

Deformable Lung Registration for Pulmonary Image Analysis of MRI and CT scans



Mattias Paul Heinrich
St Hilda's College

Supervisors:
Julia Schnabel, Mark Jenkinson and Sir Michael Brady

A thesis submitted for the degree of
Doctor of Philosophy in Engineering

Hilary Term 2013

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Abstract

Medical imaging has seen a rapid development in its clinical use in assessment of treatment outcome, disease monitoring and diagnosis over the last few decades. Yet, the vast amount of available image data limits the practical use of this potentially very valuable source of information for radiologists and physicians. Therefore, the design of computer-aided medical image analysis is of great importance to imaging in clinical practice. This thesis deals with the problem of deformable image registration in the context of lung imaging, and addresses three of the major challenges involved in this challenging application, namely: designing an image similarity for multi-modal scans or scans of locally changing contrast, modelling of complex lung motion, which includes sliding motion, and approximately globally optimal mathematical optimisation to deal with large motion of small anatomical features. The two most important contributions made in this thesis are: the formulation of a multi-dimensional structural image representation, which is independent of modality, robust to intensity distortions and very discriminative for different image features, and a discrete optimisation framework, based on an image-adaptive graph structure, which enables a very efficient optimisation of large dense displacement spaces and deals well with sliding motion. The derived methods are applied to two different clinical applications in pulmonary image analysis: motion correction for breathing-cycle computed tomography (CT) volumes, and deformable multi-modal fusion of CT and magnetic resonance imaging chest scans. The experimental validation demonstrates improved registration accuracy, a high quality of the estimated deformations, and much lower computational complexity, all compared to several state-of-the-art deformable registration techniques.

Acknowledgements

I would like to thank Julia Schnabel for her great guidance and advice for my work in this thesis. Her support has been essential in many ways: encouraging me to focus on publicising my work – as well as giving me freedom to explore my (partly haphazard) ideas. I am grateful for her efforts to get me to work with interesting people: our clinical collaborators and all the great people she gathered in our group. She has been extremely helpful in formulating my written work in a more comprehensive way. Finally, I thank her for being reassuring after occasional failures and never advising against taking enough holidays.

I thank, Mike Brady and Mark Jenkinson for their excellent co-supervision. Mark for being very thorough in going through the mathematical details and thereby helping me to gain a better understanding. Mike for sharing his immense knowledge and frequently pointing me to related concepts in the literature. I would like to thank Andrew Zisserman for being supportive throughout my D.Phil., and him and Nikos Paragios for a challenging, but enjoyable and fair viva.

I was very lucky to be able to work with really great people at our lab. Most importantly: Ivor and Manav, with whom I not only had a great time travelling, but also many extremely useful and some more comical discussions. I thank Amalia, Richard and Bartek for many interesting conversations and Julien, Monica, Ana, Jieqing, Sana, Ben, etc. for being cheerful and making the lab an enjoyable place.

I thank my parents and Johannes for their everlasting support and love, and my friends back home for not forgetting about me. But above all, I thank Julia for everything: for coming to Oxford with me so that this place feels like home, making life fun every day and assuring me what really is important.

Contents

1	Introduction	14
2	Motivation and challenges of deformable lung registration	20
2.1	Medical image acquisition	22
2.1.1	X-ray computed tomography	22
2.1.2	Magnetic resonance imaging	24
2.2	Medical image registration	26
2.3	Clinical applications	28
2.3.1	Deformable multi-modal registration for fusion	28
2.3.2	Respiratory motion estimation	29
2.4	Challenges of lung registration	29
2.4.1	Large motion of small features	30
2.4.2	Sliding motion at lung surfaces	30
2.4.3	Intensity variation due to lung compression	31
2.4.4	Multi-modal image similarity for deformable registration . .	32
3	Validation of deformable image registration	35
3.1	Description of imaging data	37

3.2	Measures for evaluation of registration	38
3.2.1	Surrogate measures for accuracy based on image intensities	39
3.2.2	Surrogate metrics for quality of deformations	39
3.2.3	Clinically relevant image-derived measures	42
3.2.4	Higher-level clinical measures	45
3.3	Evaluation using landmark localisation	45
3.3.1	Similarity-based landmark localisation	48
3.3.2	Evaluation of robustness against intensity distortions	50
3.4	Summary	52
4	Modelling of complex lung motion	54
4.1	Variational regularisation	57
4.1.1	Linear or homogenous diffusion	57
4.1.2	Robust norms	58
4.1.3	Direction-dependent diffusion regularisation	59
4.1.4	Non-local regularisation	61
4.2	Parameterisation of deformations	63
4.3	Image-adaptive regularisation using trees	65
4.4	Symmetric and diffeomorphic transforms	68
4.4.1	Novel approach to symmetric and inverse-consistent transformations	69
4.5	Summary	71
5	Spatial context for statistical similarity metrics	73
5.1	Introduction to statistical similarity metrics	78
5.2	Mutual information	79

5.2.1	Pointwise normalised mutual information	80
5.2.2	Hierarchical mutual information	83
5.2.3	Conditional mutual information	84
5.3	Textural mutual information	85
5.3.1	Textons	86
5.3.2	Cluster trees	88
5.4	Experiments and results	92
5.4.1	Parameter sensitivity	93
5.4.2	Robustness against intensity distortions	96
5.5	Summary	98
6	Multi-dimensional structural image representation	100
6.1	Overview of structural image representations	102
6.2	Entropy images	106
6.3	Structure tensor gradient orientation	107
6.3.1	Gradient orientation	108
6.3.2	Orientation based on structure tensors	110
6.4	Modality independent neighbourhood descriptor (MIND)	113
6.4.1	Derivation of MIND representations	116
6.4.2	Extension to the self-similarity context	120
6.4.3	Multi-modal similarity metric using MIND	121
6.5	Experiments and results	123
6.5.1	Parameter sensitivity	124
6.5.2	Robustness against intensity distortions	126

6.6	Discussion of multi-modal similarity metrics and representations	130
6.6.1	Outlook: Metric learning	132
7	Towards globally optimal energy minimisation	136
7.1	Optimisation for lung registration	138
7.2	Continuous optimisation for registration	142
7.2.1	Demons framework and diffemorphism	143
7.3	Gauss-Newton optimisation	145
7.3.1	Diffusion-regularised deformable registration	146
7.3.2	Coarse-to-fine image registration	148
7.3.3	Gauss-Newton for multi-modal registration	150
7.3.4	Rigid registration with Gauss-Newton optimisation	151
7.4	Evaluation of continuous optimisation	152
7.5	Efficient MRF-based discrete optimisation	155
7.5.1	Parameterisation of grid and displacements	160
7.5.2	Dense stochastic sampling	161
7.5.3	Minimum-spanning-tree	162
7.5.4	Incremental diffusion regularisation	165
7.6	Experiments using discrete optimisation	166
7.6.1	Influence of regularisation weighting α	167
7.6.2	Evaluation of the influence of our contributions	168
7.7	Physiologically motivated image registration	172
7.7.1	Experiments using hyper-labels	175
7.8	Summary	176

