlation coefficient at $x \in \mathbb{R}^3$ can be defined as:

$$\text{NCC}_{\mathcal{K}}[\mathcal{T}, \mathcal{R}](x) = \frac{\text{Cov}_{\mathcal{K}}(\mathcal{T}, \mathcal{R})(x)}{\sqrt{\text{Var}_{\mathcal{K}}(\mathcal{T})(x)} \sqrt{\text{Var}_{\mathcal{K}}(\mathcal{R})(x)}} \quad (8.22)$$

and the (squared) local correlation distance measure as:

$$\mathcal{D}^{\text{LCC}}[\mathcal{R}, \mathcal{T}, y] = 1 - |\Omega|^{-1} \int_{\Omega} \text{NCC}_{\mathcal{K}}^2[\mathcal{T}(y), \mathcal{R}](x) dx. \quad (8.23)$$

The discretization is the same as for the global NCC distance measure \mathcal{D}^{NCC} (see Sec. 8.3.2). The only difference is the convolution and the additional

integral over all local correlation coefficients. Their discretization is straight forward, but the convolution leads to significantly higher computational costs compared to the global distance measure. To accelerate the computation Weese et al. [247] evaluate the local correlations only on 10% of the image volume and Cachier et al. [25] exploit the fact that a 3D Gaussian function can be separated into three 1D Gaussian functions for each spatial direction one.

8.4 Specific Distance Measures

Although general distance measures are successful in many applications, specific distance measures, which incorporate a priori knowledge about the imaging process or image contents, potentially improve registration results. In particular ultrasound imaging is a challenging modality for image registration. Any information about the imaging process and the image contents in a given application should be used to improve the registration process. We will give different examples of specific distance measures and lay a special focus on measures for ultrasound images.

8.4.1 Consideration of Image Content

Masking of image regions is a simple method to incorporate knowledge about unwanted image content or to enhance image content with low contrast. For example if we want to register an individual patient image data set containing a brain lesion to an atlas of healthy subjects the registration method might try to shrink the lesion to make the patient data look more similar to the healthy subjects. Masking the region of the lesion can improve the registration results.

Henn et al. [86] introduce a SSD measure which ignores an image region $A \subset \Omega$:

$$\mathcal{D}[y] = \frac{1}{2} \int_{\Omega \setminus A} (\mathcal{T}(y(x)) - R(x))^2 dx \quad (8.24)$$

$$= \frac{1}{2} \int_{\Omega} \chi_A(x) (\mathcal{T}(y(x)) - R(x))^2 dx \quad (8.25)$$

where $\chi_A(x)$ is the characteristic function of A :

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in \Omega \setminus A, \\ 0 & \text{if } x \in A. \end{cases} \quad (8.26)$$

A more general masking approach is formulated by Schuhmacher et al. [211, 212]. They introduce combinations of weighting masks instead of a single binary mask to enhance the intensities of low contrasted structures or to reduce intensities, for example to hide lesions or different rectum fillings.

Another possibility to incorporate structural knowledge about the image content into the distance measure are corresponding anatomical structures (landmarks) like points, curves, surfaces or volumes in reference and template image. Hömke [89] calls those corresponding structures shape constraints and introduces distance measures which are based on distance functions on the shapes. He adds the shape distance measure to the SSD intensity distance measure. We interpret such additional distance measures as penalizers which constrain the displacement field. In Chapter 10 we present such penalizers and direct possibilities to constrain the displacement field by corresponding structures.

Masking and shape constraints usually require the identification of masks resp. structures in template and reference image. In Chapter 9 we will introduce distance measures for the case where an explicit model of structures in one of the images is available. The distance measures compare the model with the intensity values of the other image.

8.4.2 Consideration of Imaging Characteristics

The incorporation of knowledge about the imaging process and the appearance of anatomical structures in the particular imaging modalities is a further direction for application specific distance measures. All presented specific distance measures are of a common type. For the definition of this distance measure type we introduce the general operators $\Phi_{\mathcal{R}}$ and $\Phi_{\mathcal{T}}$ which perform some processing on the reference $\mathcal{R} : \Omega \rightarrow \mathbb{R}$ resp. template image $\mathcal{T} : \Omega \rightarrow \mathbb{R}$ and yield vector-valued images $\Phi_{\mathcal{R}}[\mathcal{R}] : \Omega \rightarrow \mathbb{R}^n$ resp. $\Phi_{\mathcal{T}}[\mathcal{T}] : \Omega \rightarrow \mathbb{R}^m$ on the same domain $\Omega \subset \mathbb{R}^3$ as the original images. A suitable multi-signal distance measure \mathcal{D}^{MS} is applied to the resulting vector-valued images:

$$\mathcal{D}[\mathcal{R}, \mathcal{T}, y] = \mathcal{D}^{\text{MS}}[\Phi_{\mathcal{R}}[\mathcal{R}], \Phi_{\mathcal{T}}[\mathcal{T}(y)]]. \quad (8.27)$$

Now we will show different instances of this distance measure type.

Often imaging artifacts like MRI inhomogeneities or ultrasound shadowing disturb the registration process significantly. Different methods to compensate imaging artifacts for registration purposes have been published. Penney et al. [175] for example suggest shadow removal for ultrasound registration. In some cases also the image noise is reduced before registration, in particular speckle noise in ultrasound imaging ([130]; Sec. 2.4.1). The aim of such pre-processing operations is to make the application of general distance measures possible. The pre-processing operators yield scalar-valued images $\Phi_{\mathcal{R}}[\mathcal{R}], \Phi_{\mathcal{T}}[\mathcal{T}] : \Omega \rightarrow \mathbb{R}$ on which conventional (single-signal) distance measures can be applied. For one of the images the operator can be the identity operator $\Phi[\mathcal{T}] = \mathcal{T}$.

Besides specific pre-processing also the imaging characteristics of the modalities can be considered. For example, ultrasound imaging consists mainly of

tissue specific attenuations (speckle, tissue echogenicity) and reflections at tissue interfaces (see Sec. 2.4.1). MR images do not contain such reflections, but high gradient magnitudes indicate tissue interfaces. Therefore Roche et al. [190] combine intensity and gradient magnitude information for the registration of MRI and ultrasound. They use a bivariate correlation ratio based on a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ which maps the intensities and gradient magnitudes of the CT image \mathcal{T} to the intensities containing attenuations and reflections of the ultrasound image \mathcal{R} . In this case the operators are $\Phi_{\mathcal{T}}[\mathcal{T}] = (\mathcal{T}, \|\nabla\mathcal{T}\|)$ and $\Phi_{\mathcal{R}}[\mathcal{R}] = \mathcal{R}$.

A more general approach to compensate different appearances of the same anatomical structure in different modalities is the use of pseudo-modalities. Either both original modalities are converted to a common pseudo-modality [1, 175] or a pseudo-modality is simulated from one modality which looks similar to the other modality [236, 130, 248]. Then instead of a distance measure for multimodal images like mutual information a measure for single modality images like the cross correlation coefficient can be used.

An example for a common pseudo-modality is introduced by Penney et al. [175]. They convert the intensity values of an MRI image \mathcal{T} and an ultrasound image \mathcal{R} into vessel probability images $\Phi_{\mathcal{T}}[\mathcal{T}] = P_{\mathcal{T}} : \Omega \rightarrow [0, 1]$ resp. $\Phi_{\mathcal{R}}[\mathcal{R}] = P_{\mathcal{R}} : \Omega \rightarrow [0, 1]$ and register these probability images by means of the normalized cross correlation distance measure $\mathcal{D}^{\text{NCC}}[P_{\mathcal{R}}, P_{\mathcal{T}}]$.

On the other hand Wein et al. [248, 249, 250] simulate ultrasound images from CT (\mathcal{T}) to enable the registration of an ultrasound-like pseudo-modality with an original ultrasound image \mathcal{R} . The simulation is based on the assumption that the acoustic impedance of tissue is proportional to its density (and a constant speed of sound in the tissue). The tissue density is related to the intensity values of the CT (Hounsfield units). The strength of ultrasonic reflections $T_1 : \Omega \rightarrow \mathbb{R}$ at tissue interfaces including shadowing effects is simulated by integrating reflections and transmissions along each ultrasonic scanline. In addition the echogenicity $T_2 : \Omega \rightarrow \mathbb{R}$ is simulated by a heuristic mapping of a narrow CT soft tissue intensity range onto echogenicities. As they do not know how much reflections and echogenicities contribute to the ultrasound intensities the coefficients of a linear combination of both physical effects are implicitly estimated by a bivariate correlation ratio distance measure on $\Phi_{\mathcal{T}} = (T_1, T_2)$ and $\Phi_{\mathcal{R}} = \mathcal{R}$. Therefore they call their measure linear correlation of linear combinations (LC^2).

Chapter 9

Distance Measures Incorporating Shape Information

9.1 Introduction

In the previous chapter we gave two examples how a priori knowledge about image contents (masking and shape constraints) can be incorporated into a distance measure. Here we propose “image content aware” distance measures for the special case of registration of pre- to intraoperative data. This intraoperative registration problem is asymmetric: preoperative time constraints for preprocessing the data are not tight, yet intraoperative scheduling is. The idea is to preprocess the preoperative data to get a precise model of important structures with special shapes and then register this model with the raw intensity data of the intraoperative ultrasound. We put the focus on tube-like vessel structures, which are locally cylindrical, but structures like organ surfaces, which are locally plate-like, are also possible. Such distance measures which incorporate shape information, are often hybrid measures in the sense that features/shapes are compared to intensities.

Aylward et al. [6] proposed a hybrid approach of fitting extracted features (vessels) from preoperative data rigidly and directly to ultrasound intensity data. Our approach is similar to the one of Aylward et al. [6] and we will show how their approach can be formulated in our framework.

The aim of these hybrid approaches is to increase the robustness of the registration process by incorporating model knowledge. In addition, the registration process can be accelerated significantly by reducing the evaluation of a distance measure to the location of relevant structures. The aim of this chapter is to define distance measures which are suited to compare vessel models and ultrasound intensity data robustly and more efficiently than common intensity-based distance measures.

At first we will model vessels geometrically by their center lines and radii. We also generate intensity models of the vessels to get an image that only con-

tains the vessels. The first class of distance measures compares these intensity models with the intraoperative ultrasound data by means of local correlation. The idea of the second class of distance measures is to generate filter kernels adapted to the center lines model of the vessels and then to transform the other image such that the filter responses are maximized. We will use ideas from visualization of vessels (convolution surfaces) to generate the intensity model resp. a global filter kernel model and image processing techniques for the filtering of tube-like structures.

9.2 Vessels Intensity Models

Vessels are tube-like structures, which may have non-circular cross sections but in particular liver vessels can be well approximated by circular cross-sections. Thus the vessels are well characterized by vessel center line points and a radius for each center line point. We assume to have reliable information on the center lines and local radii of the vessels extracted from the preoperative CT data by segmentation and skeletonization in the preoperative planning process (see Sec. 2.3.2 and Sec. 2.3.3).

The center lines are a set of n space curves $c_i : I_i \rightarrow \mathbb{R}^3$, $I_i \subset \mathbb{R}$ between branching points or end points of the vessels. The curves do not intersect themselves or other curves. The branching points are left such that each center line point lies exactly on one curve. Let

$$C_v = \bigcup_i \{c_i(s) | s \in I_i\} \quad (9.1)$$

be the set of points on all vessel curves c_i parameterized by open intervals I_i . The index v stands for vessels. The function $r_v : C_v \rightarrow \mathbb{R}^+$ denotes the radius and $t_v : C_v \rightarrow \mathbb{R}^3$ the tangential direction of the vessel center line points C_v . The tangents are computable via the center line curves c_i . We omit the index v for as long as there is no confusion with reference or template landmarks (r and t), resp. linear constraints C appears.

The curve model with radii information describes the *geometry* of the vessels. To describe the *appearance* of vessels in a 3D image we will generate an intensity model of the vessels. We will focus on the bright appearance of vessels on a totally black background like in power Doppler ultrasound, but the concept can be adapted easily to dark appearing vessels like in B-mode ultrasound or bright vessels on a background with constant intensity as seen in contrast-enhanced CT images. First we model the intensity profile of a vessel cross-section and then extrude this intensity profile along the vessel center lines to get the whole 3D model.

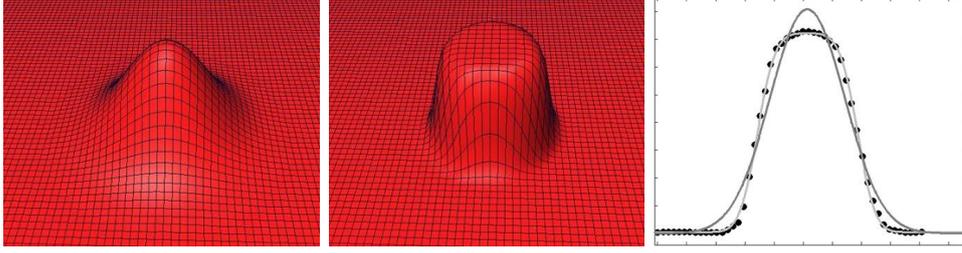


Figure 9.1 2D Gaussian (left) and Gaussian convolved cylinder (middle) intensity profile of a vessel cross-section. On the right: Gaussian (dark gray) and Gaussian convolved cylinder (light gray) profile compared to the real vessel profile in power Doppler ultrasound data (black dots).

9.2.1 2D Vessel Intensity Profiles

In an ideal image of the locally cylindrical vessel geometry a vessel cross-section is just a circular disc with a constant intensity. We ignore the actual intensity and define a more abstract model where the intensity is one and the background zero. For registration distance measures like the local correlation coefficient measure or the normalized gradient field measure the actual intensity does not matter. Thus the 2D intensity profile of a vessel cross-section with a radius parameter r in an ideal image can be defined by a binary disk function:

$$g_{\text{disk}}(x_1, x_2, r) = \begin{cases} 1 & \text{if } R \leq r \\ 0 & \text{else} \end{cases} \quad \text{with } R = \sqrt{x_1^2 + x_2^2}. \quad (9.2)$$

Real imaging systems do not have ideal imaging properties. For example, they have only a limited resolution. The point spread function (PSF) describes the response of an imaging system to a point object [74]. The point is blurred by the system and appears as a blob in the image. Mathematically the image is the convolution of objects with the point spread function.

Often Gaussians are used to approximate the intensity profiles of vessel cross-sections. As pointed out by several authors [262, 106] a Gaussian convolved cylinder profile is a better approximation for the appearance of vessels, in particular for vessels with bigger diameter (Fig. 9.1). This follows from the above assumption of an ideal disc-like intensity profile convolved with a Gaussian point spread function. The intensity profile generated by a Gaussian convolved cylinder is defined by:

$$g_{\text{convcyl}}(x_1, x_2, r, \sigma) = g_{\text{disk}}(x_1, x_2, r) * G_{\sigma}^{2D}(x_1, x_2) \quad (9.3)$$

where the function g_{disk} is defined as above and the 2D Gaussian function $G_{\sigma}^{2D}(x_1, x_2) = G_{\sigma}(x_1)G_{\sigma}(x_2)$ is defined by a product of two 1D Gaussian functions with $G_{\sigma}(x_1) = (\sqrt{2\pi}\sigma)^{-1}e^{-x_1^2/2\sigma^2}$ due to the separability property

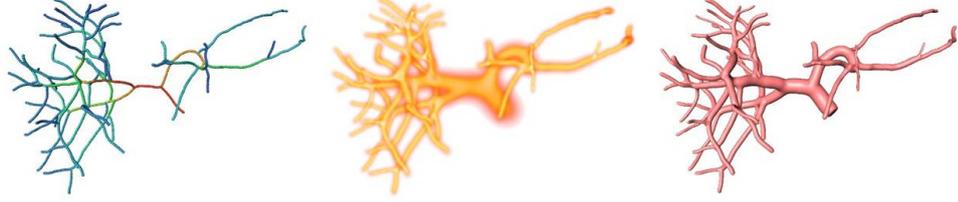


Figure 9.2 Left: Vessel center lines with color-coded radius information (red=large radii, blue = small radii). Middle: Volume rendering visualization of the intensity model. Right: Isosurface of the intensity model (convolution surface).

of the Gaussian function. Unfortunately there is no analytical closed form solution of the convolution integral, but Wörz et al. [261] give a suitable approximation.

As the 2D Gaussian is rotationally symmetric: $G_\sigma(x_1, x_2) = \tilde{G}_\sigma(R) = (2\pi\sigma^2)^{-1}e^{-R^2/2\sigma^2}$ with $R = \sqrt{x_1^2 + x_2^2}$ and also $g_{\text{disk}}(x_1, x_2)$, it follows that also g_{convcyl} is rotationally symmetric .

9.2.2 3D Intensity Model Based on Convolution

Now we have an intensity model of a vessel cross-section. The question remains how these 2D cross-section models at each center line point can be extruded to an overall 3D generalized cylinder model of the vessels. We were inspired by the approach used for the visualization of structures by convolution surfaces [17, 164]. A convolution surface $S_c \subset \mathbb{R}^3$ is an isosurface $S_c = \{x \in \mathbb{R}^3 \mid F(x) = c\}$ for a certain isovalue $c \in \mathbb{R}$ of a scalar (implicit) function $F : \mathbb{R}^3 \rightarrow \mathbb{R}$. The scalar function is generated by a convolution $F(x) = (K * C)(x) = \int_C K(s-x)ds$ of a set $C \subset \mathbb{R}^3$ of geometrical primitives (points, curves, polygons) with a kernel function $K : \mathbb{R}^3 \rightarrow \mathbb{R}$. The exact meaning of the convolution integration over the set C will be explained below. The set C is also called the “skeleton” of the visualized object. Convolution surfaces are in particular suitable for the visualization of tubular structures like vessels (see Fig. 9.2) because they are smooth and in particular smooth at the branchings of vessels [164].

In our case the vessel center lines are the skeleton of the vessels. As the center lines are infinitesimally thin, we describe them by Dirac delta functions δ like Bloomenthal [16] to get non-zero integrals when convolving them with the kernel function K . Bloomenthal [16] utilize a 3D Gaussian function as the kernel function $K = G_\sigma^{3D}(x_1, x_2, x_3)$ because of its symmetry and its

separability property: $G_\sigma^{3D}(x_1, x_2, x_3) = G_\sigma(x_1)G_\sigma(x_2)G_\sigma(x_3)$. He gives a discretization and efficient implementation of the convolution integral. The idea is to approximate the curves by linear line segments. Due to the superposition property of the convolution integral: $K * (C_1 \cup C_2) = (K * C_1) + (K * C_2)$, the convolution integral for each line segment can be computed and summed up to the overall convolution integral. Without loss of generality let a line segment with length a lie on the x_1 -axis and be defined as a product of 1D Dirac delta functions: $l(x_1, x_2, x_3) = l_{x_1}(x_1)\delta(x_2)\delta(x_3)$ with $l_{x_1}(x_1) = 1$ for $0 \leq x_1 \leq a$ and $l_{x_1}(x_1) = 0$ otherwise. In distribution theory the delta function is defined more exactly by its effect onto another function if the product of both functions is integrated. Now the convolution integral of the line segment with a 3D Gaussian kernel can be written as:

$$F_l^G(x) = \int_l G_\sigma^{3D}(s - x)ds = \int_{\mathbb{R}^3} l(y)G_\sigma^{3D}(y - x)dy. \quad (9.4)$$

Bloomenthal [16] showed that this convolution integral can be reduced to a product of a scalar *distance filter* term $G_\sigma^{2D}(x_2, x_3)$ and an *integration filter* term $\int_0^a G_\sigma(y_1 - x_1)dy_1$:

$$F_l^G(x) = G_\sigma^{2D}(x_2, x_3) \int_0^a G_\sigma(y_1 - x_1)dy_1. \quad (9.5)$$

The first term is called the distance filter because the value depends on the distance of the point x to the line segment. The integration filter smoothes the line segment along the line direction.

We use the same integration filter term as Bloomenthal, but extent the distance filter in two directions to get a better intensity modeling of real vessels. We consider the different radii along the vessels and use the Gaussian convolved cylinder intensity profile from Sec. 9.2.1 instead of the Gaussian profile leading to the kernel:

$$\tilde{K}_{\text{convcyl}}(x, r, \sigma) = g_{\text{convcyl}}(x_2, x_3, r, \sigma)G_\sigma(x_1). \quad (9.6)$$

The respective convolution integral for the line segment l is then:

$$F_l^{\text{convcyl}}(x) = \int_l \tilde{K}_{\text{convcyl}}(s - x, r(s), \sigma)ds \quad (9.7)$$

$$= g_{\text{convcyl}}(x_2, x_3, r, \sigma) \int_0^a G_\sigma(y_1 - x_1)dy_1. \quad (9.8)$$

Now the line segment has to be transformed to the position and orientation of the part of the vessel center lines it approximates. Therefore we define a local coordinate system at each center line point $s \in C_v$ by two normal directions $n_1, n_2 : C_v \mapsto \mathbb{R}^3$, $n_1(s) \perp n_2(s)$, perpendicular to the tangential direction $t(s)$. Due to the rotational symmetry of the kernel the exact choice

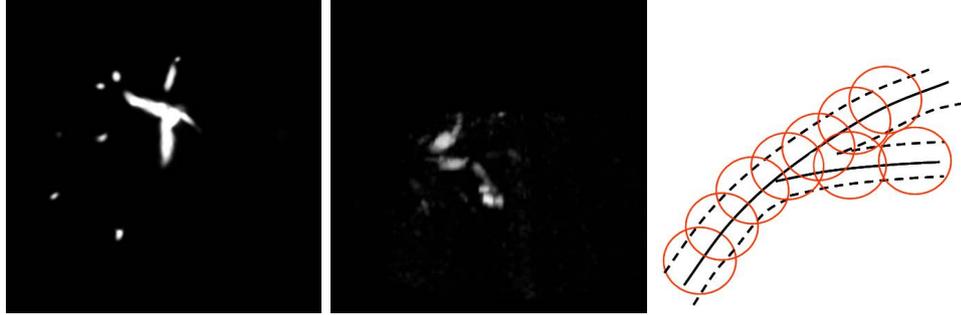


Figure 9.3 Left: Preoperative intensity model. Middle: Original intraoperative power Doppler ultrasound image. Right: For each center line point the normalized correlation coefficient (NCC) is computed for a local neighborhood.

of the normal directions n_1 and n_2 does not matter. The normal directions can be constructed from the tangential direction by means of the cross product. We choose a rotation $T^{rot} = [t(s), n_1(s), n_2(s)]$ for rotating the kernel onto a tangential direction $t(s)$ and introduce the abbreviation:

$$K_{\text{convcyl}}(s-x, r(s), t(s), \sigma) = \tilde{K}_{\text{convcyl}}([t(s), n_1(s), n_2(s)](s-x), r(s), \sigma). \quad (9.9)$$

A vessel intensity model $\mathcal{R}_{\text{mod}} : \mathbb{R}^3 \rightarrow \mathbb{R}$ based on the Gaussian convolved cylinder profile can now be defined by:

$$\mathcal{R}_{\text{mod}}(x) = \int_{C_v} K_{\text{convcyl}}(s-x, r(s), t(s), \sigma) ds. \quad (9.10)$$

In Fig. 9.2 the intensity model of liver vessels based on a center lines skeleton with radius information and an isosurface of this model representing a convolution surface is shown.

9.3 Correlation Measure Based on the Vessel Intensity Model

Template matching is an image processing method where we look for an object resp. a small part of an image which matches a small template image [23]. The template is moved over the image and the position is determined where the image region is most similar to the template. Some of the distance measures we introduced for image registration are also used to quantify the similarity of template and image region. For example, the correlation coefficient is utilized, if the contrast and brightness between the template and the searched image region differs.

We transfer the idea of template matching to the registration of the intensity model generated in the previous section to intraoperative ultrasound data. We take a local region of each part of the vessel model (a template) and compare

its intensities with the intensities of the ultrasound image in the same region. As we only modeled the relative intensity distribution of the vessels, but not the absolute intensities as they appear in the ultrasound data the correlation coefficient is a suitable measure for the comparison, because it is invariant against affine-linear intensity differences (see Sec. 8.3.2). Hence, all the local correlation coefficients along the vessel model are summed up to get a distance measure for the intensity model and ultrasound data (see Fig. 9.3). This means we get a local correlation coefficient distance measure \mathcal{D}^{LCC} like in Sec. 8.3.2, but the reference image \mathcal{R} is exchanged by the intensity model image \mathcal{R}_{mod} and the local correlation coefficients are not computed on the whole image, but only along the vessel center lines C_v :

$$\mathcal{D}^{LCC}[\mathcal{R}_{\text{mod}}, \mathcal{T}, C_v, y] = 1 - |\Omega|^{-1} \int_{C_v} \text{NCC}_K^2(\mathcal{T}(y), \mathcal{R}_{\text{mod}})(s) ds. \quad (9.11)$$

The kernel function $K : \mathbb{R}^3 \rightarrow \mathbb{R}$ can simply be a binary function representing a spherical region as shown in Fig. 9.3. The size of the spherical region is adapted to the vessel radius at the vessel center line point. From the application's view the intraoperative ultrasound data make up the reference image, but for the evaluation of the distance measure it is more efficient to take the preoperative intensity model as the reference image, such that the regions defined by the kernel function do not have to be transformed. In this case the transformation needed for the application can be computed by local inversion [41].

An interesting variant of the local correlation measure is to weight the more important vessels with a large diameter higher than those with a small one. The diameters are known from the model. The application of a multilevel strategy on the intensity model and intraoperative image data is straightforward.

9.4 Convolution Based Measures

The general idea of this class of distance measures is to use local filters, which are adapted to the size and direction of the preoperative vessel model at a given point $\mathbf{s} \in C_v$ on a vessel center line. These filters give a high response, if a tube-like structure with similar radius and direction lies at the same point in the intraoperative intensity data (see Fig. 9.4). We restrict ourselves to linear filters defined by a convolution with a kernel function $K(s, r(s), t(s)) : \mathbb{R}^3 \rightarrow \mathbb{R}$ depending on $s \in \mathbb{R}^3$. The kernel function is parameterized by the radius $r(s) \in \mathbb{R}^+$ and the tangential direction $t(s) \in \mathbb{R}^3$ of the vessel model at the center line point $s \in C_v$ as for the vessel intensity model defined in Sec. 9.2.

Now a distance measure can be defined by integrating all local filter responses of different points s on the vessel skeleton C_v :

$$\mathcal{D}^{\text{CONV}}[\mathcal{R}_{\text{mod}}, \mathcal{T}, y] = - \int_{C_v} \int_{\Omega} \mathcal{T}(y(x)) K(s - x, r(s), t(s)) dx ds \quad (9.12)$$

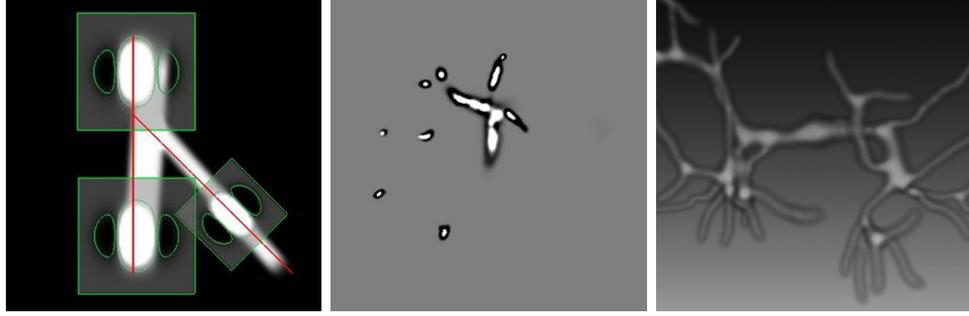


Figure 9.4 Local vessel filter kernels on vessel center lines (red) overlaid on the intraoperative image (left Fig.). Global 3D filter kernel with positive (white) and negative (black) values in a 2D slice (middle Fig.) and 3D volume rendering visualization (right Fig.)

where $\mathcal{T} : \Omega \rightarrow \mathbb{R}$ with $\Omega \subset \mathbb{R}^3$ is the intraoperative 3D ultrasound image. The intensity model \mathcal{R}_{mod} is implicitly represented by the kernel function $K(s, r(s), t(s))$. As in Sec. 9.3 the preoperative intensity model is chosen as reference image, to avoid transforming the filter kernels.

If the kernel function K is a 3D Gaussian function we get the distance measure of Aylward et al. [6] in the case that the standard deviation σ is set to the radius of the vessel model.

Note that the convolution with a filter kernel can be interpreted as the correlation between the mirrored filter kernel and an image. The difference to the correlation coefficient lies in the local normalization.

Although we apply our method to tube-like features, the framework is general and we expect that it works also for other features, e.g., with a locally plate-like shape [45] like the surface of the liver.

9.4.1 Reformulation to a Global Filter Kernel

Since for each point on the vessel center lines a 3D convolution has to be performed the run-time for a straightforward evaluation of the distance measure might be similar to common intensity-based distance measures. But fortunately the convolutions for each center line point with local filter kernels can be combined to one convolution with a *global* filter kernel by exchanging the order of the integration:

$$\begin{aligned} \mathcal{D}^{\text{CONV}}[\mathcal{T}, \mathcal{R}_{\text{mod}}, y] &= - \int_{C_v} \int_{\Omega} \mathcal{T}(y(x)) K(s - x, r(s), t(s)) dx ds \\ &= - \int_{\Omega} \mathcal{T}(y(x)) \int_{C_v} K(s - x, r(s), t(s)) ds dx. \end{aligned} \quad (9.13)$$

The global filter kernel $P(x) = \int_{C_v} K(s - x, r(s), t(s)) ds$ can be determined preoperatively (see Fig. 9.4). During the intraoperative registration process only the cross correlation $\int_{\Omega} \mathcal{T}(y(x)) P(x) dx$ of the global kernel and the intraoperative data \mathcal{T} has to be computed. This is an essential conclusion for the efficient use of the distance measure in the context of our intraoperative registration method. As the global kernel is almost zero at a certain distance to the vessel center lines, the distance measure does not need to be evaluated for those parts.

We only consider kernel functions which are rotationally symmetric around the vessel axis and Gaussian along the axis. Therefore we define a local coordinate system $u = [t, n_1, n_2](x - s)$ similar to the intensity model in Sec. 9.2.2. The kernels can be written in the new coordinate system as

$$K(\mathbf{u}) = K_{\text{dist}} \left(\sqrt{u_1^2 + u_2^2} \right) G_{\sigma}(u_3). \quad (9.14)$$

The structure of the kernels is the same as for the intensity model in Sec. 9.2.2. The global kernel function P can be efficiently implemented exactly like the vessel intensity model by a piecewise linear approximation of the center lines (see Sec. 9.2.2).

9.4.2 Vessel Filters Kernels

Now we are searching for local filter kernels, which give high responses at a vessel center line point in the reference image if at this same point in the currently transformed template image a tube-like structure is present with a similar radius and direction as the vessel model in the reference image. Several filter kernels have been published which rely on differential operators to calculate partial derivatives. To make the numerical determination of derivatives more robust to noise the images are usually smoothed by a Gaussian filter (Gaussian convolution) before differential operators are applied. Due to the commutative property between the derivative operators and the Gaussian convolution the image can equivalently be convolved by derivatives of Gaussians.

Laplacian Kernel

A well-known vessel filter which is based on second derivatives of Gaussians was published by Frangi et al. [68]. They analyze the eigenvalues $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ of the Hessian matrix \mathbf{H} for each voxel. The eigenvector \mathbf{v}_1 corresponds to λ_1 points in the direction of the vessel. For bright vessels on a dark background the eigenvalues have the property: $\lambda_1 \approx 0$ and $\lambda_1 \ll \lambda_2 \approx \lambda_3$. Frangi et al. [68] define a scalar valued “vesselness” function depending on this property. Because the radii of the vessels are unknown, the vesselness response is

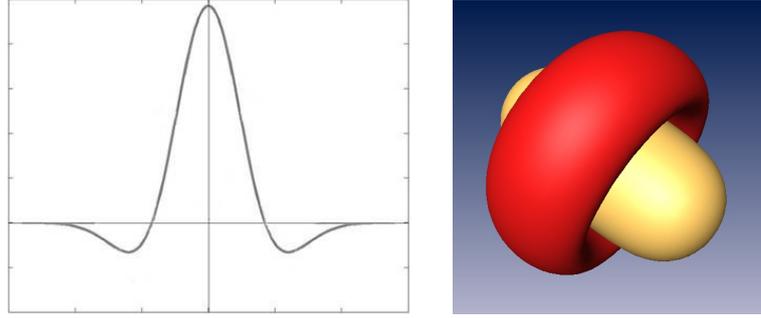


Figure 9.5 Left: Gaussian second derivative resp. profile of 3D Laplacian kernel. Right: Isosurface visualization of 3D Laplacian kernel with positive values (yellow) inside the vessel and negative values (red) outside the vessel.

calculated at multiple scales by computing the Hessian with Gaussian derivatives at multiple scales. At every voxel the vesselness value with the highest response is selected and the corresponding scale represents the radius of the vessel.

In the registration setting however, radii and directions of the vessels are already known from the preoperative model. Instead of computing the direction and radius via the Hessian matrix the filter kernel can directly be steered by the known parameters. Motivated by the vesselness filters of Frangi et al. we define a filter kernel based on the sum of the Gaussian second derivatives in the two normal directions (see Fig. 9.5). This results in a Laplacian filter in the normal plane which is Gaussian weighted in the vessel direction. The second Gaussian derivative in the x_1 direction at a point $x = (x_1, x_2, x_3)$ is:

$$G_{x_1x_1}(x, \sigma) = \left(\frac{x_1^2}{\sigma^4} - \frac{1}{\sigma^2} \right) G(x, \sigma). \quad (9.15)$$

The kernel has to be transformed to the position of a center line point \mathbf{y} and orientation of the local coordinate system $u = [t, n_1, n_2](x - y)$. This yields the following filter kernel:

$$K_L(u, \sigma) = G_{n_1n_1}(u, \sigma) + G_{n_2n_2}(u, \sigma) = L(\sqrt{u_1^2 + u_2^2}, \sigma)G(u_3, \sigma) \quad (9.16)$$

where $L(r, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \left(\frac{r^2}{\sigma^4} - \frac{2}{\sigma^2} \right)$ shows rotational symmetry. The kernel K_L is of the general form defined in Eq. 9.14. It is important to note that such kinds of filters have to be normalized appropriately to reach scale invariance and σ has to be chosen correctly such that the response is maximal for a given radius r . Because the Laplacian kernel has a zero level set at $\sqrt{u_1^2 + u_2^2} = \sqrt{2}\sigma$ the correct choice is $\sigma = (\sqrt{2})^{-1}r$. In addition, the distance measure (not the kernel) is set to 0 for values ≤ 0 to avoid negative correlations.

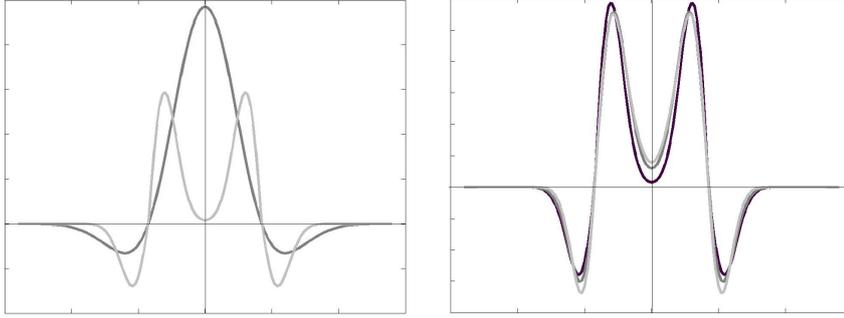


Figure 9.6 Left: Profile of Laplacian kernel K_L (dark gray) and difference of Gaussian first derivatives K_D (light gray). Right: Difference of Gaussian first derivatives kernel K_D (black) compared to kernel K_K of Krissian (dark gray) and Laplacian of convolved cylinder (light gray).

Difference of Gaussian First Derivatives Kernel

Frangi et al. [68] assumed a vessel model with a Gaussian intensity profile for the cross-sections. As we discussed in Sec. 9.2.1 a better model is a Gaussian convolved cylinder. The following kernels are better suited for this vessel model.

Krissian et al. [106] propose a more radius-sensitive kernel based on Gaussian first derivatives G_{v_α} in all radial directions $v_\alpha \in \mathbb{R}^3, \|v_\alpha\| = 1$ in the normal plane of a vessel integrated on a circle with radius r : $K_K(u, r, \sigma) = \frac{1}{2\pi} \int_{\alpha=0}^{2\pi} \int G_{v_\alpha}(u + rv_\alpha, \sigma) d\alpha$.

Since there is no analytical closed form for this integral we suggest a similar and easily computable kernel by rotating first derivatives of 1D Gaussians instead of 2D Gaussians, which is of the type defined in Eq. 9.14:

$$K_D(u, r, \sigma) = \left[G' \left(\sqrt{u_1^2 + u_2^2} + r, \sigma \right) - G' \left(\sqrt{u_1^2 + u_2^2} - r, \sigma \right) \right] G(u_3, \sigma). \quad (9.17)$$

Normalizing the kernel to $\int K_D(u, r, \sigma) du = 0$ yields a similar kernel to the one of Krissian et al. (Fig. 9.6). Compared to the Laplacian filter the kernel is concentrated more on the boundary of a vessel (see Fig. 9.6) leading to a higher radius sensitivity steerable by the parameter σ . Note, that the kernel K_D is similar to the Laplacian of a Gaussian convolved cylinder (Fig. 9.6).