8 Experimental evaluation on clinical scans	179
8.1 Deformable registration of inhale-exhale CT scans	181
8.1.1 Comparison of optimisation strategies and image representations	182
8.2 Deformable multi-modal registration of CT and MRI scans	188
8.2.1 Chest CT and MRI of patients with empyema	189
8.3 Summary	193
9 Conclusion and outlook	195
9.1 Spatial context for robust similarity	196
9.2 Efficient graph-based optimisation	197
9.3 Outlook	199
9.3.1 US-MRI registration for neurosurgery	199
9.3.2 Detail-preserving sparse image representations	202
9.3.3 Marginal distributions for segmentation propagation	206
9.4 Summary	208
A List of publications	210
A.1 Co-authored publications	211
A.2 Awards and prizes	212
Bibliography	213

List of Figures

1.1	Graphical outline of thesis, symbolising different parts of a registration framework.	15
2.1	Axial slices of MRI and CT scans with legend of the anatomy.	24
2.2	Illustration of the challenges of deformable CT lung registration.	33
3.1	Jacobian maps for registrations with different complexity.	42
3.2	Examples of anatomical measurements for lungs.	43
3.3	Visible Human Dataset used for landmark localisation experiment.	47
3.4	Overview of the proposed landmark evaluation experiment.	49
3.5	Example of landmark localisation in the inhale-exhale CT dataset.	50
3.6	Simulated intensity distortions of MRI T1 and T2 scans of the VHD.	51
4.1	Illustration of sliding motion for CT scan during respiration.	56
4.2	Influence of robust penalty functions on deformation fields.	60
4.3	Different orders b of (1D) B-spline basis functions	64
4.4	Image-adaptive regularisation using a minimum-spanning-tree.	65
4.5	Example of Prim's algorithm to find the minimum-spanning-tree.	67
4.6	Inverse consistency error of iterative inversion method.	70

5.1	Example of corresponding slices of CT and MRI scans.	74
5.2	Categorisation of image similarity metrics based on their underlying principles: statistical, structural and contextual information.	77
5.3	Visual comparison of statistical similarity metrics: cross-correlation, correlation ratio and mutual information.	79
5.4	Joint intensity distributions of MRI and CT images.	81
5.5	Overview of texton clustering for textural mutual information.	87
5.6	Illustration of the concept of cluster forests for textural MIn.	91
5.7	Sensitivity of parameter choice of statistical similarity metrics for landmark localisation.	95
5.8	Cumulative error distribution of landmark localisation error for statistical similarity metrics.	97
5.9	Landmark localisation error for multi-modal scans with intensity distortions using statistical similarity metrics.	98
6.1	Local entropy images for MRI and CT slices.	107
6.2	Orientation estimation using gradients or the structure tensor. . . .	112
6.3	Concept for the use of MIND as multi-modal similarity metric. . . .	115
6.4	Spatial search region of MIND using different sampling strategies. .	119
6.5	Concept of self-similarity context compared to MIND with 6-NH. .	121
6.6	Hamming distance for similarity evaluations of MIND.	122
6.7	Sensitivity of parameter choice for structural image representations on landmark localisation error.	125
6.8	Localisation error of MIND with different self-similarity layouts with increasing image noise.	127

6.9	Cumulative error distributions of landmark localisation error for structural image representations.	128
6.10	Landmark localisation error for multi-modal images with intensity distortions using structural image representations.	129
6.11	Cumulative error distribution of landmark localisation for best performing methods.	131
7.1	Comparison of continuous and discrete optimisation principles. . . .	139
7.2	Visual outcome of deformable registration for inhale-exhale CT scan pair using Gauss-Newton optimisation and MIND similarity. . . .	153
7.3	Evolution of registration error of various continuous optimisation strategies with increasing iteration count.	156
7.4	Complexity of deformations and computation times for different continuous optimisation methods.	156
7.5	Flow-chart of individual steps of our MRF-based deformable registration approach deeds	159
7.6	Example of minimum-spanning-tree for coronal lung CT slice. . . .	163
7.7	Concept of lower envelope computation for pair-wise potential calculation with subpixel offsets in belief propagation.	166
7.8	Influence of regularisation weighting and number of stochastic samples on registration accuracy.	168
7.9	Cumulative distribution of landmark error for discrete optimisation framework, including different parts of our technical contributions. .	170
7.10	Example registration result using deeds , comparing the obtained deformation fields of random and image-adaptive spanning trees. .	171

7.11 Demonstration of simultaneous lung ventilation estimation using deeds compared to the standard Jacobian approach.	176
8.1 Visual comparison of difference images and deformation magnitude after registration of lung CT for different approaches.	183
8.2 Comparison of registration error for lung CT registration between continuous and discrete optimisation.	184
8.3 Visual comparison of registration result of inhale-exhale CT for gsyn and deeds	185
8.4 Quantitative evaluation of accuracy for CT/MRI registration. . . .	191
8.5 Example of deformable CT/MRI registration using NMI or MIND. .	192
9.1 Visual example for deformable MRI-US registration.	201
9.2 Registration error and computation time for MRI-US registration. .	202
9.3 Example of sparse image representation using super pixels.	203
9.4 Visual outcome of supervoxel-based registration of lung CT scans. .	205
9.5 Example of segmentation propagation using uncertainty estimates. .	207
9.6 Resulting segmentation overlap score for label propagation using marginal distributions.	208

List of Tables

7.1	Overview of registration results (computation time, complexity of deformations, and landmark error) for different parts of discrete optimisation framework deeds	169
8.1	Overview of registration results for CT dataset, including deformation complexity and computation time.	187

Chapter 1

Introduction

The importance of medical imaging for diagnosis, monitoring and treatment of disease has steadily risen over the last decades. The resolution, contrast and dimensionality of medical scans is constantly improving, but this comes at the cost of an increasing the amount of data to be assessed by researchers and clinicians. Especially for high-dimensional and multi-modal data, automated analysis tools are required to extract the most useful information. Additionally, computerised image analysis methods are more repeatable and not prone to intra- and inter-observer inconsistencies, which makes them very suitable for deriving quantitative measures. Image registration, the process of estimating a spatial transformation relating corresponding anatomical and/or functional locations between scans, is a versatile tool for several analysis tasks, including motion correction of 4D sequences, fusion of multi-modal scans, atlas-based segmentation and measuring longitudinal change.

Figure 1.1 gives a graphical outline of the components of the deformable registration method presented in this thesis. A brief overview of the individual chapters

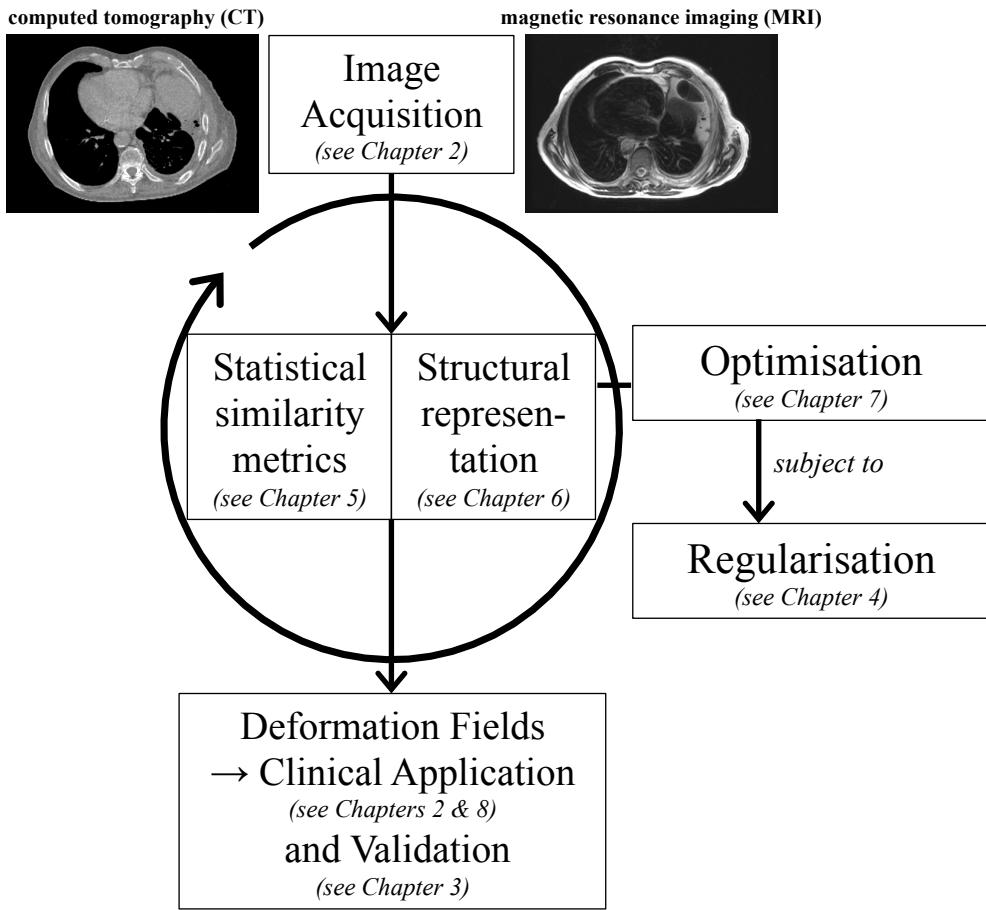


Figure 1.1: Graphical outline of this thesis, which symbolises the different parts of a registration framework (image similarity, regularisation and optimisation) within the clinical context of medical image analysis (image acquisition and clinical application).

is given below, along with the research contributions in this thesis.

Chapter 2 introduces two clinical applications of lung registration, deformable multi-modal fusion and respiratory motion estimation, which are of particular interest in this work. Thereafter, the challenges involved are discussed, which arise due to imperfections in image acquisition, such as noise, artefacts and bias field, as well as the complex motion of the underlying anatomy and physiology. These challenges are only partially resolved by state-of-the-art registration methods, which

is why deformable image registration remains a very active research area.

Chapter 3 deals with the task of determining registration accuracy, which is an important aspect for the development of medical image registration algorithms and for comparison or benchmarking of different methods. In the absence of ground truth information of the real motion, clinically relevant anatomical features, e.g. landmarks, surfaces and volumetric segmentations are labelled manually for a number of scans by clinical experts and used to evaluate automatic methods in terms of accuracy. Other (complementary) metrics, which are used to assess and compare different registration methods is the quality of the obtained deformation fields, e.g. the complexity of deformations, singularities in the motion fields and their inverse-consistency.

Chapter 4 describes concepts for the modelling of the complex respiratory motion of the lungs, which consists of both smooth elastic-like deformations and discontinuous sliding at the interface between the lungs and rib cage. Previous work, on regularisation functionals, which enable directional-dependent smoothing of the deformation fields, is presented. However, there are limitations of these methods, e.g. the dependence on accurate segmentation of potential sliding interfaces. We present a novel approach using an image-derived minimum-spanning-tree to connect control-points in a parametric transformation model, which results in a simple, yet accurate model of lung motion [Heinrich et al., 2012d]. Additionally, a modular method to obtain diffeomorphic and symmetric transformation is presented, which can be used regardless of the employed optimisation strategy and not only removes the potential bias of the choice of order of scans, but also improves registration accuracy.

Chapter 5 discusses statistical forms of defining image similarity across scans.

Similarity metrics are the main driving force of deformable registration. While their definition is straightforward within the same modality, where intensity differences can be directly evaluated, it remains a very challenging task across modalities. A common approach to this problem, the maximisation of mutual information (MI), aims at deriving a statistical relationship between intensity distributions from multi-modal scans and to at minimising their joint entropy. However, since this procedure is based on the assumption that a single global relationship between intensities exists, it will fail in the presence of image distortions or large initial misalignment. A novel approach called textural mutual information is introduced in Sec. 5.3 [Heinrich et al., 2012a], which incorporates spatial context into the computation of mutual information in order to improve the robustness and accuracy of multi-modal similarity measures.

Chapter 6 describes novel concepts for structural image representation. This is a complementary approach to statistical similarity measures and has theoretical and practical advantages over mutual information based methods in terms of the arising optimisation problem, especially when dealing with highly complex motion. The key idea is to find an alternative image representation, which is independent of modalities and can therefore be compared across scans using intensity differences. After discussion of a recent method, which uses a scalar valued representation based on local entropy calculation, two novel approaches to multi-dimensional structural image representation are introduced. The structure-tensor based orientation measure in Sec. 6.3 uses local gradient orientation as a feature [Heinrich et al., 2011c]. The main contribution in this chapter is the derivation of the modality independent neighbourhood descriptors (MIND) [Heinrich et al., 2012b]. MIND is derived from the idea of image self-similarity and enables the

construction of a highly discriminative image representation, and is shown to be independent of modality and robust against image distortions and large initial misalignments.

Chapter 7 deals with the challenges of the optimisation of the registration cost functions based on the current approaches. The improvements of moving from locally defined gradient descent optimisation (demons approaches), over globally regularised Gauss-Newton optimisation driven by local gradients, towards globally optimal graph-based discrete optimisation are discussed. A new highly efficient graph-based registration approach is introduced, which employs very dense sampling of possible displacements [Heinrich et al., 2012d]. Along with the increased accuracy of the deformable registration of inhale-exhale CT lung scans, the contributions made in this chapter also enable a greatly improved computational efficiency. This enables the exploration of higher-dimensional problems, in which not only the geometric transformation parameters are estimated, but also physiological parameters, e.g. lung ventilation (see Sec. 7.7) [Heinrich et al., 2013a].

Chapter 8 presents experimental results for the discussed clinical applications using two datasets, inhale-exhale CT scans and multi-modal MRI and CT volumes. The findings are validated and evaluation with manually annotated anatomical landmarks. The novel contributions made in this thesis are compared to state-of-the-art methods in terms of registration accuracy.

Chapter 9 concludes this thesis with discussing the prospects and remaining challenges of the field of medical image registration in general and gives potential future directions for research in this area. A combination of discrete optimisation and MIND is applied to near real-time deformable ultrasound-MRI registration for

the use in image-guided neurosurgery. Sparse image representation have a great potential for the use in high-resolution, multi-dimensional medical image analysis, as they may provide a better trade-off between efficiency and accuracy. An approach to edge- and detail-preserving image representation for motion estimation with discrete optimisation will be presented [Heinrich et al., 2013b]. Another application for which discrete optimisation offers potential benefits is for segmentation propagation. Here, the use of not only the most probable but the full distribution of many possible transformations can be used to improve segmentation accuracy.

Chapter 2

Motivation and challenges of deformable lung registration

The aim of this chapter is to discuss the potentials, current limitations and challenges of deformable lung registration with respect to its clinical applications. It provides a brief introduction to the acquisition of clinical lung scans, the general framework and terminology used in the literature of deformable image registration, and common applications to pulmonary image analysis.