9.4.3 Conclusion

In the last sections we show how a priori knowledge about the shape of anatomical structures can be incorporated into the distance functional. The main idea is to use local filter kernels, which fit the local shape of anatomical structures

in the reference image and give a high filter response in the presence of the same shape in the template image. We focus on tube-like shapes like vessels and present different filter kernels for such structures. We show that the kernels can be divided into two classes. The first class (Laplacian kernels) assume a Gaussian intensity profile of the vessels the second class (difference of first derivatives) a Gaussian convolved cylinder intensity profile. Although we developed the convolution based distance measure D^{CONV} in particular for tube-shaped vessels the idea of the measure is quite general and can for example also be adapted to the locally plate-like shaped liver surface, which is partially imaged in B-mode ultrasound data.

Positive properties of D^{CONV} are that its evaluation is efficient and that in all filter kernels the standard deviation of the Gaussian can be used for a multi-scale approach. A first validation of D^{CONV} is performed in Chapter 14 for the registration of preoperative data and intraoperative 3D ultrasound data of the liver.

Chapter 10

Possibilities of Constraining Displacement Fields

10.1 Introduction

In the preceding parts, different possibilities were shown and developed how a priori knowledge about the image classes and contents can be incorporated into the distance measure \mathcal{D} comparing two images. Another starting point for prior knowledge incorporation is to restrict the admissible transformations y by known additional properties, such as volume preservation, rigidity of anatomical structures or known anatomically corresponding points, so-called landmarks. The next two sections explain how to deal with such additional constraints in the given variational framework and give some important examples of such constraints. The focus in the following chapters will be on landmark constraints.

10.2 Constraints in Variational Image Registration

There are three general possibilities to add constraints to the variational image registration framework (3.4). The most used and often easiest to implement is to add a penalizing term \mathcal{P} to the registration functional, which only depends on the transformation y and not directly on the images R and T :

$$\mathcal{J}[y] = \mathcal{D}[T(y), R] + \alpha \mathcal{S}[y - y^{\text{ref}}] + \beta \mathcal{P}[y] \xrightarrow{y} \min. \quad (10.1)$$

For example, let \mathcal{P} be $\mathcal{P}^{\text{VP}}[y] = \int_{\Omega} |\det(\nabla y) - 1| dx$, to enhance volume preservation. The Jacobian determinant $\det(\nabla y)$ gives the local volume change of y . The term $\mathcal{P}^{\text{VP}}[y]$ penalizes deviations of the Jacobian determinant from unity. A Jacobian determinant of one means complete volume preservation. The drawback of this kind of constraint is that all local changes are integrated in one single value and globally weighted to the rest of the functional via the parameter β . This means that there is no local control on the

volume change and significant changes might occur in some regions. In addition, the choice of β depends on the particular image pair and is not intuitive. These kinds of constraints can be called *soft constraints* or *global constraints*.

An alternative possibility is to formulate an optimization problem with equality constraints \mathcal{C} depending on y :

$$\begin{aligned} \mathcal{J}[y] = \mathcal{D}[T(y), R] + \alpha \mathcal{S}[y - y^{\text{ref}}] & \xrightarrow{y} \min, \\ \text{subject to } \mathcal{C}[y] & = 0. \end{aligned} \quad (10.2)$$

For the volume preservation example, the constraints \mathcal{C} might be defined as: $\mathcal{C}^{\text{VP}}[y](x) = \det(\nabla y(x)) - 1 = 0$, for all $x \in \Omega$. These constraints ensure volume preservation in any part of the image domain. Instead of penalizing unwanted solutions the constraints rule them out completely. Thus the constraints are also called *hard constraints* or *local constraints*.

This kind of exact equality constraints is often too restrictive for many applications. Usually there is only an estimate of quantitative properties available. Instead of equality constraints, inequality constraints can also be used:

$$\begin{aligned} \mathcal{J}[y] = \mathcal{D}[T(y), R] + \alpha \mathcal{S}[y - y^{\text{ref}}] & \xrightarrow{y} \min, \\ \text{subject to } \mathcal{C}[y] & \geq 0. \end{aligned} \quad (10.3)$$

In case of the volume preservation example, the volume might be allowed to change locally in a given range: $\kappa_m(x) \leq \mathcal{C}^{\text{VP}}[y](x) \leq \kappa_M$, for all $x \in \Omega$ and with non-negative compressibility functions $\kappa_m \leq \kappa_M$. These constraints can be reformulated into $\mathcal{C}^{\text{VP}}[y] - \kappa_m \geq 0$ and $\kappa_M - \mathcal{C}^{\text{VP}}[y] \geq 0$ to fit the generic formulation in equation 10.3. Using equality or inequality hard constraints is a powerful tool to incorporate additional knowledge about the transformation into the registration process, although the numerical solution is demanding. For example, in the case of local equality volume constraints there is one constraint for each voxel.

10.3 Examples for Constraints

Registration methods with constraints are an active field of research, motivated by prior knowledge about the mechanical behavior of anatomical structures or user supplied additional information on the transformation like corresponding landmarks.

The volume preservation constraint discussed in the previous section reflects the incompressible behavior of soft tissue for small deformations and short time periods. In particular in non-rigid registration of pre- and postcontrast MR images of the breast it is indispensable to enforce volume preservation. Without this constraint the volume of contrast-enhancing lesions would significantly shrink after registration, because the optimizer would try

to get rid of a contrast-enhanced lesion, which is not visible in the pre-contrast image resulting in strong differences in the distance measure. A global penalizer approach using parametric B-spline based registration has been investigated by Rohlfing et al. [192]. Instead of the above mentioned penalizer $\mathcal{P}^{\text{VP}}[y] = \int_{\Omega} |\det(\nabla y) - 1| dx$ the logarithm of the Jacobian $\mathcal{P}^{\text{VPlog}}[y] = \int_{\Omega} |\log(\det(\nabla y))| dx$ was used to symmetrically weight a local expansion and compression. Guaranteed local volume preservation and guaranteed limited volume change in a nonparametric constrained registration framework was introduced by Haber et al. [78, 79].

Another important property of anatomical structures is their elasticity resp. rigidity. Bones are very rigid, organs and other soft tissues less so; they behave approximately like elastic materials. One way to handle images containing structures with different elasticities is to assign spatially varying Lamé parameters into the elastic regularizer (see Sec. 6.3.1, based on a segmentation of the structures [94]). A more direct way to ensure the rigidity on given structures is kept under y is to penalize the local deviations from linearity, orthogonality and orientation preservation of y [149]. Soft constraints approaches using such a penalizer have been published for parametric [136, 222] as well as nonparametric [148, 149] image registration. In addition a formulation with equality constraints has been introduced by Haber et al. [76].

If the deformation fields y are analyzed further, e.g., in deformation-based morphometry [4] or for the construction of statistical shape models [39], it is particularly important to have smooth and invertible (bijective) deformation fields. Such smooth and invertible transformations are called diffeomorphisms. For diffeomorphic deformation fields every point in one image has exactly one corresponding point in the other image. Although many registration techniques try to determine smooth deformation fields foldings might and actually do occur. In particular, for inter-patient registration with large deformation fields explicit control or avoidance of folding is often implemented. For example Christensen et al. [34] estimate the forward (y) and reverse (\hat{y}) transformation while constraining these transforms to be inverse to each other ($y = \hat{y}^{-1}$). Another solution is to compose a large diffeomorphic transformation from small diffeomorphic transformations ($y = y_n \circ \dots \circ y_1$) such as Rueckert et al. [198] proposed for free-form deformations based on B-splines. A good starting point to ensure transformation regularity and bijectivity is the local Jacobian determinant ($\det(\nabla y)$). Some registration algorithms monitor the size of the Jacobian determinant and re-initialize or stop the registration process, if its value gets to small [32, 147]. A way to avoid the incidence of small values for the Jacobian determinant is to add inequality constraints on the Jacobian determinant as introduced in the context of volume preservation (see 10.2 and [79]).

An obvious possibility to incorporate a priori knowledge about the deformation field y is to directly ascertain some deformation vectors based on in-

10. POSSIBILITIES OF CONSTRAINING DISPLACEMENT FIELDS

teractively or automatically defined corresponding points or other geometrical features. Such landmark-based registration approaches are the focus of the next two chapters.

Chapter 11

Parametric Landmark Registration

11.1 Introduction

A simple yet important and widely used way to constrain the possible transformations y is to provide some prominent points, so called point landmarks, in one image and the corresponding points in the other image which have to be mapped onto each other by y . This means that the transformation (displacement vectors) on those points is already known. The registration task is to compute the displacements at locations between the landmarks. In case of rigid registration in 3D the whole transformation is already determined by three corresponding pairs of points, but for non-rigid registration a suitable deformation behavior on regions without landmarks has to be modeled. The well-known thin-plate spline approach models deformations as the bending of a thin-plate. Besides the thin-plate spline approach a more realistic model for human tissue based on linear elasticity theory is used here (Gaussian elastic body splines). The rigid and affine-linear as well as both mentioned non-rigid transformations can be parameterized allowing efficient registration schemes.

In clinical practice the interactive or semi-automatic localization of anatomical landmarks is always prone to errors. We show and develop ways how to deal with those uncertainties and give a more flexible and general landmark concept. Quasi-landmarks, which are precisely defined only in some directions extend the point landmark concept. For example, corresponding vessel segments can be identified, where the landmark position is clearly defined perpendicular to the vessel, but not in the direction of the vessel. To provide methods using such application specific landmarks allows for more intuitive and efficient interactive landmark definition.

Conventional landmark registration only use distance measures $\mathcal{D}^{\text{LM}}[y, r, t]$ determining the deviations between corresponding landmark positions, which are presented by r and t . No distance measure $\mathcal{D}[\mathcal{T}(y), \mathcal{R}]$, depending directly on the underlying image data, is considered. This strong reduction of information to only some points of the whole image data sets only leads to

satisfactory non-rigid registration results in dedicated applications. Combinations of landmark and intensity information are presented in the subsequent chapter.

In the following two chapters we assume that landmark locations and uncertainty estimations are already given. In part IV - Applications to Liver Surgery, we give examples how those localizations and estimations can be defined in clinical applications.

11.2 Point Landmarks in Clinical Applications

Point landmarks are in widely used in computer assisted surgery. They are an alternative possibility to register two image data sets as well as to register a preoperative image data set to a patient on the operating table [51]. Point data arise from two sources: prominent anatomical points and artificial fiducial markers. The first approach is also called markerless pair-point registration and the second marker-based pair-point registration [51]. A non-invasive way to define point landmarks is to use exterior or interior anatomical points like the tip of the nose, the corners of the eyes, the ear cartilage or interior points like the tips of the frontal, temporal and occipital horns of the ventricular system in the brain or branchings of vessels. The accurate extraction of such anatomical points is a challenging image processing task and often interactively supported. In addition, in the region of interest a sufficient number of anatomical points have to be available. In marker-based registration, markers are attached to the patient before preoperative imaging. This is a logistic challenge in clinical routine. They are designed to be clearly identifiable in images of different modalities (see for example [142] and references in [51]). In addition, they can be located precisely with navigated pointers on the patient. Different types of markers have been developed to solve the problem of attaching the markers to the patient: skin-markers using adhesives, bone-mounted markers and special fixtures like splints attached to the dentition. Skin-markers are easy to use, but their accuracy is limited due to skin mobility or swelling of the underlying tissue. The highest accuracies are reached with bone-mounted markers like titanium screws drilled into the cranium under local anesthesia. An attempt to achieve the accuracy of bone-mounted markers without their invasiveness are assemblies holding all fiducial markers, which are rigidly attached to the patient at the dentition for instance. A more detailed description and discussion of artificial markers can be found in [51]. Here we will concentrate on anatomical landmarks, because for abdominal applications artificial landmarks are only applicable with a big effort. Some attempts to use skin-markers have been reported ([153] and [158]), but compared to neurosurgery exterior landmarks on the abdomen provide only coarse information about displacements inside the body due to deformability of the inner organs.

The extraction of anatomical landmarks can be divided into the detection and the localization process [196]. Detection means, a (point-like) feature has to be found in an image. The localization implies the estimation of the accurate position of the feature. An elaborate overview and analysis on the performance of detection and localization operators by means of geometric and intensity-based landmark models is given by Rohr [196]. For clinical applications the automatic detection of corresponding landmark pairs is seldom feasible. A suitable strategy is to perform the detection interactively and to refine the position by means of automatic localization operators. Wörz et al. [259] for example present parametric intensity models for a precise localization of a landmark inside a given region of interest. A more detailed view on the extraction of landmarks in the liver is given in the clinical application Chapters 14 and 15.

11.3 Notations

In this section we define general landmark registration schemes and stipulate the notation. Let $r_j, t_j \in \Omega \subset \mathbb{R}^3$, $j = 1, \dots, N_{LM}$ be corresponding point landmarks in the reference resp. template image. N_{LM} is the number of landmark pairs. The aim of the registration is to find a transformation y , which maps corresponding points onto each other:

$$y(r_j) = t_j, \quad j = 1, \dots, N_{LM}. \quad (11.1)$$

The role of reference and template landmarks is exchanged here (it might be expected $y(t_j) = r_j$) to be consistent with the definition of y in the Eulerian framework (see Chapter 3). As y maps from the domain of R to the domain of T also the landmarks as parts of the particular domains have to be mapped in this direction. The equations 11.1 can be interpreted as *interpolation* conditions. We are looking for a transformation y , which interpolates the displacement vectors defined at the landmark positions.

As the localization of point landmarks is always afflicted with uncertainties the *interpolation* conditions should be modified to *approximation* conditions:

$$y(r_j) \approx t_j, \quad j = 1, \dots, N_{LM}. \quad (11.2)$$

The quality of an approximation is usually quantified by an adequate norm resp. a distance measure for the landmark pairs. Before going into more detail we first clarify some notations.

11.3.1 Representation of Landmarks

For the numerical treatment of landmark registration problems it is important to have suitable notations and representations of the landmark pairs.

The first possibility is to arrange the landmarks in matrices $\bar{r}, \bar{t} \in \mathbb{R}^{N_{LM} \times 3}$, where the j -th row contains the coordinates of landmark $r_j \in \mathbb{R}^3$ respectively $t_j \in \mathbb{R}^3$:

$$\bar{r} = \begin{pmatrix} r_1^T \\ \vdots \\ r_{N_{LM}}^T \end{pmatrix} \quad \text{and} \quad \bar{t} = \begin{pmatrix} t_1^T \\ \vdots \\ t_{N_{LM}}^T \end{pmatrix}. \quad (11.3)$$

This representation is often used for rigid landmark registration as we will show later.

The second possible notation is to stack all landmark positions one after the other into one big vector $\hat{r}, \hat{t} \in \mathbb{R}^{3N_{LM}}$. Formally this stacking can be described by the vec -operator applied to the landmark matrix $\bar{r}, \bar{t} \in \mathbb{R}^{N_{LM} \times 3}$, which stacks the elements of a matrix *column-wise*. Therefore for our purpose the vec -operator has to be applied to the transposed landmark matrix:

$$\hat{r} = \text{vec}(\bar{r}^T) \quad \text{and} \quad \hat{t} = \text{vec}(\bar{t}^T). \quad (11.4)$$

This notation is used, e.g., by Rohr et al. for interpolating and approximating thin-plate spline landmark registration [195].

The third possibility is to stack all landmark coordinates again into one big vector $r, t \in \mathbb{R}^{3N_{LM}}$, but in order to be consistent with Chapter 4.2 first all x_1 -coordinates of all landmarks than all x_2 - and x_3 -coordinates are stacked into the big vector. Formally this stacking is described by the vec -operator again:

$$r = \text{vec}(\bar{r}) \quad \text{and} \quad t = \text{vec}(\bar{t}). \quad (11.5)$$

This is a particularly helpful representation, because it can directly be used in the general optimization framework (see 5.1) and is consistent to the notations of the discrete grids in Chapter 4.2.

We will use the third notation (Eg. 11.5) predominately in the following sections and chapters, but at some places the matrix notation will be advantageous (11.3) for rigid registration and the second representation (11.4) for the derivation of some weighted schemes.

The two big vector notations can easily be transformed to each other by permutation matrices. To transform \hat{r} into r a permutation matrix $P_{\hat{r} \rightarrow r} \in \mathbb{R}^{3N_{LM}} \times \mathbb{R}^{3N_{LM}}$ is applied:

$$r = P_{\hat{r} \rightarrow r} \hat{r} \quad \text{with} \quad P_{\hat{r} \rightarrow r} = \begin{pmatrix} I_{N_{LM}} \otimes (1 & 0 & 0) \\ I_{N_{LM}} \otimes (0 & 1 & 0) \\ I_{N_{LM}} \otimes (0 & 0 & 1) \end{pmatrix} \quad (11.6)$$

where $I_{N_{LM}}$ is the identity matrix of dimension $\mathbb{R}^{N_{LM}}$ and \otimes the Kronecker product operator. For the inverse transformation from r to \hat{r} just the inverse

of the permutation matrix has to be applied:

$$\hat{r} = P_{r \rightarrow \hat{r}} r \quad \text{with} \quad P_{r \rightarrow \hat{r}} = P_{\hat{r} \rightarrow r}^{-1} = P_{\hat{r} \rightarrow r}^T. \quad (11.7)$$

We specify for both vector representations, how the notations $y(\bar{r})$ and $y(r)$ for the transformed reference landmarks have to be interpreted. Let $y(\hat{r})$ be defined as:

$$y(\hat{r}) := (y(r_1)^T, \dots, y(r_{N_{LM}})^T)^T \quad (11.8)$$

and $y(r)$ as:

$$y(r) = (y_1(r_1), \dots, y_1(r_{N_{LM}}), y_2(r_1), \dots, y_2(r_{N_{LM}}), y_3(r_1), \dots, y_3(r_{N_{LM}}))^T. \quad (11.9)$$

11.4 Principle Landmark Registration Schemes

Corresponding landmark pairs can be seen as constraints at some points of the sought deformation field. All three possibilities of constraint variational image registration as presented in Chapter 10 are also suitable for landmark registration and will be explained in the next subsections: global constraints by a penalizer, equality constraints and inequality constraints. Rigid landmark registration takes on a special position because no constraints on the landmarks but on the allowed transformations (rigidity) are defined.

11.4.1 Distance Measures for Landmarks

Before coming to the three general landmark registration schemes we define a distance measure determining the deviations between corresponding landmark positions. Usually the sum of the Euclidean distances of the landmark pairs is used for that purpose. But since in clinical practice localization of landmarks is always prone to errors, we use a more general distance measure:

$$\mathcal{D}^{\text{LM}}[y, r, t] = \|y(r) - t\|_{LM}^2 = \sum_{j=1}^{N_{LM}} \|y(r_j) - t_j\|_{LM}^2 \quad (11.10)$$

where $\|\cdot\|_{LM}$ denotes a suitable norm. In the simplest case this is the Euclidean norm $\|\cdot\|_2$, but also a weighted norm $\|\cdot\|_{LM}$ considering the aforementioned localization uncertainties will be presented in the following section 11.5.

If the landmark distance measure is applied as a penalizer in the integral registration functional $\mathcal{J}[y]$ a slightly different formulation using the point evaluation or Dirac δ functional is needed:

$$\mathcal{P}^{\text{LM}}[y, r, t] = \sum_{j=1}^{N_{LM}} \|\delta_{r_j}[y] - t_j\|_{LM} \quad (11.11)$$

with $\delta_{r_j}[y] = y(r_j)$ (for details, see [147]).

11.4.2 Equality Landmark Constraints

As mentioned at the beginning of this section, landmark registration can be seen as an interpolation problem. To get a smooth transformation function $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, which interpolates the vector-valued template landmark positions $t_j \in \mathbb{R}^3$ at the given nodes $r_j \in \mathbb{R}^3$, a regularizer \mathcal{S} is optimized subject to the interpolating equality landmark constraints:

$$\begin{aligned} \mathcal{J}[y] = \mathcal{S}[y] &\xrightarrow{y} \min, \\ \text{subject to } y(r) - t &= 0. \end{aligned} \quad (11.12)$$

Parametric solutions of the optimization problem exist for particular regularizers \mathcal{S} . Two different regularizers \mathcal{S} will be presented in the following chapters leading to interpolating thin-plate (iTPS; 11.8.1) and interpolating Gaussian elastic body splines (iGEBS; 11.9.1) as analytical solutions of the optimization problem.

11.4.3 Approximative Landmark Registration with Explicit Transformation Constraints

In clinical practice, the localization of the landmarks is given only approximately (landmarks with localization uncertainties). This means no exact matching of corresponding landmarks can be expected and therefore exact interpolation conditions are too strict. In particular, interpolation conditions for rigid and affine registration can generally only be fulfilled for very few landmarks with uncertainties, because there are only few degrees of freedom. For example, in the case of rigid transformations three landmarks can theoretically be mapped exactly onto each other although their localization is inexact. In the case of non-rigid transformations the number of degrees of freedom is usually high enough to find interpolating transformations. But nevertheless localizations of the landmarks are uncertain, here approximative solutions, which consider the localization uncertainty, are more suitable.

By means of the landmark distance measure (11.10) approximative landmark registration schemes can be developed. The search space for the transformation y is restricted to a parametric space \mathcal{M} , typically of low dimension. The transformations are explicitly defined by linear combinations of some basis functions q_k : $y(x) = \sum \gamma_k q_k(x)$. Suitable choices for \mathcal{M} are based on rigid, affine linear or spline transformations. We define an approximative landmark registration scheme to be minimizing the landmark distance measure over the restricted transformation space \mathcal{M} :

$$\begin{aligned} \mathcal{J}[y, r, t] = \mathcal{D}^{\text{LM}}[y, r, t] &\xrightarrow{y} \min, \\ \text{subject to } y &\in \mathcal{M}. \end{aligned} \quad (11.13)$$

On the restricted transformation search space the distance measure does usually not drop to zero.

11.4.4 Landmark Distances as a Global Penalizer

Instead of defining an explicit search space the transformations can be characterized by a regularizer \mathcal{S} . The regularizer and the landmark distance penalizer \mathcal{P}^{LM} (11.11) can be summed up to a registration functional:

$$\mathcal{J}[y] = \mathcal{P}^{\text{LM}}[y, r, t] + \lambda_A \mathcal{S}[y] \xrightarrow{y} \min \quad (11.14)$$

with a weighting parameter $\lambda_A \in \mathbb{R}^+$, which controls the relation between the smoothness of the transformation y via \mathcal{S} and the approximation quality induced by the penalizer \mathcal{P} . Interestingly, for some regularizers \mathcal{S} there are again analytical solutions of the optimization problem: for example approximating thin-plate (aTPS; Sec. 11.8.2) and Gaussian elastic body splines (aGEBS; Sec. 11.9.2).

11.4.5 Inequality Landmark Constraints

A further possibility to relax landmark interpolation to approximation conditions are inequality hard constraints:

$$\begin{aligned} & \mathcal{J}[y] = \mathcal{S}[y] \xrightarrow{y} \min, \\ \text{subject to } & \|y(r_j) - t_j\|_{LM}^2 \leq \text{tol}_j^2, \quad j = 1, \dots, N_{LM} \end{aligned} \quad (11.15)$$

with $\text{tol}_j \in \mathbb{R}^+$ being the tolerances defined as radii of distance spheres around the landmarks. A solution of the optimization problem is only feasible, if the deformed landmark $y(r_j)$ is inside a sphere of radius tol_j around the corresponding template landmark t_j . If we use a weighted landmark norm $\|\cdot\|_{LM}$, this sphere is deformed to an ellipsoid leading to anisotropic tolerances. Such a weighted landmark distance norm will be presented in the next section 11.5.

The constrained optimization problem can be reformulated to the generic constrained optimization problem presented in Chapter 5:

$$\begin{aligned} & \mathcal{J}[y] = \mathcal{S}[y] \xrightarrow{y} \min, \\ \text{subject to } & \text{tol}_j^2 - \|y(r_j) - t_j\|_{LM}^2 \geq 0, \quad j = 1, \dots, N_{LM}. \end{aligned} \quad (11.16)$$

We are not aware of any publication suggesting such a landmark constrained registration scheme. In Chapter 12 we will show how it can be numerically solved as a special case of the more general combined landmark and intensity registration approach.

11.5 Landmark Localization Uncertainties

The localization of landmarks is always prone to errors no matter if they are defined manually or semi-automatically. There are two general ways how to deal with such localization uncertainties. The first one is to analyze the resulting registration accuracy from the localization uncertainties. Localization

uncertainty for landmarks is also called the "fiducial landmark error" (FLE). The seminal work of Fitzpatrick et al. [65] establishes the statistical relationship between the expected squared errors in the localization of landmarks (FLE) and the target registration error (TRE) at an arbitrary point. This error estimation is explained in more detail in Chapter 13.

The second approach is to consider the uncertainties already in the registration process by weighting the influence of certain landmarks or in particular directions. The hope is, that higher uncertainties in some regions can be compensated by trusted information from other regions. This means the aim is to deal with the localization uncertainties such that as much information as possible can be used. This approach will be the focus for the rest of the chapter.

The idea is to consider the covariance of landmark localization errors in a weighted distance measure (Mahalanobis distance) as described by Rohr [196]. It is assumed that the covariance of each landmark is at least approximately known. The main problem is to combine the covariances of reference and corresponding template landmarks into one common weighting. In the general case of anisotropic covariances their orientation has also to be transformed by y to combine the uncertainty informations of reference and template landmarks in a common coordinate space. We will see that by introducing assumptions on the anisotropies and transformations y the weighted landmark distance measures can be defined, such that they are sufficient for many practical applications. We start with an examination of the landmark covariance combination and then introduce the weighted landmark distance measure.

11.5.1 Combining Covariances of Landmark Uncertainties

Usually, we can assume additive noise on the landmark localization. Let \tilde{t}_j, \tilde{r}_j be the true landmark positions and $\Delta t_j, \Delta r_j$ the errors. Then the measured landmark position is:

$$t_j = \tilde{t}_j + \Delta t_j, \quad r_j = \tilde{r}_j + \Delta r_j \quad (11.17)$$

where the errors are assumed to have zero mean $E(\Delta t_j) = 0, E(\Delta r_j) = 0$ and are independent and identically distributed with covariance matrices $\text{Var}(t_j) = E(\Delta t_j \Delta t_j^T) = \Sigma_{t_j} \in \mathbb{R}^3 \times \mathbb{R}^3, \text{Var}(r_j) = E(\Delta r_j \Delta r_j^T) = \Sigma_{r_j} \in \mathbb{R}^3 \times \mathbb{R}^3$.

Covariance matrices are positive semi-definite symmetric matrices and every positive semi-definite symmetric matrix defines a covariance matrix. We assume positive definite instead of semi-positive definite matrices because in practical situations there is no direction where the localization uncertainty is exactly 0.

In the following we will call landmarks with localization uncertainties "Quasi-Landmarks" [195], because they are not point landmarks with an exact position.

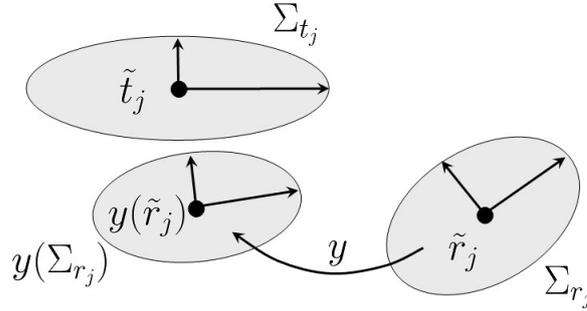


Figure 11.1 The landmark uncertainties are described by the covariances $\Sigma_{r_j}, \Sigma_{t_j}$ in different coordinate systems. In the general case, the orientation of the covariance of the reference landmark is changed by the transformation y .

Definition 11.1 (Quasi-Landmark). A *Quasi-Landmark* is a point $p \in \Omega \subset \mathbb{R}^3$ with a semi-positive definite matrix $\Sigma_p \in \mathbb{R}^{3 \times 3}$ describing the localization uncertainty of the point p .

Generally localization errors occur on the template as well as on the reference landmarks. This means the two covariance matrices for each landmark pair have to be combined. Inserting the measured template ($t_j = \tilde{t}_j + \Delta t_j$) and transformed measured reference landmark position ($y(\tilde{r}_j + \Delta r_j)$) into the landmark interpolation condition ($t_j = y(r_j)$) leads to

$$t_j = y(\tilde{r}_j + \Delta r_j) + \Delta t_j. \quad (11.18)$$

The problem is, that the covariance matrices of reference and template landmark rely on different coordinate systems. After transformation by y the orientation of the covariance of the reference landmark Σ_{r_j} is changed (see Fig. 11.1). If we take a rotation matrix $Q \in \mathbb{R}^{3 \times 3}$ as a simple example for y (leading to $t_j = Qr_j$), the Gauss-Markov theorem states that the best linear unbiased estimator (BLUE) for Q is given by the minimization of the Mahalanobis distance [52, 165]:

$$\mathcal{D}[Q] = \sum_{j=0}^{N_{LM}} (t_j - Qr_j)^T (Q\Sigma_{r_j}Q^T + \Sigma_{t_j})^{-1} (t_j - Qr_j) \quad (11.19)$$

with the covariance matrices $\Sigma_{t_j}, \Sigma_{r_j}$ for the landmarks t_j and r_j . We have to minimize with respect to the group of rotations $SO(3)$. Only iterative solutions have been published for this problem [52, 165].

We will now show that the general case where the reference covariance matrix has to be transformed is not needed in the applications we do have in mind. The following derivation is mainly taken from the book of Rohr [196].

First we assume the artificial case, where no transformation ($y = \text{id}$) between reference and template landmarks is necessary, i.e., they have been measured in the same coordinate system. Then the noisy landmark conditions 11.18 read as:

$$t_j = \text{id}(\tilde{r}_j + \Delta r_j) + \Delta t_j = \tilde{r}_j + \Delta_j \quad (11.20)$$

where $\Delta_j = \Delta r_j + \Delta t_j$ is the joint error of reference and template landmark. The corresponding joint covariance matrix reads as:

$$\begin{aligned} \Sigma_j &= E(\Delta_j \Delta_j^T) \\ &= E((\Delta r_j + \Delta t_j)(\Delta r_j + \Delta t_j)^T) \\ &= E(\Delta r_j \Delta r_j^T) + E(\Delta t_j \Delta t_j^T) + E(\Delta r_j \Delta t_j^T) + E(\Delta t_j \Delta r_j^T) \\ &= E(\Delta r_j \Delta r_j^T) + E(\Delta t_j \Delta t_j^T) \\ &= \Sigma_{r_j} + \Sigma_{t_j} \end{aligned} \quad (11.21)$$

where we used $E(\Delta t_j \Delta r_j^T) = E(\Delta r_j \Delta t_j^T) = 0$ because the noise of the landmarks has been assumed to be uncorrelated. So the combined covariance matrix Σ is just the sum of the covariance matrices of corresponding reference and template landmarks.

Next, we consider the case where the covariance matrices have been measured in different coordinate systems but are related by an affine-linear transformation $y(x) = Ax + v$ with $A \in \mathbb{R}^3 \times \mathbb{R}^3, v \in \mathbb{R}^3$. If we replace y in the general interpolation equation 11.18 by the affine-linear transformation we get:

$$t_j = A(\tilde{r}_j + \Delta r_j) + v + \Delta t_j. \quad (11.22)$$

The joint error is then $\Delta_j = A\Delta r_j + \Delta t_j$ and it follows for the joint covariance matrix:

$$\begin{aligned} \Sigma_j &= E(\Delta_j \Delta_j)^T \\ &= E((A\Delta r_j + \Delta t_j)(A\Delta r_j + \Delta t_j)^T) \\ &= E(A\Delta r_j (A\Delta r_j)^T) + E(\Delta t_j \Delta t_j^T) \\ &= AE(\Delta r_j \Delta r_j^T)A^T + E(\Delta t_j \Delta t_j^T) \\ &= A\Sigma_{r_j}A^T + \Sigma_{t_j}. \end{aligned} \quad (11.23)$$

Thus the covariance matrices of the reference landmarks have to be transformed first ($A\Sigma_{r_j}A^T$) and then added to the covariance matrices of the corresponding template landmarks to get the combined covariance matrices.

If a general non-rigid registration transformation y is sought, this formula can be applied. First, the unknown affine-linear transformation is approximately determined based on the landmark pairs ignoring the localization

uncertainty information. Second, the affine part A is used to compute the combined covariance matrices Σ_j and third, the combined matrices are included into a non-rigid landmark registration scheme, for example by using the weighted landmark distance measure from above.

If an approximate calculation of the affine-linear transformation is not accurate enough or a rigid registration transformation y is sought directly, which considers the localization uncertainties, the combination formula from above cannot be applied. For general covariance matrices the reference covariance matrices have to be transformed by the sought transformation y leading to challenging registration problems (see 11.19). Fortunately, in many practical problems the shapes of the error ellipsoids of a corresponding landmark pair are very similar. The reason for that is that usually the directions of the eigenvectors and the shape of the covariances correspond to the geometry of the underlying structures. For example, the localization uncertainty of landmarks at vessel segments is high in the direction of the vessel segments and low perpendicular to the segments. This holds for the reference and the corresponding template landmark because they are located at the same anatomical structure. After rigid registration even the orientations of the error ellipsoids resp. the directions of the eigenvectors of the covariance matrices are similar. Therefore we will neglect the covariance matrices of the reference landmarks more or less and only take the covariance matrices of the template landmarks into account. The global size resp. the eigenvalues of the reference covariance matrices can be added to the template covariance matrices if needed. In Chapter 15, the choice of the covariance matrices in a concrete application will be presented. Exceptions are scenarios where an additional error source is present in one of the landmark sets, which is not caused by the geometry of the underlying anatomical structure. Image to patient registration is an example where in one landmark set the additional localization uncertainty caused by the tracking system is present. This changes the shape of the covariance matrices for this set.

11.5.2 Weighted Landmark Distance Measure

In the previous section we showed how covariance matrices for the joined localization uncertainties of reference and template landmarks can be computed by combining the single covariance matrices of corresponding landmarks. The uncertainty information can be considered in a weighted landmark distance measure. The Mahalanobis distance known from statistics and introduced by Mahalanobis in 1936 [138] is an obvious weighted distance measure:

$$D^{\text{WLM}}[y] = \sum_{j=1}^{N_{LM}} \|y(r_j) - t_j\|_{W_j}^2 = \sum_{j=1}^{N_{LM}} (y(r_j) - t_j)^T W_j (y(r_j) - t_j) \quad (11.24)$$

with $W_j = \Sigma_j^{-1}$ being the inverse of the combined covariance matrix of the corresponding landmarks t_j and r_j . This means the higher the location uncertainty (high error variance) the lower the weighting. By definition the covariance matrix is at least positive semi-definite. For the covariances of our landmarks we assume positive definite matrices. Thus the weighting matrices are also positive definite, because the inverse of a positive definite matrix is again positive definite.

Rohr et al. [195] introduced a notation of the weighted landmark distance measure, where all landmarks are stacked successively into big vectors \hat{r} and \hat{t} (see 11.4). They put the weight matrices $W_j \in \mathbb{R}^3 \times \mathbb{R}^3$ onto the diagonal of a big block-diagonal weight matrix $\hat{W} \in \mathbb{R}^{3N_{LM}} \times \mathbb{R}^{3N_{LM}}$:

$$\hat{W} = \begin{pmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_{N_{LM}} \end{pmatrix} \quad (11.25)$$

with

$$W_j = \begin{pmatrix} w_{11j} & w_{12j} & w_{13j} \\ w_{21j} & w_{22j} & w_{23j} \\ w_{31j} & w_{32j} & w_{33j} \end{pmatrix}. \quad (11.26)$$

This notation can be converted into our vector notation r and t of landmarks sorted by spatial coordinate directions (see Eq. 11.5) by the permutation matrix P defined in Equation 11.6. With the abbreviation $P = P_{r \rightarrow \hat{r}}$ and $\hat{x} = Px$ we get

$$\hat{x}^T \hat{W} \hat{x} = (Px)^T \hat{W} (Px) = x^T P^T \hat{W} P x = x^T W x. \quad (11.27)$$

This means the weighted scalar product $\hat{x}^T \hat{W} \hat{x}$ can be rewritten by $x^T W x$. The weight matrix \hat{W} is converted into W by multiplying the permutation matrix from left and right: $W = P^T \hat{W} P$. The resulting matrix $W \in \mathbb{R}^{3N_{LM} \times 3N_{LM}}$ looks like:

$$W = \begin{pmatrix} w_{111} & & 0 & w_{121} & & 0 & w_{131} & & 0 \\ & \ddots & & & \ddots & & & \ddots & \\ 0 & & w_{11N_{LM}} & 0 & & w_{12N_{LM}} & 0 & & w_{13N_{LM}} \\ \hline w_{211} & & 0 & w_{221} & & 0 & w_{231} & & 0 \\ & \ddots & & & \ddots & & & \ddots & \\ 0 & & w_{21N_{LM}} & 0 & & w_{22N_{LM}} & 0 & & w_{23N_{LM}} \\ \hline w_{311} & & 0 & w_{321} & & 0 & w_{331} & & 0 \\ & \ddots & & & \ddots & & & \ddots & \\ 0 & & w_{31N_{LM}} & 0 & & w_{32N_{LM}} & 0 & & w_{33N_{LM}} \end{pmatrix}.$$

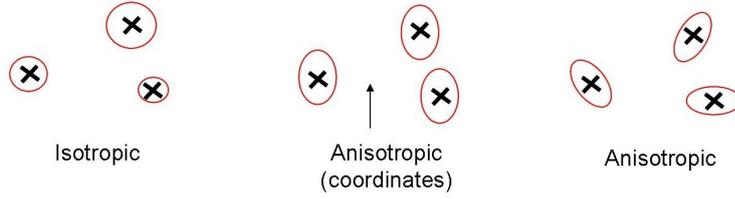


Figure 11.2 Three different types of landmark uncertainties: isotropic (error spheres), anisotropic in coordinate direction (arrow) and arbitrarily anisotropic (error ellipsoids).