This chapter will introduce some of the fundamental principles of medical image acquisition and medical image analysis. The major applications and challenges for clinical use of automated deformable motion estimation (or registration) of lung scans will be discussed. Medical imaging entails the formation of images from the inside of humans, primarily in a non-invasive way. It is one of the key contributions to the enormous improvement of medical diagnosis and treatment during the 20th century, because quantitative image analysis enables clinicians to make more objective decisions. The benefits of medical imaging have driven the medical

understanding of human anatomy, physiology and subsequently diseases to a new level and therefore have made substantial improvements in treatment efficacy.

Tomography extends the use of medical imaging from planar 2-dimensional images to volumetric 3D scans. It depends heavily on mathematical models and computational algorithms for the reconstruction of a 3D image from several local measurements or projections. Medical imaging modalities can be roughly divided into ionising and non-ionising techniques. While ultrasound (US) and magnetic resonance imaging (MRI) scans are non-ionising and considered to be harmless, the radiation exposure of X-ray computed tomography (CT) or positron emission tomography (PET) can be accumulate and potentially cause cell mutations. Nevertheless, both CT and PET can provide valuable information (e.g. better spatial resolution and metabolic information respectively) that is not available in MRI and US. In this thesis the focus lies on structural images and dynamic sequences of CT and MRI, which will be discussed in detail in Secs. 2.1.1 and 2.1.2.

The medical speciality radiology aims to extract clinically useful information from these images. The main task is to detect and monitor abnormalities, such as tumours, and make decisions regarding diagnosis and / or further treatment. Real-time images can also directly guide a surgical intervention or image-guided radiotherapy (IGRT). The increasing amount, improved detail and higher dimensionality of available medical scans have led to the emergence of the field of **medical image analysis**. The aim of analysis techniques is to provide the radiologist with additional high-level information directly derived from the images. These high throughput, repeatable, and accurate quantitative image-based measurements can greatly assist the human observer. An explanation for the advantages of computerised analysis is the steadily increasing processing power and the limited ability

of humans to comprehensively assess three- or four-dimensional data. Example applications of medical image analysis are the automatic delineation (image segmentation) of organs or solid tumours, motion estimation and / or correction and the detection of localised volume changes between two scans (image registration). This thesis focusses on deformable registration of lung scans. A short introduction into the terminology of medical image registration is given in Sec. 2.2, more detailed discussions will follow in Chapters 4 to 7. In Sec. 2.3 we discuss some potential clinical applications, including deformable multi-modal fusion for diagnosis and treatment planning and lung ventilation estimation (with potential use for assessment of breathing disorders). The major methodological challenges for these tasks will be discussed in Sec. 2.4, motivating the novel contributions made during this thesis in the following Chapters.

2.1 Medical image acquisition

2.1.1 X-ray computed tomography

Computer Tomography was introduced into clinical practice in the 1970s [Smith and Webb, 2010, Chapter 2]. A CT scan is acquired by taking several X-ray projections of the patient while rotating the X-ray tube and the detector array around him or her. The densities and attenuation coefficients of tissues are different, thus when utilising a geometric reconstruction algorithm (commonly filtered backprojection) a 3D reconstruction of the anatomical structures can be achieved. Modern scanners can capture up to 16 slices (with thicknesses of less than 0.5 mm) simultaneously within half a second, making CT a very fast and very precise medical

imaging technology. CT scans are routinely used for imaging the chest, the heart, the abdomen and the pelvis. Advantages of CT compared to other modalities are the speed of acquisition, which results in low per-scan costs, high spatial resolution, excellent dense tissue contrast, low distortion due to motion during the scan and the direct usability of attenuation correction based on CT images in IGRT and nuclear medicine image reconstruction. CT scans are expressed in Hounsfield units (HU), which are independent of scanner manufacturer and imaging sequence, making them a direct quantitative measurement. The drawbacks of CT scans are the comparatively poor soft tissue contrast and the radiation exposure caused by the X-rays, which can in high doses (e.g. accumulated by excessive repetitive scans) cause DNA damage or induce cancer [Brenner and Hall, 2007].

Dynamic CT imaging

Dynamic or 4D CT is used to assess motion during the respiratory cycle. For the planning of image-guided radiotherapy of lung cancer patients, motion estimated from 4D-CT can be used to improve the margins of the gated radiation [Weiss et al., 2007]. Spatio-temporal CT sequences are usually acquired during free breathing in a scanner in cine-mode connected to a respiration-monitoring system, which enables the reconstruction of discrete temporal frames. In Sec. 7.7 inhale and exhale pairs of 4D-CT sequences are used for lung ventilation estimation. This regional functional assessment of lung functionality can be used both for diagnostic tasks, e.g. for patients suffering from chronic obstructive pulmonary disease (COPD) or asthma, and radiotherapy to define tumour margins, which avoid radiation of well-functioning lung tissue.

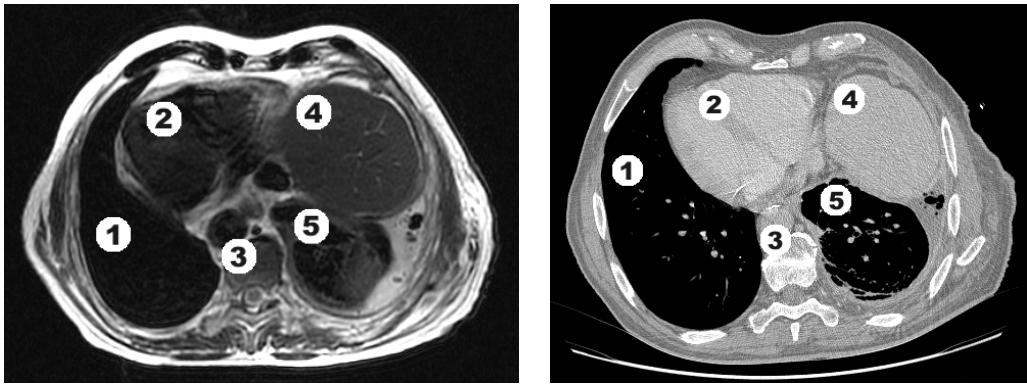


Figure 2.1: Two axial slices of 3D MRI (left) and CT (right) chest scans, showing (1) left lung, (2) heart, (3) spine, (4) liver and (5) right lung.

2.1.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) was developed in the 1970s and uses non-ionizing radio frequency (RF) signals to acquire images [Smith and Webb, 2010, Chapter 5]. The main physical principle of MRI derives from the fact that the magnetic moment (spin) of water (or lipid) protons can be aligned by an external stationary magnetic field B_0 in z-direction. Depending on the field strength the protons precess at the Larmor frequency $f = \gamma B_0 / (2\pi)$, where γ is a constant (gyromagnetic ratio). If an additional short RF pulse with exactly that frequency is applied (orientated perpendicular to B_0), the net magnetisation is flipped out of alignment with the static field by a certain flip angle α . This results in a time-varying magnetic flux, which induces a current in the receiver coil – the measured MR signal (dependent on the proton density). After the RF pulse has been switched off, the magnetisation in z-direction slowly returns to the equilibrium state with a (spin-lattice) relaxation time T_1 and the magnetisation in x- and y-direction returns to 0 with a smaller time constant (spin-spin) T_2 . The specific chemical environment of the water protons leads to different relaxation times and

can therefore yield excellent soft tissue contrast, which can be manipulated by the flip angle and RF pulse strength. The spatial localisation of the MR signal is determined by a frequency sampling strategy. In addition to the stationary magnetic field, three gradient fields are applied, so that there is a spatially varying field strength distribution over the region of interest. Changing the applied gradient fields over the course of acquisition can be regarded as sampling in the frequency domain (so called "k-space"), consequently an inverse Fourier transform is sufficient to reconstruct the image. The advantages of MRI scans are the very high soft tissue contrast (see Figure 2.1 (a)), the possibility of functional imaging (diffusion, perfusion) and the acquisition is considered to be harmless. The drawbacks are much longer scanning times (up to 15 minutes for typical sequences with high spatial resolution), distortion due to inhomogeneities in the magnetic field and image artefacts due to patient motion during the scanning period. In order to avoid motion artefacts during acquisition, in particular for lung scans without respiratory gating, the acquisition time has to be substantially reduced (to the time of one breath-hold) causing a substantial deterioration of image quality and scan resolution.