The weighted landmark distance measure (11.24) can be rewritten with this notation as:

$$D^{\text{WLM}}[y, r, t] = (y(r) - t)^T W (y(r) - t). \quad (11.28)$$

This is a suitable notation for the general optimization framework and will be used for example in chapter 12.

The localization uncertainties can be classified into different types of uncertainty (see Fig. 11.2): isotropic in all spatial directions (error spheres), anisotropic in coordinate directions and general anisotropic concerning the local coordinate system of each landmark (error ellipsoids).

We begin with a special isotropic case. If the covariance matrices are all identity matrices $\Sigma_j = I_3, j = 1, \dots, N_{LM}$ the weight matrix is also an identity matrix $W = I_{3N_{LM}}$ and the landmark distance measure is just the Euclidean distance. In the general isotropic case there is one weight for each landmark pair: $a_1, \dots, a_{N_{LM}}$. The covariance matrices are different diagonal matrices for different landmarks, but with all values on the diagonal identical: $\Sigma_j = \frac{1}{a_j} I_3, a_j \in \mathbb{R}^+, j = 1, \dots, N_{LM}$. In this case the landmarks are weighted relative to each other. The big weight matrix W is also diagonal with elements: $w_{111} = w_{221} = w_{331} = a_1, \dots, w_{11N_{LM}} = w_{22N_{LM}} = w_{33N_{LM}} = a_{N_{LM}}$.

In the general anisotropic case, the errors in different coordinate directions are correlated, i.e., the covariance matrices Σ_j are not diagonal. But it is assumed that there is no correlation between *different* landmark pairs leading to a block diagonal weight matrix W (see equation 11.25). This is the case we will use most frequently. Anisotropic localization uncertainties in coordinate directions (different weights for different spatial coordinates) is a special case for anisotropic weighting. For each coordinate direction in 3D space one weight a_1, a_2, a_3 is determined. The weights are distributed on the diagonal of W : $w_{111} = \dots w_{11N_{LM}} = a_1, w_{221} = \dots w_{22N_{LM}} = a_2, w_{331} = \dots w_{33N_{LM}} = a_3$. All other elements of W are zero.

In the next sections we will present unweighted rigid and non-rigid landmark based registration schemes and their weighted extensions.

11.6 Rigid Landmark Registration

Landmark registration with rigid transformations is widely used in computer assisted surgery. As the shape of bones does not change between different image acquisitions of the same patient rigid transformations are suitable to transform preoperative images in neurosurgery or orthopedics to the real-world patient geometry in the OR. The big advantage of such low-dimensional transformations is that only few landmarks are needed to determine a rigid transformation unambiguously. Theoretically only three landmark pairs define a unique rigid transformation. Different representations of rigid transformations exist mainly for the rotation (see section 7.3). Usually the rotation and translation are computed separately.

For the conventional unweighted rigid landmark registration there is a closed form solution, but for the weighted case only heuristics and iterative solutions have been published. We introduce a straight forward numerical algorithm based on a parameterization of the rigid transformation and a Gauss-Newton optimization.

11.6.1 Unweighted Rigid Landmark Registration

The first formulations and solutions of the rigid landmark registration problem can be found in Psychometrics (quantitative Psychology). They call it the "Orthogonal Procrustes Problem" and, at first, only the rotational part of a rigid transformation was taken into account. The terminology¹ is due to Hurley and Cattell [92]. There, the rotation is represented as an orthogonal (3×3) matrix, hence the term Orthogonal Procrustes problem. The Orthogonal Procrustes problem is a constrained optimization problem or to be more precise a constrained least squares problem. In vector norm notation the problem reads as:

$$\begin{aligned} \sum_j \|Q^{\text{rot}} r_j + v - t_j\|_2^2 & \xrightarrow{Q^{\text{rot}}, v} \min, \\ \text{subject to } Q^{\text{rot}} & \in SO(3) \end{aligned} \quad (11.29)$$

with $Q^{\text{rot}} \in \mathbb{R}^3 \times \mathbb{R}^3$ a matrix representing the rotation and $v \in \mathbb{R}^3$ the translation. The *rot* tag is left for ease of notation: $Q = Q^{\text{rot}}$. $SO(3)$ is the special orthogonal group, meaning Q is an orthogonal matrix ($Q^T Q = I_3$) and $\det(Q) = +1$.

In psychometrics, they usually prefer a formulation in matrix norm notation

¹Procrustes (the stretcher) is a figure from Greek Mythology, who had an iron bed, on which he invited passerby to lie down. Then he fitted his victim into his bed by either stretching him, if he was too short or chopping off his legs if he was too tall.

is preferred:

$$\begin{aligned} & \|\bar{r}Q + \mathbf{1}_{N_{LM}}v^T - \bar{t}\|_F^2 \stackrel{Q,v}{\rightarrow} \min, \\ & \text{subject to } Q^T Q = I_3, \\ & \det(Q) = 1 \end{aligned} \quad (11.30)$$

with \bar{r}, \bar{t} being the matrices containing the landmarks row-wise as defined in section 11.3 and $\mathbf{1}_{N_{LM}} = (1, 1, \dots, 1)^T$ being a vector with all elements being ones. The used Frobenius matrix norm is defined as $\|A\|_F = \sqrt{\text{tr}(AA^T)}$ where tr is the matrix trace.

Hurley and Cattell only introduced the unconstrained linear least squares problem (no orthogonality constraint, no translation and scaling), which is equivalent to an affine landmark registration. The first solutions for the orthogonal Procrustes problem were given by Green [73] for full rank matrices and by Schönemann [208, 209] for general matrices, which may be also rank-deficient. Farrel and Stuelbnagel fixed the problem of reflections. Schönemann and Carrol [210] extended the orthogonal Procrustes problem also to translations and global scaling and called it Extended Orthogonal Procrustes Problem. They described the centering of the data to separate the determination of rotation and translation as will be described below. The Orthogonal Procrustes Problem was rediscovered by Horn [90] and Arun et al. [2] for applications in photogrammetry. In this community the problem is called "Absolute Orientation Problem". The general solution considering the reflection problem was rediscovered by Umeyama [231].

Here we give the common closed-form solution by means of centering the data like Schönemann and the singular value decomposition (SVD). One derivation can be found for in [238]. At first, the means of the reference and template landmark sets are translated onto each other. Therefore let $A = \bar{r} - \mathbf{1}_{N_{LM}}r_m^T$ and $B = \bar{t} - \mathbf{1}_{N_{LM}}t_m^T$ be the centered landmark sets where r_m and t_m are the mean vectors of r_j and $t_j, j = 1, \dots, N_{LM}$. Now the rotation matrix Q is given by solving

$$\min_Q \|AQ - B\|_F^2 \quad \text{subject to} \quad Q \in \text{SO}(3). \quad (11.31)$$

The Frobenius norm can be rewritten to

$$\|AQ - B\|_F^2 = \text{tr}((AQ - B)^T(AQ - B)) \quad (11.32)$$

$$= \text{tr}(Q^T A^T A Q) - 2\text{tr}(AQB^T) + \text{tr}(B^T B) \quad (11.33)$$

$$= \|A\|_F^2 - 2\text{tr}(AQB^T) + \|B\|_F^2 \quad (11.34)$$

because $\text{tr}(Q^T A^T A Q) = \text{tr}(AQQ^T A^T)$ and $Q^T Q = I_3$. Since the first and last term are constant with regard to Q , the term $\text{tr}(AQB^T) = \text{tr}(B^T A Q)$ has to be maximized to minimize the whole expression. Let USV^T be a singular

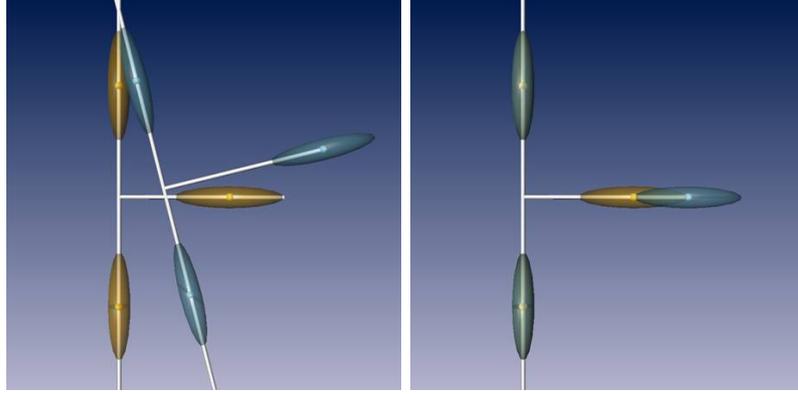


Figure 11.3 Anisotropic landmark uncertainties at vessel segments. Before registration (left) and after (right).

value decomposition of $B^T A$, then

$$\text{tr}(B^T A Q) = \text{tr}(U S V^T Q) = \text{tr}(S V^T Q U) = \text{tr}(S Z) = \sum_{i=1}^3 \sigma_{ii} z_{ii} \quad (11.35)$$

where $Z = V^T Q U$. Since Z is orthonormal, any $z_{ij} \leq 1$ for all $i, j = 1, 2, 3$. This means the sum in equation 11.35 is maximized, if $Z = I_3$ leading to $Z = I_3 = V^T Q U$ and the solution Q_* is given by $Q_* = V U^T$. To ensure that Q_* is a rotation and not a reflection a positive determinant has to be enforced:

$$Q_* = V \cdot \text{diag}([1, 1, \det(V \cdot U^T)]) \cdot U^T. \quad (11.36)$$

As we centered the landmarks to get the optimal rotation Q_* , the translation v is determined by:

$$v = t_m - Q_* \cdot r_m. \quad (11.37)$$

11.6.2 Weighted Rigid Landmark Registration

There are different applications, where anisotropic landmark localization uncertainties occur motivating the use of weighted rigid landmark registration techniques: a) the anisotropic position error of CRFs (Coordinate Reference Frames) used in navigation systems [7, 254], b) the pre-registration of weighted non-rigid TPS and GEBS schemes (see sections 11.8 and 11.9), and c) interactive initial registration of liver vessels. The last application concerning the interactive landmark registration is motivated by the fact that vessel segments can often be identified easier and can better be localized than the branching points resp. that the number of identifiable corresponding segments is higher than the number of identifiable corresponding branching points. The high

localization uncertainty of segment landmarks in the direction of the vessels has to be taken into account in the registration process. In Fig. 11.3 rigid segment landmark registration with anisotropic uncertainties is illustrated.

In section 11.5 we explained that we can differentiate between anisotropic localization uncertainties of the *coordinates* and the general anisotropic localization uncertainty for *each single landmark*.

Coordinates with Anisotropic Localization Uncertainties

Rigid registration with anisotropic coordinate uncertainties is equivalent to the Weighted Orthogonal Procrustes Problem (WOPP) from Psychometrics. The weighting matrix is just a 3×3 matrix describing the confidence in the three spatial coordinates. In contrast to the weighted landmark distance introduced in Section 11.5.2, now the weighting matrix $\tilde{W} \in \mathbb{R}^{3 \times 3}$ is multiplied from the *right* instead from the left to the difference between template and transformed reference landmarks:

$$\left\| (\bar{r}Q^{\text{rot}} + \mathbf{1}_{N_{LM}}v^T - \bar{t})\tilde{W} \right\|_F^2 \xrightarrow{Q^{\text{rot}}, v} \min \quad \text{subject to } Q^{\text{rot}} \in \text{SO}(3). \quad (11.38)$$

We choose to use the matrix norm notation from psychometrics again because it is commonly used in the literature (see for example [8]). There is no known closed-form solution like in the unweighted case and only two iterative algorithms, one by Koschat and Swayne [104] and one by Chu and Trendafilov [35]. Batchelor and Fitzpatrick [8] propose a modified version of the iterative algorithm of Koschat and Swayne. In contrast to Koschat and Swayne who only cover rotations, they also consider translations and showed, that by translating the centroids of the landmarks onto each other first the rotation can be determined and then the translation is computable by means of the resulting rotation like in the unweighted case.

We already showed in Section 11.5.2 that the anisotropic coordinate weighting can be formulated as a special case of the general anisotropic weighted landmark problem, which will be discussed in the next section. It will be presented an iterative algorithm, which actually also solves this special case.

General Anisotropic Localization Uncertainties for Single Landmarks

If each single landmark features an individual anisotropic localization uncertainty, a more general weighted landmark registration scheme is needed. Let the weighting matrix W be defined as in Section 11.5.2 and let the localization uncertainty covariance matrices of corresponding landmark pairs already be combined as discussed in Section 11.5.1. We use the parameterized form of a rigid transformation: $y(\gamma^{\text{rig}}, r) = Q(r)q(\gamma^{\text{rig}})$ with $Q(r) \in \mathbb{R}^{3N_{LM} \times 9}$, $\gamma^{\text{rig}} \in \mathbb{R}^6$ and the non-linear mapping $q(\gamma^{\text{rig}}) \in \mathbb{R}^{9 \times 1}$ defined in Equation 7.9. The matrix $Q(r) = I_3 \otimes (\bar{r}, \mathbf{1}_{N_{LM}})$ is defined like the matrix $Q(\mathbf{x})$ for a grid \mathbf{x} in-

roduced in Section 7.3. We omit the tag *rig* at the parameter vector $\gamma = \gamma^{\text{rig}}$ for ease of notation.

Then the weighted rigid landmark registration problem is defined by the following optimization problem

$$\mathcal{D}^{\text{WLM}}(\gamma) = \frac{1}{2} \|Q(r)\gamma - t\|_W^2 = \frac{1}{2} ((Q(r)\gamma - t)^T W (Q(r)\gamma - t)) \xrightarrow{\gamma} \min. \quad (11.39)$$

Balachandran et al. [7] call it the "Anisotropic Orthogonal Procrustes Problem". For this problem, too, no closed-form solution has been published so far. Balachandran et al. solve the problem only approximately with a heuristic algorithm. First they determine an approximative rotation based on the unweighted formulation, then the weighted centroids are computed and the actual rotation is solved for the demeaned version using the weighted centroids. At last the translation is computed by means of the rotation and the weighted centroids. There is a quite complicated *iterative* algorithm (see [165]) for solving the anisotropic orthogonal Procrustes problem.

In the following we present our own straight forward iterative solution by a Gauß-Newton optimization scheme. Therefore the weighted registration problem is represented by an outer function ϕ and a residual function \tilde{r} :

$$\mathcal{D}^{\text{WLM}}(\gamma) = \frac{1}{2} \|Q(r)q(\gamma) - t\|_W^2 = \phi(\tilde{r}(\gamma)) \xrightarrow{\gamma} \min \quad (11.40)$$

where $\phi(x) = \frac{1}{2}x^T x$ and $\tilde{r}(\gamma) = Q(r)q(\gamma) - t$. Since the residual function \tilde{r} is nonlinear in γ a numerical (iterative) optimization approach is taken. For a Gauß-Newton method the analytical derivatives of residual and outer function with respect to the parameters γ are needed:

$$d_\gamma \tilde{r}(\gamma) = Q(r) \cdot d_\gamma q(\gamma), \quad d_\gamma \phi(\tilde{r}(\gamma)) = \tilde{r}(\gamma), \quad d_\gamma^2 \phi(x) = 1. \quad (11.41)$$

The derivative $d_\gamma q(\gamma)$ of the non-linear parameter mapping q has already been given in Sec. 7.3.

For the determination of the Gauß-Newton search direction p_k in iteration k the linear system $H^G f(x_k)p_k = -\nabla f(x_k)$ has to be solved (see Sec. 5.3). Here the approximative Hessian and the gradient of the objective function are:

$$H_G = d\tilde{r}^T d^2 \phi(\tilde{r}) d\tilde{r}, \quad \nabla \mathcal{D}^{\text{WLM}} = d\tilde{r}^T \tilde{r}. \quad (11.42)$$

Now the Gauß-Newton equations for the weighted rigid registration problem can be set up:

$$d\tilde{r}(\gamma_k)^T d\tilde{r}(\gamma_k)p_k = -d\tilde{r}(\gamma_k)^T \tilde{r}(\gamma_k) \quad (11.43)$$

where γ_k is the parameter vector from the last iteration. More details concerning the Gauß-Newton optimization algorithm can be found in the Section 5.3 about unconstrained optimization.

An alternative to the unconstrained optimization problem based on a parameterization of the rigid transformation is to directly optimize the constrained optimization problem with respect to the rigid transformation $y(x) = Q^{\text{rot}}x + v$:

$$\mathcal{D}^{\text{WLM}}(y) \xrightarrow{y} \min \quad \text{subject to} \quad Q^{\text{rot}} \in \text{SO}(3). \quad (11.44)$$

Fig. 11.4 shows the landmark based registration result using an unweighted (conventional) and a weighed landmark distance measure for a simulated situation. The optimization is based on a SQP-algorithm for constrained problems (see Sec. 5.4). The orthogonality constrained $Q^T Q = I_3$ was used instead of $Q \in \text{SO}(3)$. The determinant condition $\det(Q) = +1$ is only checked at the end of the optimization process.

11.7 Affine Linear Landmark Registration

If besides rotation and translation scaling and shearing in a landmark registration problem have to be considered, the sought transformation y is modeled by an affine linear function $y = Ax + v$ as defined in section 7.2. Affine linear landmark registration is much easier to solve than rigid registration because the orthogonality constraint can be omitted.

11.7.1 Unweighted Affine Linear Landmark Registration

Let $y^{\text{aff}}(\gamma^{\text{aff}}, r) = Q(r)\gamma^{\text{aff}}$ with $\gamma^{\text{aff}} \in \mathbb{R}^{12}$ be the parameterized affine linear function as defined in Section 7.2. We omit the *aff* tag at the parameters and just write $\gamma = \gamma^{\text{aff}}$. Then the unweighted affine linear landmark registration problem can be formulated as:

$$\mathcal{D}^{\text{LM}}(\gamma) = \frac{1}{2} \|Q(r)\gamma - t\|^2 = \frac{1}{2} ((Q(r)\gamma - t)^T (Q(r)\gamma - t)) \xrightarrow{\gamma} \min. \quad (11.45)$$

The unweighted affine linear landmark registration problem is just an unconstrained linear least squares problem. By differentiation of the objective function with respect to γ and setting this derivative to zero we get the normal equations:

$$d_{\gamma} \mathcal{D}^{\text{LM}}(\gamma) = Q(r)^T ((Q(r)\gamma - t)) = 0, \quad (11.46)$$

$$\iff Q(r)^T Q(r)\gamma = Q(r)^T t. \quad (11.47)$$

By solving this linear system we get the parameters γ_* minimizing the objective function and thus the landmark registration problem is solved by the affine linear transformation $y_*^{\text{aff}} = Q(r)\gamma_*$.

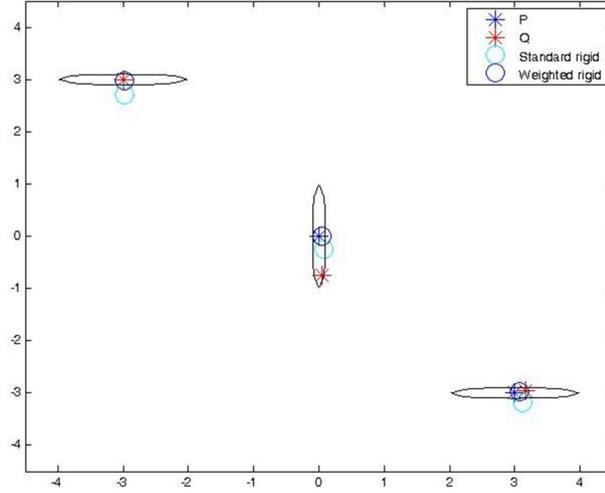


Figure 11.4 Let the true positions of the landmarks Q^{true} be exactly the same as the landmark P (blue stars). The landmarks Q^{true} are now disturbed by simulated anisotropic localization uncertainties illustrated by the ellipses. The resulting landmarks are denoted by Q (red stars). Then a rigid transformation with the translation $(-1.0, 0.5)$ and a rotation about the origin of 30 degrees is simulated and applied to the landmarks P (not illustrated). Now the transformation which transforms the landmarks P back to the landmarks Q is determined by the unweighted and weighted rigid landmark registration methods. Due to the localization uncertainties the landmarks P cannot be mapped back to their original position (blue stars) by the unweighted approach (bright blue circles). But if the uncertainties are considered by a weighted approach the original positions can be reproduced almost exactly (dark blue circles). The original positions of P are exactly the same positions as Q^{true} .

11.7.2 Weighted Affine Linear Landmark Registration

Like in rigid landmark registration anisotropic localization uncertainties of the landmarks can also be considered by a weighted landmark distance measure in affine linear landmark registration. Let the weighting matrix W be defined as in Section 11.5.2 and the affine linear transformation be parameterized as in the previous section. Then the weighted affine linear landmark registration problem is defined as:

$$\mathcal{D}^{\text{WLM}}(\gamma) = \frac{1}{2} \|Q(r)\gamma - t\|_W^2 = \frac{1}{2} ((Q(r)\gamma - t)^T W (Q(r)\gamma - t)) \xrightarrow{\gamma} \min. \quad (11.48)$$

We omit the dependence from the reference landmarks r of the matrix $Q = Q(r)$ for ease of notation. As in the unweighted case the first derivative of the distance measure is determined and set to zero. For the computation of $d_\gamma \mathcal{D}^{\text{WLM}}$ the distance measure is rearranged to

$$\frac{1}{2}(Q\gamma - t)^T W(Q\gamma - t) = \frac{1}{2}(\gamma^T Q^T W Q \gamma - 2t^T W Q \gamma + t^T W t). \quad (11.49)$$

By setting the derivative of the weighted distance measure to zero we now get the weighed normal equations:

$$d_\gamma \mathcal{D}^{\text{WLM}}(\gamma) = Q^T W Q \gamma - t^T W Q = 0, \quad (11.50)$$

$$\iff Q^T W Q \gamma = Q^T W t. \quad (11.51)$$

The matrix $Q^T W Q$ is symmetric, because the weight matrix W is symmetric: $(Q^T W Q)^T = Q^T W^T Q = Q^T W Q$. This property was used for the first term of the derivative to get: $(Q^T W Q)^T \gamma = Q^T W Q \gamma$. By solving the weighted normal linear equations we get the parameters γ_* minimizing the objective function as in the unweighted case and thus the landmark registration problem is solved by the affine linear transformation $y_*^{\text{aff}} = Q(r)\gamma_*$.

11.8 Thin-Plate Splines (TPS)

Affine-linear transformations possess six degrees of freedom more than rigid transformations and thus offer global non-rigid transformations. But the degrees of freedom are still quite few and local non-rigidities can not be handled. Thin-plate splines are a well-known class of locally non-rigid transformations often used for landmark-based registration of soft-tissues in computer assisted surgery. We will briefly summarize *interpolating* thin-plate splines and show how they can be extended to *approximating* thin-plate splines enabling the incorporation of anisotropic landmark localization uncertainties. In addition, the penalizer-based approximating thin-plate splines are further extended to a novel approach based on inequality landmark constraints.

Thin-plate splines [50, 243] are smooth functions where the smoothness is controlled by a functional (energy) on the derivatives of the spline function similar to the regularizers on the transformation in nonparametric image registration. In fact, the thin-plate spline energy is closely related to the curvature regularizer (see [150], p.123). The thin-plate spline functional smoothes each coordinate function y_i of an arbitrary (but differentiable) transformation $y = (y_1, y_2, y_3)^T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ separately:

$$\mathcal{S}^{\text{TPS}}[y] = \sum_{i=1}^3 \mathcal{S}^{\text{tps}}[y_i]. \quad (11.52)$$

The functionals on the coordinate functions $y_i : \mathbb{R}^3 \rightarrow \mathbb{R}$ are defined as:

$$\mathcal{S}^{\text{tps}}[y_i] = \int_{\mathbb{R}^3} \text{tr}(\nabla^2 y_i \cdot \nabla^2 y_i) dx \quad (11.53)$$

where $\text{tr}(A)$ is again the trace of a matrix A and $\nabla^2 y_i$ the Hessian matrix of y_i . The functional can be expanded to:

$$\mathcal{S}^{\text{tps}}[y_i] = \int_{\mathbb{R}^3} (\partial_{11}y)^2 + (\partial_{22}y)^2 + (\partial_{33}y)^2 + 2((\partial_{12}y)^2 + (\partial_{13}y)^2 + (\partial_{23}y)^2) dx. \quad (11.54)$$

The TPS-functional $\mathcal{S}^{\text{TPS}}[y]$ describes the linearized version of the bending energy of a thin plate. In principle the TPS-functional can be formulated for arbitrary dimensions and also higher order [195]. Here the order is 2 because only second derivatives are involved.

11.8.1 Interpolating TPS

Interpolating thin-plate splines were introduced by Bookstein [20] to describe non-rigid deformations (warps). He minimized the thin-plate spline functional with respect to the transformation $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ while ensuring the landmark interpolation conditions by equality constraints:

$$\mathcal{J}^{\text{iTPS}}[y] = \mathcal{S}^{\text{TPS}}[y] \xrightarrow{y} \min \quad \text{subject to} \quad y(r) - t = 0. \quad (11.55)$$

This means we are looking for the smoothest function with respect to the TPS-functional, which exactly maps each reference landmark to its corresponding template landmark.

A nice property of this constrained minimization problem is that its solution is representable by a parameterization $y^{\text{tps}}(\gamma^{\text{tps}})$ of the transformation. The parameters $\gamma^{\text{tps}} = (\gamma^1, \gamma^2, \gamma^3)^T \in \mathbb{R}^{3(N_{LM}+4)}$ are the solutions of three independent linear systems. For a detailed derivation see [147]. Since the coordinate functions y^i are separately smoothed by the thin-plate spline functional, the corresponding parameters γ^i can be determined independently to get $y^i = y^i(\gamma^i)$. The parameterization is given by a linear combination of shifted radial basis functions (RBF) $\rho_j(x) = \|x - r_j\|$ around the reference landmarks r_j and a polynomial term describing the affine-linear part of the transformation:

$$y^i(x) = \sum_{k=1}^{N_{LM}} c_k^i \|x - r_k\| + a_0^i + a_1^i x_1 + a_2^i x_2 + a_3^i x_3 \quad (11.56)$$

where $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$.

Now this parameterization of y^i is inserted into the landmark interpolation condition $t_j = y(r_j)$ for landmark j :

$$t_j^i = y^i(r_j) = (\|r_j - r_1\| \quad \dots \quad \|r_j - r_{N_{LM}}\|) \begin{pmatrix} c_1^i \\ \vdots \\ c_{N_{LM}}^i \end{pmatrix} + \begin{pmatrix} 1 & r_1^j & r_2^j & r_3^j \end{pmatrix} \begin{pmatrix} a_0^i \\ a_1^i \\ a_2^i \\ a_3^i \end{pmatrix} \quad (11.57)$$

where t^i are the components of the template landmarks in coordinate direction i resp. the i -th column of the landmark matrix $t^i = \bar{t}_{:,i}$.

The parameters $\gamma_c^i = (c_1^i, \dots, c_{N_{LM}}^i)^T$, $\gamma_a^i = (a_0^i, a_1^i, a_2^i, a_3^i)^T$ are now summarized to one parameter vector $\gamma^i = (\gamma_c^i, \gamma_a^i)$. Together with the matrix $A = (\mathbf{1}_{N_{LM}} \bar{r})$ for the affine part and the matrix $B = [\|r_j - r_k\|]_{j,k}$ for the locally non-rigid part the interpolation condition for all landmarks can be written as:

$$t^i = B\gamma_c^i + A\gamma_a^i. \quad (11.58)$$

If we add the boundary conditions $A^T \gamma_c^i = 0$ to the interpolation conditions we get the linear system:

$$Q^{\text{tps}}(r)\gamma^i = \begin{pmatrix} B & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} \gamma_c^i \\ \gamma_a^i \end{pmatrix} = \begin{pmatrix} t^i \\ 0 \end{pmatrix} \quad (11.59)$$

where the system matrix $Q^{\text{tps}}(r)$ depends on the reference landmarks r but is equivalent for all components y^i . The boundary conditions $A^T \gamma_c^i = 0$ ensure that the locally non-rigid part of the transformation is zero at infinity.

The solution of this linear system provides the parameters $\gamma^i = (\gamma_c^i, \gamma_a^i)$ for the i -th component function y^i of the transformation of an arbitrary point $x \in \mathbb{R}^3$ (see Eq. 11.56).

In 3D the minimal number of landmarks needed to determine a thin-plate spline transformation is $N_{LM} = 5$. The null space of the thin-plate spline functional \mathcal{S}^{TPS} is spanned by the affine linear part of the parameterized transformation because only second derivatives are considered by the functional. For an affine linear transformation the minimal number of landmarks is $n = 4$. This means for a thin-plate spline registration at least one additional landmark is needed compared to an affine linear registration. In addition not all landmarks may lie on a single line or plane.

The popularity of thin-plate spline registration is due to the facts that the resulting transformation function is smooth and derivatives of any order exist, that no free parameters are involved, which need tuning, that there is a closed-form solution and that a physical interpretation is given.

11.8.2 Approximating TPS

Instead of formulating an interpolation problem based on the thin-plate spline functional $\mathcal{S}^{\text{TPS}}[y]$ and equality constraints for the matching of the landmarks we can formulate an approximation problem based on an unconstrained penalizer approach:

$$\mathcal{J}^{\text{aTPS}}[y] = \mathcal{D}^{\text{LM}}[y] + \theta \mathcal{S}^{\text{TPS}}[y] \xrightarrow{y} \min. \quad (11.60)$$

Now the reference landmarks are not exactly mapped to their corresponding template landmarks anymore but the sum of the distances between them is kept as small as possible. The landmark distance measure:

$$\mathcal{D}^{\text{LM}}[y] = \|y(r) - t\|_{LM}^2 \quad (11.61)$$

was already defined in Equation 11.10. The weighting parameter θ can be tuned by the user to emphasize the smoothing functional \mathcal{S}^{TPS} against the landmark distance penalizer \mathcal{D}^{LM} . It turns out that the higher θ the smoother the resulting transformation but the lower the approximation quality at the landmark pairs.

Interestingly, there is a very similar analytic solution for this *approximative* thin-plate spline problem as for the *interpolating* thin-plate spline problem. The same type of parameterized transformation $y^i(\gamma^i)$ as defined in Equation 11.56 solves the approximation problem [50, 243, 147, 117]. The only difference is the additional term $\theta I_{N_{LM}}$ in the resulting linear system:

$$Q^{\text{tps}}(r)\gamma^i = \begin{pmatrix} B + \theta I_{N_{LM}} & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} \gamma_c^i \\ \gamma_a^i \end{pmatrix} = \begin{pmatrix} t^i \\ 0 \end{pmatrix}. \quad (11.62)$$

Until now we used Euclidean distances for corresponding landmarks in the landmark distance measure \mathcal{D}^{LM} . But in contrast to interpolating thin-plate splines also weighted distances can reasonably be used in the approximative formulation leading to the weighted distance measure (see Sec. 11.5.2):

$$\mathcal{D}^{\text{WLM}} = \|y(r) - t\|_W = (y(r) - t)^T W (y(r) - t). \quad (11.63)$$

This distance measure allows the incorporation of landmark localization uncertainties into the registration scheme.

A separation into independent linear systems for the transformation components y^i is not possible anymore. Nevertheless the derivation of an analytic solution for the weighted approximative TPS registration problem is still possible [245]. First we show the analytic solution using the landmark and weight matrix notation of Rohr et al. [195, 194] and then we convert this notation to our standard notation.

Rohr et al. [195, 194] stacked the landmarks into long vectors $\hat{r}, \hat{t} \in \mathbb{R}^{3N_{LM}}$ and defined the block diagonal weight matrix $\hat{W} \in \mathbb{R}^{3N_{LM} \times 3N_{LM}}$ where each

of the blocks is an individual weight matrix $W_j \in \mathbb{R}^{3 \times 3}$ for each pair of landmarks (see Sec. 11.3). In addition they stacked the searched parameters into long vectors: $\hat{\gamma}_c = (c_1^1, c_1^2, c_1^3, \dots, c_{N_{LM}}^1, c_{N_{LM}}^2, c_{N_{LM}}^3) \in \mathbb{R}^{3N_{LM}}$ and $\hat{\gamma}_a = (a_0^1, a_0^2, a_0^3, \dots, a_3^1, a_3^2, a_3^3)$. Then they replaced the matrices A and B in the linear system for the solution of the unweighted approximating TPS problem (Eq. 11.62) by the matrices:

$$A_{\hat{W}} = (A_{ij} \cdot I_3) = A \otimes I_3 \quad \text{and} \quad B_{\hat{W}} = (B_{ij} \cdot I_3) = B \otimes I_3. \quad (11.64)$$

Now the parameters $\hat{\gamma}_c, \hat{\gamma}_a$ for the analytic solution of the weighted approximating TPS problem are determined by the linear system:

$$\begin{pmatrix} \hat{B}_W + \theta \hat{W} & \hat{A}_W \\ \hat{A}_W^T & 0 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_c \\ \hat{\gamma}_a \end{pmatrix} = \begin{pmatrix} \hat{t} \\ 0 \end{pmatrix}. \quad (11.65)$$

The notation and solution by Rohr et al. can be converted easily to our preferred notation. Let the landmarks be put coordinate-wise into the long vectors t and r like in Sec. 11.3, and the weighting matrix W be defined as in Sec. 11.5.2. Let the parameters also be ordered coordinate-wise like for interpolating TPS (see Sec. 11.8.1): $\gamma_c = (\gamma_c^1, \gamma_c^2, \gamma_c^3)$ and $\gamma_a = (\gamma_a^1, \gamma_a^2, \gamma_a^3)$. Then the matrices A and B have to be replaced by:

$$A_W = I_3 \otimes A \quad \text{and} \quad B_W = I_3 \otimes B \quad (11.66)$$

where the order of the factors in the Kronecker products is exchanged compared to the notation of Rohr et al. Now the solution of the weighted approximating TPS problem for the converted notation is the solution of the following linear system:

$$\begin{pmatrix} B_W + \theta W & A_W \\ A_W^T & 0 \end{pmatrix} \begin{pmatrix} \gamma_c \\ \gamma_a \end{pmatrix} = \begin{pmatrix} t \\ 0 \end{pmatrix}. \quad (11.67)$$

11.8.3 Inequality Constrained TPS

Instead of constraining the transformation by equality constraints as for interpolating TPS or by using a penalizer as for approximating TPS it is also possible to define an approximative scheme by *inequality* constraints:

$$\begin{aligned} \mathcal{J}[y] = & \quad \mathcal{S}^{\text{TPS}}[y] && \xrightarrow{y} \min \\ \text{s.t.} & \quad \mathcal{C}^{\text{isoLM}}[y] && \leq \text{tol}_{LM} \end{aligned}$$

with

$$\mathcal{C}^{\text{isoLM}}[y] = \begin{pmatrix} \|y(r_1) - t_1\|^2 \\ \vdots \\ \|y(r_{N_{LM}}) - t_{N_{LM}}\|^2 \end{pmatrix}$$

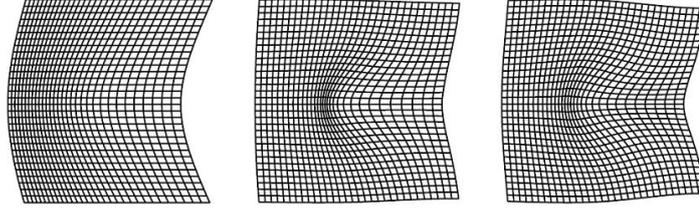


Figure 11.5 A single landmark is displaced horizontally resulting in different deformation fields using TPS (left) as well as GEBS for Poisson's ratios of $\nu = 0$ (middle) and $\nu = 0.49$ (right). It can be seen that for TPS deformations occur in horizontal direction only whereas for GEBS horizontal and vertical deformations are recognizable (courtesy of Stefan Wörz, Biomedical Computer Vision Group, DKFZ Heidelberg).

and

$$tol_{LM} = (tol_1^2, \dots, tol_{N_{LM}}^2)^T, \quad tol_j \in \mathbb{R}_0^+ . \quad (11.68)$$

The idea of this approach is to determine the smoothest transformation with respect to the TPS functional \mathcal{S}^{TPS} which keeps the transformed landmarks r_j inside error spheres around the corresponding landmarks t_j . This is a more intuitive approach than the weighting between the TPS functional and a landmark distance penalizer where the weighting factor cannot be intuitively chosen as in the approximating TPS scheme from the previous section. The practical disadvantage of the inequality constrained approach is that no analytical solution is known. But the constrained optimization problem can be solved by a numerical scheme presented in a later chapter about a combined landmark and image intensity registration approach (see Chap. 12). If the image distance measure is left in this combined approach, only the regularizer and the constraints remain, which leads exactly to the above inequality constrained TPS problem. In addition, we also consider anisotropic localization uncertainties in Chapter 12.

11.9 Gaussian Elastic Body Splines (GEBS)

The TPS functional represents a relatively coarse deformation model, because transverse contraction does not lead to longitudinal dilation (see also [264]). In comparison, Gaussian elastic body splines (GEBS) introduced by Kohlrausch et al. [103] are derived from the Navier-Lamé equation, which describes the deformation of homogeneous elastic materials and takes cross-effects between contraction and dilation into account (see Fig. 11.5):

$$\mu \Delta u + (\lambda + \mu) \nabla(\text{div} u) + f = 0 \quad (11.69)$$

where $u : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is only the displacement field and not the whole transformation $y(x) = x + u(x)$. The parameters $\lambda, \mu \in \mathbb{R}^+$, also called Lamé constants, describe the material properties. The Lamé constants are related to Poisson's ratio $\nu = \lambda/(2\lambda + 2\mu)$. The Navier-Lamé equation models the deformation u of an elastic body resulting from the application of some forces $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. Some insights into the physical properties of an elastic body modeled by the Navier equations are given in [147]. The Navier equations are the Euler-Lagrange equations of the linearized elastic potential $\mathcal{S}^{\text{elas}}$ introduced in Sec. 6.3.1 and used as the elastic regularizer for nonparametric registration. In case of nonparametric image registration the forces in the Navier equation stem from the distance measure for *image* pairs. In the following we will show how these forces can be generated by distances between *landmark* point pairs. For landmark registration with Gaussian elastic body splines, Gaussian smoothed forces around the template landmarks are used. At first, we assume the simple case of a Gaussian force field $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ around the origin:

$$f(x) = cg(\|x\|) = c \frac{1}{(\sqrt{2\pi}\sigma)^3} e^{-\frac{\|x\|^2}{2\sigma^2}} \quad (11.70)$$

where $c \in \mathbb{R}^3$ represents the strength and the direction of the forces and the standard deviation $\sigma \in \mathbb{R}^+$ of the Gaussian function $g : \mathbb{R} \rightarrow \mathbb{R}$ controls the spatial influence of the forces.

For such a force field f an analytic solution of the Navier equation can be derived [103]:

$$u(x) = \mathbf{G}(x)c \quad (11.71)$$

where $c \in \mathbb{R}^3$ and $\mathbf{G} : \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$ is a matrix-valued function defined as:

$$\begin{aligned} \mathbf{G}(x) = & \left(\frac{\alpha^2 + \sigma_f^2}{r^3} \text{erf}(\hat{r}) - \beta \frac{e^{-\hat{r}^2}}{r^2} \right) \mathbf{I}_3 \\ & + \left(\frac{r^2 - 3\sigma_f^2}{r^5} \text{erf}(\hat{r}) + 3\beta \frac{e^{-\hat{r}^2}}{r^4} \right) \mathbf{x}\mathbf{x}^T \end{aligned} \quad (11.72)$$

with $\hat{r} = r/(\sqrt{2}\sigma_f)$, $\alpha = 3 - 4\nu$, $\beta = \sigma_f \sqrt{2/\pi}$ and the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi$. \mathbf{I}_3 denotes the 3×3 identity matrix and ν is Poisson's ratio $\nu = \lambda/(2\lambda + 2\mu)$, $0 \leq \nu \leq 0.5$.

In the next section we will shift the matrix-valued function \mathbf{G} from the origin to the landmarks and use the resulting functions as basis functions for a parameterized displacement field which solves the interpolating and approximating Gaussian elastic body splines registration problem.

Interestingly, thin-plate splines can be motivated similarly if we use Dirac point forces instead of Gaussian forces for the biharmonic partial differential

equation (PDE) instead for the Navier-Lamé PDE. Then the basis functions of the thin-plate splines are fundamental solutions of the biharmonic equation [196]. For a mathematically precise formulation the theory of distributions is needed to define the Dirac points forces. Fundamental solutions are solutions of the resulting distributional PDE.

11.9.1 Interpolating GEBS

In the previous section, we considered only a single Gaussian force $f(x) = cg(\|x\|)$. Now we associate each landmark pair with a Gaussian force $f_j(x) = c_j g(\|x\|)$ and superpose all forces to an overall force field $\sum_{j=1}^{N_{LM}} f_j(x)$. Due to the linearity of the Navier-Lamé equation the solution of the Navier-Lamé equation for this superposed force field is a superposition of N_{LM} GEBS basis functions [103, 264]. Therefore we define the parametric GEBS transformation as:

$$y^{\text{GEBS}}(x) = x + \sum_{k=1}^{N_{LM}} \mathbf{G}(x - r_k) c_k \quad (11.73)$$

where $r_k \in \mathbb{R}^3$ are the reference landmarks and $c_k \in \mathbb{R}^3$ are the coefficients we are looking for. In contrast to thin-plate splines there is no affine part in this transformation. Therefore an affine linear transformation is determined and applied first and then the GEBS transformation only compensates the nonlinear displacements.

Substituting the transformation y^{GEBS} in the interpolation conditions $y(r_j) = t_j$ yields the following linear system:

$$y^{\text{GEBS}}(r_j) = r_j + \sum_{k=1}^{N_{LM}} G(r_j - r_k) c_k = t_j, \quad j = 1, \dots, N_{LM} \quad (11.74)$$

$$\iff \sum_{k=1}^{N_{LM}} G(r_j - r_k) c_k = t_j - r_j, \quad j = 1, \dots, N_{LM} \quad (11.75)$$

With $\gamma = (c_1^T, \dots, c_{N_{LM}}^T)^T$ and the landmark notation $\hat{t}, \hat{r} \in \mathbb{R}^{3N_{LM}}$ we rewrite this linear system in matrix notation:

$$Q^{\text{GEBS}}(\hat{r})\gamma = \hat{t} - \hat{r} \quad (11.76)$$

where the system matrix $Q^{\text{GEBS}} \in \mathbb{R}^{3N_{LM} \times 3N_{LM}}$ is defined as:

$$Q^{\text{GEBS}}(\hat{r}) = \begin{pmatrix} G(r_1 - r_1) & \cdots & G(r_1 - r_{N_{LM}}) \\ \vdots & \ddots & \vdots \\ G(r_{N_{LM}} - r_1) & \cdots & G(r_{N_{LM}} - r_{N_{LM}}) \end{pmatrix}. \quad (11.77)$$

The solution of this system provides the coefficients for the parametric transformation y^{GEBS} which solves the interpolating GEBS registration problem.

11.9.2 Approximating GEBS

As for thin-plate splines an approximating approach has also been introduced for Gaussian elastic body splines, which allows the consideration of landmark localization uncertainties [264]. The approximating GEBS landmark registration is represented by the following optimization problem:

$$\mathcal{J}^{\text{aGEBS}}[y] = \mathcal{D}^{\text{GWLM}}[y] + \theta \mathcal{S}^{\text{elas}}[y] \xrightarrow{y} \min \quad (11.78)$$

where $\theta \in \mathbb{R}^+$ weights the regularizer $\mathcal{S}^{\text{elas}}$ against the landmark distance measure $\mathcal{D}^{\text{GWLM}}[y]$, which is defined below. The elastic potential $\mathcal{S}^{\text{elas}}$ is exactly defined as in Sec. 6.3.1, where it is used as the elastic regularizer in nonparametric registration. We already mentioned in the introduction about GEBS that the Navier-Lamé equation is the associated PDE to the elastic potential.

Wörz et al. [264] propose the following weighted landmark distance measure:

$$\mathcal{D}^{\text{GWLM}}[y] = \int \sum_{j=0}^{N_{LM}} g(\|x - r_j\|) (t_j - y(x))^T W_j (t_j - y(x)) \, dx \quad (11.79)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is again the Gaussian function and W_j are the weight matrices resulting from the localization uncertainties of the landmarks (see Sec. 11.5.2).

The interesting point is that Wörz et al. [264] determined that the transformation y which minimizes the optimization problem (Eq. 11.78) can be represented by exactly the same parametric transformation as for the interpolating GEBS problem based on the same matrix-valued basis functions \mathbf{G} (Eq. 11.72):

$$y(x)^{\text{GEBS}} = x + \sum_{k=1}^{N_{LM}} \mathbf{G}(x - r_k) c_k. \quad (11.80)$$

Although both interpolating and approximating GEBS rely on the same basis function \mathbf{G} the resulting linear system of equations to determine the linear coefficients c_k of the parametric transformation differs. In contrast to interpolating GEBS, the linear system for approximating GEBS includes additional sums of Gaussian forces weighted by the weight matrices W_j . For a detailed description of the linear system see [264].

11.9.3 Inequality Constrained GEBS

As for thin-plate splines we introduce an inequality landmark constraints registration scheme, which is related to GEBS:

$$\mathcal{J}[y] = \begin{array}{ll} \mathcal{S}^{\text{elas}}[y] & \xrightarrow{y} \min, \end{array} \quad (11.81)$$

$$\text{s.t.} \quad \mathcal{C}^{\text{isoLM}}[y] \leq \text{tol}_{LM} \quad (11.82)$$

with $\mathcal{C}^{\text{isoLM}}[y]$ and tol_{LM} exactly defined as for TPS (see Sec. 11.8.3).

The main idea of this approach is the same as for the inequality constrained TPS: the determination of the smoothest transformation which keeps the transformed landmarks r_j inside error spheres around the corresponding landmarks t_j . But here the smoothness is defined with respect to the elastic functional $\mathcal{S}^{\text{elas}}$ instead of the TPS functional \mathcal{S}^{TPS} . The advantages (more intuitive, no additional weighting parameter) are the same as for the TPS approach. The practical disadvantage is that no analytical solution is known. But the constrained optimization problem can be solved with an approach similar to the TPS by a numerical scheme presented in a later chapter about a combined landmark and intensity registration approach (see Chap. 12).

Chapter 12

Combining Landmarks and Intensity-based Registration

12.1 Introduction

All parametric landmark registration schemes presented in the previous Chapter 11 have one disadvantage in common. The original images are used only to define the landmarks (features). The actual registration process is based exclusively on those landmarks meaning that all image contents not represented in those landmarks are completely ignored.

There are different possibilities to combine landmark and intensity information to a joint registration formulation leading to unconstrained (with penalizer), equality constrained and inequality constrained optimization problems. The main idea is to minimize the (image) distance measure while controlling the distances of corresponding landmarks.

12.2 Landmark Distances Penalizer

An obvious possibility to combine landmark and intensity information is to add a penalizer $\mathcal{P}^{\text{LM}}[y]$ measuring the distances between corresponding landmarks to the standard registration functional consisting of the distance measure \mathcal{D} and the smoother \mathcal{S} :

$$\mathcal{J}[y] = \mathcal{D}[R, T(y)] + \alpha \mathcal{S}[y - y_0] + \beta \mathcal{P}^{\text{LM}}[y, r, t] \xrightarrow{y} \min$$

with an additional weighting factor $\beta \in \mathbb{R}^+$.

The penalizer \mathcal{P} can be the (unweighted) squared Euclidean distance penalizer $\mathcal{P}^{\text{LM}}[y, r, t] = \sum_{j=1}^{N_{\text{LM}}} \|\delta_{r_j}[y] - t_j\|_2^2$ (see Eq. 11.11), but also a scheme using a weighted penalizer has been published [263]. We still have an unconstrained optimization problem. An optimize-then-discretize approach via Euler-Lagrange equations was published in [58, 57]. For a discretize-then-optimize strategy we only have to discretize the penalizer as the discretization

of the distance measure and the regularizer are already known. Then the combined landmark registration problem can be solved by the same Gauß-Newton optimization scheme as for the standard nonparametric registration problem [171]. The discretization of the penalizer will be presented in section 12.5.

12.3 Equality Landmark Constraints

As the landmark misfit is controlled by the sum of the individual landmark distances in the penalizer approach, the distance of individual landmarks can still be high after optimization. In addition, the penalizer approach suffers from parameter tuning (weighting factor β). Therefore we formulate the combined landmark and intensity registration problem as an equality constrained optimization problem:

$$\mathcal{J}[y] = \mathcal{D}[y] + \alpha\mathcal{S}[y - y_0] \xrightarrow{y} \min, \quad (12.1)$$

$$s.t. \quad \mathcal{C}^{\text{eqLM}}[y] = 0 \quad (12.2)$$

where $\mathcal{C}^{\text{eqLM}} = y(r) - t$ represents the landmark interpolation constraints $y(r) = t$. This equality constrained formulation guarantees that corresponding landmarks are exactly mapped onto each other while the minimal value for the intensity distance measure \mathcal{D} and the regularizer \mathcal{S} is sought. An interpretation of this registration approach is that the deformation y is known and fixed on the landmarks while the deformations in the space between the landmarks are smoothed by the regularizer and driven by the intensity information measured by the distance measure. By omitting the distance measure $\mathcal{D} = 0$ and setting the regularizer to the thin-plate spline functional $\mathcal{S} = \mathcal{S}^{\text{TPS}}$ we get exactly a thin-plate spline registration problem, which is based on the landmarks alone.

Numerical solutions for the equality landmark constrained optimization problem will be proposed in section 12.6.1.

12.4 Inequality Landmark Constraints

The equality landmark constraints introduced in the previous section are often too restrictive because interactive landmark definition is always prone to errors. Thus we are now presenting a new approach by considering individual landmark localization inaccuracies. Isotropic landmark localization uncertainties can be modeled by error spheres around the landmarks. The equality constraints are loosened to inequality constraints, which enables the landmarks to move inside the error spheres but guarantees that corresponding landmarks have only a restricted distance to each other defined by some tolerances, e.g. the radius of the error spheres. Then the error spheres can be

interpreted as tolerance spheres. The landmark constraint registration problem with isotropic tolerances for the landmarks distances [172] is formulated as

$$\begin{aligned} \mathcal{J}[y] = & \mathcal{D}[y] + \alpha \mathcal{S}[y - y_0] \xrightarrow{y} \min, \\ \text{s.t.} \quad & \mathcal{C}^{\text{isoLM}}[y] \leq \text{tol}_{LM} \end{aligned}$$

with

$$\mathcal{C}^{\text{isoLM}}[y] = \begin{pmatrix} \|y(r_1) - t_1\|^2 \\ \vdots \\ \|y(r_{N_{LM}}) - t_{N_{LM}}\|^2 \end{pmatrix}$$

and

$$\text{tol}_{LM} = (\text{tol}_1^2, \dots, \text{tol}_{N_{LM}}^2)^T, \quad \text{tol}_j \in \mathbb{R}_0^+ . \quad (12.3)$$

The constraints are reformulated to $\text{tol}_{LM} - \mathcal{C}^{\text{isoLM}}[y] \geq 0$ to match the constraints formulation of the general inequality constrained optimization problem 5.3 with $c(x) \geq 0$. Choosing tol_{LM} as a zero vector leads to the same solution as the equality constraint problem from the previous section.

As the localization uncertainty of a landmark might deviate in different directions (anisotropic errors) we extend the tolerance spheres to ellipsoidal tolerance volumes for each landmark pair by the weighted norm (see also Sec. 11.5.2):

$$\|y(r_j) - t_j\|_{W_j}^2 = (y(r_j) - t_j)^T W_j (y(r_j) - t_j). \quad (12.4)$$

As the weighting matrices W_j already contain the tolerances as a global scaling we formulate the inequality constraints as:

$$\mathcal{C}^{\text{WLM}}[y] \leq \mathbf{1}_{N_{LM}} \quad (12.5)$$

with

$$\mathcal{C}^{\text{WLM}}[y] = \begin{pmatrix} \|y(r_1) - t_1\|_{W_1}^2 \\ \vdots \\ \|y(r_{N_{LM}}) - t_{N_{LM}}\|_{W_{N_{LM}}}^2 \end{pmatrix}.$$

Again the inequality constraints have to be rewritten to match the general inequality constraints $c(x) \geq 0$ leading to the following inequality landmark constraint registration problem with anisotropic tolerances:

$$\begin{aligned} \mathcal{J}[y] = & \mathcal{D}[y] + \alpha \mathcal{S}[y - y_0] \xrightarrow{y} \min, \\ \text{s.t.} \quad & \mathbf{1}_{N_{LM}} - \mathcal{C}^{\text{WLM}}[y] \geq 0. \end{aligned}$$

For the determination of the weighting matrices W_j see the discussion in section 11.5.1 on how to combine covariance matrices representing the localization uncertainties of corresponding landmark pairs and see section 14.4 on how to set the covariance matrices in a practical application. Setting $W_j = I_3$ for all $j = 1, \dots, N_{LM}$ leads to the isotropic case from above.

12.5 Discretization of Landmark Constraints

The discretization of the distance measures and regularizers has already been described. Only the discretization of the equality and inequality landmark constraints is left.

12.5.1 Equality Landmark Constraints

The main problem of the discretization of the landmark constraints (or the penalizer) is that we have a discrete transformation \mathbf{y} given on a grid $\mathbf{x}(\Omega, m)$ for the evaluation of the distance measure and regularizer. But the landmarks do not necessarily lie on grid points. Thus the transformed positions $y(r_j)$ of the landmarks r_j are approximated by tri-linear interpolation of the discrete transformation values on neighbored grid points.

Linear interpolation in the 1D case of a scalar value $y(x')$ at the position x' from the values \mathbf{y}_p and \mathbf{y}_{p+1} on two neighbored grid points $\mathbf{x}_p \leq x' < \mathbf{x}_{p+1}$ is just a weighted averaging of these two neighbored values:

$$y(x') \approx (1 - \xi) \cdot \mathbf{y}_p + \xi \cdot \mathbf{y}_{p+1} \quad (12.6)$$

where $\xi = x' - \mathbf{x}_p$. This means with a vector containing just two non-zero elements $c_j = (0, \dots, 0, (1 - \xi), \xi, 0, \dots, 0)$ we can approximate the transformation y at the position $x' = r_j$ by $y(r_j) \approx c_j^T \mathbf{y}$ where \mathbf{y} is the discrete transformation.

In three dimensions three linear interpolations (tri-linear) are performed, one in each coordinate direction for each of the three components of the transformation $y = (y_1, y_2, y_3)^T$. Due to the order of the coordinates in the long vector notation of the grid $\mathbf{x}(\Omega, m)$ and the transformation \mathbf{y} (see Sec. 4.2) the two non-zero elements of the interpolation coefficient vector $c_j \in \mathbb{R}^n$ are at according positions ($n = m_1 \cdot m_2 \cdot m_3$). Then each transformation component y_i is approximated by $y_i(r_j) \approx c_j^T \mathbf{y}_{(i-1)n+1:i \cdot n}$ for $i = 1, 2, 3$ where the interpolation coefficients c_j for each component $y_i(r_j)$ stay the same. Remember that first all y_1 -components, then all y_2 - and y_3 -components are stored in the discrete transformation vector $\mathbf{y} \in \mathbb{R}^{3n}$. As each of the three components of the transformation $y = (y_1, y_2, y_3)^T$ has to be interpolated we get the following

approximation of the transformation y for a *single* landmark r_j :

$$y(r_j) \approx (I_3 \otimes c_j^T) \mathbf{y} = \begin{pmatrix} c_j^T & & \\ & c_j^T & \\ & & c_j^T \end{pmatrix} \mathbf{y} \in \mathbb{R}^3. \quad (12.7)$$

Now the interpolation coefficients $c_1, \dots, c_{N_{LM}} \in \mathbb{R}^n$ for *all* N_{LM} transformed landmarks $r_1, \dots, r_{N_{LM}}$ are collected into one matrix \tilde{C} with

$$\tilde{C} = \begin{pmatrix} c_1^T \\ \vdots \\ c_{N_{LM}}^T \end{pmatrix} \in \mathbb{R}^{N_{LM} \times n}. \quad (12.8)$$

Together with equation 12.7 the transformation of *all* landmarks can be approximated by means of the coefficient matrix $C = (I_3 \otimes \tilde{C}) \in \mathbb{R}^{3N_{LM} \times 3n}$:

$$y(r) \approx C \mathbf{y} = (I_3 \otimes \tilde{C}) \mathbf{y} = \begin{pmatrix} \tilde{C} & & \\ & \tilde{C} & \\ & & \tilde{C} \end{pmatrix} \mathbf{y} \in \mathbb{R}^{N_{LM}}. \quad (12.9)$$

Now the landmark equality constraints can be approximated by the following discrete (linear) constraints:

$$C^{\text{eqLM}} = y(r) - t \approx C \mathbf{y} - t =: C^{\text{eqLM}}. \quad (12.10)$$

Note that the coefficient matrix C depends on the location of the landmarks r and that $C \mathbf{y}$ describes the location of the transformed landmarks $y(r)$.

For the SQP strategy for constrained optimization problems (see Sec. 5.4.1) the first derivative of the discrete constraints with respect to the the transformation \mathbf{y} is needed:

$$\nabla_{\mathbf{y}} C^{\text{eqLM}}(\mathbf{y}) = C. \quad (12.11)$$

12.5.2 Inequality Landmark Constraints

Now we discretize the *inequality* landmark constraints. The unweighted and weighted case are derived in parallel. The discretization of the constraints vectors $C^{\text{isoLM}}, C^{\text{WLM}} \in \mathbb{R}^{N_{LM}}$ consists of two main steps: first the constraints are rewritten such that the long vector notation of the landmarks (r, t) can be used consistently and then the discretization of the transformed reference landmarks $y(r_j)$ from the previous section is used again.

If we define the difference vectors $v_j = y(r_j) - t_j \in \mathbb{R}^3$ the constraints vectors look like:

$$C^{\text{isoLM}} = \begin{pmatrix} \|v_1\|^2 \\ \vdots \\ \|v_{N_{LM}}\|^2 \end{pmatrix} \in \mathbb{R}^{N_{LM}}, \quad C^{\text{WLM}} = \begin{pmatrix} \|v_1\|_{W_1}^2 \\ \vdots \\ \|v_{N_{LM}}\|_{W_{N_{LM}}}^2 \end{pmatrix} \in \mathbb{R}^{N_{LM}}.$$

Now the difference vectors are stacked into one long vector $v = y(r) - t = (v_x^1, \dots, v_x^{N_{LM}}, v_y^1, \dots, v_y^{N_{LM}}, v_z^1, \dots, v_z^{N_{LM}})^T \in \mathbb{R}^{3N_{LM}}$. The constraints vectors have a special structure: They take the difference vector $v \in \mathbb{R}^{3N_{LM}}$ and summarize three elements of the vector at a time to one value by determining the weighted or unweighted scalar product of the three elements. This means a vector of dimension $3N_{LM}$ is reduced to a vector of dimension N_{LM} . This reduction is performed by the matrix

$$A = \begin{pmatrix} I_{N_{LM}} & I_{N_{LM}} & I_{N_{LM}} \end{pmatrix} \in \mathbb{R}^{N_{LM} \times 3N_{LM}}. \quad (12.12)$$

As we need the squared components of v we set up the term:

$$\text{diag}(v) \cdot v \in \mathbb{R}^{3N_{LM}} \quad (12.13)$$

where $\text{diag}(v)$ is a $(3N_{LM} \times 3N_{LM})$ -matrix with v on the main diagonal. Now the constraints vectors can be rewritten to:

$$\mathcal{C}^{\text{isoLM}} = A \text{diag}(v) \cdot v \in \mathbb{R}^{N_{LM}}, \quad \mathcal{C}^{\text{WLM}} = A \text{diag}(v) \cdot W \cdot v \in \mathbb{R}^{N_{LM}}. \quad (12.14)$$

The difference vector $v = y(r) - t$ is discretized like in the previous section by $v \approx C\mathbf{y} - t$ leading to the discrete (weighted) constraints vector:

$$\mathcal{C}^{\text{WLM}}[y] \approx \mathcal{C}^{\text{WLM}}(\mathbf{y}) = A \text{diag}(C\mathbf{y} - t) \cdot W \cdot (C\mathbf{y} - t). \quad (12.15)$$

The discretization of the *unweighted* inequality constraints $\mathcal{C}^{\text{isoLM}}$ looks the same with $W = I_{3N_{LM}}$.

Again the first derivative of the constraints with respect to \mathbf{y} is needed for the optimization scheme:

$$\nabla_{\mathbf{y}} \mathcal{C}^{\text{WLM}}(\mathbf{y}) = 2A \text{diag}(C\mathbf{y} - t) \cdot W \cdot C. \quad (12.16)$$

12.6 Algorithms for Landmark Constraint Registration

12.6.1 Equality Landmark Constraints

Registration with equality landmark constraints can be solved by either eliminating the constraints and solving an unconstrained problem or directly solving the constrained problem.

Due to the linearity of the constraints, we are able to eliminate the constraints and reformulate problem 12.2 as an unconstrained minimization problem [170, 121, 77]. The main idea is to decompose the displacement into a special part y_{spec} that fulfills the constraint function $Cy_{\text{spec}} = t$ and a homogeneous part y_{hom} that fulfills $Cy_{\text{hom}} = 0$. One possible specific solution is a Thin-Plate Spline $y_{\text{spec}} = y^{\text{TPS}}$, which interpolates the landmarks. Thus y^{TPS} satisfies the landmark constraints. Furthermore, each homogeneous solution

can be represented as a linear combination of a basis of the Null space of C . Thus $y_{\text{hom}} = Zw$, with Z containing the basis vectors of the Null space of C and w the coefficient vector. Hence we get a parameterization of y in terms of the parameters w :

$$y(w) = y_{\text{hom}} + y_{\text{spec}} = Zw + y^{\text{TPS}}. \quad (12.17)$$

Henceforth, the optimization is performed via varying the parameters w and the resulting unconstrained optimization problem reads:

$$\begin{aligned} \min_w \quad J(w) &= D(w) + \alpha S(w) \\ &= D(R, T(y(w))) + \alpha S(y(w) - y^{\text{ref}}). \end{aligned}$$

For the solution of the unconstrained problem again a standard Gauß-Newton scheme can be applied.

As an alternative to the presented constraint elimination a direct method has also been proposed using the SQP algorithm [170] for constrained optimization problems. Here we will use the augmented Lagrangian method with a Gauß-Newton type approximation of the Hessian as described in Section 5.4. The objective function is defined like in the unconstrained registration problem (see Sec. 6.4) and again a multi-level strategy is used (see Sec. 6.5). The only additional ingredient besides the optimization scheme are the discretized constraints, which are already linear and their derivatives.

12.6.2 Inequality Landmark Constraints

There is no essential difference between the unweighted and weighted inequality landmark constrained registration problem. Like for the equality constraint case in the previous section the objective function is given as in the unconstrained problem and a multi-level strategy is used. Again the discrete constraints are the only additional ingredient and also linear. The algorithm for the general inequality constrained optimization problem is described in section 5.4.2.

Part III

Applications to Liver Surgery

Chapter 13

Assessment of Registration Algorithms

Before we present three different scenarios in computer assisted liver surgery (Chapters 14 and 15) where the developed registration techniques can be applied reasonably we will discuss the assessment of registration algorithms and introduce new methods for the measurement of the accuracy, which are particularly suitable for the registration of the liver with its interior vessel structures.

The achievable accuracy is the outstanding criterion for the assessment of registration methods. A high registration accuracy is essential for the precise multi-modal surgery planning, as well as for the generation of exact atlases and contributes significantly to the overall accuracy of intraoperative navigation and mechatronic assistance systems. The resulting alignment of reference and template image does not have to be perfect, but has to be adequate for the medical question. Often a high accuracy implies a long runtime and partially additional effort for the interactive incorporation of expert knowledge like the definition of landmark positions or parameter tuning. The criteria runtime and effort are of particular importance for time critical registration tasks like the transfer of preoperative to intraoperative image data. A further criterion closely associated with accuracy is the robustness of a method. A method is called robust, if it is possible to reach reliably high accuracy despite a changed starting point or artifacts in the image data. While accuracy and robustness concern the *effectiveness* of a registration algorithm, runtime and effort concern its *efficiency*.

In the following sections we will define and discuss these assessment criteria more precisely: accuracy, robustness, runtime and effort. We will also present concepts and quantitative methods for the assessment. The aspect of accuracy will be most elaborated, because without sufficient accuracy the other aspects are of less importance.

13.1 What Exactly is Accuracy and How Can It Be Determined?

At first it is indispensable to discuss the meaning of the term registration accuracy. A clear definition is necessary for a fair comparison of different approaches and to accomplish trust into registration methods resp. to understand their limits.

Besides the definition of the term accuracy also the reason for the accuracy determination is important. Fitzpatrick [66] differentiates, if the accuracy is investigated for a class of image pairs or for one concrete image pair. In the first case the intention is, whether a registration method is successfully applicable for a particular clinical problem in general. As even in a class of similar images the variations are complex between the images, this will be statistical investigations. This means the registration accuracy for a future image pair is predicted based on a well defined class of image pairs. Such claims are prone to statistical uncertainties, which can sometimes be estimated themselves. In the second case considering the accuracy of one concrete image pair it is intended to assure the safety of an individual patient. The clinician wants to know, if the registration accuracy is below a given threshold. In clinical routine this information can usually only be captured by visual control or indirect measurements. For example, for navigation systems the coincidence of real positions of a navigated pointer tip at anatomical landmarks and the corresponding positions of the virtual tip visualized in the image data on the screen can be checked.

13.1.1 Target Registration Error (TRE)

Now we come back to the question of how the accuracy concerning registration results can be defined. A registration algorithm produces an optimal transformation with regard to an objective function. This transformation usually consists of a distance measure and a smoothing term. Although the transformation is optimal with respect to the objective function, this does not mean it is *ideal* with respect to the underlying anatomical structures. An ideal transformation maps every point of the template image onto its corresponding anatomical point in the reference image. It is perfectly accurate. The computed optimal transformation is the best guess given the available information, but it is rarely ideal. This means with an ideal transformation an anatomical target point in the template image is exactly mapped on to its location in the reference image. If the computed optimal transformation is not ideal, the anatomical target point will be displaced some distance from the actual location in the reference image. This displacement is called **target registration error (TRE)**. The TRE is the commonly accepted measure of the registration accuracy [142, 66, 62]. We now give a mathematical formulation of the important TRE. Let $y^{\text{ideal}}, y^{\text{opt}} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be the ideal and the

13.1. What Exactly is Accuracy and How Can It Be Determined?

computed optimal transformation. Then the TRE at a point $x \in \mathbb{R}^3$ is given by a vector-valued function $TRE : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. The error vector $TRE(x)$ is given by

$$TRE(x) = y^{\text{opt}}(x) - y^{\text{ideal}}(x) \quad \text{for all } x \in \mathbb{R}^3. \quad (13.1)$$

In many publications not the individual error components in all three spatial directions are reported, but only the absolute value $\|TRE(x)\|$ of the error vector (magnitude of the vector field). A high registration accuracy is given by a low TRE.

The TRE vector field can be visualized directly or characterized by descriptive statistics. For example, the average error and the standard deviation of the error magnitude over the whole image or partly in clinically relevant regions like the surgical target structures or regions of risk can be determined. Typically the statistical description of the TRE is restricted to points inside or on the rim of lesions, which are intended to be surgically removed or to regions of functional brain activity, which are investigated for diagnostic purposes.

In some applications, e.g., intraoperative navigation systems, it is important for the surgeon to know that the registration error is below a critical limit of about 2-3 mm, i.e., we need to know the *maximal* error. The specification of such extreme values is difficult and often not meaningful, because the maximum and minimum of a statistical distribution are hardly reproducible in particular in small sample sets. In contrast to the mean or the standard deviation, which converge to a reliable value with increasing sample size (law of large numbers), the value of the maximum is monotonously increasing with the number of samples. While above a certain sample size there are only minor changes in the value of the mean, the maximum depends on and further increases with the sample size. A stable alternative to the maximum are quantiles, for example, the 95% quantile gives the error value for which 95% of all sample values are smaller than this error value.

In general it is important to be aware that the registration accuracy varies considerably over the image region as Woods [258] pointed out. This has to be kept in mind, if the whole accuracy vector field or distribution is summarized by a few parameters by descriptive statistics. For a local region around the surgical target structure such a summarization might be useful, but for the whole image region it is often not.

In most cases registration results are presented in a joint visualization of reference and transformed template image (image fusion) to enable a direct comparison of the image contents. This visual control does not allow a quantitative determination of the accuracy but even small deviations at the borders of anatomical structures are noticed by human observers as Wong et al. [256] and also Fitzpatrick et al. [63] have shown.

The visualization of the displacement field or parameter images from this field, e.g., magnitude of local volume change, are at least as important as the

image fusion visualization. It is possible to get identically transformed images, which have been achieved by totally different displacement fields. The reason is that differences can only be recognized at transitions between structures and not in homogeneous image regions. Thus the plausibility of the displacement field should be checked in all cases.

13.1.2 Methods to Determine the TRE

The big problem with determining the registration accuracy is that the ideal transformation, which is needed to compute the TRE, is never actually known for clinical data. This ideal displacement field is also named the *ground truth* transformation. The term “ground truth” originates from remote sensing to detect and classify objects on Earth. Informations directly gathered at the ground by terrain sensing like soil samples, vegetation etc. are used to evaluate a classification achieved by images from the surface of the earth taken from planes or satellites.

Simulations

Computer simulations are an exception, where the ideal transformation for the determination of the registration accuracy is known. Artificial movements or deformations of structures are simulated, applied to an image and the registration algorithm has to find back its way to the starting position and shape of the structures. The deviations of the computed image point locations to their original locations can be evaluated. In this case the ideal transformation is just the inverse of the simulated transformation $y^{\text{ideal}} = (y^{\text{sim}})^{-1}$. That means the target registration error is $TRE = y^{\text{opt}} - y^{\text{ideal}} = y^{\text{opt}} - (y^{\text{sim}})^{-1}$. The TRE can be rewritten by applying first the simulated transformation: $TRE \circ y^{\text{sim}} = y^{\text{opt}} \circ y^{\text{sim}} - (y^{\text{sim}})^{-1} \circ y^{\text{sim}} = y^{\text{opt}} \circ y^{\text{sim}} - \text{id}$. Hence the inverse of the simulated transformation has not to be computed. We simply determine the deviation from the identity.

Either real clinical image data or artificial structures (so-called software or digital phantoms) are deformed by y^{sim} . Some methods for generating deformations are presented in Sec. 13.2.3. The degree of realism of the simulated deformation is vague while the modeling of rigid transformations is straight forward. Care has to be taken only for the generation of a set of random rigid transformations with an intended distribution. How we can generate uniformly distributed rigid transformation samples and rigid transformation samples distributed in a region around a given reference transformation, will be described in Sec. 13.3. Although general properties of the registration method can be investigated very well using simulations under controlled conditions, the direct transfer of the results to real clinical data is often problematic. To increase the degree of realism it is possible to use estimated real deformations instead of complete artificial deformations. The estimation can be performed by a

registration based on manually defined anatomical landmarks in clinical data pairs [186]. These estimated deformations are applied to an image in the same manner as the purely artificially generated ones above.

The estimated deformations from clinical data pairs are potentially realistic, but the image contents and characteristics of the deformed image are identical (up to the deformations) to the original image. Images in clinical routine which are acquired at different time points or even with different modalities feature significantly different image characteristics and even contents in some regions due to noise, artifacts, other acquisition parameters or other imaging properties in different modalities. The registration is more complicated in such cases compared to deformed but identical images. One possibility to get more realistically deformed images (even simulating different modalities) is to simulate the images themselves and not only the deformation. The idea is to deform only the geometry (a model) of the anatomical structures and to simulate the appearance of the geometry for a given imaging modality. A well-known example is the *Digital Brain Phantom* of the database BrainWeb [38, 5], which contains realistically simulated MRI image data of this phantom. BrainWeb has already been used in hundreds of publications. Also other imaging modalities like PET and SPECT can be computed on this digital phantom by means of simulators like PET-SORTEO [184] or SimSET [82]. For ultrasound the simulation software Field II [93] is available.

Physical Phantoms

Physical phantoms are one possibility to consider real imaging properties like image distortions or noise. These phantoms range from simple cube-like plastic blocks to simplified anatomical models. Usually metal balls or screws (depending on the imaging modality) are mounted as reference points at or in the physical phantom such that a reference transformation y^{ref} can be determined which approximates the ideal transformation y^{ideal} quite accurately if the number of reference points is high enough. The disadvantage of physical phantoms is that like in the case of simulations and digital phantoms realistic non-rigid transformations, as they occur in clinical applications, are difficult to achieve. Another disadvantage of physical phantoms is that the contrast of the phantom image is often higher and imaging artifacts are rare resp. less strong. A special class of physical phantoms are animal or human preparations, which enjoy a certain popularity in the accuracy evaluation of surgical navigation systems, because they seem to be close to medical reality. However, the physiological changes after the death must not be underestimated. For instance a perfused organ possesses a different elastic behavior than an anatomical specimen. The use of specimen makes sense in the context of bony structures. An alternative are experiments on living animals. Theoretically it is possible to implant reference markers also in living animals, but the effort of animal trials is high and the anatomy is often not sufficiently comparable

to humans.

Reference Structures on Real Clinical Images

Registration accuracies, which are determined on simulations and phantoms, have only a limited validity for the accuracies on real patient data. Simulations and phantoms are always a model of reality where it might happen that exactly those factors, which limit the registration accuracy, were not modeled. The realism from simulations over physical phantoms to clinical patient data increases, but the verifiability decreases (see Fig. 13.1). Questions about the sensitivity of registration methods regarding particular aspects of the images (like noise etc.) can be investigated in a dedicated manner with simulations and partially with phantoms. Though the complexity and realism is limited the ideal transformation y^{ideal} (ground truth) is known perfectly or in the case of phantoms an accurate approximation of the ideal transformation can be determined at least at some points. In contrary, for clinical image data realistic properties and a realistic complexity are given, but the ideal transformation y^{ideal} is usually not or only approximately known on some regions or points.

Thus we are looking for possibilities to determine the accuracy of registration results on real clinical patient data without knowing the ground truth. That means we need references which approximate the ground truth well. The reference with the highest known accuracy yet is denoted as the *gold standard*. The generation and establishment of a gold standard on patient data is a challenge. Only in rare cases images with artificial landmarks as references are available. One example how such a gold standard can be determined, which is based on bone implanted screws, will be described in Sec. 13.2.2. In most cases, anatomical landmarks or structures defined by experts are the gold standard for the evaluation of the registration accuracy on patient data. We will discuss such gold standards in more detail in Sec. 13.2.3 and introduce a reference standard based on vessel trees in Sec. 13.4.

13.1.3 Analytical Estimation of the TRE

Until now there is only one registration scenario where the TRE has been estimated analytically: rigid landmark registration. Further details for rigid landmark registration are presented in Sec. 13.2.1. The problem with intensity-based registration is that for an analytical determination of the TRE a good model for the image pairs is needed, but the variations of medical images are high and complex. For non-rigid registration a feasible deformation model has to be known for the computation of the TRE.