Dynamic contrast enhanced MRI

The recent development of dynamic MRI image sequences have had an great impact in cancer imaging. In addition to the structural information dynamic sequences are often useful to achieve sufficient image contrast and more detailed information about tumours and their microenvironment, in particular their vascular systems [OConnor et al., 2011]. Dynamic contrast enhanced (DCE) imaging uses a baseline structural scan, followed by a time-series of scans acquired after in-

jection of a contrast agent (CA), which allows to obtain a functional time-activity curve for each voxel. The pharmacokinetics exhibited by these curves enable the extraction of physiological parameters based on an approximated model function.

2.2 Medical image registration

Due to its great potential clinical benefits and remaining challenges, medical image registration has become a large and active field of research, over the past decades. An introduction can be found in [Hajnal et al., 2001], a comprehensive review including recent methodological approaches in [Sotiras et al., 2013] and an experimental comparison of pulmonary image registration methods in [Murphy et al., 2011b]. The main mathematical and algorithmic challenges for deformable registration are due to the problem being under-constrained, non-linear, and non-convex. The terminology of image registration algorithms usually contain three important aspects to model and solve the given problem:

- a transformation and/or regularisation model, which imposes a prior on permissible deformations based on a physically or mathematically motivated model
- a similarity term, which measures the (dis)similarity of images during alignment based on their intensities, geometric features or higher-level information (e.g. segmentation labels)
- an optimisation method to find a local (or global) minimum of an energy function consisting of dissimilarity term and regularisation penalty.

[Holden, 2008] presents an overview of commonly used transformation models for medical image registration. In Chapter 4, regularisation approaches from the literature are reviewed and alternatives are proposed and discussed, which are necessary in order to accurately model (and sufficiently constrain) the complex respiratory motion. Finding a suitable similarity metric, that is discriminative to different underlying anatomical structures and also robust to image intensity and geometric distortions (caused by the different physical phenomena of medical image acquisition) is perhaps the most challenging problem of medical image registration. In Chapters 5 and 6 novel approaches are introduced for statistical and structural image similarity metrics, which address these challenges by incorporating spatial context into the formulation. Local gradient-based methods, such as gradient descent, the conjugate gradient method, or Newton-like methods are often used to optimise the energy function of a registration problem (an overview of gradient-based optimisation for parametric medical image registration can be found in [Klein et al., 2007]). Due to the non-convexity of the problem, most approaches only find local minima of the energy function, especially in the presence of large motions. The limitations of available processing time makes a (global) search over all possible local minima impractical. A new approach, which obtains a computationally tractable, but still globally optimal solution of a registration problem will be presented in Chapter 7 based on a graphical Markov random field (MRF) model.

2.3 Clinical applications

Two particular clinical applications are addressed in this thesis, deformable multi-modal registration for image fusion and respiratory motion estimation. The current challenges for deformable registration are highlighted in Sec. 2.4.

2.3.1 Deformable multi-modal registration for fusion

As mentioned in Sec. 2.1 there has been a great amount of development of new (and improved) medical imaging modalities within recent years. For the clinical tasks of diagnosis and monitoring of diseases, in particular tumours, multiple modalities carry complementary relevant informations. As mentioned before, MRI excels in soft tissue contrast and its variability to use different imaging sequences (T1, T2, proton density, etc.). Yet, it lacks the high spatial (and temporal) resolution of CT and its good contrast for dense tissue (bones, etc.). Contrarily, CT has a poor soft tissue contrast. Multi-modal fusion, the combination of multiple modalities based on multi-modal registration, therefore has potentially great clinical impact. Manually aligning images for multi-modal fusion is very time-consuming, not well repeatable, and may be less accurate in areas with low visible contrast. Multi-modal registration can also be employed for the alignment of a high-quality planning scan and a lower quality pre- or intra-operative scan for image-guided interventions (surgery or radiotherapy). Here, a manual alignment is not feasible due to the time-constraints, and often only a rough alignment based on external markers is used, which could substantially reduce the treatment efficacy.

2.3.2 Respiratory motion estimation

There are several clinically relevant applications for intra-patient respiratory motion estimation. First, for longitudinal monitoring of lung tumours or nodules different breathing levels at the distinct imaging sessions need to be compensated for to make an accurate visual or automatic comparison of the different time-points [Staring et al., 2009b]. Lung ventilation or pulmonary function can be estimated using deformable registration of two (or more) CT scans acquired at different respiratory levels (usually within the same session) to assess and understand breathing disorders, e.g. COPD, or in general for patients undergoing IGRT. Registration of 4D-CT scans can also potentially be used to estimate a patient-specific motion model to improve the dose-painting and thus treatment efficacy of gated or intensity-modulated radiotherapy for lung cancer patients [Guerrero et al., 2005] in order to obtain more accurate tumour margins and avoid radiation of well-functioning lung tissue.

2.4 Challenges of lung registration

In the following, the main challenges for deformable registration of single- and multi-modal lung scans are discussed. Non-rigid registration algorithms aim to solve an ill-posed, non-convex optimisation problem with several million degrees of freedom. There are four particular challenges in lung registration: large motions of small features and sliding motions between organs, changing image contrast due to compression (in CT), non-functional intensity mapping across different modalities (here MRI and CT).

2.4.1 Large motion of small features

Motion within the lungs can often be larger than the scale of the features (vessels and airways), see Fig. 2.2 for a visualisation. This can and does cause a registration algorithm getting trapped in a local minimum, and may lead to an erroneous registration. Local minima are frequently encountered in lung registration. Most deformable registration algorithms (23 out of 24 algorithms in a recent comparison study on pulmonary CT registration [Murphy et al., 2011b]) use continuous optimisation, which is particularly susceptible to local minima. Multi-resolution schemes can help alleviate this non-convexity. However, this still requires the corresponding anatomical structures to be partially overlapping, which is not the case for many features during breathing motions. In [Brox et al., 2008a] a hybrid approach consisting of both a local intensity-based and sparse descriptor matching has been introduced within a variational framework. An alternative approach to avoid local minima is the use of discrete optimisation, which is usually formulated on a Markov random field (MRF). Discrete optimisation offers numerous advantages, in particular greater control over the displacement space, to overcome these limitations. However, the space \mathcal{L} of possible displacements needs to be quantised, leading to many more degrees of freedoms. A computationally very efficient approach for MRF-based deformable registration based on a minimum-spanning-tree graph model and a stochastic dense displacement sampling is presented in Sec. 7.5.