	Artificial Simulation	Physical Phantom	Clinical Data
Knowledge about y^{ideal}	High	Middle	Low
Realism of Deformation	Low	Low to Middle	High
Realism of Appearance	Middle	Low to Middle	High
Realism of Geometry	High	Low to Middle	High

Figure 13.1 Assessment of different general approaches to determine the registration error.

13.2 Previous Work

13.2.1 Accuracy of Rigid Landmark Based Registration

For rigid landmark registration there is an analytically derived estimation of the target registration error (TRE) for all image points. The accuracy of landmark based registration depends of course on the localization accuracy of the landmarks. The localization inaccuracy is also called fiducial localization error (FLE) [65]. For image-to-patient registration the highest localization accuracy is reached with special artificial markers, which are screwed into the bone and are very well discriminable in different modalities [142].

Due to the localization inaccuracies of the landmarks (FLE) the probability is very low that there is a rigid registration, which maps corresponding landmarks exactly onto each other. Instead, the landmark registration algorithms we presented in Section 11.6 determine a rigid transformation, which minimizes the squared distance between corresponding landmark pairs. The remaining sum of squared distances at the end of the minimization process is also called fiducial registration error (FRE) and is often reported as the registration “accuracy” of navigation systems. But Fitzpatrick [61] showed that FRE and TRE are uncorrelated. Theoretically the FRE can be small while the important TRE is big.

Fitzpatrick et al. derived the expected value of the TRE squared analytically [65] as well as the whole distribution of the TRE [64] for each space point depending on the FLE, the number of the landmark pairs N and the geometrical configuration of the landmarks. They assumed that the FLE statistics for all points are identical, isotropic, and have a zero-mean normal distribution. They used the standard unweighted landmark distance measure for their derivation (see Sec. 11.6). Danilchenko and Fitzpatrick [43] treated the case of

non-identical, anisotropic, but still zero-mean normally distributed fiducial localization errors. For further information and literature references about TRE for general FLE distributions see the introduction of [213]. Danilchenko and Fitzpatrick additionally derived the expected FRE squared and TRE squared for a *weighted* landmark distance measure (see Sec. 11.6) incorporating information about anisotropic landmark localization uncertainties (FLE). They also provide a MATLAB code for the computation of FRE and TRE.

The TRE varies over the image region. The registration error is smallest in the center of the landmarks and is increasing with increasing distance to the center. The increase is not the same in each direction, but ellipsoidal. The three main axes of the ellipsoid are determined by the spatial landmark configuration, which is a very significant influencing factor on the accuracy. West et al. show the spatial distribution of the TRE for different landmark configurations in neurosurgery in their publication [253].

13.2.2 Accuracy of Rigid Intensity Based Registration

There is no analytical estimation of the TRE for rigid intensity based registration, but a very accurate gold standard based on bone-implanted screws [142, 252]. The accuracy of this gold standard was analyzed very accurately by Maurer et al. [142]. The gold standard was used in **The Retrospective Image Registration Evaluation Project** (RIRE) of the Vanderbilt university (often abbreviated as the Vanderbilt project). In this project, image data of the three modalities CT, MRI and PET were used from patients, which underwent navigated neurosurgery. Bone markers were implanted in all patients before image acquisition. Afterwards the appearance of the markers was removed from the images to enable a blind study. Important brain regions were identified by neurosurgeons and the centers of those *volumes of interest* (VOI) were used for the error analysis. The TRE was computed by determining the deviations between the positions of the VOI center points after application of the computed rigid transformations and the gold standard transformation based on the bone markers. Unfortunately such databases like the one at the Vanderbilt university are very rare.

The accuracy and robustness of a registration method depends on several factors, such as image content, imaging modality, imaging artifacts, image noise, distance measure, optimization algorithm, implementation details, etc. Even if a perfect accuracy determination method is available, it is difficult to identify the influence of a particular factor on the accuracy due to the complex interdependences between the numerous factors. The distance measure is one of the factors that highly influences the registration accuracy. Often the performance of a distance measure is evaluated by the resulting registration accuracy including the influences of all the other factors or by plotting distance measure values for transformations in the neighborhood of a gold

standard reference transformation [80]. Ideally, the plot of the distance measure shows that the optimal value of the distance measure occurs exactly and distinctively for the reference transformation, no local minima are near the optimal value where an optimizer can be trapped, and capture range is big. If an initial transformation lies inside the capture range around the reference transformation the optimization process will find the reference transformation (see also Sec. 13.5.1).

For a set of clinical data the plot of the distance measure values will usually not be ideal. The question is how this qualitative evaluation can be transferred to an objective quantitative evaluation. Skerl et al. [218] propose a protocol for the quantitative evaluation of distance measures. They sample the space of possible transformations by a set of N randomly selected lines passing the reference transformation. On each line the distance measure is evaluated for M uniformly spaced sample transformations. The length of the lines is defined by a maximal distance R from the reference transformation. Methods for the generation of uniformly distributed directions for the definition of the lines are presented in Sec. 13.3. Skerl et al. use three translation and three Euler angle parameters to define a rigid transformation. This means the space of rigid transformations has six dimensions and each sample point is represented by a six-dimensional parameter vector. The template image is transformed by each of the $N \times M$ sample transformations (including the reference transformation) and the distance measure between the transformed template image and the reference image is computed. The resulting distance measure values are the basis for the quantification of five different properties of the distance measure: accuracy (ACC), distinctiveness of global minimum (DO), capture range (CR), number of maxima (NOM), risk of nonconvergence (RON). The mathematical definition of the properties is given in [218].

The protocol of Skerl et al. offers the possibility to investigate the performance of a distance measure (and its implementation) without the influences of a concrete registration process.

13.2.3 Accuracy of Non-Rigid Registration

As described in Section 13.2.1 and 13.2.2 for rigid landmark based resp. rigid intensity based registration there are already accepted methods for the determination of the accuracy. The accuracy analysis of non-rigid registration algorithms is significantly more complicated. Different evaluation methods have been proposed, but until now no established gold standard exists.

Often artificial deformations are simulated to show that an algorithm is able to compensate at least certain deformations [194, 109]. Those deformations usually have only little in common with real anatomical deformations. A little bit more realistic deformations are generated by manually moving single landmarks and computing landmark based registration results to get artifi-

cially transformed points for all image points [186]. Schnabel et al. [206] try to increase the degree of realism by means of biomechanical simulations.

The construction of deformable physical phantoms is significantly more difficult than rigid phantoms. In addition, more landmarks have to be inserted into the phantom as references for the accuracy determination. For rigid transformations only three landmarks are theoretically enough to compute the transformation at any point. One example for an elastic phantom is the *truth cube* [100]. The truth cube is a cube made of silicone rubber with edges of 8 cm length filled with 343 teflon spheres of 1.6 mm in diameter. CT scans of the truth cube with and without deformations induced by pressing a plate resp. a sphere onto the upper side are available in the internet [100]. The identification of the sphere landmarks is also available. The cube is intended as a physical standard for modeling soft tissue, but it is also suitable for accuracy analysis of non-rigid registration methods [260]. Cash et al. [28] have built an artificial liver phantom, which is also made from silicone rubber and contains teflon spheres. Real porcine livers with embedded sphere landmarks are also imaginable, but they are not long-living and reproducible.

The determination of the registration accuracy on real clinical image data is inevitable as discussed in Sec. 13.1.2. Usually anatomical point landmarks are defined manually in reference and template image by medical experts to determine the registration accuracy. Such points are, for example, branchings of vessels [185, 237, 200]. The problem is that the position of the point landmarks can only be set and reproduced with a restricted accuracy by the same expert at different time points (intra-rater reliability) as well as by different experts (inter-rater reliability). In addition, the accuracy evaluation is only possible in areas where reference landmarks can be identified. Despite those potential difficulties this is currently the most convincing strategy for the validation of non-rigid registration methods at least in areas near the reference landmarks.

Corresponding curves on organ boundaries, surfaces or whole regions are representing a bigger part of the images than single anatomical points. The matching of corresponding segmented regions can be measured with the same methods as used in the accuracy assessment of segmentation algorithms [33]. Curves and surfaces are sometimes easier and more accurately identifiable than points. However the information which point on one curve/surface corresponds to which point on the other curve/surface is missing. Simply taking the nearest point underestimates the real distance and thus the registration error. In Sec. 13.4 we will present a validation method where point correspondences on vessel center lines (curves) between vessel branching points and metrics, which consider the localization uncertainties of the points, will be defined.

As in the case of rigid registration the direct investigation of distance measures is important to identify the potentials to decrease registration errors and to assess their performance. Non-rigid transformations possess significantly

higher degrees of freedom compared to rigid transformations. Therefore a qualitative visualization of distance function values in the neighborhood of the optimum as in the rigid case (see Sec. 13.2.2) is difficult. Skerl et al. [219] have extended their protocol for a quantitative analysis of distance measures from rigid to non-rigid transformations. They measure exactly the same five different properties as in the rigid case: accuracy, distinctiveness, capture range, number of maxima, risk of nonconvergence. The only difference is that the sample points in the high-dimensional parameter space represent simulated non-rigid transformations. Skerl et al. [219] simulate local deformations by systematically displacing a set of control points of B-spline transformations. For each control point a set of N probing lines with randomly selected displacement directions is generated. Each line is sampled on M points leading to $N \times M$ different displacement vectors for each control point. The resulting non-rigid transformations are applied to the template image and the distance measure values of the transformed template images to the reference image are computed. A detailed description can be found in [219].

13.3 Simulating Rigid Transformations

Simulations are a suitable approach to study the behavior of registration algorithms as well as to compare different algorithms. We discussed the advantages and disadvantages of simulations in Sec. 13.1.2. Here we deal with the simulation of *rigid* transformations. Rigid transformations can easily be generated by one of the parameterizations presented in Sec. 7.3. The challenge is to generate random rigid transformations with a given distribution, which will be usually a uniform distribution. While uniformly distributed translations are fairly easy to produce, rotations are not. We will use some ideas from sphere point picking and directional statistics to generate random translations and rotations. We often only need random translations and rotations in a given restricted range, for example around a known ground truth transformation. We will use the von Mises-Fisher distribution, which is an analogon of the Gaussian distribution on the sphere, to generate rigid rotations around a given ground truth rotation.

There are two general approaches to generate random samples from a given distribution: a regular deterministic sampling of the configuration space and a random sampling according to a given distribution. Rigid transformations can be described by a parameterization with a 6D parameter space. Therefore, a regular deterministic sampling would create a very high number of samples (N^6) even for a small number of sample points (N) in each of the 6 coordinate directions. Therefore we prefer random sampling to get a good overview of the configuration space.

13.3.1 Sampling from a Given Distribution

In this section, two methods are shown to generate values of a random variable $X : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$, which are distributed according to a given probability distribution. Usually only uniformly distributed numbers are provided by pseudo-random number generators on the computer. From this simple distribution more complex distributions can be obtained by the following general methods.

Inverse Transform Sampling. A general possibility is the *Inverse Transform Sampling* also called the inverse probability integral transform [46]. The main idea is to invert the cumulative distribution function (cdf) of a random variable X analytically or numerically. The inverse cdf is also called quantile function.

The sampling method is based on the following theorem. Let $F_X = P[X \leq x] \forall x \in \mathbb{R}$ be the cumulative distribution function (cdf) of the continuous random variable $X : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$ with an inverse F^{-1} . If U is a random variable uniformly distributed on $[0, 1]$, then $F^{-1}(U)$ has the distribution function F . The proof is straightforward:

$$\begin{aligned} &P[F^{-1}(U) \leq x] \\ &= P[U \leq F(X)] \quad (\text{monotonic } F \text{ applied on both sides}) \\ &= F(X), \quad \text{since } P[U \leq y] = y \text{ if } U \text{ is uniformly distributed.} \end{aligned}$$

Now we generate uniformly distributed samples, apply the inverse cumulative distribution function F^{-1} and get samples, which are distributed according to F . If no analytical inverse is known, the non-linear equation $F(X) = U$ has to be solved numerically leading to an approximate sampling instead of an exact sampling method.

The inverse transform sampling strategy will be used to sample the von Mises-Fisher distribution to generate anisotropic random directions (13.3.2) and uniformly distributed rotations (13.3.3).

Rejection Sampling. An alternative general sampling method is rejection sampling also called acceptance-rejection or accept-reject method [46]. Instead of sampling directly from a distribution an envelope distribution is used, where sampling is easier and samples outside the envelope are rejected.

Let X be a random vector with density $f : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$, and let U be an independent, uniformly distributed, scalar random variable on $[0, 1]$. We assume that a dominating density $g : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$ and a constant $c \in \mathbb{R}^+$ exist such that $f(x) \leq cg(x)$ for all $x \in \Omega$. Then random variates with density f can be obtained by the following algorithm:

1. Generate random variables X with density g on \mathbb{R}^N and U with uniform distribution on $[0, 1]$.

2. Check if $U < f(X)/(cg(X))$. If this holds, X is a sample of the distribution f . If not, reject X and try another sample (step 1).

We give a short explanation that the algorithm works. The $(N+1)$ -dimensional random variable $(X, Ucg(X))$ is uniformly distributed in the area below the curve/graph of cg in \mathbb{R}^{N+1} . Then $(X, cUg(X))$ is also uniformly distributed below the graph of f and the N -dimensional projection X of the samples must have density f . Detailed proofs are given by Devroye [46]. The challenge of rejection sampling is to find an efficient envelope function. Efficient means that the rate of rejected samples is preferably low in particular in higher dimensions.

We will use this sampling strategy in the next section for uniformly sampled translations.

13.3.2 Random Translations and Directions

We introduce two methods to generate uniformly distributed translations inside a box and inside a sphere. Each translation will have a random direction and a random (but constrained) length. In addition, we will present methods providing uniformly and non-uniformly distributed random directions with a fixed given length.

Uniformly Distributed Translations Inside a Box. A very simple method to generate uniformly distributed translations inside a box (cuboid) of size $(X_{\text{size}}, Y_{\text{size}}, Z_{\text{size}})$ is described by Kuffner [108]. Assume we have a pseudo-random number function $rand()$ which returns uniformly distributed values on the interval $[0, 1)$. Independent random values along each axis scaled by the size of the box in the respective direction lead to uniformly distributed translations inside the given box:

$$(x, y, z) = (X_{\text{size}}rand(), Y_{\text{size}}rand(), Z_{\text{size}}rand()). \quad (13.2)$$

Uniformly Distributed Translations Inside a Sphere. As the corners of a box are further away than the centers of the faces of the box some directions are over-represented. A box is not rotation-invariant in contrast to a sphere. Therefore translations sampled inside a sphere are not biased to certain directions. To generate uniformly sampled translations inside a sphere the described rejection sampling method is used.

The idea is to take uniformly distributed points inside a unit cube and reject all points which are outside a unit sphere. The efficiency of the sampling method is $\pi/6$ independent of the radius ($V_{\text{Sphere}}(d) = \frac{\pi}{6}d^3$).

This sampling technique is also applicable to higher dimensions with samples inside hypercubes and hyperspheres, but with increasing dimension the probability of rejection also increases and the efficiency decreases.

Uniformly Distributed Random Directions. The motivation for generating random directions is to get translations with a defined length. Random directions are also needed by some methods for the generation of random rotations (see Sec. 13.3.3).

A field where random directions are implicitly generated is sampling on the surface of a sphere, which is also called *sphere point picking*. Each random point on the sphere represents a random direction. We briefly describe three of the many different sphere point picking methods:

1. The first sphere point picking method is simple. At first uniformly distributed samples inside a sphere are generated by rejection sampling as described above and the samples are projected from inside the sphere onto its surface by normalizing the length of the sample directions.
2. An elegant method for sphere point picking is to use 3D Gaussian random variables and divide each resulting random vector by its length. This leads to a uniform distribution on the unit sphere because the Gaussian distribution is spherical symmetric. It is also a simple method for hypersphere picking in higher dimensions [141]. In particular for higher dimensions this is more efficient than rejection sampling.
3. An obvious method to get random points on the unit sphere is to take the spherical parameterization $x = \cos \theta \sin \phi$, $y = \sin \theta \sin \phi$, $z = \cos \phi$, for the two angles θ and ϕ uniformly distributed on $[0, 2\pi)$ resp. $[0, \pi]$. The problem is that the resulting points are not uniformly distributed on the sphere. The reason is that the infinitesimal area element on the unit sphere is $dA = \sin \phi d\theta d\phi$. This means the area element depends on the angle ϕ leading to a higher point density at the poles compared to the equator. A similar effect occurs for random rotations parameterized by Euler angles (see Sec. 13.3.3 and Fig. 13.3). It is possible to correct this effect. To get uniformly distributed points the joint probability $p(\theta, \phi) = p_\theta(\theta)p_\phi(\phi)$ has to be the product of the independent probability densities for θ and ϕ : $p_\theta(\theta) = \frac{1}{2} \sin \theta$ and $p_\phi(\phi) = \frac{1}{2\pi}$ [21]. We use inverse transform sampling (Sec. 13.3.1) to generate point samples according to this distribution. The cumulative distribution function (cdf) F for θ is:

$$F(\theta) = \frac{1}{2} \int_0^\theta \sin(x) dx = \frac{1}{2}(1 - \cos(\theta)). \quad (13.3)$$

The inverse of the cdf is $F^{-1}(y_1) = \cos^{-1}(2y_1 - 1)$. In Sec. 13.3.1 we showed that if y_1 is uniformly distributed on $[0, 1)$, then $\theta = F^{-1}(y_1)$ is distributed according to F . A similar derivation for ϕ leads to $\phi = 2\pi y_2$, where y_2 is uniformly distributed on $[0, 1)$.

Non-Uniformly Distributed Random Directions. We need often samples from a region around a given reference direction rather than from the

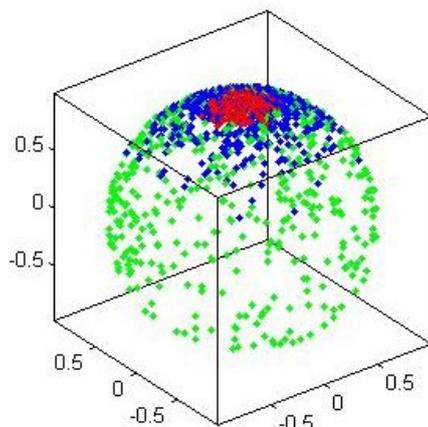


Figure 13.2 Points on the sphere sampled from three different von Mises-Fisher distributions with $\kappa = 1$ (green), $\kappa = 10$ (blue), $\kappa = 100$ (red) and the mean direction $(0,0,1)$.

whole space of possible directions (whole sphere surface). We mentioned in the introduction that the von Mises-Fisher distribution can be seen as a Gaussian-like distribution on a sphere (see Fig.13.2). It is an *isotropic* distribution around a mean direction μ with a concentration parameter κ leading to circular contours on the sphere surface. The concentration parameter κ is comparable to the standard deviation parameter σ of a Gaussian distribution. The von Mises-Fisher probability function on the 2D sphere S^2 in 3D space is:

$$f_3(x; \mu, \kappa) = C_3(\kappa) \exp(\kappa \mu^T x) \quad (13.4)$$

with $\kappa \geq 0$, $\|\mu\| = 1$ and the normalization constant $C_3(\kappa)$ is equal to

$$C_3(\kappa) = \frac{\kappa}{4\pi \sinh \kappa} = \frac{\kappa}{2\pi(e^\kappa - e^{-\kappa})}. \quad (13.5)$$

The bigger the concentration parameter κ , the higher the concentration of the distribution (the lower the dispersion). The distribution is unimodal for $\kappa > 0$ and uniform for $\kappa = 0$.

Now we show how to sample from this density to get points on the sphere which are von-Mises-Fisher distributed. Ulrich [230] derived a general result, but for the 3D case an easier solution is available [60]. A three-dimensional random vector X has von Mises-Fisher distribution if and only if

$$X = ((1 - W^2)^{\frac{1}{2}}V, W)^T \quad (13.6)$$

where V is a uniformly distributed two-dimensional unit vector and W is a scalar random variable in the range $[-1, 1]$ with the density $f(\omega) = c_\kappa^{-1}e^{\kappa\omega}$.

The normalizing constant is given by $c_\kappa = \frac{2}{\kappa} \sinh(\kappa) = \frac{2}{\kappa} \frac{e^\kappa - e^{-\kappa}}{2}$. The uniformly distributed vector V can be easily obtained by parameterizing $V = (\cos \theta, \sin \theta)$ and θ uniformly distributed on the interval $(0, 2\pi]$.

We use inverse transform sampling (see Sec. 13.3.1) to sample W . The cumulative distribution function F_W of W with density f is given by:

$$F_W(t) = P[W \leq t] = c_\kappa^{-1} \left[\frac{e^{\kappa t}}{\kappa} - \frac{e^{-\kappa t}}{\kappa} \right]. \quad (13.7)$$

The inverse F_W^{-1} of the distribution function (quantile function) is given by

$$F_W^{-1}(y) = \frac{1}{\kappa} \log(e^{-\kappa} + \kappa c_\kappa y). \quad (13.8)$$

If we take a uniformly distributed random variable $Y \sim U[0, 1]$, then is $W \equiv F_W^{-1} F_W$. Now we have all the ingredients to sample the von Mises-Fisher distribution by using Equation 13.6 and the described distributions for V and W .

Wood [257] improved the algorithm of Ulrich [230]. See also [268] for an application in probabilistic white matter fiber tracking based on DTI.

13.3.3 Random Rotations

For the validation of rigid registration methods often random rotations are generated simply by uniformly distributed Euler angles over the whole or a restricted angle range. But the resulting random rotations are *not* uniformly distributed over the space $SO(3)$ of rotations or a restricted part of the space (see Fig. 13.3). We will show how the Euler angles have to be distributed to get uniformly distributed random rotations and present alternative methods for the generation of such random rotations. We will also explain how *non-uniformly* unimodally distributed random rotations can be simulated around a given reference rotation.

Uniformly Distributed Random Rotations

In Chapter 7.3 we have seen that there are several representations of rotations in 3D by different parameterizations. Based on three of these parameterizations we will present methods to generate uniformly distributed rotations. A fourth method, called Arvo's method, will also be sketched.

Axis and Angle Method. One would think that uniformly distributed random rotation axes with corresponding uniformly distributed random rotation angles lead to uniformly distributed random rotations. Brannon [21] gives a qualitative reason and illustration why this is *not* the case. A simple standard criterion for uniformity is that the distribution is invariant under arbitrary rotations. Miles [146] derives an invariant probability density, which produces uniformly distributed rotations. This probability density implies that

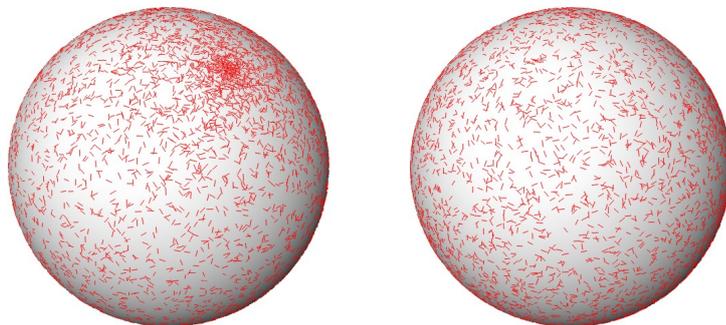


Figure 13.3 Rotations are visualized as a rotated sphere with a rotated tangent on the sphere [146, 108]. The left figure shows non-uniformly distributed (biased against poles) rotations generated by uniformly distributed Euler angles and the right figure shows uniformly distributed rotations generated by an improved Euler angle method.

the rotation axis has to be uniformly distributed on the sphere S^2 which can be achieved by the methods presented in Sec. 13.3.2 and the rotation angle α has to be distributed according to $p_\alpha(\alpha) = \frac{2}{\pi} \sin^2 \frac{\alpha}{2}$. The distribution for the rotation angle can be sampled by inverse transform sampling (see Sec. 13.3.1). The cumulative distribution function is $F_\alpha(\alpha) = \int_0^\alpha p_\alpha(x) dx = \frac{1}{\pi}(\alpha - \sin \alpha) = y$. Unfortunately the inverse $F^{-1}(y)$ cannot be analytically determined. Thus the equation $y = \frac{1}{\pi}(\alpha - \sin \alpha)$ has to be solved numerically for α , which is the drawback of the method. If y is uniformly sampled on $[0, 1)$, then α is distributed according to $p_\alpha(\alpha)$.

Euler Angles Method We mentioned in the introduction that uniformly sampled Euler angles (see Sec. 7.3) do not lead to uniformly distributed rotations (see Fig. 13.3). Similar to the axis and angle method a distribution for the Euler angles has been derived which leads to uniformly distributed rotations [146, 21]: $p_{(\theta, \phi, \gamma)} = p_\theta(\theta)p_\phi(\phi)p_\gamma(\gamma) = \frac{1}{2\pi} \frac{1}{2} \sin \phi \frac{1}{2\pi} = \frac{1}{8\pi^2} \sin \phi$. This distribution can be sampled by inverse transform sampling. The cumulative distribution function $F_\phi(\phi) =: y$ for ϕ and its inverse $F_\phi^{-1}(y) = \cos^{-1}(2y - 1)$ are the same as for the uniform sampled directions in Sec. 13.3.2. The other two angles θ and γ have to be distributed uniformly on $[0, 2\pi)$. A detailed algorithm for uniform sampling of rotations based on Euler angles is described by Kuffner [108].

Quaternions Method. Shoemake [216] introduced a sampling method based on quaternions which is also used by Kuffner [108]. The main idea is to generate uniformly distributed points on the unit hypersphere S^3 in *four* dimensions leading to unit quaternions which are uniformly distributed on the space of rotations $SO(3)$. Shoemake [216] presents a method for the generation of uniformly sampled points on S^3 . Some of the sphere picking methods for

S3 in Sec. 13.3.2 can easily be extended to S3. It is also possible to use a non-uniform distribution on S3 to generate random rotations around a reference rotation. For example the von Mises-Fisher distribution is also available for n-dimensional spheres.

Arvo's Method. The idea of Arvo [3] is to rotate an object vertically at the north pole of a sphere about a uniformly random amount and then rotate the axis of the north pole to a random position on the sphere. Arvo gives a method for the random rotation of the axis. The resulting rotation samples are uniformly distributed within $SO(3)$ [3]. If we restrict the range of the random input variables, we get uniformly distributed rotations within given limits.

Non-Uniformly Distributed Rotations

An extension of the von Mises-Fisher distribution for non-uniformly distributed directions is the von MisesFisher *matrix* distribution [47, 101] which can be used to construct non-uniform distributions over the space $SO(3)$ of rotation matrices. It has been shown by Prentice [183] that the von Mises-Fisher matrix distribution on $SO(3)$ is equivalent to the Bingham distribution on S3, which is defined in [13]. The points on S3 can again be interpreted as quaternions. A sampling method for the Bingham distribution can be found in [257].

13.4 Reference Standard Based on Vessel Trees

We discussed the importance of corresponding reference structures in clinical images for the accuracy determination of (non-rigid) registration methods in Sec. 13.1.2 and Sec. 13.2.3. In particular for non-rigid registration problems it is not possible to identify for each point in a clinical template image its corresponding point in the clinical reference image. Thus the validation is restricted to some important anatomical structures that can be identified in real clinical data.

What are the potential anatomical structures in CT and ultrasound images of the liver which can be used for defining the corresponding reference standard features? The first possibility is to take the target structure directly, e.g., the tumor in oncological liver resections. But tumors are usually only rough references because tumor boundaries are often difficult to delineate in particular in ultrasound images. The centers of small and good discriminable tumors might be suitable references. The liver surface is a second possibility to define reference standard features. The liver surface is quite smooth with only very few point-like features and some curve features. In preoperative CT data the whole liver surface is imaged and explicitly given from operation planning. In intraoperative ultrasound images only parts of the posterior liver

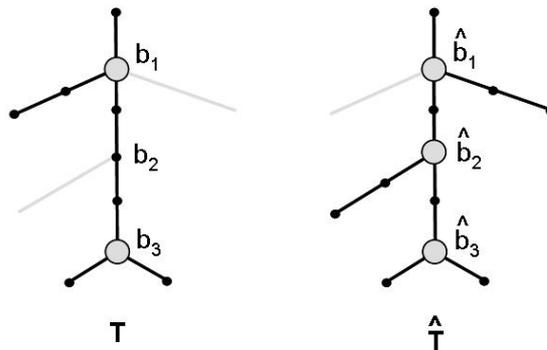


Figure 13.4 Reference T and template tree \hat{T} are minors of the underlying anatomical supertree (gray).

surface are identifiable. In postoperative CT data parts of the original surface are missing due to the resection. In addition the liver surface is clinically less important than the vessels which are good identifiable features and clinically very important. Thus we focus on the vessels.

Usually corresponding anatomical landmarks on vessels manually identified by experts are used [186, 237, 200] as reference points for the evaluation of the registration accuracy. We present a method, which also uses manually identified corresponding vessel branchings in both modalities, but where an automatic algorithm then determines whole corresponding substructures (graph minors) of the vessel trees while verifying their consistency [120].

To evaluate a non-rigid registration result the vessels in the template data are deformed with the computed transformation and a metric measures how close corresponding vessel parts are in the reference data after the transformation. We define four different evaluation metrics for vessel structures. The main challenge is to consider the inaccuracies of the defined vessel correspondences in the metrics.

We will use the vessel tree correspondence method and the metrics for the evaluation of non-rigid liver registration in Chapter 14 and Chapter 15.

13.4.1 Definition of Vessels Correspondences

We assume that a segmentation and center line extraction of the vessel trees from two different data sets (e.g. CT and 3D ultrasound) is available (see also Sec. 2.3.2). In our case the center lines consist of linearly interpolated 3D node point sequences. Based on those nodes the vessels can be represented as a tree (Fig. 13.4). Let $T = (V, E)$ and $\hat{T} = (\hat{V}, \hat{E})$ be directed trees for the reference resp. model data. All edges point in the direction of the given

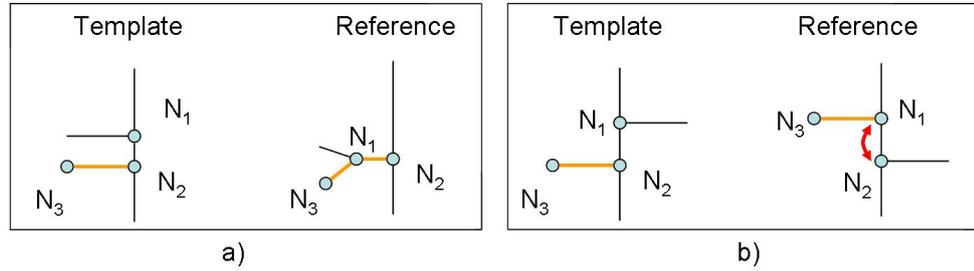


Figure 13.5 Caused by inaccuracies: a) edges might be contracted near very close branching nodes, or b) the order of very close branching nodes might be exchanged.

root nodes $r \in V$ and $\hat{r} \in \hat{V}$. The node subsets $B \subseteq V$ and $\hat{B} \subseteq \hat{V}$ contain all branching nodes (node degree $\delta(v) > 2$). Let $P(v, w)$ be the unique path from node v to node w . We call a path between two branching nodes or a branching and an end node ($\delta(v) = 1$) a (vessel) segment path.

The trees T, \hat{T} are not identical and only parts of the underlying anatomical vessel tree (supertree) can be identified in both modalities. There is neither a subtree isomorphism of T, \hat{T} to a common supertree nor a homeomorphism. Due to missing branches a segment path in T might be represented by several segment paths in \hat{T} and vice versa (see Fig. 13.5a). In addition the order of branching points or other topological changes might occur (see Fig. 13.5b) caused by inaccuracies in the image processing pipeline. To get a subtree isomorphism the contraction of some segment paths to one single node is necessary. This means a supertree contains T and \hat{T} as a minor (see Fig. 13.4). Intuitively, a graph G is a minor of a graph H , if G can be obtained from H by a series of vertex/edge deletions and edge contractions.

As the consistent assignment of branching nodes and segment paths is only heuristically solvable based on geometric properties [29], we interactively define a subset of corresponding branching points $(b_i, \hat{b}_i), i = 1, \dots, n$, $b_i \in B, \hat{b}_i \in \hat{B}$. Let B_{corr}, \hat{B}_{corr} contain all b_i resp. \hat{b}_i . To ease the interaction, a mouse click can be set onto the branching of the vessel surface and the nearest branching point on the center line graph is determined automatically.

Determination of Corresponding Segment Paths

For two given vessel center line trees T, \hat{T} and a subset of corresponding branching node pairs $(b_i, \hat{b}_i), i = 1 \dots n$ the “CorrespondingPaths” algorithm (see Fig. 13.6) determines corresponding path pairs (P_i, \hat{P}_i) . As the directed path to the root is unique, the algorithm starts from each of the n corresponding branching point pairs and ends, if another corresponding branching point

```

1: All nodes in  $V$  and  $\hat{V}$  are unmarked.
2: for  $n = 1$  to  $n$  do
3:   Let  $v := b_i, \hat{v} := \hat{b}_i$  {Start with corres. node pair}
4:   repeat
5:     Mark  $v, v := parent(v)$ 
6:   until  $v$  is already marked or  $v \in B_{corr}$  or  $v$  is root  $r$ 
7:   repeat
8:     Mark  $\hat{v}, \hat{v} := parent(\hat{v})$ 
9:   until  $\hat{v}$  is already marked or  $\hat{v} \in \hat{B}_{corr}$  or  $\hat{v}$  is root  $\hat{r}$ 
10:  if  $v \in B_{corr} \wedge \hat{v} \in \hat{B}_{corr}$ , but  $(v, \hat{v})$  is NOT a corres. pair then
11:    Set  $P_i := \emptyset$  and  $\hat{P}_i := \emptyset$ 
12:    print Resolve topological inconsistency at paths starting at  $(b_i, \hat{b}_i)$ 
13:  else
14:    Set  $P_i := P(b_i, v)$  and  $\hat{P}_i := P(\hat{b}_i, \hat{v})$ 
15:  end if
16: end for

```

Figure 13.6 CorrespondingPaths Algorithm.

or a point, which has already been visited, has been reached. The second termination criterion is necessary because not all branching points might have been interactively assigned. It is important to check, if the points reached are an assigned branching point pair in order to detect topological or assignment errors. Most of the inconsistencies can be resolved by automatic contraction of very short segment paths. Remaining problems are eliminated interactively. Based on the resulting path pairs each model path \hat{P}_i is reparameterized according to the reference path P_i . This means the relative distances between successive reference nodes are transferred to the length of the model path. Now each node on a reference line corresponds to a point on the model line. Let $(p_j, \hat{p}_j), j = 1, \dots, m$ be the resulting corresponding point pairs on all segments. The correspondences produced by this algorithm are much better than simple closest point correspondences as illustrated in Fig. 13.7. An alternative algorithm for correspondence determination on vessel center lines (coronary arteries) was introduced by Schaap et al. [202]. They sample the center lines equidistantly and introduce valid correspondences defined by ordered sets of connections. Then they search a valid correspondence minimizing the Euclidean length of all connections by means of Dijkstra's algorithm.

13.4.2 Evaluation Metrics on the Vessel Reference Standard

Until now we have defined corresponding point sets on the vessel center lines in reference and template image. For the evaluation of a registration result we need metrics which quantify the differences of corresponding vessel parts after registration. Some of the metrics will be used on liver reference vessels in Chapter 14 and 15.

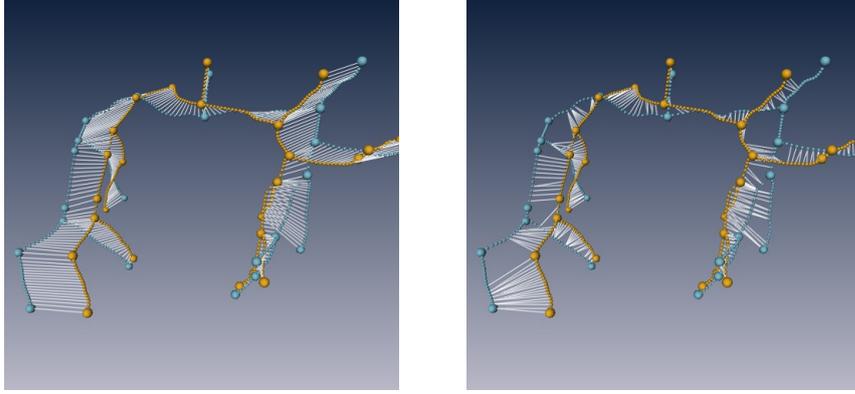


Figure 13.7 The proposed parametrization between branching points produces much better correspondences (left) than common closest point correspondences (right).