2.4.2 Sliding motion at lung surfaces

Most registration algorithms include prior knowledge about the smoothness of deformations into the optimisation process to avoid physically implausible folding or

gaps in the deformation field. The smoothness constraint typically assumes homogeneous motion and can be part of the transformation model or used as a penalty (regularisation) term. This assumption is violated in the case of a sliding motion between two objects (at their boundary), which naturally occurs during respiration (a further detailed illustration of this will be shown in Fig. 4.1 in Chapter 4). A homogeneous smoothness prior at sliding surfaces causes the registration to be inaccurate. Several authors address this problem by masking out the background objects that follow a different motion (e.g. a "motion mask" is used in [Vandemeulebroucke et al., 2012]). Two separate registrations are then performed for the foreground and background objects. However, this requires an optimal fusion of the two resulting motion fields. In [Schmidt-Richberg et al., 2012a], a direction-dependent regularisation is proposed that is based on an automatically detected mask. [Baluwala et al., 2013] decouple the tangential and normal components of the force field within an elastic registration. In Sec. 4.1 more approaches for discontinuity-preserving regularisation of the motion are reviewed. A regularisation penalty based on the modified L_p norm presented was proposed by us in [Heinrich et al., 2010b] (see Sec. 4.1.2). In Sec. 4.3 a different graph structure, namely an intensity-derived minimum-spanning-tree, is introduced which effectively models sliding preserving motion.

2.4.3 Intensity variation due to lung compression

A local change in lung volume is expressed in CT scans as a relative difference in the corresponding Hounsfield units within the breathing cycle. The change in density (and image intensity) can be problematic for deformable registration

if a one-to-one intensity mapping is assumed. Similarity metrics that assume a globally linear relationship (e.g. cross-correlation) or statistical dependency (e.g. mutual information), cannot resolve for the locally varying contrast. Recently, so called mass-preserving similarity terms have been introduced by [Yin et al., 2009] and [Gorbunova et al., 2012]. In Sec. 7.7 an alternative solution is proposed, based on the discrete MRF-based registration framework presented in Sec. 7.5. An additional fourth label dimension is introduced into the discrete label space, which represents a local multiplicative intensity variation for each control point in the graph. The advantage of this approach is that we can simultaneously estimate a dense motion field and a regularised map of local density change. This directly provides a ventilation image of the lung functionality.

2.4.4 Multi-modal image similarity for deformable registration

When considering scans from different modalities, the intensity variations are more severe than for the aforementioned CT scans. There is usually no functional mapping between the intensities across modalities. Therefore statistical metrics, such as mutual information (MI) [Maes et al., 1997] have been widely used for multi-modal registration. However, the application of MI for non-rigid registration has proven to be very difficult, for example several disadvantages (caused by interpolation, initial misalignment, etc.) have been discussed in [Pluim et al., 2000] and [Haber and Modersitzki, 2007]. In addition, locally varying intensity variations, due to bias fields in MRI and lung compression in CT, violate the assumption of a global statistical dependence and make the use of more sophisticated MI variants

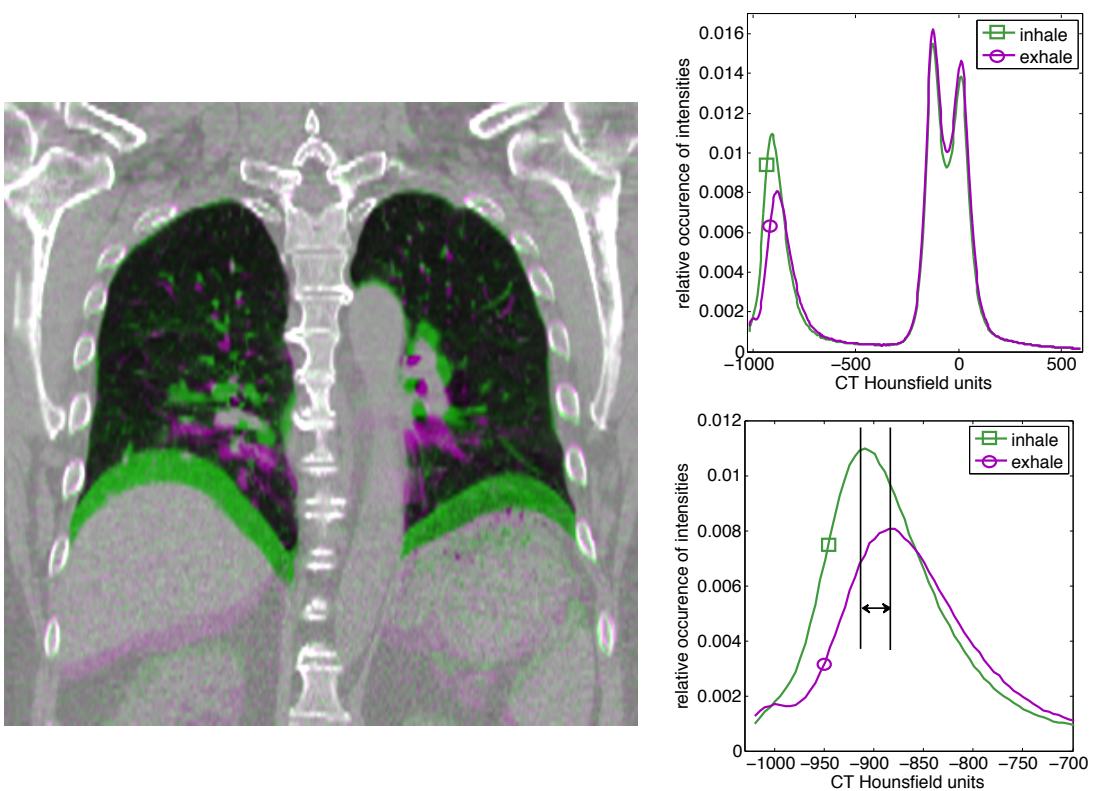


Figure 2.2: Illustration of some of the challenges of deformable lung registration. Overlay of inhale (green) and exhale (magenta) phase of 4D-CT scan [Castillo et al., 2009]. The motion of the diaphragm is roughly 20 mm in this case, which is larger than the thickness of vessels within the lungs. The histograms of the HU values of inhale (green) and exhale (magenta) scan are shown on the right. The magnification of values between -1000 and -700 HU shows a shift of intensities (≈ 30 HU) between breathing phases due to compression.

necessary. Few examples for deformable multi-modal registration can be found in the literature: [Ou et al., 2011] present visual results for MRI to histology registration, [Loeckx et al., 2007] report improved volume overlap for the rectum in CT/MRI registration, and [Rivaz and Collins, 2012] evaluate ultrasound to MRI registration with manual landmarks. The work of [D’Agostino et al., 2003; Mellor and Brady, 2005; Wachinger and Navab, 2012] aims at multi-modal deformable registration, however, only presents results for synthetically deformed images. In

Chapter 5 we address the unsolved challenges of multi-modal similarity by making use of spatial context for the definition of image similarity. We present an approach to include information of small image patches to formulate a textural mutual information in Sec. 5.3 [Heinrich et al., 2012a]. Based on the concept of image self-similarity we devise a new multi-dimensional image representation in Sec. 6.4: the modality independent neighbourhood descriptors [Heinrich et al., 2012b].