Distance of Corresponding Points on Center Lines: The simplest metric is the average distance between corresponding point pairs:

$$M_{dist}(T, \hat{T}) := \frac{1}{m} \sum_{j=1}^m \|p_j - \hat{p}_j\|. \quad (13.9)$$

Weighted Point Distance: The corresponding center line points are generated with some location uncertainties. The uncertainties along the lines are usually higher than perpendicular to the lines. In particular location uncertainties of the branching points lead to translations of the correspondences along the lines. Let Σ_j be an estimated covariance matrix of the anisotropic localization uncertainty for the point pair (p_j, \hat{p}_j) . Based on these weighting matrices a weighted point distance measure can be defined (see Sec. 11.5.2):

$$M_{weighted}(T, \hat{T}) := \frac{1}{m} \sum_{j=1}^m (p_j - \hat{p}_j)^T \Sigma_j^{-1} (p_j - \hat{p}_j). \quad (13.10)$$

Directional Deviation Metric: From the vessel center lines in each corresponding point pair (p_j, \hat{p}_j) (except in the branching points) a normalized tangential vector pair (d_j, \hat{d}_j) can be computed. The sine of the angle between the direction vectors is a measure for the deviation of the two directions. As the cross product (\times) is related to the sine an evaluation metric can be defined by:

$$M_{dir}(T, \hat{T}) = \frac{1}{m} \sum_{j=1}^m \|d_j \times \hat{d}_j\|. \quad (13.11)$$

Segment-Wise Relative Volume Overlap: A metric that is less dependent on the center line extraction is the relative volume overlap of corresponding vessel segments. Such overlap measures are very popular in the evaluation of segmentation methods [84]. Each vessel voxel is assigned to the closest segment path point. For all points on corresponding segment paths P_i, \hat{P}_i the assigned voxels define the volumes V_i, \hat{V}_i of the particular segments. The the relative overlap measure ist defined as:

$$M_{vol}(T, \hat{T}) := \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{|V_i \cap W_i|}{|V_i \cup W_i|} \right). \quad (13.12)$$

This measure is only sensitive to small deviations, but does not differentiate bigger deviations. In particular for small corresponding vessels which do not overlap there is no difference in the metric whether the vessels are very close to or very far from each other. Therefore Crum et al. [42] introduced a correspondence tolerance parameter $\tau \in \mathbb{R}_0^+$. Regions are considered overlapping if they lie within a distance of τ . A further interesting variant is presented by Schaap et al. [202]. They consider only the overlap of clinically relevant parts of the vessels. An easy measure for the relevance is the diameter of a vessel; the bigger the diameter the more important the vessel.

13.5 Further Assessment Criteria

13.5.1 Robustness

It is expected that an image processing method produces reliable results with an acceptable accuracy at least for a given image class. Hence a registration method has to be tested for a significant number of clinical image data samples, where artifacts and image qualities are contained, which typically occur in clinical practice. The sensitivity of the resulting registration accuracy to disturbances like noise and artifacts may be investigated explicitly by simulations. Also the sensitivity of the results to starting position and chosen parameters needs to be considered.

13.5.2 Effort and Runtime

The ultimate aim of the development and application of a registration algorithm is an automatic method. On clinical images this is often not feasible without decreasing the accuracy and robustness. Thus it may be better to reach a sufficiently accurate result with a higher interactive effort than an inaccurate result with a fully automatic method. Interactive specification of prior knowledge or the manual adjustment of parameters is only reasonable for medical experts and acceptable in clinical routine, if the interaction is intuitive resp. a needed parameter change is easily understandable.

A frequently occurring question is how long a registration process lasts and it is equally often answered with “it depends”. Most of the optimization algorithms are iterative methods (see Chapter 5). They are performed until a stop criterion is reached. This means the runtime of a registration process depends on these stopping criteria. The runtime of registration methods, for which closed-form solutions exist, is usually significantly shorter. The used hardware is of course crucial. The processor performance as well as the size of the main memory have an influence on the runtime. Most of the registration algorithms can at least in parts be parallelized, which results in significant runtime reduction. No expensive special parallel computers are necessary for this as modern standard PCs already feature multicore processors. With the efficient implementation of a registration method, which considers the potential of multicore processors, the runtime can be improved accordingly.

In general, rigid registration methods are very fast because only few parameters have to be optimized. The runtime of schemes which are based on corresponding landmarks depends on the number of landmarks which is typically low (between 4 and 10). Thus very low runtimes of few seconds are possible. The runtime for the application of the determined transformation depends on the number of voxels in the template image because the transformation has to be applied to every single voxel (interpolation). But usually the computation is below one minute. This is basically also true for non-rigid landmark registration where the application of the transformation is more computationally expensive but still between one and two minutes.

The main computation time for intensity-based registration methods is needed for the evaluation of the objective function (with derivatives) and the solution of large linear systems for the computation of the search direction for the next optimization step. With the presented multi-level strategies (see Sec. 6.5) and efficient optimization algorithms (see Sec. 5) very significant accelerations can be achieved even for non-rigid registration problems.

Chapter 14

Transfer of Preoperative Models to Intraoperative 3D Ultrasound Data

14.1 Introduction

An important issue in computer-assisted surgery of the liver is a fast and reliable transfer of preoperative resection plans (see Sec. 2.3.3) to the intraoperative situation. We already presented navigation systems based on intraoperative 3D ultrasound which display the position and orientation of surgical instruments in relation to liver vessels and tumors imaged in the ultrasound data (see Sec. 2.4.2 and Sec. 2.4.3). The problem is to match the planning data, derived from preoperative CT images, with these intraoperative 3D ultrasound images. As the liver deforms significantly in the intraoperative situation non-rigid registration is necessary.

There are only few works published regarding CT/MRI-to-ultrasound registration. Rigid methods have been presented, which are either intensity-[190, 220] or feature-based [179, 175]. Usually the liver vessels serve as features, because of their easy identification in CT/MRI and ultrasound data, in particular in power Doppler ultrasound. Extensions of such vessel-based approaches to non-rigid transformations are described in [115, 114, 186, 185]. These methods, however, suffer from the problem that vessels cannot be extracted automatically from ultrasound data with high accuracy and at high speed. Alternatively, hybrid approaches [6] fit preoperatively extracted features directly to the intraoperative image data. In liver surgery those features are already available from surgery planning. The new distance measures we presented in Sec. 9 allow an extension of these hybrid methods to non-rigid transformations [118].

Non-rigid landmark registration using thin-plate splines (see Sec. 11.8) has been applied to MRI-ultrasound registration [72]. We already argued in Sec. 12 that pure landmark schemes only take into account information at the landmark positions and ignore the complete intensity information of all other image

points. In Chapter 12 we presented different possibilities to combine intensity based registration and some interactively chosen landmark pairs.

At first we add landmark interpolation conditions as equality constraints to the non-parametric registration problem to guarantee that the corresponding landmarks are exactly mapped onto each other (see Sec. 12.3). We apply the approach to clinical data. As the interactive localization of point landmarks is always prone to errors, we apply the inequality constrained optimization scheme from Sec. 12.4, which guarantees corresponding landmarks to be at most a given distance apart from each other after registration, in a second experiment. We use isotropic tolerances for landmarks at vessel branchings and anisotropic tolerances at vessel segments between branchings as the localization uncertainties deviate in different directions at vessel segments. We start with a description of the clinical data.

14.2 Specification of Clinical Image Data

14.2.1 Preoperative CT Data

A detailed description of CT imaging of the liver is given in Sec. 2.3.1. Here we specify the imaging parameters of the used clinical CT data. For each patient a triphasic helical single-source 64-slice multidetector computed tomography (MDCT) scan of the abdomen (LightSpeed VCT; General Electric Medical Systems, Milwaukee, WI) was acquired. The MDCT was performed after intravenous mechanical injection of 120 ml nonionic iodinated contrast medium (iodine, 370 mg/ml; Ultravist 370; Schering, Berlin, Germany) at a flow rate of 4 ml/s. Bolus tracking was used for an early arterial phase (upper abdomen) to optimize contrast filling of the vessels. This resulted in a scan delay of approximately 18 seconds. Delays of 20 and 60 seconds from the beginning of the early arterial scan were used for the portal venous (upper abdomen) and late venous phase (entire abdomen) scans, respectively. The collimation was set to 64×1.25 mm, with a gantry rotation time of 0.7 seconds. The table feed was 13.75 mm/rotation for the arterial and portal venous phase and 35.0 mm/rotation for the venous phase. Tube current and voltage were set to 350 mA and 120 kV for the arterial and portal venous phase, and to 280 mA and 120 kV for the venous phase, respectively. Images were reconstructed with a slice thickness of 1.25 mm.

14.2.2 3D Ultrasound

The intraoperative 3D ultrasound images were acquired directly on the liver in the open abdomen by the Voluson 730 machine already described in Sec. 2.4.1. A mechanically swept 3D probe containing a curved array transducer for abdominal applications was used which was wrapped in a sterile drape. The original image geometry is given in torus coordinates (see Sec. 2.4.1). The

original data are transferred to the navigation computer via DICOM and reformatted to 0.5 mm isotropic voxels (see Sec. 2.4.1). B-mode and power Doppler ultrasound images are acquired synchronously. Thus both ultrasound modalities are naturally registered.

14.3 Non-Rigid Registration with Equality Landmark Constraints

The non-rigid registration approach with landmark interpolation conditions was formulated in Sec. 12.3. We use the distance measure normalized gradient (NGF) field distance measure (see Sec. 8.3.5) and the elastic regularizer (see Sec. 6.3).

The discretization of the resulting equality constrained optimization problem is given in Sec. 12.5.1. The discretized optimization problem is solved by the algorithm described in Sec. 12.6.1, which eliminates the constraints and solves the resulting unconstrained optimization problem by a standard Gauß-Newton scheme. Finally, to avoid local minima and to speed up convergence we also use a multilevel approach via the parameter of a smoothing spline as described in Sec. 6.5.

The approach and the evaluation on clinical data is based on our publication Lange et al. [121].

14.3.1 Landmark Identification at Vessel Branchings

Only a small number of corresponding point landmarks (usually five to six, rarely up to ten) can be identified interactively in the available time window in the OR. Thus, efficient and intuitive interaction mechanisms are required to support landmark placement. For contrast-enhanced CT and power Doppler ultrasound images of the liver corresponding vessel branchings are a natural choice for point landmarks. In the preoperative CT data all branching points are computed in advance, based on the center lines of the segmented vessels. In order to interactively select a point landmark in CT data, the vessels are visualized as surfaces. When the user clicks on the surface near a vessel branching, then the nearest predetermined branching point is chosen automatically.

Unfortunately, reliable vessel segmentation from US images has not yet been achieved in a robust and accurate fashion. Therefore, landmarks in the intraoperative US data have to be identified interactively in the intensity data. One way to improve the definition of the intraoperative landmarks could be to click only approximately near the corresponding vessel branching and then automatically fit a vessel branching model to the intensity data at this location [261]. This, however, is subject to future research.

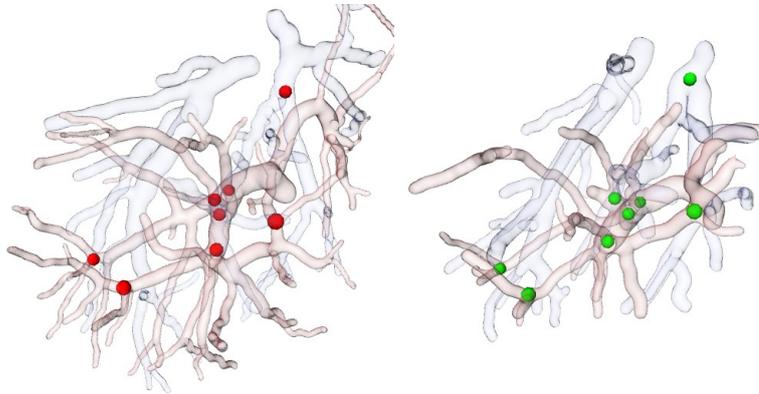


Figure 14.1 Selected corresponding landmarks at vessel branchings from CT data (left Fig.) and ultrasound data (right Fig.).

14.3.2 Results

A qualitative and quantitative validation on clinical liver data sets (Sec. 14.2) of three different patients has been performed. Only the portal venous phase of the preoperative CTs was used for registration. The portal veins are of high contrast in all cases, but the hepatic veins were hardly visible in case one and three and significantly lower in case two compared to the portal veins. Only the power Doppler ultrasound images were considered, the B-mode images were ignored. Eight landmarks at branching points of the portal veins were chosen interactively (Fig. 14.1) lasting five to ten minutes.

In all three cases a rigid registration was performed first to get a better initialization for the non-rigid registration and to show improvements by non-rigid compared to rigid registration. Next, a thin-plate spline (TPS) registration (see Sec. 11.8) was performed based on the eight landmark pairs. Finally, the combined approach was applied starting with the TPS displacement vector field as a specific solution of the landmark constraints (see Sec. 12.6.1).

The combined approach started with a grid spacing for the displacement vector field of 6 mm and ended with a finest spacing of 3 mm and 32^3 grid points. The resulting displacement field was interpolated to the original resolution of 0.5 mm. The run time of the optimization was approximately 10 min on an Intel Core Duo processor, with 1.83 GHz and 2 GB RAM using MATLAB 7.6. The algorithm has not yet been optimized for runtime.

The maximal landmark difference after combined registration is below 0.3 mm (reached for case one). These differences are due to the linear interpolation of the displacements at the landmarks from the neighboring grid points (see Sec. 12.5) with a grid spacing of 3 mm. At the same time the distance measure reduces by 73% compared to TPS registration.

In Fig. 14.3 three differently oriented image planes illustrate the perfor-

Table 14.1 Mean (std) distance between corresponding points on vessel center lines (in mm).

	Rigid	TPS	Combined
Case 1	4.4(\pm 1.8)	2.5(\pm 2.0)	2.6(\pm 2.0)
Case 2	4.6(\pm 1.9)	4.6(\pm 3.4)	3.6(\pm 3.4)
Case 3	4.7(\pm 2.3)	5.1(\pm 4.2)	4.9(\pm 4.3)

Table 14.2 Ratio of corresponding points on vessel center lines above 3 mm distance (in %).

	rigid	TPS	Combined
Case 1	82%	27%	30%
Case 2	76%	61%	41%
Case 3	75%	59%	54%

mance of the combined landmark-intensity registration approach in comparison to a rigid transformation and a TPS warping based on the landmarks alone. It is clearly visible that a rigid transformation is not satisfactory thus justifying a non-rigid method. The combined method improves the TPS warping.

For the quantitative validation we used dense corresponding points on center lines of portal and hepatic veins (see Sec. 13.4 for detailed information). The distribution of the distances on the center lines are visualized in Fig. 14.2. The statistics of the distances is summarized in Tab. 14.1 and 14.2.

The mean distances between corresponding points after rigid registration are in the range of 4.4 to 4.7 mm and 75% to 82% of the points are above 3 mm distance. This means there are significant deformations left after rigid registration. For case one, TPS as well as the combined method reduce the distances considerably, but there is virtually no difference between TPS and combined method. In this case, however, the TPS registration already leads to good results. There is only little space left for improvements in this case. For case two, TPS only slightly improves upon the rigid registration, yet the combined method improves significantly. For case two, TPS and combined method do not improve the rigid registration if looking at the mean distance, but the ratio of distances above 3 mm reduces, for the combined method more as for the TPS. Looking at the visualizations of the distances some vessel parts are very close (below 1 mm) after combined registration although the registration is based on the intensity images and not on the center lines. In case two, there are two parts of the hepatic veins, whose distance is large after rigid registration and still increases after TPS registration. It seems that the combined approach cannot compensate the large deviations that exist after

14. TRANSFER OF PREOPERATIVE MODELS TO INTRAOPERATIVE 3D ULTRASOUND DATA

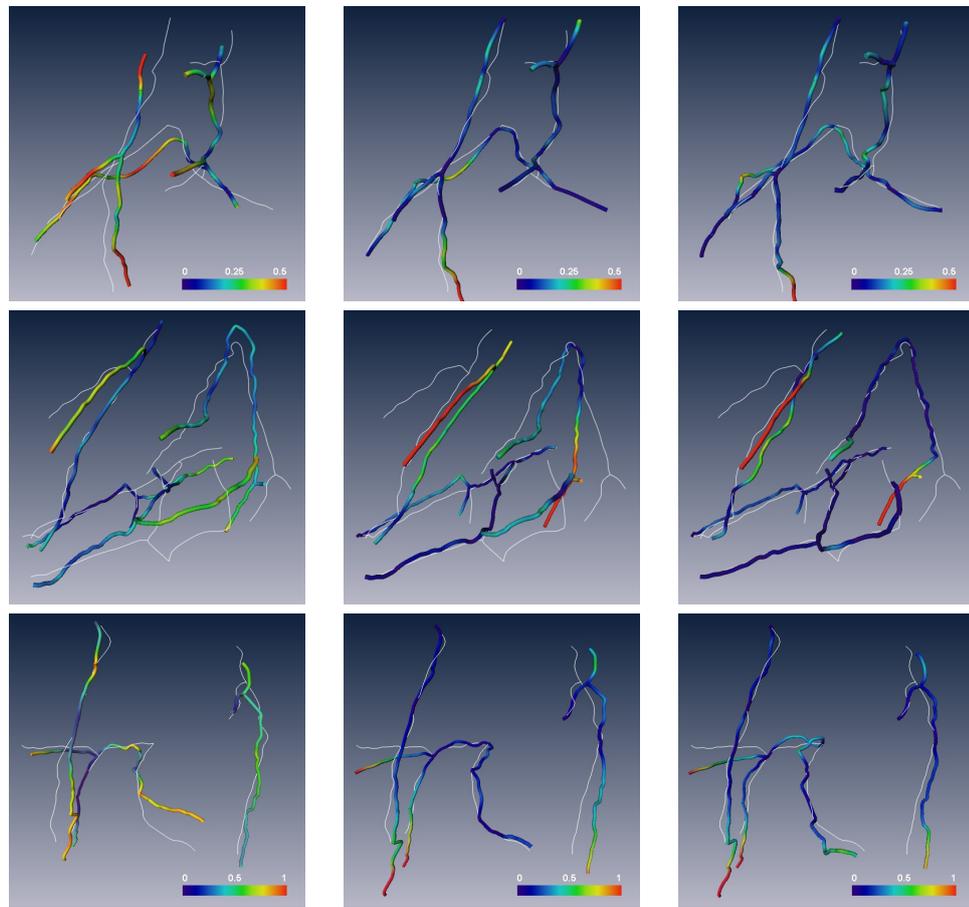


Figure 14.2 Distribution of distances for corresponding points on vessel center lines. The distances are color-coded onto the registered model center lines. The thin white lines represent the reference center lines. Each row illustrates one clinical case. The first column shows the distributions after rigid, the second column after TPS and the last column after combined registration.

TPS warping or are even induced by it.

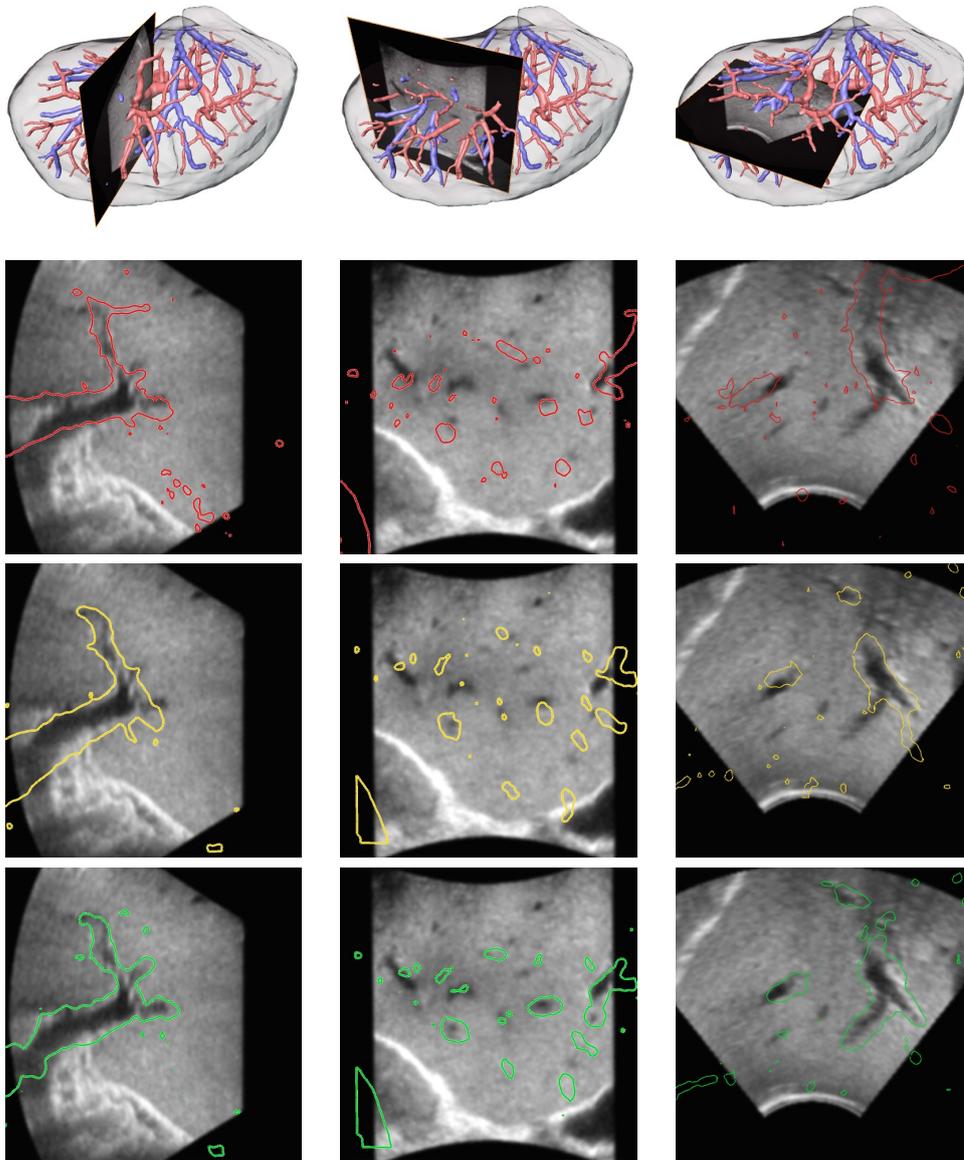


Figure 14.3 Registration results for case one. First row: Position of three different slice orientations in relation to liver model from CT data. The second, third and fourth row illustrate the transformed CT data as isolines in the B-Mode ultrasound data. Second row: After rigid registration. Third row: After thin-plate spline registration. Fourth row: After combined nonlinear landmark-intensity registration.

14.3.3 Discussion and Conclusion

The main difficulties in our specific application are the quality of the US data with low contrast, high noise and artifacts, like shadowing in B-Mode ultrasound or motion artifacts in power Doppler ultrasound. In addition, the contrast of the hepatic veins is low in the portal venous phase of the CTs, and the contrast of the portal veins is low in the hepatic venous phase.

We have applied a combined landmark-intensity registration approach to clinical image pairs of portal venous phase CT and 3D power Doppler ultrasound data. The advantage of the method is that a priori knowledge – provided in terms of few landmarks – guides the registration process, and reduces the number of local minima. In contrast to incorporating the landmarks via a penalizer term no additional parameter has to be tuned in this constrained optimization formulation. The landmarks are guaranteed to match each other for each landmark pair. In case of a penalizer, however, the sum of the landmark distances is minimized through a trade-off with the distance measure and the regularizer. Thus, the distances of single landmark pairs might still be high after registration.

Although the qualitative and quantitative validation results are promising, some challenges remain. The first challenge is the low – or often non-existing – contrast of hepatic veins in the portal venous phase of the CTs. A possible solution is to generate high contrast intensity models containing portal and hepatic veins (see Sec. 9.2 and [118]). In a preliminary experiment we used such intensity models for cases one and three. In both cases we got an improved mean distance (case 1: $2.5(\pm 1.5)$ mm, case 3: $4.4(\pm 3.9)$ mm).

A second challenge are inaccuracies in the localization of the landmarks. In the presented approach no landmark errors are assumed. This leads to distorted vessels in some regions, because the endpoints of the vessels are forced to the inaccurate landmark positions. Incorporating landmarks as soft constraints is not an adequate solution to this problem, because the distance of corresponding landmarks is minimized only globally and not locally, as mentioned above. A better solution is the introduction of local tolerances as will be explained in the next section.

The third challenge has been revealed in section 14.3.2. TPS might provide a bad starting value in regions far away from landmarks. The combined registration process might not be able to compensate the initial displacements in these regions. Other interpolating functions like GEBS (see Sec. 11.9.1 or [103]) are promising alternatives.

14.4 Non-Rigid Registration with Inequality Landmark Constraints

As the interactive localization of point landmarks is always prone to errors, we generalize the combined landmark and intensity approach to an inequality constrained optimization scheme, which guarantees corresponding landmarks to be at most a given distance apart from each other after registration. As the localization uncertainties might deviate in different directions we introduce a landmark constrained registration scheme with anisotropic tolerances (error ellipsoids), which can be used for the registration of vascular structures.

The continuous formulation of the landmark constrained approach with anisotropic tolerances, its discretization and the optimization algorithm are described in Sec. 12.4, Sec. 12.5.2 and Sec. 12.6.2. This approach yields a higher flexibility in defining landmark pairs. For example instead of using point landmarks at corresponding liver vessel branchings, which are usually difficult to identify in 3D, also landmarks between two corresponding branchings (vessel segments) can be used, which are often easier to identify. With the latter landmarks the localization uncertainty is high along the vessel, but low perpendicular to it (see Fig. 14.4). Thus in addition to landmarks at vessel branchings with isotropic localization uncertainties we will integrate landmarks at segments with anisotropic localization uncertainties into the nonrigid registration framework. In the following the two types of landmarks are called *branching* and *segment* landmarks.

14.4.1 Determination of Weighting Matrices

The covariance of the anisotropic localization uncertainty can be modeled via their eigenvalues and eigenvectors. The first eigenvector $v_1^j \in \mathbb{R}^3$ points in the direction of the corresponding vessel at the position of the landmark, the other two eigenvectors $v_2^j, v_3^j \in \mathbb{R}^3$ are perpendicular to it. As the localization uncertainty is high in the direction of the vessel and low perpendicular to it depending on the area of the vessel cross section the eigenvalues are chosen as $\lambda_1^j = 5r_j^2, \lambda_2^j = r_j^2, \lambda_3^j = r_j^2$, with r_j being a radius estimation of the vessel at landmark j . The radii are already available due to the preoperative modeling of the vessels for the surgery planning process. With $D_j = \text{diag}(\lambda_1^j, \lambda_2^j, \lambda_3^j)$ and $V_j = (v_1^j, v_2^j, v_3^j)$ the uncertainty matrices are defined as:

$$\Sigma_j = V_j^T D_j V_j. \quad (14.1)$$

The weighting matrices are then the inverse of the covariance matrices: $W_j = \Sigma_j^{-1}$. The approach can be seen as a generalization of the isotropic tolerance method described in [172]. To handle such isotropic tolerances the matrix Σ is chosen as $\Sigma_j = a_j^2 I_3$, with $a_j \in \mathbb{R}^+$ being the radius of an error sphere around landmark j .

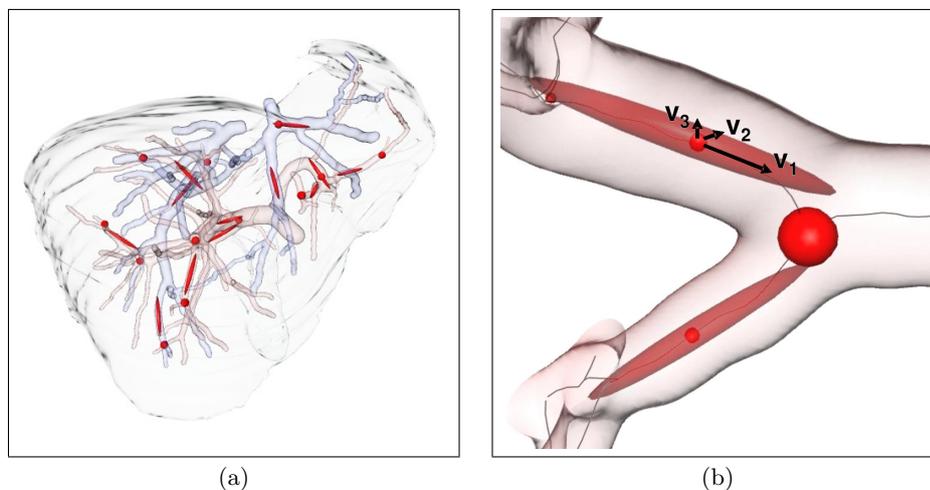


Figure 14.4 a) Different types of landmarks at vessel branchings (isotropic error spheres) and segments (anisotropic error ellipsoids). b) A detailed view with two segment and one branching landmark. The eigenvectors v_1, v_2, v_3 of the anisotropic error ellipsoid of one segment landmark are shown.

It is assumed, that after a rigid pre-registration corresponding vessel segments approximately point in the same direction, such that the two covariance matrices of corresponding landmarks can be added to one joint covariance matrix for each landmark pair. The inverse of the joint covariance matrix is the weighting matrix. An elaborated discussion of this issue can be found in Sec. 11.5.1.

14.4.2 Results

We show the effectiveness of the proposed registration scheme on an illustrative example, which is a simplified geometry of portal veins in a real liver CT containing only the biggest vessels. We determine a realistic deformation based on a combined intensity and landmark registration with equality constraints of clinical CT and 3D ultrasound data. The landmarks have been defined on vessel branchings. This deformation is applied to the example image to get an artificial template. The first row of Fig. 14.5 shows the vessels extracted from the template and reference image as well as the landmarks and vessel center lines. The points on the center lines are displaced by 5 ± 2 mm. Six landmark pairs have been chosen interactively on the vessel segments. Landmark 3 has been moved by 10 mm in the direction of the vessel.

We chose the curvature regularizer $\mathcal{S}^{\text{curv}}$ (see Sec. 6.3) with y^{ref} being the initial displacement field and the sum of squared differences \mathcal{D}^{SSD} distance measure (see Sec. 6.2) because in our example the vessels appear bright com-

pared to the background in template and reference image. We used a multi-level and multi-resolution strategy for the images T, R and the displacement field y . We started on a grid with 7.5 mm spacing and refined twice to a final spacing of 1.9 mm. The original resolution of the images was $1 \times 1 \times 1 \text{ mm}^3$.

As can be seen in the second and third row of Fig. 14.5 the isotropic tolerance at landmark three is too restrictive to compensate the displacement of the landmark, but the anisotropic tolerance is suitable for compensation while keeping the restrictive tolerance perpendicular to the vessel.

14.4.3 Discussion and Conclusion

The contribution of this work is a modeling of a combined landmark and intensity registration approach as an inequality constrained optimization problem guaranteeing that each reference landmark lies within an error ellipsoid around the corresponding template landmark at the end of the registration process. In contrast to Wörz et al. [263] the anisotropically weighted landmark differences are not added as a penalizer to the registration functional, but as hard inequality constraints. In addition a direct optimization scheme has been implemented instead of an alternating optimization scheme.

Vessel segment landmarks with anisotropic localization uncertainties are a promising alternative and/or extension to vessel branching landmarks with isotropic localization uncertainties. They offer an additional flexibility for the interactive definition of landmarks on vessel trees allowing for an intuitive and efficient registration workflow. The first results on an illustrative but realistic example are promising to an extent that the next step will be a thorough and quantitative investigation on a significant number of clinical data sets from patients, which underwent computer assisted liver surgery.



Figure 14.5 In the first column reference (gray beige) and template (red) vessels are shown. The second row shows the error ellipsoids around template landmarks, the position of reference landmarks and the vessel center lines. In the first row the original deformed vessels and landmark positions are presented. Landmark 3 is displaced in the direction of the vessel. In the second row we present registration results using isotropic tolerances around the landmarks and in the third row results using anisotropic tolerances.

14.5 Non-Rigid Registration with Shape Distance Measure

In the Sections 14.3 and 14.4 we incorporated a priori knowledge about the sought registration *transformation* by constraining the transformation at some landmark pairs on vessel segments or branchings. Now we apply the shape distance measure based on a global filter kernel, which we defined in Sec. 9.4, to incorporate a priori knowledge about the *image content*. The liver vessel center lines and their radii extracted during the planning process from preoperative CT data (see Sec. 2.3.2) represent a priori knowlegde about the image content, which we also expect in the intraoperative 3D ultrasound data. Based on this information a global filter kernel (see Sec. 9.4.1) is generated using local Laplacian vessel filter kernels K_L defined in Eq. 9.16. The distance measure $\mathcal{D}^{\text{CONV}}$ defined in Eq. 9.13 is the response of this global filter applied to the intraoperative ultrasound data. The better the transformed ultrasound data fit the structures modeled into the global filter kernel the higher the filter response and the lower the distance measure.

In the following we summarize first results, which were published in [118, 119].

14.5.1 Results

In order to qualitatively validate the proposed distance measure $\mathcal{D}^{\text{CONV}}$ we use it in a multilevel B-spline scheme [199, 191] (without effective multi-resolution strategy) to register artificially deformed data. Vessel center lines are extracted with radii from real intraoperative 3D power Doppler ultrasound data. These center lines are deformed by a realistic B-spline deformation generated by a feature-based registration approach [115, 114] and thereby the center line points are shifted by $4.5(\pm 2.9)$ mm on average and maximally 9.6 mm (see Fig. 14.6 a). The global kernel is determined on the deformed center lines and rigidly (Fig. 14.6 b) resp. nonrigidly (Fig. 14.6 c) registered. The deformation is substantially reduced and the original state is recovered well from a visual point of view. We quantify the resulting deviations from the original and the registered vessels by computing the distance of corresponding center line points. After rigid registration a deviation of 3.3 (+/-0.2) mm on average and a maximum of 7.7 mm is left. After non-rigid registration the deviation is reduced to 1.0 (+/- 0.4) mm on average and a maximum of 2.3 mm. It cannot be expected that the original state can be perfectly reproduced by the registration algorithm, since segmentation, skeletonization and radius computation lead to certain inaccuracies.

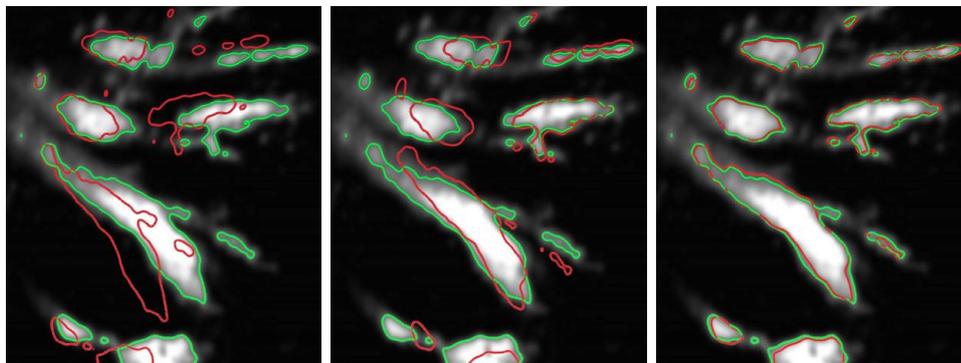


Figure 14.6 Power Doppler ultrasound data of liver vessels with a) artificially deformed, b) rigidly and c) non-rigidly registered vessels.

14.5.2 Discussion and Conclusion

The shape distance measure $\mathcal{D}^{\text{CONV}}$ based on a global vessel filter kernel has been successfully applied to the non-rigid registration of a realistic vessel tree model to a clinical 3D power Doppler ultrasound data set of the liver. But the result is very preliminary. Instead of using the parametric B-spline approach we will integrate this measure and the local correlation based measure described in Sec. 9.3 into the non-parametric image registration framework we introduced in Sec. 6.4.

For a detailed evaluation the properties of the distance measure will first be investigated directly by the protocols of Skerl et al. [218, 219] for simulated rigid and non-rigid transformations, which we discussed in Sec. 13.2.2 and Sec. 13.2.3. The simulation of rigid transformations is explained and discussed in Sec. 13.3.

Then the whole registration process including distance measure, regularizer and optimization algorithm will be evaluated on simulated rigid and non-rigid transformations (Sec. 13.1.2). The advantage of the registration of models and image data is that also changes of the model geometry and topology can easily be simulated by changing radii or by removing vessel branches. The sensitivity of the distance measure to such changes occurring in real data can then be investigated.

In a final step the distance measure and registration process have to be evaluated on clinical image data pairs. Therefore we will use the reference standard based on dense corresponding vessel center line points, which was defined in Sec. 13.4 and proved its value already in Sec. 14.3, Sec. 14.4 and Chap. 15.

In conclusion we showed that the distance measure works in principle and we have developed all tools for a thorough investigation and evaluation.

Chapter 15

Quantitative Postoperative Control

15.1 Introduction

The validation of new computer-assisted surgical procedures in oncological liver surgery like preoperative planning (Chapter 2.3) and intraoperative navigation (Chapter 2.4) is challenging. The main question is, how accurate a resection plan has been implemented in the operating room. Besides complete tumor removal it is very important to resect exactly the planned parts of the liver vessels to ensure blood supply and drainage of the remaining liver tissue. The comparison of planned (virtual) and resected (real) tissue volumes [129] is a popular, but only coarse and unspecific validation possibility, because on the one hand the determination of the resected volume is imprecise due to blood loss and on the other hand equal volumes do not imply that virtual and real parts of the liver correspond. In an extreme case the volumes might be exactly equal, but the liver parts do not overlap at all. A more detailed validation can be obtained by determining the remaining vessel parts based on 3D models extracted from pre- and postoperative CT data (see Fig. 15.1).