Chapter 3

Validation of deformable image registration

Evaluation of the quality of a registration algorithm on clinical data against a ground truth metric is a very important step in developing new methodologies. The objective of this chapter is to discuss metrics, which are suitable for the validation of the deformable registration methods presented in this thesis. For pulmonary image analysis anatomical or geometric landmarks are one of the most important evaluation criteria.

The task of validation of deformable registration is far from being trivial and several (sometimes controversial) evaluation criteria have been proposed in the past. Evaluation of registration quality primarily has two different goals: assessing the robustness and accuracy of a proposed method for a given clinical tasks, and benchmarking of several methods in order to rank them. The former is addressed in almost every publication introducing a new clinical application of image

registration. Benchmarking of different deformable registration methods has only recently started: first, for the deformable inter-subject registration of brain MRI scans by [Klein et al., 2009] and thereafter for intra-subject lung registration of CT scans by [Murphy et al., 2011b]. In Sec. 3.2.1 we discuss a number of surrogate metrics, clinically relevant image-derived metrics and higher-level clinical metrics, which are frequently used for evaluation. Often, no clinically meaningful ground truth is obtained and only surrogate metrics are used to validate a new registration technique. Note that, using a mock registration tool, which exploits the insensitivity of surrogate metrics, [Rohlfing, 2012] demonstrates why these might lead to unreliable results and should therefore be avoided. The ability of our presented new similarity metrics and image representations to robustly align scans from different modalities is presented in Chapter 8 and compared to other metrics.

This chapter is organised in the following way. First, the imaging data used in this thesis is described in Sec. 3.1. Second, surrogate measures based on intensities and properties of deformations are discussed in Sec. 3.2.1 and 3.2.2. Third, measures based on clinically relevant anatomical features, including landmarks, volumetric, and surface segmentations are presented. For lung motion estimation, a particular focus of validating registration methods lies on expert-annotated anatomical landmarks. A disadvantage of the previous two groups of measures is that they usually require a dense deformation field after registration. It is therefore challenging to isolate the individual effects of different parts of the registration method (similarity term, transformation model, optimisation). In Sec. 3.3 we present an alternative evaluation procedure: regional landmark localisation, which is used in Chapters 5 and 6 to evaluate different similarity metrics. Here,

the similarity cost is only locally computed in order to localise a geometric or anatomical landmark within a specific search region.

3.1 Description of imaging data

In this thesis three different datasets are used to evaluate the contributions presented in this thesis for the driving clinical applications of this work and compare them to state-of-the-art methods. First, the visible human dataset (VHD) [Ackerman, 1998] will be used in Chapter 5 and 6 to study different similarity metrics. This dataset¹ consists of different MRI modalities (T1-, T2- and PD-weighted) and scans were acquired post-mortem, which means that there is intrinsically no motion present. Section 3.3 describes the experiments, which are performed to compare different similarity metrics (using these scans). Second, a number of CT scans with respiratory motion are used in Chapters 7, 5 and 6 to compare different optimisation strategies and again similarity metrics. Estimating and compensating for respiratory motion is an important area of research with applications in diagnosis (of lung functionality and breathing disorders) and image-guided radiotherapy. The results for this dataset will be presented in Secs. 7.4, 7.6 and Chapter 8, and compared against other state-of-the-art methods, which have been applied to the same scans. Since these scans are from the same modality, the main focus is on the use of transformation model, regularisation, and optimisation. Third, we used volumetric images of eleven patients suffering from empyema, a lung disease, who were scanned for diagnostic purposes by our collaborators with both MRI and CT. Different scanning protocols were employed for these clinical datasets. The CT vol-

¹The VHD is available from the National Library of Medicine with a licence agreement.

umes include scans with contrast, without contrast, and a CTPA (CT Pulmonary Angiogram) protocol. For the MRI scans, both T1-weighted and T2-weighted FSE-XL sequences within a single breath-hold were employed. All patients suffered from empyema, a lung disease characterised by infection of the pleura and excess fluid within the pleural space. The extra fluid may progress into an abscess and additionally, cause the adjacent lung to collapse and/or consolidate. Both modalities are useful for detecting this pathology, but because the patients are scanned in two different sessions and at different levels of breath-hold, there are non-rigid deformations, which makes it difficult for the clinician to relate the scans. The quality of the MRI scans is comparatively poor, due to motion artefacts, bias fields and a slice thickness of around 8 mm.

3.2 Measures for evaluation of registration

The Retrospective Registration Evaluation Project (RREP) study [West et al., 1997] is an important example of a gold-standard evaluation and benchmarking of registration accuracy. A number of volumetric brain scans from different modalities (CT, MRI and PET) from the same patient had to be rigidly registered. The gold standard transformations were obtained by the organisers using implanted fiducial markers and their appearance had been removed (or disguised) before distributing the data. The accuracy in terms of target registration error (TRE) was then evaluated for all algorithms in order to rank the participating algorithms. For non-rigid deformations, which occur for example during respiratory motion or longitudinal studies, no such gold standard exists and other measures for the evaluation of registration accuracy have to be found.

3.2.1 Surrogate measures for accuracy based on image intensities

The simplest surrogate measure for registration accuracy is the image similarity after registration. For single-modal registrations, the mean squared error (equivalent to SSD) and for scans from different modalities or with changes in contrast, mutual information are often used. However, these are also very popular similarity metrics, which are widely used to drive the registration itself, so the results will be biased. Additionally, over-fitting of the data (in principle aligning image noise) would be favoured by these metrics, if no additional metric is used to ensure a certain smoothness of the obtained deformations. [Rohlfing, 2012] presents a slightly ironic discussion of this issue. A *completely useless registration tool* (CURT) is introduced, which merely sorts the intensities in the fixed and moving image and derives a deformation field by assigning voxels across images based on their sorting index. Even though this results in a meaningless transformation, it seemingly outperforms state-of-the-art registration methods based on similarity-based surrogate metrics. Therefore, similarity-based metrics should be avoided for evaluating registration accuracy. They might, however, be useful for choosing parameter settings.

3.2.2 Surrogate metrics for quality of deformations

As discussed above, similarity-based agreement of images after alignment is a necessary but not sufficient requirement for a successful registration. A second category of surrogate metrics attempts to evaluate the quality of obtained transformations. The inverse consistency error (ICE) introduced by [Christensen and

Johnson, 2001] evaluates how much the result of a registration algorithm is affected by the order of target and moving image (I_t and I_m respectively). Given the forward and backward transforms ϕ_F and ϕ_B with respective displacement fields \mathbf{u}^F and \mathbf{u}^B the ICE and its mean $\overline{\text{ICE}}$ are defined as:

$$\overline{\text{ICE}} = \|\phi_F \circ \phi_B\|^2 \text{ and } \text{ICE}(\mathbf{x}) = \|\mathbf{u}^F(\mathbf{x}) + \mathbf{u}^B(\mathbf{x} + \mathbf{u}^F(\mathbf{x}))\|^2 \quad (3.1)$$

where \circ indicates the composition of two transformations. The use of a symmetric transformation approach can reduce inverse inconsistency. For many applications (especially intra-patient registration) the physical plausibility of transformations may be known a priori. Therefore the occurrence of singularities in the motion field, which results in implausible folding, can be used to assess the quality of the registration. The Jacobian $Jac(\mathbf{u})$ of the transformation $\phi = \mathbf{Id} + \mathbf{u}$ (with identity transform \mathbf{Id} and displacement field $\mathbf{u} = [u, v, w]^T$) was defined by e.g. [Rey et al., 2002] as:

$$Jac(\mathbf{u}) = \det(\mathbf{Id} + \nabla \mathbf{u}) = \begin{vmatrix} \frac{\partial u}{\partial x} + 1 & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} + 1 & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} + 1 & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} + 1 \end{vmatrix} \quad (3.2)$$

The Jacobian gives an intuitive measure of the local deformation properties. Voxels for which the Jacobian has a value greater than 1 are expanded, the ones for which $Jac < 1$ contracted, and voxels for which $Jac < 0$ disappear (hence there is a singularity in the motion field). The fraction of voxels with negative Jacobians is a popular surrogate measure for registration quality, e.g. in [Murphy et al., 2011b].