The aim is to visualize and quantify those vessel parts, which have been resected as planned and those, which should have been preserved, but have been removed nevertheless [12] (Fig. 15.2). Due to deformations of the liver between pre- and postoperative CT acquisitions non-rigid registration algorithms are needed. The challenge of this intra-patient registration task is due to the fact that significant parts of the liver are missing in the postoperative images. Although some publications exist dealing with liver registration (see Chapter 14) only few articles have been published on non-rigid registration of pre- and postoperative image data. In some publications image data of the brain before and during resp. after resection of tumors [55, 103] is used. In the case of the liver only the evolution of tumors of the same patient based on two different image acquisitions, but not the resection of tumors has been considered [30]. The algorithm of Charnoz et al. [30] finds corresponding liver vessel center lines via tree matching, but until now the method has been validated

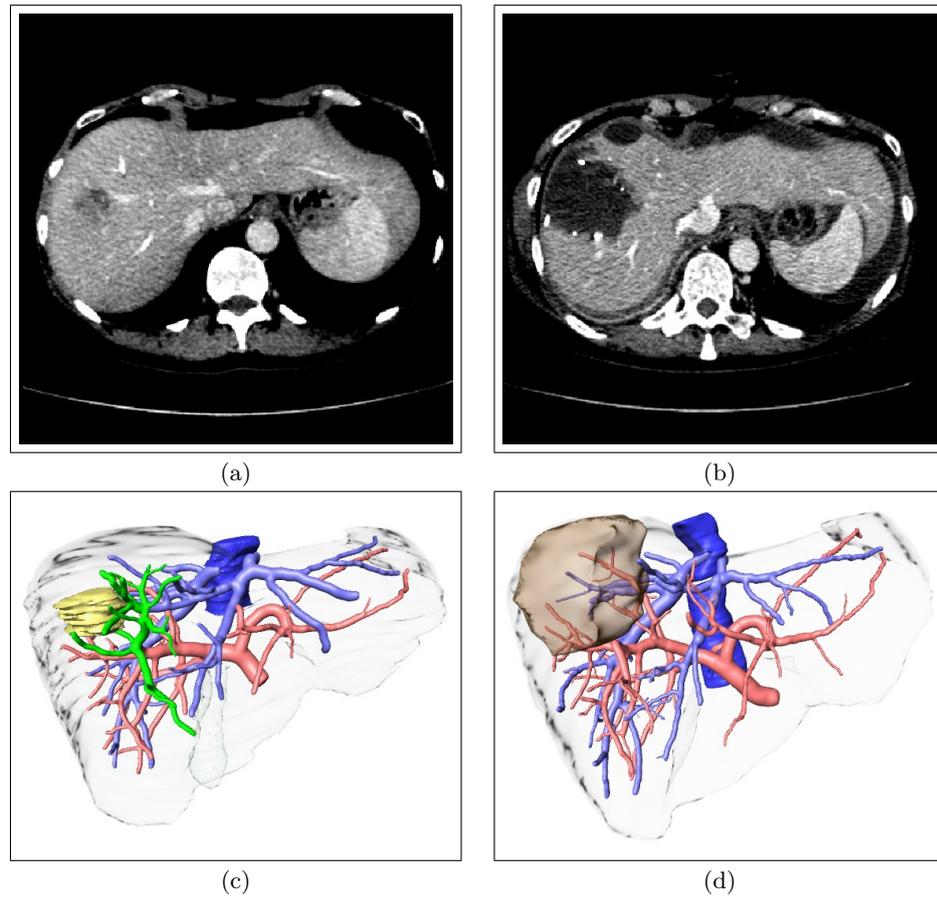


Figure 15.1 A preoperative (a) and postoperative (b) contrast-enhanced CT slice of the liver. The portal and hepatic veins appear bright, the tumor on the left and the resected part appear dark. Around the resected part bright spots are observable indicating clips. In c) a 3D surface model of the liver vessels and the tumor extracted from preoperative CT data is shown; d) the according 3D surface model after tumor resection.

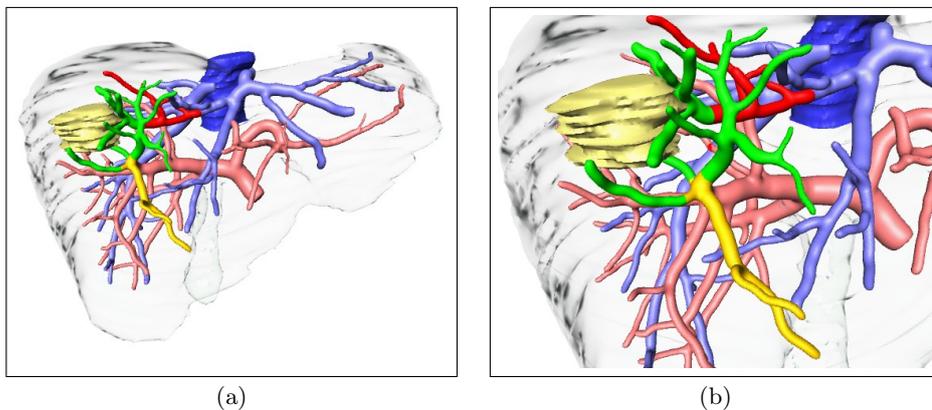


Figure 15.2 a) Again the preoperative 3D model is shown, but the parts of the vessels, which have been resected as planned (green) and unnecessarily (red) have been determined by a registered postoperative 3D model. In b) a closer look around the tumor can be seen.

only on one clinical data set. Our approach is based on interactively chosen corresponding point landmarks using different interpolation and approximation schemes based on splines. Besides the natural choice of landmarks at vessel branchings we introduce a special kind of landmarks adapted to vessel segments. This is a joint work with Stefan Wörz and Karl Rohr and has partially been published elsewhere [126, 127]. The used landmark registration methods are explained in detail in Chap. 11.

15.2 Interactive Vessel Registration

The task is to register preoperative with postoperative CT data. This is a monomodal application and the CT data are of high quality. The issue is, that the liver and the surrounding tissue are not only deformed, but also parts of the liver are missing after the resection. This means some image parts of the preoperative image data are missing in the postoperative data.

The main structures which are identifiable inside the liver are vessels. The vessels (portal veins and hepatic veins) were segmented semi-automatically in pre- and postoperative image data. Thereafter their center lines and local radii have been computed automatically.

As in Sec. 14.4 we interactively identify point landmark pairs with isotropic localization errors and segment landmarks along vessels between two branchings. With the latter landmarks the anisotropic localization uncertainty is high along the vessel, but low perpendicular to it (see Fig. 14.4). The covariance matrices and the resulting weighting matrices W_j for the branching and

segment landmarks are defined as in Sec. 14.4.1. The interactively defined landmark pairs with specified isotropic and anisotropic localization uncertainties are used to register non-rigidly the pre- and postoperative 3D models with interpolating thin-plate splines (iTPS) described in Sec. 11.8.1, interpolating Gaussian elastic body splines (iGEBS) described in Sec. 11.9.1 and approximating Gaussian elastic body splines (aGEBS) described in Sec. 11.9.2.

15.3 Results

The different registration methods were compared on clinical pre- and postoperative contrast-enhanced CT data sets of 13 different patients, which have undergone oncological liver resections. In Fig. 15.3 pre- and postoperative 3D models of five out of the 13 patients are shown. If possible resected volumes are marked in red in the postoperative models. In the last column a comparison of pre- and postoperative vessels is shown on the preoperative model. The color scheme is the same as used in the introduction 15.1 and Fig. 15.2.

For validation we used dense point correspondences on the vessel center lines and a weighted validation metric as described in Section 13.4 and [120]. In the last column of Fig. 15.4 dense point correspondences for five out of the 13 patients are shown on the preoperative vessel center lines. The used vessel center line points are marked in red and the corresponding points of the postoperative vessels are marked in green and connected by yellow lines. The location of the postoperative (green) points are shown after rigid registration. It can be seen that significant differences are left after rigid registration. For the quantitative evaluation we used a weighted Euclidean distance between corresponding points, where the weighting in the direction of the vessels is set to zero such that only the distance perpendicular to the vessels is measured.

Five sets of landmarks for each patient were interactively chosen as follows:

1. as many as possible branching landmark pairs (r_j^B, t_j^B) ,
2. a comparable number of segment landmarks on vessel segments between the branching landmarks (r_j^S, t_j^S) ,
3. a combination of the branching and segment landmarks (r_j^C, t_j^C) ,
4. a reduced number of 12 branching landmark pairs (r_j^{B12}, t_j^{B12}) and
5. a reduced number of 12 segment landmark pairs (r_j^{S12}, t_j^{S12}) .

In Fig. 15.4 the used branching and segment landmarks on the portal veins of five out of 13 patients are shown. On average over all 13 patients $32(\pm 16)$ branching and $65(\pm 32)$ segment landmarks were used.

The iTPS and iGEBS non-rigid registration approaches have been applied to landmark sets 1 and 4 which only include branching landmarks, but the aGEBS approximation was applied to all five landmark sets. The resulting average weighted Euclidean distances for each patient are shown in Fig. 15.5. We obtained a weighted distance of 4.9 mm after rigid registration averaged over all patients. iGEBS and aGEBS using branching landmarks decrease the average distance down to 1.7 and 1.5 mm, resp. aGEBS using segment landmarks and iTPS using branching landmarks yield comparable results and lead to 1.4 mm average distance. By using a combination of segment and branching landmarks the best results of 1.0 mm average distance were achieved. In Fig. 15.6 for one patient the remaining differences after rigid (4.7 mm on average), iTPS on branching landmarks (1.5 mm on average) and aGEBS registration on branching and segment landmarks (1.0 mm on average) are visualized.

Because interactive determination of landmarks is tedious and time-consuming we like to define only a necessary number of landmarks. Therefore we also performed a validation based on a reduced set of 12 landmarks (see Fig. 15.5). In this case we obtained 2.5 mm for iGEBS, the same distance of 2.3 mm for iTPS and aGEBS on branching landmarks and the best result of 2.0 mm for aGEBS on segment landmarks. A combination of the twelve branching and the twelve segment landmarks was not investigated, because doubling such a small number of landmarks would not be a fair comparison.

15.4 Conclusions

Interpolating and approximating landmark-based schemes were presented for non-rigid registration of pre- and postoperative CTs of the liver. Besides point landmarks at vessel branchings, where the localization uncertainties were assumed to be zero (exact localization) resp. to be isotropic, also landmarks along tube-like vessel segments with anisotropic localization uncertainties (error ellipsoids) were introduced. Five different sets of landmarks were validated on 13 clinical image data pairs using interpolating TPS, interpolating GEBS, and approximating GEBS. With both types of landmarks (branchings and segments) comparable registration accuracies can be obtained and a combination of both improves the accuracy. For a low number of landmarks segment landmarks are even superior. Segment landmarks with anisotropic localization uncertainties are a promising alternative and/or extension to branching landmarks. They offer an additional flexibility in interactive landmark registration allowing an intuitive and efficient registration workflow. Based on the registration results an automatic algorithm can be developed to identify parts of the vessel trees, which have been removed during the surgical procedure. This allows a visual (see Fig. 15.3) and quantitative [12] assessment of liver resections.

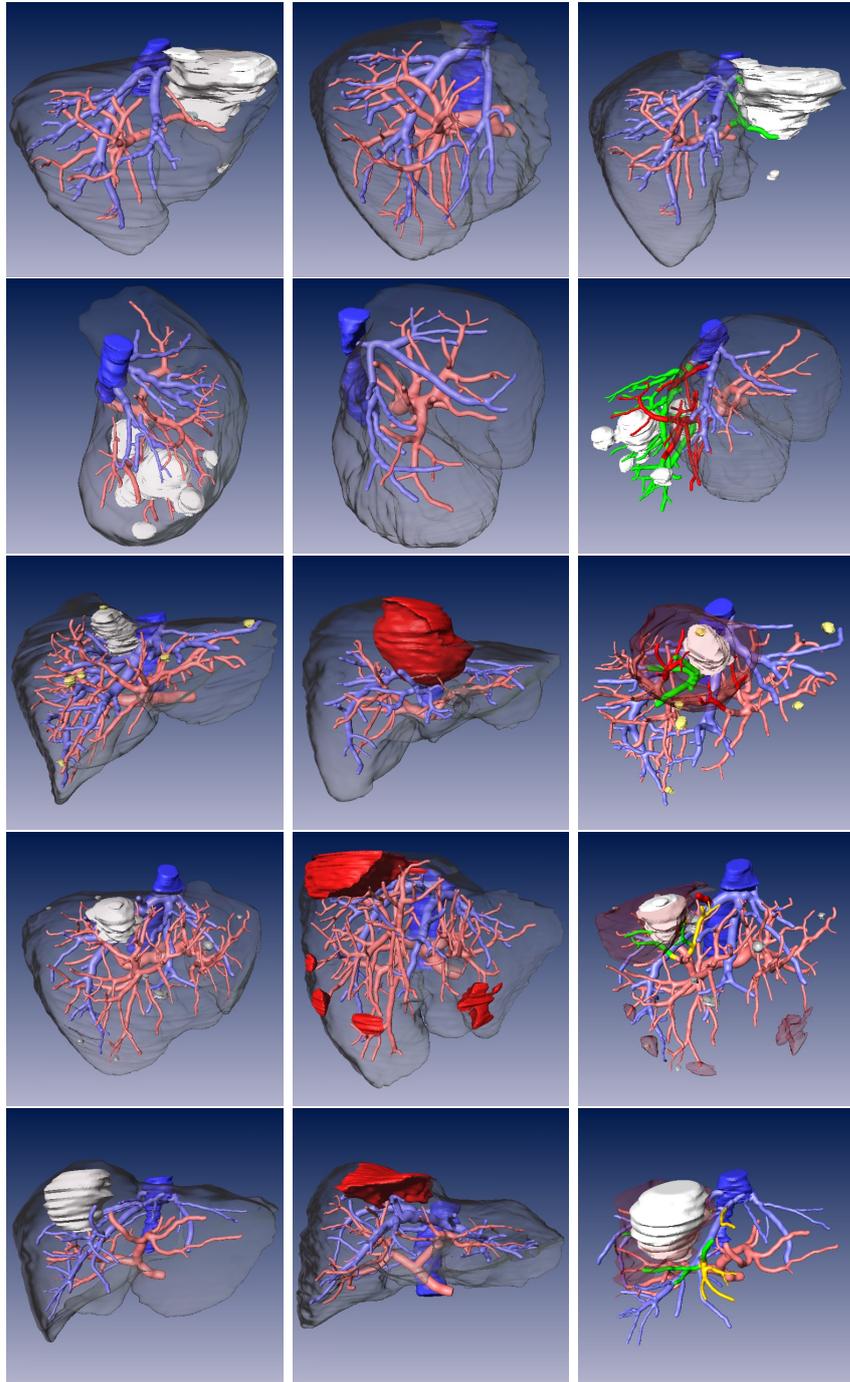


Figure 15.3 A 3D model from preoperative (first column) and postoperative (second column) CT data from five out of 13 patients are shown. In the last column the vessels are marked on the preoperative model according to their status after resection with the color scheme described in Fig. 15.2.

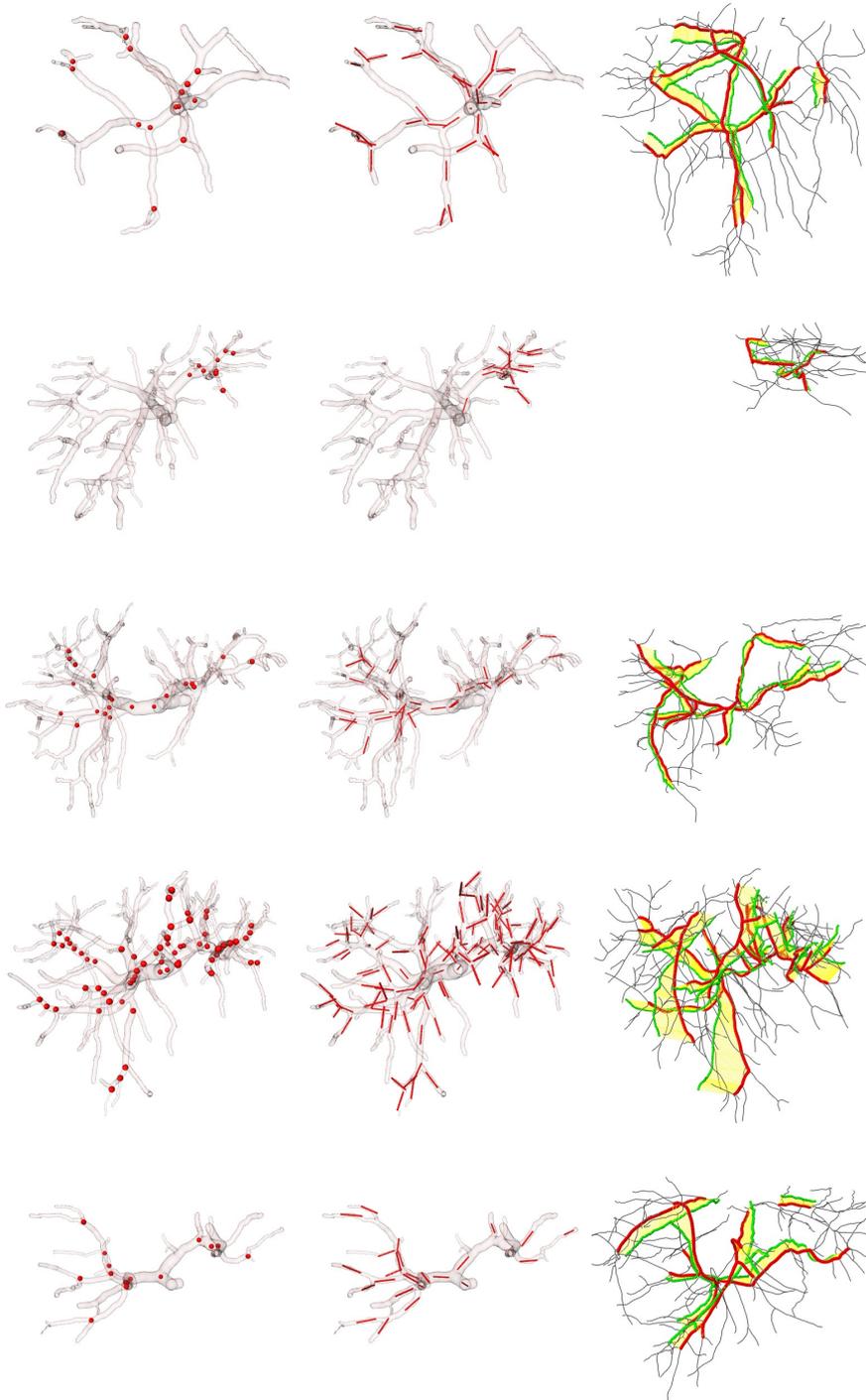


Figure 15.4 The chosen branching (first column) and segment (second column) landmarks on the portal veins of five out of 13 overall patients as well as the parts of the vessel center lines used for validation are illustrated (last column).

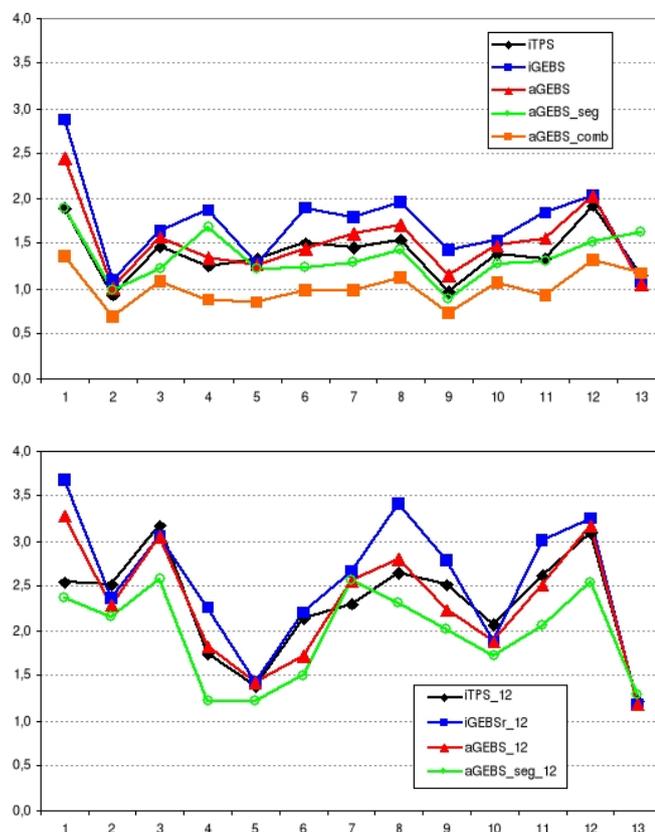


Figure 15.5 Mean weighted Euclidean distances (in mm) at vessel center lines of all 13 patients for *all* (upper row) and a *reduced set* (lower row) of 12 landmark pairs after registration.

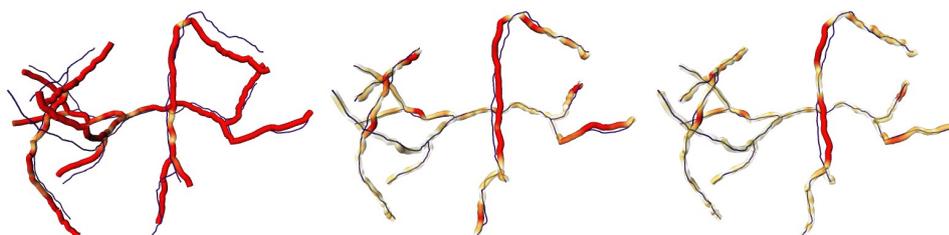


Figure 15.6 Color-coded differences on corresponding vessel center line points of case seven after rigid (left), non-rigid iTPS on branching landmarks (middle), and non-rigid aGEBS registration on branching and segment landmarks (right). Differences of 0 mm are white, between 0 and 3 mm are yellow resp. orange and of 3 mm and above are red.

Conclusions

In the Chapters 14 and 15 we have presented solutions for two important types of non-rigid image registration problems arising in computer assisted liver surgery:

- the transfer of preoperative images and planning models onto intraoperative 3D ultrasound data needed for the accurate implementation of a preoperative planning by means of a navigation system,
- the comparison of pre- and postoperative vessel models for the quantitative assessment of computer assisted planning and intraoperative navigation systems.

We have shown that the solutions fulfill the requirements necessary for being used in a clinical environment in the future. First validation trails are already under way. On a small sample set of image data from a real clinical setting intraoperative registration with a clinically relevant accuracy was already reached. The registration of pre- and postoperative vessel models was proved to be very accurate on a significant number of clinical data sets.

These non-rigid registration problems are so demanding that general registration algorithms are likely to fail. We do not think that even in the future there will only one excellent algorithm which solves all registration problems arising from very different applications. One result of this thesis is that the modeling of a priori application specific knowledge into a general and flexible registration framework is a successful strategy. Such a priori knowledge is or can be made available in different forms. For example, if explicit anatomical models with certain shapes like the liver vessels are on hand this knowledge about the image content can be considered in the registration distance measure. Explicit application specific information about the registration transformation is another form of a priori knowledge.

The basic challenge is to model the a priori knowledge mathematically consistent into a non-parametric image registration framework while ensuring that the registration process stays efficient which is in particular important for intraoperative registration.

We developed a distance measure which evaluates the response of a local vessel filter at each point on the vessel center lines extracted from preoperative data. The sum of all these filter responses is maximized assuming a high response in the presence of a vessel in the intraoperative data. This basic idea can be extended to other shapes like plates or corners by means of according filter kernels. We discovered that the distance measure can be reformulated such that the local vessel filter kernels can be integrated preoperatively and then the resulting global kernel just has to be multiplied with the intraoperative data. This allows an efficient computation of the distance measure during the intraoperative registration process. A byproduct of this development is a mathematical model of realistic vessel intensities, which can be used directly in a distance measure or for simulations in the evaluation process.

Another starting point for prior knowledge incorporation is to restrict the admissible transformations by known additional properties, such as anatomically corresponding points. In clinical practice the interactive or semi-automatic localization of anatomical landmarks is always prone to errors. Therefore we show and develop ways how to deal with these uncertainties and use a more flexible and general landmark concept, which also contains landmark types at structures with no point-like shapes. The focus is on landmarks at tube-shaped vessel segments with anisotropic localization uncertainties. We systematically show how for different classes of transformations anisotropic landmarks can be integrated into pure landmark registration schemes. In particular we establish a higher flexibility and accuracy of non-rigid algorithms for the registration of pre- and postoperative images.

We also combine intensity and landmark information by incorporating landmark constraints into the non-parametric image registration framework. The landmarks are integrated as hard constraints ignoring localization uncertainties as well as constraints with anisotropic tolerances for the consideration of localization uncertainties. This is mathematically modeled as equality resp. inequality constrained optimization problems and solved by the according efficient algorithms. The effectiveness of both schemes is shown for the non-rigid registration of preoperative data to intraoperative 3D ultrasound images of the liver. From a methodological point of view the combined approach has a lot of potential for different clinical applications by offering the possibility to incorporate additional a priori knowledge provided by a human expert into a non-rigid registration process.

Fully automatic CT/MRI to 3D US registration as well as fully automatic pre- to postoperative liver CT registration is still an open problem and only few papers have been published concerning these issues. In order to progress towards a better understanding of the associated problems as well as to provide intermediate clinical solutions, additional manual interaction can be helpful to provide a priori knowledge, as we have shown in this thesis. This interaction may be considered a drawback, because it is time-consuming and

user-dependent. However, for the practical integration into the intraoperative navigation system the interaction will be reduced to a minimum through additional image processing and intuitive interaction techniques for the landmark definition.

As all available data sets (CT portal venous, CT late venous, B-Mode ultrasound, power Doppler ultrasound) provide complementary information, a consideration of all sets might further improve the registration process. We already integrated the portal and late venous phase in the shape distance measure by using extracted portal resp. hepatic veins from different phases. The integration of both ultrasound modes requires the development of multi-signal distance measures. Additional information can be incorporated intra-operatively into the registration process by a navigated instrument. With the tip of the instrument it is possible to acquire points and curves on the liver surface which can be used as landmarks in addition to vessel branching and segment landmarks. The extension of the proposed registration schemes to these plate-like and line-like features is possible.

Clinical validation of non-rigid registration is a challenging and application-specific problem in itself. Another important contribution of this thesis is the systematic review of methods for the determination of the resulting registration accuracy and of simulation techniques for uniformly resp. non-uniformly distributed rigid transformations for validation purposes. In addition a reference standard was introduced for computing the registration accuracy on clinical data containing vessel trees. The reference standard is based on dense corresponding point sets on the vessel center lines and proved its value for the validation of our registration methods for computer assisted liver surgery.

Overall we have developed innovative components for medical image registration and embedded them mathematically and algorithmically into an existing flexible framework for non-rigid registration.

Bibliography

- [1] A Andronache, M von Siebenthal, G Székely, and Ph Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Med Imag Anal*, 12(1):3–15, 2008.
- [2] K Arun, T Huang, and S Blostein. Least-squares fitting of two 3-D point sets. *IEEE T Pattern Anal*, 9(5):698–700, 1987.
- [3] J Arvo. *Graphics Gems III*, chapter Fast random rotation matrices, pages 117–120. Academic Press Professional, 1992.
- [4] J Ashburner, C Hutton, R Frackowiak, I Johnsrude, C Price, and K Friston. Identifying global anatomical differences: Deformation-based morphometry. *Hum Brain Mapp*, 6:348–357, 1998.
- [5] B Aubert-Broche, AC Evans, and L Collins. A new improved version of the realistic digital brain phantom. *Neuroimage*, 32:138–145, 2006.
- [6] SR Aylward, J Jomier, S Weeks, and E Bullitt. Registration and analysis of vascular images. *Int J Comput Vision*, 55(2-3):123–138, 2003.
- [7] R Balachandran, JM Fitzpatrick, and RF Labadie. Fiducial registration for tracking systems that employ coordinate reference frames. In *Proc. of Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display*, volume 5744, 2005.
- [8] PG Batchelor and JM Fitzpatrick. A study of the anisotropically weighted procrustes problem. In *Proc. of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2000.
- [9] M Baumhauer, T Simpfendorfer, BP Müller-Stich, D Teber, CN Gutt, J Rassweiler, HP Meinzer, and I Wolf. Soft tissue navigation for laparoscopic partial nephrectomy. *Int J CARS*, 3(3-4):307–314, 2008.
- [10] S Beller, M Hünerbein, S Eulenstein, T Lange, and PM Schlag. Feasibility of navigated resection of liver tumors using multiplanar visualization of intraoperative 3D ultrasound data. *Ann Surg*, 246(2):288–294, 2007.

- [11] S Beller, M Hünerbein, T Lange, S Eulenstein, B Gebauer, and PM Schlag. Image-guided surgery of liver metastases by 3D ultrasound - based optoelectronic navigation. *Brit J Surg*, 94(7):866–875, 2007.
- [12] S Beller, T Lange, E Eulenstein, L Beller, S Chopra, O Dudeck, and PM Schlag. 3D elaboration of postoperative CT data after liver resection: technique and utility. *Int J CARS*, 3(6):581–589, 2008.
- [13] C Bingham. An antipodally symmetric distribution on the sphere. *Ann Stat*, 2(6):1201–1225, 1974.
- [14] W Birkfellner. *Computerassistierte Chirurgie*, chapter Lokalisierungssysteme, pages 177–189. Elsevier, 2010.
- [15] M Birth, M Kleemann, P Hildebrand, and HP Bruch. Intraoperative online navigation of dissection of the hepatic tissue - a new dimension in liver surgery. In *Proc. of CARS*, pages 770–774, 2004.
- [16] J Bloomenthal. *Skeletal Design of Natural Forms*. PhD thesis, University of Calgary, 1995.
- [17] J Bloomenthal and K Shoemake. Convolution surfaces. *Comp Graph*, 25(4):251–256, 1991.
- [18] HG Bock, E Kostina, and JP Schlöder. Numerical methods for parameter estimation in nonlinear differential algebraic equations. In *Proc. of GAMM Mitt.*, volume 30, pages 376–408, 2007.
- [19] MM Bonsato, A Staubert, CR Wirtz, V Tronnier, and S Kunze. Initial experience with an ultrasound-integrated single-rack neuronavigation system. *Acta Neurochir*, 143(11):1127–1132, 2001.
- [20] FL Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE T Pattern Anal*, 11(6):567–585, 1989.
- [21] RM Brannon. A review of useful theorems involving proper orthogonal matrices referenced to three dimensional physical space. Technical Report NM 87185-0820, Computational Physics and Mechanics, Sandia National Laboratories Albuquerque, 2002.
- [22] C Broit. *Optimal registration of deformed images*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania, 1981.
- [23] R Brunelli. *Template Matching Techniques in Computer Vision*. John Wiley and Sons, 2009.
- [24] T Buzug. *Computed Tomography - From Photon Statistics to Modern Cone-Beam CT*. Springer, 2008.

-
- [25] P Cachier and X Pennec. 3D non-rigid registration by gradient descent on a gaussian-windowed similarity measure using convolutions. In *Proc. of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pages 182–189, 2000.
- [26] B Cady, G Steele, RL Jenkins, and et al. Surgical margin in hepatic resection for colorectal metastasis: a critical and improvable determinant of outcome. *Ann Surg*, 227:566–571, 1998.
- [27] DM Cash, MI Miga, SC Glasgow, BM Dawant, LW Clements, Z Cao, RL Galloway, and WC Chapman. Concepts and preliminary data toward the realization of image-guided liver surgery. *J Gastrointest Surg*, 11(7):844–859, 2007.
- [28] DM Cash, TK Sinha, WC Chapman, H Terawaki, BM Dawant, RL Galloway, and MI Miga. Incorporation of a laser range scanner into image-guided liver surgery: surface acquisition, registration, and tracking. *Med Phys*, 30(7):1671–1682, 2003.
- [29] A Charnoz, V Agnus, G Malandain, S Nicloulau, M Tajine, and L Soler. Design of robust vascular tree matching: Validation on liver. In *Proc. of IPMI 2005*, volume 3566 of *LNCS*, pages 443–455, 2005.
- [30] A Charnoz, V Agnus, G Malandain, L Soler, and M Tajine. Tree matching applied to vascular system. volume 3434 of *Graph-Based Representations in Pattern Recognition*, pages 183–192. Springer, 2005.
- [31] SS Chopra, J Rump, SC Schmidt, F Streitparth C Seebauer, G Schumacher, I Van der Voort, and U Teichgräber. Imaging sequences for intraoperative MR-guided laparoscopic liver resection in 1.0-T high field open MRI. *Eur Radiol*, 19:2191–2196, 2009.
- [32] GE Christensen. *Deformable shape models for anatomy*. Phd thesis, Sever Institute of Technology, Washington University, 1994.
- [33] GE Christensen, X Geng, JG Kuhl, J Bruss, TJ Grabowski, IA Pirwani, MW Vannier, JS Allen, and H Damasio. Introduction to the non-rigid image registration evaluation project (NIREP). In *Proc. of WBIR*, volume 4057 of *LNCS*, pages 128–135, 2006.
- [34] GE Christensen and HJ Johnson. Consistent image registration. *IEEE T Med Imaging*, 20(7):568–582, 2001.
- [35] MT Chu and NT Trendafilov. On a differential equation approach to weighted or orthogonal procrustes problem. *Stat Comput*, 8:125–133, 1998.

- [36] U Clarenz, M Droske, S Henn, M Rumpf, and K Witsch. Computational methods for nonlinear image registration. In *Proc. of Mathematical Methods for Registration and Applications to Medical Imaging, Mathematics in Industry*, volume 10, 2006.
- [37] A Collignon. *Multi-modality medical image registration by maximization of mutual information*. PhD thesis, Catholic University of Leuven, 1998.
- [38] DL Collins, AP Zijdenbos, V Kollokian, JG Sled, NJ Kabani, CJ Holmes, and AC Evans. Design and construction of a realistic digital brain phantom. *IEEE T Med Imaging*, 17(3):463–468, 1998.
- [39] TF Cootes, S Marsland, CJ Twining, K Smith, and CJ Taylor. Group-wise diffeomorphic non-rigid registration for automatic model building. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 316–327, 2004.
- [40] C Couinaud. *Le Foie: Études anatomiques et chirurgicales [The Liver: Anatomical and Surgical Studies]*. Masson, 1957.
- [41] WR Crum, O Camara, and DJ Hawkes. Methods for inverting dense displacement fields: Evaluation in brain image registration. In *Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4791 of *LNCS*, pages 900–907. Springer, 2007.
- [42] WR Crum, LD Griffin, DL Hill, and DJ Hawkes. Zen and the art of medical image registration: correspondence, homology, and quality. *Neuroimage*, 20(3):1425–1437, 2003.
- [43] A Danilchenko and JM Fitzpatrick. General approach to error prediction in point registration. In *Proc. of SPIE Medical Imaging*, volume 7625 of *SPIE*, 2010.
- [44] RS Dembo and T Steihaug. Truncated-newton algorithms for large-scale unconstrained optimization. *Math Program*, 26:190–212, 1983.
- [45] M Descoteaux, M Audette, K Chinzei, and K Siddiqi. Bone enhancement filtering: Application to sinus bone segmentation and simulation of pituitary surgery. *Comput Aided Surg*, 11(5):247–255, 2006.
- [46] L Devroye. *Non-Uniform Random Variate Generation*, chapter 2, General principles in random variate generation, pages 27–82. Springer-Verlag, 1986.
- [47] TD Downs. Orientation statistics. *Biometrika*, 59:665–676, 1972.
- [48] M Droske and M Rumpf. A variational approach to non-rigid morphological registration. *SIAM J Appl Math*, 64(2):668–687, 2004.

-
- [49] TJ Dubinsky, C Cuevas, MK Dighe, O Kolokythas, and JH Hwang. High-intensity focused ultrasound: Current potential and oncological applications. *Am J Roentgenol*, 190(1):191–199, 2008.
- [50] J Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Analyse Numérique*, 10(12):5–12, 1976.
- [51] G Eggers, J Mühling, and R Marmulla. Image-to-patient registration techniques in head surgery. *J Oral Maxil Surg*, 35:1081–1095, 2006.
- [52] RSJ Estépar, A Brun, and CF Westin. Robust generalized total least squares iterative closest point registration. In *Proc. of MICCAI*, volume 3216 of *LNCS*, pages 234–241, 2004.
- [53] Leonhard Euler. *Theoria motus corporum solidorum seu rigidorum*. Cornell University Library, 1765.
- [54] JH Fasel, D Selle, CJ Evertsz, F Terrier, HO Peitgen, and P Gailloud. Segmental anatomy of the liver: poor correlation with CT. *Radiology*, 206(1):151–156, 1998.
- [55] M Ferrant, A Nabavi, B Macq, PM Black, FA Jolesz, R Kikinis, and SK Warfield. Serial registration of intraoperative MR images of the brain. *Med Imag Anal*, 6(4):337–359, 2002.
- [56] B Fischer and J Modersitzki. Fast diffusion registration. In *Proc. of Inverse Problems, Image Analysis, and Medical Imaging*, volume 313, pages 117–129, 2002.
- [57] B Fischer and J Modersitzki. Combination of automatic non-rigid and landmark based registration: the best of both worlds. In *Proc. of SPIE Medical Imaging 2003: Image Processing*, volume 5032, pages 1037–1048, 2003.
- [58] B Fischer and J Modersitzki. Combining landmarks and intensity driven registrations. In *Proc. of GAMM*, volume 3, pages 32–35, 2003.
- [59] B Fischer and J Modersitzki. Curvature based image registration. *J Math Imaging Vis*, 18(1):81–85, 2003.
- [60] NI Fisher, T Lewis, and BJJ Embleton. *Statistical analysis of spherical data*. Cambridge University Press, 1993.
- [61] JM Fitzpatrick. Fiducial registration error and target registration error are uncorrelated. In *Proc. of SPIE, Vol. 7261, 726102*, 2009.
- [62] JM Fitzpatrick. The role of registration in accurate surgical guidance. *Proc Inst Mech Eng H*, 224(5):607–622, 2010.

- [63] JM Fitzpatrick, DL Hill, Y Shyr, J West, C Studholme, and CR Maurer. Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain. *IEEE T Med Imaging*, 17:571–585, 1998.
- [64] JM Fitzpatrick and JB West. The distribution of target registration error in rigid-body point-based registration. *IEEE T Med Imaging*, 20(9):917–927, 2001.
- [65] JM Fitzpatrick, JB West, and CR Maurer. Predicting error in rigid-body point-based registration. *IEEE T Med Imaging*, 17(5):694–702, 1998.
- [66] M Fitzpatrick. *Medical Image Registration*, chapter Detecting Failure – Assessing Success, pages 117–142. CRC Press, 2001.
- [67] Y Fong, J Fortner, RL Sun, and et al. Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. *Ann Surg*, 230:309–318, 1999.
- [68] AF Frangi, WJ Niessen, KL Vincken, and MA Viergever. Multiscale vessel enhancement filtering. In *Proc. of MICCAI*, pages 130–137, 1998.
- [69] BB Frericks, FC Caldarone, B Nashan, DH Savellano, G Stamm, TD Kirchoff, HO Shin, A Schenk, D Selle, W Spindler, J Klempnauer, HO Peitgen, and M Galanski. 3D CT modeling of hepatic vessel architecture and volume calculation in living donated liver transplantation. *Eur Radiol*, 14(2):326–333, 2004.
- [70] V Gibbs, D Cole, and A Sassano. *Ultrasound Physics and Technology: How, Why and When*. Churchill Livingstone, Elsevier Health Sciences, 2009.
- [71] PE Gill, W Murray, and MH Wright. *Practical optimization*. Academic Press, 1981.
- [72] DG Gobbi, RM Comeau, and TM Peters. Ultrasound/MRI overlay with image warping for neurosurgery. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 106–114, 2000.
- [73] BF Green. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17:429–440, 1952.
- [74] R Grimmer, J Krause, W Karolczak, R Lapp, and M Kachelrieß. Assessment of spatial resolution in CT. In *Proc. of IEEE Medical Imaging Conference, Dresden, Germany*, 2008.
- [75] A Gronningsaeter, A Kleven, S Ommedal, TE Aarseth, T Lie, F Lindseth, T Langø, and G Unsgård. Sonowand, an ultrasound-based neuronavigation system. *Neurosurgery*, 47(6):1373–1379, 2000.

-
- [76] E Haber, S Heldmann, and J Modersitzki. A computational framework for image-based constrained registration. *Linear Algebra Appl*, 431(3-4):459–470, 2009.
- [77] E Haber, S Heldmann, and J Modersitzki. A scale-space approach to landmark constrained image registration. In *Proc. of the Second Int Conf on Scale Space Methods and Variational Methods in Computer Vision (SSVM)*, LNCS, pages 1–12, 2009.
- [78] E Haber and J Modersitzki. Numerical methods for volume preserving image registration. *Inverse Probl*, 20:1621–1638, 2004.
- [79] E Haber and J Modersitzki. Image registration with guaranteed displacement regularity. *Int J Comput Vision*, 71:361–372, 2007.
- [80] E Haber and J Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Methods Inf Med*, 46(3):292–299, 2007.
- [81] J Harms, M Bartels, H Bourquain, HO Peitgen, T Schulz, T Kahn, J Hauss, and J Fangmann. Computerized CT-based 3D visualization technique in living related liver transplantation. *Transplant P*, 37(2):1059–1062, 2005.
- [82] R Harrison, S Vannoy, D Haynor, S Gillispie, M Kaplan, and T Lewellen. Preliminary experience with the photon history generator module of a public-domain simulation system for emission tomography. *IEEE Nucl Sci Symp Conf R*, 2:1154–1158, 1993.
- [83] WR Hedrick and CL Peterson. Image artifacts in real-time ultrasound. *J Diagn Med Sonog*, 11:300–308, 1995.
- [84] T Heimann, B van Ginneken, M Styner, Y Arzhaeva, V Aurich, C Bauer, A Beck, and et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE T Med Imaging*, 28(8):1251–1265, 2009.
- [85] S Henn. A Levenberg-Marquardt scheme for nonlinear image registration. *Bit*, 43(4):743–759, 2003.
- [86] S Henn, L Hömke, and K Witsch. Lesion preserving image registration with applications to human brains. In *Proc. of Pattern Recognition, 26th DAGM Symposium*, volume 3175 of LNCS, pages 143–154, 2004.
- [87] D Hogemann, G Stamm, H Shin, KJ Oldhafer, HJ Schlitt, D Selle, and HO Peitgen. Individual planning of liver surgery interventions with a virtual model of the liver and its associated structures. *Radiologe*, 40(3):267–273, 2000.

- [88] M Holden, LD Griffin, and DLG Hill. Multi-dimensional mutual information image similarity metrics based on derivatives of linear scale space. In *Proc. of the APRS Workshop on Digital Image Computing*, 2005.
- [89] Lars Hömke. *A multigrid method for elastic image registration with additional structural constraints*. PhD thesis, Heinrich-Heine Universität Düsseldorf, 2006.
- [90] BKP Horn. Closed-form solution of absolute orientation using unit quaternions. *J Opt Soc Am A*, 4:629–642, 1987.
- [91] BKP Horn and BG Schunck. Determining optical flow. *Artif Intell*, 17:185–203, 1981.
- [92] JR Hurley and RB Cattell. The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behav Sci*, 7:258–262, 1962.
- [93] JA Jensen. *Imaging of Complex Media with Acoustic and Seismic Waves*, chapter Ultrasound Imaging and its modelling, pages 135–166. Topics in Applied Physics. Springer Verlag, 2002.
- [94] S Kabus, A Franz, and B Fischer. Spatially varying elasticity in image registration. *Methods Inf Med*, 46:287–291, 2007.
- [95] D Kainmüller, T Lange, and H Lamecker. Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In *Proc. of MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, pages 109–116, 2007.
- [96] W Kalender. *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications*. MCD Verlag, 2000.
- [97] W Kalender. X-ray computed tomography. *Phys Med Biol*, 51:R29–R43, 2006.
- [98] W Kalender and A Polacin. Physical performance characteristics of spiral CT scanning. *Med Phys*, 18(5):910–915, 1991.
- [99] W Kalender, W Seissler, E Klotz, and P Vock. Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology*, 176:181–183, 1990.
- [100] AE Kerdok, SM Cotin, MP Ottensmeyer, AM Galea, RD Howe, and SL Dawson. Truth cube: Establishing physical standards for soft tissue simulation. *Med Imag Anal*, 7(3):283–291, 2003.

-
- [101] CG Khatri and KV Mardia. The von Mises-Fisher matrix distribution in orientation statistics. *J Roy Stat Soc B Met*, 39(1):95–106, 1977.
- [102] JJ Koenderink. The structure of images. *Biological Cybernetics*, 1984.
- [103] J Kohlrausch, K Rohr, and HS Stiehl. A new class of elastic body splines for nonrigid registration of medical images. *J Math Imaging Vis*, 23:253–280, 2005.
- [104] MA Koschat and DF Swayne. A weighed procrustes criterion. *Psychometrika*, 56(2):229–239, 1991.
- [105] O Kripfgans, J Rubin, A Hall, M Gordon, and B Fowlkes. Measurement of volumetric flow. *J Ultras Med*, 25(10):1305–1311, 2006.
- [106] K Krissian, G Malandain, N Ayache, R Vaillant, and Y Troussel. Model-based detection of tubular structures in 3D images. *Comput Vis Image Und*, 80(2):130–171, 2000.
- [107] K Krissian, CF Westin, R Kikines, and K Vosburgh. Oriented speckle reducing anisotropic diffusion. *IEEE T Image Process*, 16(5):1412–1424, 2007.
- [108] JJ Kuffner. Effective sampling and distance metrics for 3D rigid body path planning. In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA)*, volume 4, pages 3993–3998, 2004.
- [109] J Kybic and M Unser. Fast parametric elastic image registration. *IEEE T Med Imaging*, 12(11):1427–1441, 2003.
- [110] H Lamecker and T Heimann. *Computerassistierte Chirurgie*, chapter Bildsegmentierung, pages 63–81. Urban und Schwarz Verlag, Elsevier GmbH, 2010.
- [111] H Lang. Technik der Leberresektion - Teil I. Leberanatomie und Operationsplanung. *Chirurg*, 78:761–774, 2007.
- [112] H Lang. Technik der Leberresektion - Teil II. Operationsdurchführung. *Chirurg*, 78:849–866, 2007.
- [113] H Lang, A Radtke, M Hindennach, T Schroeder, NR Frühauf, M Malago, H Bourquain, HO Peitgen, KJ Oldhafer, and CE Broelsch. Impact of virtual tumor resection and computer-assisted risk analysis on operation planning and intraoperative strategy in major hepatic resection. *Arch Surg*, 140(7):629–638, 2005.

- [114] T Lange, S Eulenstein, M Hünnerbein, H Lamecker, and PM Schlag. Augmenting intraoperative 3D ultrasound with preoperative models for navigation in liver surgery. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3217 of *LNCS*, pages 534–541. Springer, 2004.
- [115] T Lange, S Eulenstein, M Hünnerbein, and PM Schlag. Vessel-based non-rigid registration of MR/CT and 3D ultrasound for navigation in liver surgery. *Comput Aided Surg*, 8(5):228–240, 2003.
- [116] T Lange, S Eulenstein, S Kraft, and PM Schlag. Electromagnetic anchor tracking for soft tissue navigation. In *Proc. of Jahrestagung der Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, pages 211–213. Der andere Verlag, 2010.
- [117] T Lange, M Hünnerbein, S Eulenstein, S Beller, and PM Schlag. Development of navigation systems for image-guided laparoscopic tumor resections in liver surgery. In *Minimally Invasive Tumor Therapies*, volume 167 of *Recent Results in Cancer Research*, chapter 2, pages 13–36. Springer, 2006.
- [118] T Lange, M Hünnerbein, S Eulenstein, S Beller, PM Schlag, and H Lamecker. A distance measure for non-rigid registration of geometrical models to intensity data. *Int J Comp Ass Rad Surg*, 2 (Suppl 1):S204–S206, 2007.
- [119] T Lange, H Lamecker, M Hünnerbein, S Eulenstein, S Beller, and PM Schlag. A new class of distance measures for registration of tubular models to image data. In *Proc. of Bildverarbeitung für die Medizin (BVM), Informatik aktuell*,, pages 101–105, 2007.
- [120] T Lange, H Lamecker, M Hünnerbein, S Eulenstein, S Beller, and PM Schlag. Validation metrics for non-rigid registration of medical images containing vessel trees. In *Proc. of Bildverarbeitung für die Medizin (BVM)*, pages 82–86, 2008.
- [121] T Lange, N Papenberg, S Heldmann, J Modersitzki, B Fischer, H Lamecker, and PM Schlag. 3D ultrasound-CT registration of the liver using combined landmark-intensity information. *Int J Comp Ass Rad Surg*, 4(1):79–88, 2009.
- [122] T Lange, N Papenberg, S Heldmann, J Modersitzki, B Fischer, and PM Schlag. 3D ultrasound-CT alignment of the liver using combined landmark-intensity registration. In *International Journal of Computer Assisted Radiology and Surgery*, 3 (Suppl 1), pages S107–S108, 2008.

-
- [123] T Lange, N Papenberg, J Olesch, B Fischer, and PM Schlag. Landmark constrained non-rigid image registration with anisotropic tolerances. In *IFMBE Proceedings 25/IV World Congress of Medical Physics and Biomedical Engineering*, page 2238 ff, 2009.
- [124] T Lange, T Wenckebach, H Lamecker, M Seebass, M Hünerbein, S Eulenstein, B Gebauer, and PM Schlag. Registration of different phases of contrast-enhanced CT/MRI data for computer-assisted liver surgery planning: Evaluation of state-of-the-art methods. *Int J Med Robot Comp*, 1(3):6–20, 2005.
- [125] T Lange, T Wenckebach, H Lamecker, M Seebass, M Hünerbein, S Eulenstein, and PM Schlag. Registration of portal and hepatic venous phase of MR/CT data for computer-assisted liver surgery planing. In *Proc. of Computer Assisted Radiology and Surgery (CARS)*, volume 1281 of *Int. Congress Series*, pages 768–772. Elsevier, 2005.
- [126] T Lange, S Wörz, K Rohr, and PM Schlag. Experimental comparison of landmark-based methods for 3D elastic registration of pre- and postoperative liver CT data. In *Proc. of SPIE Medical Imaging: Visualization, Image-guided Procedures and Modeling*, 2009.
- [127] T Lange, S Wörz, K Rohr, and PM Schlag. Landmark-based 3D elastic registration of pre- and post-operative liver CT data - an experimental comparison. In *Proc. of Bildverarbeitung für die Medizin (BVM)*, pages 107–111, 2009.
- [128] K Lehmann, JP Ritz, S Valdeig, A Schenk, Ch Holmer, HO Peitgen, HJ Buhr, and BB Frericks. Portal vein segmentation of a 3D-planning system for liver surgery - in vivo evaluation in a porcine model. *Ann Surg Oncol*, 15(7):1899–1907, 2008.
- [129] AJ Lemke, MJ Brinkmann, T Schott, SM Niehues, U Settmacher, P Neuhaus, and R Felix. Living donor right liver lobes: preoperative CT volumetric measurement for calculation of intraoperative weight and volume. *Radiology*, 240(3):736–742, 2006.
- [130] A Leroy, P Mozer, Y Payan, and J Troccaz. Rigid registration of free-hand 3D ultrasound and CT-scan kidney images. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3216 of *LNCS*, pages 837–844. Springer, 2004.
- [131] D Lesage, ED Angelini, I Bloch, and G Funka-Lea. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Med Imag Anal*, 13(6):819–845, 2009.

- [132] T Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *J Appl Stat*, 21(2):224–270, 1994.
- [133] T Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Dordrecht, 1994.
- [134] D Lindner and J Meixensberger. *Computerassistierte Chirurgie*, chapter Intraoperativer Ultraschall in der Neurochirurgie, pages 387–395. Elsevier, 2010.
- [135] F Lindseth, T Langø, J Bang, and T Hernes. Accuracy evaluation of a 3D ultrasound-based neuronavigation system. *Comput Aided Surg*, 7:197–222, 2002.
- [136] D Loeckx, F Maes, D Vandermeulen, and P Suetens. Nonrigid image registration using free-form deformations with local rigidity constraint. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3216 of *LNCS*, pages 639–646. Springer, 2004.
- [137] WE Lorensen and HE Cline. Marching cubes: A high resolution 3D surface construction algorithm. volume 21 of *SIGGRAPH*, pages 163–169, 1987.
- [138] PC Mahalanobis. On the generalized distance in statistics. In *Proc. of the National Institute of Sciences of India*, volume 2, pages 49–55, 1936.
- [139] L Maier-Hein, A Tekbas, A Seitel, F Pianka, SA Müller, S Satz, S Schawo, B Redeleff, R Tetzlaff, AM Franz, BP Müller-Stich, I Wolf, HU Kauczor, BM Schmied, and HP Meinzer. In vivo accuracy assessment of a needle-based navigation system for CT-guided radiofrequency ablation of the liver. *Med Phys*, 35(12):5385–96, 2008.
- [140] JBA Maintz, PA van den Elsen, and MA Viergever. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Med Imag Anal*, 1(2):151–161, 1996.
- [141] G Marsaglia. Coosing a point from the surface of a sphere. *Ann Math Stat*, 43(2):645–646, 1972.
- [142] CR Maurer, JM Fitzpatrick, MY Wang, RL Galloway, and RJ Maciunas. Registration of head volume images using implantable fiducial markers. *IEEE T Med Imaging*, 16:447–462, 1997.
- [143] E Meijering, WJ Niessen, and MA Viergever. Quantitative evaluation of convolution-based methods for medical image interpolation. *Med Imag Anal*, 5:111–126, 2001.

-
- [144] HP Meinzer, M Thorn, and C Cardenas. Computerized planning of liver surgery - an overview. *Comput Graph*, 26(4):569–576, 2002.
- [145] L Mercier, T Langø, F Lindseth, and L Collins. A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound Med Biol*, 31(4):449–471, 2005.
- [146] RE Miles. On random rotations in R^3 . *Biometrika*, 52(3/4):636–639, 1965.
- [147] J Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, 2004.
- [148] J Modersitzki. Image registration with local rigidity constraints. In *Proc. of Bildverarbeitung für die Medizin*, pages 444–448, 2007.
- [149] J Modersitzki. FLIRT with rigidity - image registration with a local non-rigidity penalty. *Int J Comput Vis*, 76(2):153–163, 2008.
- [150] J Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. SIAM, 2009.
- [151] JH Moltz, L Bornemann, JM Kuhnigk, V Dicken, E Peitgen, S Meier, H Bolte, M Fabel, HC Bauknecht, M Hittinger, AK Kiessling, M Pusken, and HO Peitgen. Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans. *IEEE J Sel Topics Signal Proc*, 3(1):122–134, 2009.
- [152] J Montagnat, M Sermesant, H Delingette, G Malandain, and N Ayache. Anisotropic filtering for model-based segmentation of 4D cylindrical echocardiographic images. *Pattern Recogn Lett*, 24(4–5):815–828, 2003.
- [153] R Mårvik, T Langø, GA Tangen, JO Andersen, JH Kaspersen, B Ystgaard, E Sjølie, R Fougner, HE Fjøsne, and TA Nagelhus Hernes. Laparoscopic navigation pointer for three-dimensional image-guided surgery. *Surg Endosc*, 18(8):1242–1248, 2004.
- [154] D Mucha, B Kosmecki, and T Krüger. *Computerassistierte Chirurgie*, chapter Das Konzept der Navigation, pages 163–175. Elsevier, 2010.
- [155] A Nabavi, PM Black, DT Gering, CF Westin, V Mehta, RS Perolizzi, M Ferrant, SK Warfield, N Hata, RB Schwartz, WM Wells, R Kikinis, and FA Jolesy. Serial intraoperative MR imaging of brain shift. *Neurosurgery*, 48:787–798, 2001.
- [156] SG Nash. A survey of truncated-newton methods. *J Comput Appl Math*, 124:45–59, 2000.

- [157] TR Nelson, DH Pretorius, A Hull, M Riccabona, and MS Sklansky. Sources and impact of artifacts on clinical three-dimensional ultrasound imaging. *Ultrasound Obstet Gynecol*, 16:374–383, 2000.
- [158] SA Nicolau, A Garcia, X Pennec, L Soler, and N Ayache. An augmented reality system to guide radio-frequency tumour ablation. *Comput Animat Virt W*, 16(1):1–10, 2005.
- [159] A Nilsson. Artefacts in sonography and Doppler. *Eur Radiol*, 11:1308–1315, 2001.
- [160] C Nimsky, O Ganslandt, and P Hastreiter. Intraoperative compensation for brain shift. *Surg Neurol*, 43(4):357–364, 2001.
- [161] C Nimsky, O Ganslandt, B Von Keller, J Romstock, and R Fahlbusch. Intraoperative high-field-strength MR imaging: implementation and experience in 200 patients. *Radiology*, 233:67–78, 2004.
- [162] C Nimsky, O Ganslandt, H Kober, M Buchfelder, and R Fahlbusch. Intraoperative magnetic resonance imaging combined with neuronavigation: a new concept. *Neurosurgery*, 48:1082–1091, 2001.
- [163] J Nocedal and SJ Wright. *Numerical Optimization*. Springer, 1999.
- [164] S Oeltze and B Preim. Visualization of vascular structures: Method, validation and evaluation. *IEEE T Med Imaging*, 24(4):540–548, 2005.
- [165] N Ohta and K Kanatani. Optimal estimation of three-dimensional rotation and reliability evaluation. In *Proc. of European Conference on Computer Vision (ECCV)*, volume 1406 of *LNCS*, pages 175–187. Springer, 1998.
- [166] KJ Oldhafer, D Hogemann, G Stamm, R Raab, HO Peitgen, and M Galanski. 3-dimensional (3-D) visualization of the liver for planning extensive liver resections. *Chirurg*, 70(3):233–238, 1999.
- [167] KJ Oldhafer, GA Stavrou, G Prause, HO Peitgen, TC Lueth, and S Weber. How to operate a liver tumor you cannot see. *Langenbecks Arch Surg*, 394(3):489–494, 2009.
- [168] J Olesch, N Papenberg, T Lange, M Conrad, and B Fischer. Matching CT and ultrasound data of the liver by landmark constrained image registration. In *Proc. of SPIE Medical Imaging: Visualization, Image-guided Procedures and Modeling*, volume 7261, pages online pre-publication, 2009.
- [169] Janine Olesch. Parameterschätzung in Differentialgleichungen. Master’s thesis, Institut für Mathematik, Universität zu Lübeck, 2008.

-
- [170] N Papenberg. *Ein genereller Registrierungsansatz mit Anwendung in der navigierten Leberchirurgie*. PhD thesis, Institute of Mathematics, University of Lübeck, 2008.
- [171] N Papenberg, T Lange, J Modersitzki, PM Schlag, and B Fischer. Image registration for CT and intra-operative ultrasound data of the liver. In *Proc. of SPIE Medical Imaging: Visualization, Image-guided Procedures, and Modeling*, volume 6918, page 691808, 2008.
- [172] N Papenberg, J Olesch, T Lange, PM Schlag, and B Fischer. Landmark constrained non-parametric image registration with isotropic tolerances. In *Bildverarbeitung für die Medizin (BVM)*, pages 122–126, 2009.
- [173] TM Pawlik, CR Scoggins, D Zorzo, and et al. Effect of surgical margin status on survival and site of recurrence after hepatic resection for colorectal metastases. *Ann Surg*, 241:715–722, 2005.
- [174] X Pennec, P Cachier, and N Ayache. Tracking brain deformations in time sequences of 3D US images. *Pattern Recogn Lett*, 24(4-5):801–813, 2003.
- [175] GP Penney, JM Blackall, MS Hamady, T Sabharwal, A Adam, and DJ Hawkes. Registration of freehand 3D ultrasound and magnetic resonance liver images. *Med Imag Anal*, 8(1):81–91, 2004.
- [176] M Peterhans, A vom Berg, B Dagon, D Inderbitzin, C Baur, D Candinas, and S Weber. A navigation system for open liver surgery: design, workflow and first clinical applications. *Int J Med Robot Comp*, 7:7–16, 2011.
- [177] JP Pluim, JB Maintz, and MA Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE T Med Imaging*, 19(8):809–814, 2000.
- [178] JP Pluim, JB Maintz, and MA Viergever. Mutual-information-based registration of medical images: a survey. *IEEE T Med Imaging*, 22(8):986–1004, 2003.
- [179] BC Porter, DJ Rubens, JG Strang, J Smith, S Totterman, and KJ Parker. Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. *IEEE T Med Imaging*, 20(4):354–359, 2001.
- [180] RW Prager, AH Gee, GM Treece, and L Berman. Sensorless freehand 3D ultrasound using regression of the echo intensity. *Ultrasound in Medicine and Biology*, 29(3):437–446, 2003.

- [181] RW Prager, RN Rohling, AH Gee, and L Berman. Rapid calibration for 3-D freehand ultrasound. *Ultrasound Med Biol*, 24(6):855–869, 1998.
- [182] B Preim, H Bourquian, D Selle, HO Peitgen, and KJ Oldhafer. Resection proposals for oncologic liver surgery based on vascular territories. In *Proc. of Computer Assisted Radiology and Surgery (CARS)*, pages 353–358. Springer, 2002.
- [183] MJ Prentice. Orientation statistics without parametric assumptions. *J Roy Stat Soc B Met*, 48(2):214–222, 1986.
- [184] A Reilhac, C Lartizien, N Costes, S Sans ans C Comtat, R Gunn, and A Evans. PET-SORTEO: a Monte Carlo-based simulator with high count rate capabilities. *IEEE T Nucl Sci*, 51(11):46–52, 2004.
- [185] I Reinertsen, M Descoteaux, K Siddiqi, and DL Collins. Validation of vessel-based registration for correction of brain shift. *Med Imag Anal*, 11(4):374–388, 2007.
- [186] I Reinertsen, F Lindseth, G Unsgård, and DL Collins. Clinical validation of vessel-based registration for correction of brain-shift. *Med Imag Anal*, 11(6):673–684, 2007.
- [187] DE Robinson, LS Wilson, and G Kossoff. Shadowing and enhancement in ultrasonic echograms by reflection and refraction. *J Clin Ultrasound*, 9:181–188, 1981.
- [188] A Roche, G Malandain, and N Ayache. Unifying maximum likelihood approaches in medical image registration. *Int J Imag Syst Tech*, 11(1):71–80, 2000.
- [189] A Roche, G Malandain, X Pennec, and N Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 1496, pages 1115–1124, 1998.
- [190] A Roche, X Pennec, G Malandain, and N Ayache. Rigid registration of 3-D ultrasound with MR images: A new approach combining intensity and gradient information. *IEEE T Med Imaging*, 20(10):1038–1049, 2001.
- [191] T Rohlfing. *Multimodale Datenfusion für die bildgesteuerte Neurochirurgie und Strahlentherapie*. PhD thesis, Technische Universität Berlin, Fachbereich Informatik, 2000.
- [192] T Rohlfing, CR Maurer, DA Bluemke, and MA Jacobs. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE T Med Imaging*, 22(6):730–741, 2003.

-
- [193] RJ Rohling. *3D Freehand Ultrasound: Reconstruction and Spatial Compounding*. PhD thesis, University of Cambridge, Department of Engineering, 1998.
- [194] K Rohr, M Fornefett, and H Stiehl. Spline-based elastic image registration: integration of landmark errors and orientation attributes. *Comput Vis Image Und*, 90:153–168, 2003.
- [195] K Rohr, H Stiehl, R Sprengel, T Buzug, J Weese, and M Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE T Med Imaging*, 20(6):526–534, 2001.
- [196] Karl Rohr. *Landmark-Based Image Analysis*. Springer, 2001.
- [197] DJ Rubens, S Bhatt, S Nedelka, and J Cullinan. Doppler artifacts and pitfalls. *Ultrasound Clinics*, 1:79–109, 2006.
- [198] D Rueckert, P Aljabar, R Heckemann, J Hajnal, and A Hammers. Diffeomorphic registration using b-splines. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4191 of *LNCS*, pages 702–709. Springer-Verlag, 2006.
- [199] D Rueckert, LI Sonoda, C Hayes, DL Hill, MO Leach, and DJ Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE T Med Imaging*, 18(8):712–721, 1999.
- [200] D Sarrut, V Boldea, S Miguet, and C Ginestet. Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. *Med Phys*, 33(3):605–617, 2006.
- [201] M Sato, I Bitter, M Bende, A Kaufmann, and M Nakajima. TEASAR: Tree-structure extraction algorithm for accurate and robust skeletons. In *Proc. of the 8th Pacific Graphics Conference on Computer Graphics and Application (PACIFIC GRAPHICS-00)*, pages 281–289, 2000.
- [202] M Schaap, CT Metz, T van Walsum, AG van der Giessen, AC Weustink, NR Mollet, and et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Med Imag Anal*, 13(5):701–714, 2009.
- [203] A Schenk, G Prause, and HO Peitgen. Efficient semiautomatic segmentation of 3D objects. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 1935 of *LNCS*, pages 186–195. Springer-Verlag, 2000.
- [204] PM Schlag, T Benhidjeb, and C Stroszczinski. Resection and local therapy for liver metastases. *Best Pract Res Clin Gastroenterol*, 16(2):299–317, 2002.

- [205] PM Schlag and U Stein, editors. *Regional Cancer Therapy*. Humana Press, 2007.
- [206] JA Schnabel, C Tanner, AD Castellano-Smith, A Degenhard, MO Leach, DR Hose, DL Hill, and DJ Hawkes. Validation of nonrigid image registration using finite-element methods: application to breast MR images. *IEEE T Med Imaging*, 22(2):238–247, 2003.
- [207] IJ Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. *Quart Appl Math*, 4:45–99, 1946.
- [208] PH Schönemann. *A solution of the orthogonal Procrustes problem with applications to orthogonal and oblique rotation*. PhD thesis, University of Illinois, 1964.
- [209] PH Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.
- [210] PH Schönemann and RM Carroll. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35(2):245–255, 1970.
- [211] H Schumacher, B Fischer, and A Franz. Weighted non-rigid image registration. Technical Report, Institute of Mathematics, University of Lübeck, 2005.
- [212] H Schumacher, A Franz, and B Fischer. Weighted medical image registration with automatic mask generation. In *Proc. of SPIE Medical Imaging 2006: Image Processing*, volume 6144, pages 740–747, 2006.
- [213] A Seginer. Rigid-body point-based registration: The distribution of the target registration error when the fiducial registration errors are given. *Medical Image Analysis*, 15:397–413, 2011.
- [214] D Selle. *Analysis of vessel structures in medical volume data for the computer-aided planning of surgical interventions*. PhD thesis, University of Bremen, 2000.
- [215] D Selle, B Preim, A Schenk, and HO Peitgen. Analysis of vasculature for liver surgical planning. *IEEE T Med Imaging*, 21(11):1344–1357, 2002.
- [216] K Shoemake. *Graphics Gems III*, chapter Uniform Random Rotations, pages 117–120. Academic Press, 1992.
- [217] Ken Shoemake. *Graphics Gems IV*, chapter Euler angle conversion, pages 222–229. Morgan Kaufmann, 1994.

-
- [218] D Skerl, B Likar, and F Pernus. A protocol for evaluation of similarity measures for rigid registration. *IEEE T Med Imaging*, 25(6):779–791, 2006.
- [219] D Skerl, B Likar, and F Pernus. A protocol for evaluation of similarity measures for non-rigid registration. *Med Imag Anal*, 12:42–54, 2008.
- [220] PJ Slomka, J Mandel, D Downey, and A Fenster. Evaluation of voxel-based registration of 3-D power Doppler ultrasound and 3-D magnetic resonance angiographic images of carotid arteries. *Ultrasound Med Biol*, 27(7):945–955, 2001.
- [221] L Soler, H Delingette, G Malandain, J Montagnat, N Ayache, C Koehl, O Dourthe, B Malassagne, M Smith, D Mutter, and J Marescaux. Fully automatic anatomical, pathological and functional segmentation from CT scans for hepatic surgery. *Comput Aided Surg*, 21(11):1344–1357, 2001.
- [222] M Staring, S Klein, and JPW Pluim. Nonrigid registration using a rigidity constraint. In *Proc. of SPIE Medical Imaging 2006: Image Processing*, volume 6144, pages 355–364, 2006.
- [223] C Stroszczynski, editor. *Minimally Invasive Tumor Therapies*, volume 167 of *Recent Results in Cancer Research*. Springer-Verlag, 2006.
- [224] JS Suri, K Setarehdan, and SS Singh, editors. *Advanced Algorithmic Approaches to Medical Image Segmentation*. Advances in Pattern Recognition. Springer-Verlag, 2002.
- [225] G ter Haar and C Coussios. High intensity focused ultrasound: Physical principles and devices. *Int J Hyperther*, 23(2):89–104, 2007.
- [226] F Terrier, M Grossholz, and CD Becker, editors. *Spiral CT of the Abdomen*. Springer Verlag, 1999.
- [227] ST Torp-Pedersen and L Terslev. Settings and artefacts relevant in colour/power Doppler ultrasound in rheumatology. *Ann Rheum Dis*, 67:143–149, 2008.
- [228] C Trantakis, J Meixensberger, D Lindner, G Strauß, G Grunst, A Schmidtgen, and S Arnold. Iterative neuronavigation using 3D ultrasound. A feasibility study. *Neurol Res*, 24:666–670, 2002.
- [229] TA Tuthill, JF Krücker, JB Fowlkes, and PL Carson. Automated three-dimensional US frame positioning computed from elevational speckle decorrelation. *Radiology*, 209(2):575–282, 1998.

- [230] G Ulrich. Computer generation of distributions on the m-sphere. *J Roy Stat Soc C Appl*, 33(2):158–163, 1984.
- [231] S Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE T Pattern Anal*, 13(4):376–380, 1991.
- [232] Robert-Koch-Institut und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., editor. *Krebs in Deutschland 2003-2004. Häufigkeiten und Trends*. Berlin, 2008.
- [233] M Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Proc Mag*, 16(6):22–38, 1999.
- [234] M Unser, A Aldroubi, and M Eden. On the asymptotic convergence of B-spline wavelets to gabor functions. *IEEE T Inform Theory*, 38(2):864–872, 1992.
- [235] G Unsgård, S Ommedal, T Muller, A Gronningsaeter, and TA Nagelhus Hernes. Neuronavigation by intraoperative three-dimensional ultrasound: initial experience during brain tumor resection. *Neurosurgery*, 50(4):804–812, 2002.
- [236] P van den Elsen, EJ Pol, TS Sumanaweera, PF Hemler, S Napel, and JR Adler. Grey value correlation techniques used for automatic matching of CT and MR brain and spine images. In *Proc. of SPIE Visualization in Biomedical Computing*, volume 2359, 1994.
- [237] J Vandemeulebroucke, D Sarrut, and P Clarysse. The POPI-model, a point-validated pixel-based breathing thorax model. In *Proc. of the XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, 2007.
- [238] T Viklands. *Algorithms for the weighted orthogonal procrustes problem and other least squares problems*. PhD thesis, Umeå University, Sweden, 2006.
- [239] P Viola. *Alignment by maximization of mutual information*. PhD thesis, Massachusetts Inst. Technol., Boston, MA, 1995.
- [240] PA Viola and WM Wells. Alignment by maximization of mutual information. In *Proc. of ICCV*, Fifth International Conference on Computer Vision, pages 16–23. IEEE, 1995.
- [241] TJ Vogl, R Straub, S Zangos, MG Mack, and K Eichler. MR-guided laser-induced thermotherapy (LITT) of liver tumours: experimental and clinical data. *Int J Hyperther*, 20(7):713–724, 2004.

-
- [242] RF Wagner, SW Smith, and JM Sandrik. Statistics of speckle in ultrasound B-scans. *IEEE T Son Ultrason*, 30(3):156–163, 1983.
- [243] G Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.
- [244] G Wang and M Vannier. Optimal pitch in spiral computed tomography. *Med Phys*, 24(10):1635–1639, 1997.
- [245] Y Wang. Smoothing spline models with correlated random errors. *J Am Stat Assoc*, 93(441):341–348, 1998.
- [246] Y-P Wang and S Lee. Scale-space derived from B-splines. *IEEE T Pattern Anal*, 20(10):1040–1055, 1998.
- [247] J Weese, P Rösch, T Netsch, T Blaffert, and M Quist. Gray-value based registration of CT and MR images by maximization of local correlation. In *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, pages 656–663. Springer-Verlag, 1999.
- [248] W Wein. *Multimodel Integration of Medical Ultrasound for Treatment Planning and Interventions*. PhD thesis, Technische Universität München, 2007.
- [249] W Wein, S Brunke, A Khamene, MR Callstrom, and N Navab. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. *Med Imag Anal*, 12:577–585, 2008.
- [250] W Wein, O Kutter, A Aichert, D Zikic, A Kamen, and N Navab. Automatic non-linear mapping of pre-procedure CT volumes to 3D ultrasound. In *Proc. of IEEE Int Symp on Biomedical Imaging*, 2010.
- [251] WM Wells, P Viola, H Atsumi, S Nakajima, and R Kikinis. Multi-modal volume registration by maximization of mutual information. *Med Image Anal*, 1(1):35–51, 1996.
- [252] J West, JM Fitzpatrick, MY Wang, BM Dawant, CR Maurer, RM Kessler, RJ Maciunas, and et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr*, 21(4):554–566, 1997.
- [253] JB West, JM Fitzpatrick, SA Toms, CR Maurer, and RJ Maciunas. Fiducial point placement and the accuracy of point-based, rigid body registration. *Neurosurgery*, 48(4):810–817, 2001.
- [254] JB West and CR Maurer. Designing optically tracked instruments for image-guided surgery. *IEEE T Med Imaging*, 23(5):533–545, 2004.

- [255] AP Witkin. Scale-space filtering. In *Proc. of the 8th Int. Joint Conf Art Intell*, pages 1019–1022, 1983.
- [256] JCH Wong, C Studholme, DJ Hawkes, and MN Maisey. Evaluation of the limits of visual detection of image misregistration in a brain fluorine-18 fluorodeoxyglucose PET-MRI study. *Eur J of Nucl Med*, 24:642–650, 1997.
- [257] ATA Wood. Simulation of the von Mises-Fisher distribution. *Commun Stat Simul Comput*, 23(1):157–164, 1994.
- [258] RP Woods. *Handbook of medical imaging: processing and analysis*, chapter Validation of Registration Accuracy, pages 491–497. Academic Press, 2000.
- [259] S Wörz and K Rohr. Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models. *Med Imag Anal*, 10:41–58, 2006.
- [260] S Wörz and K Rohr. Physics-based elastic image registration using splines and including landmark localization uncertainties. In *Proc. of MICCAI (2)*, LNCS, pages 678–685, 2006.
- [261] S Wörz and K Rohr. Segmentation and quantification of human vessels using a 3-D cylindrical intensity model. *IEEE T Image Process*, 16(8):1194–2004, 2007.
- [262] S Wörz and K Rohr. Spline-based elastic image registration using solutions of the Navier equation. In *Proc. of SPIE Medical Imaging: Image Processing*, 2007.
- [263] S Wörz and K Rohr. Hybrid physics-based elastic image registration using approximating splines. In *Proc. of SPIE Medical Imaging 2008: Image Processing*, 2008.
- [264] S Wörz and K Rohr. Physics-based elastic registration using non-radial basis functions and including landmark localization uncertainties. *Comput Vis Image Und*, 111(3):263–274, 2008.
- [265] CJ Wray, AM Lowy, JB Mathews, and et al. The significance and clinical factors associated with a subcentimeter resection of colorectal liver metastases. *Ann Surg Oncol*, 12:374–380, 2005.
- [266] Y Yu and ST Acton. Speckle reducing anisotropic diffusion. *IEEE T Image Process*, 11(11):1260–1270, 2002.
- [267] S Zachow, H Hahn, and T Lange. *Computerassistierte Chirurgie*, chapter Computerassistierte Chirurgieplanung, pages 119–149. Urban und Schwarz Verlag, Elsevier GmbH, 2010.

- [268] F Zhang, ER Hanock, C Goodlett, and G Gerig. Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling. *Med Imag Anal*, 13:5–18, 2009.
- [269] H Zhang, F Banovac, R Lin, N Glossop, BJ Wood, D Lindisch, E Levy, and K Cleary. Electromagnetic tracking for abdominal interventions in computer aided surgery. *Comput Aided Surg*, 11(3):127–136, 2006.