The Jacobian has also been widely used in clinical studies for neurodegenerative diseases such as Alzheimer’s [Fox et al., 1996]: in tensor-based morphometry (TBM) [Hua et al., 2008] the values are used as a feature for classification.

However, different registration algorithms with the same fraction (or complete absence) of negative Jacobians can still exhibit a large variation in complexity or smoothness in their resulting deformations. The weighting between the similarity and regularisation term has to be usually set manually within the optimisation framework of a method. In [Ou et al., 2012] the aggressiveness of different registration algorithms is measured using the maximum range of Jacobians. It is found that most methods yield a better registration accuracy based on anatomical segmentations (see next section) when a lower regularisation weighting is used (leading to a larger range of Jacobians). [Yeo et al., 2008] found a similar effect for atlas-based segmentation, however in their work a variation over weighting parameters gave an optimum value. Therefore a fair comparison across different registration strategies should always include a quantitative measure of the complexity of deformations. [Leow et al., 2007] suggests using the standard deviation (std) of Jacobian values to measure the deformation complexity, where lower values correspond to smoother transformations. Figure 3.1 shows the Jacobian maps of an inhale-exhale lung registration using two different regularisation weightings. Although both transformations are free from singularities, the second one shows a substantially higher complexity in terms of $\text{std}(\text{Jac})$.

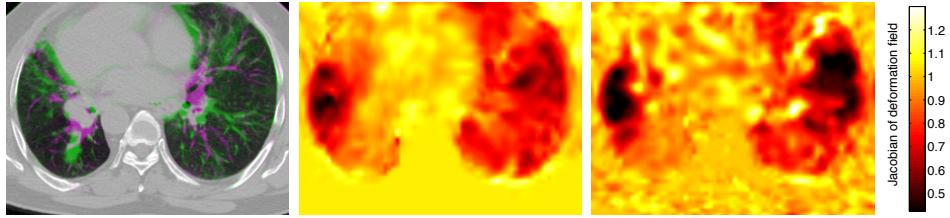


Figure 3.1: Comparison of Jacobian maps for registrations with different complexity. Expansion is indicated with bright, compression with dark colours. Left: Overlay of coronal slice of reference exhale (magenta) and moving inhale (green) image of 4D-CT sequence. Centre: Smoother transformation with $\text{std}(\text{Jac}) = 0.10$. Right: More aggressive transformation with $\text{std}(\text{Jac}) = 0.15$. Both transformations have no singularities (i.e. $\text{Jac} \geq 0$).

3.2.3 Clinically relevant image-derived measures

To overcome the problems related to surrogate measures, clinically relevant anatomical measures have to be taken into account. Here, an anatomical feature is manually annotated in both scans. After registration, the manual annotations from the moving image are transformed into the space of the fixed scan using the estimated deformation field. The discrepancy between the automatically transferred and the manually defined features is used as the registration error (or accuracy).

Comparing different volumetric labelings (segmentations) is commonly done using the Dice coefficient κ , which yields 1 for perfect overlap and 0 for non-overlapping structures. The NIREP dataset for evaluation of inter-subject brain registration has been made publicly available by [Christensen et al., 2006]. It uses 32 manually segmented small anatomical regions of MRI scans of 16 healthy subjects. Similar datasets have been used in the study by [Klein et al., 2009], which aimed at comparing and ranking different non-linear brain registration methods.

For thoracic CT registration the segmentation of the whole lungs is possible, however due to the well-defined boundary most methods are able to align the

lungs almost perfectly and there is almost no discrimination between different algorithms. This can be seen in the results and conclusions of [Murphy et al., 2011b]. In this case, the results are additionally skewed, because the lung segmentations are provided with the data and have been used by most of the participants in the study to drive the registration. For large objects the Dice coefficient is not sufficiently sensitive. Manual or automatic segmentations of smaller anatomical volumes within the lungs (fissures, lobes, and segments) is very challenging ([Van Rikxoort et al., 2009] reports an accuracy of 75%). The fissure alignment error provides a reasonably good measure of registration [Murphy et al., 2011b].

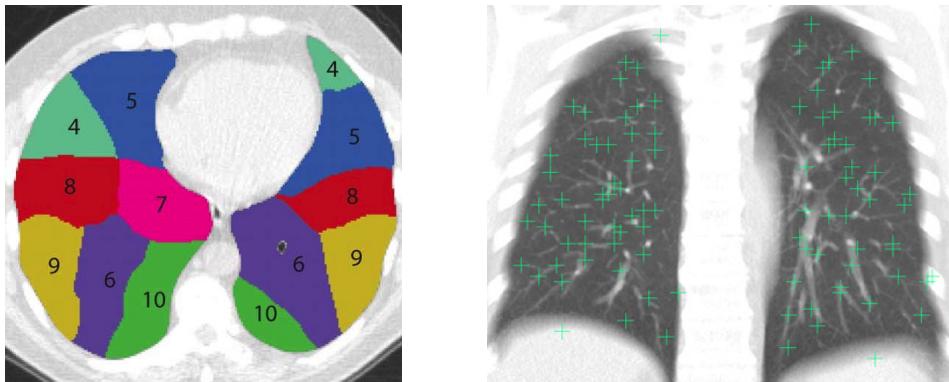


Figure 3.2: Examples of anatomical measurements for lungs. Left: Automatically segmented lung segments shown on axial CT slice from [Van Rikxoort et al., 2009]. Right: Semi-automatically detected landmarks on a maximum-intensity projection (coronal) view of a thoracic CT scan from [Murphy et al., 2011b].

Based on the results of [Murphy et al., 2011b] and the discussion of [Rohlfing, 2012], we can conclude that for the specific application of lung registration, landmark-based validation offers the best opportunity for a robust comparison of registration performance. The task of annotating landmarks in a pair of scans can be divided into two steps: automatic landmark detection and manually establishing a landmark correspondence [Murphy et al., 2011a]. The objective of the first

step is to find a number of well-distributed locations, which are sufficiently distinctive from their surroundings (i.e. landmarks), in order to be matched by a human observer. We will later present a method of automatic landmark detection in Sec. 3.3.1. The second step requires the manual interaction of an (ideally trained) observer in order to find the corresponding landmark in the second scan. For intra-patient registration, we would expect each anatomical feature to be present in both scans, but changes in the field of view or pathological changes might not always guarantee correspondence. The task is very time-consuming especially for multi-modal datasets, where the appearance of corresponding features may be very different. For this reason [Murphy et al., 2011a] complemented this step with automatic landmark matching, using a thin-plate-spline (TPS) transformation based on previous manual point matches and a local block-matching (with SSD similarity). This enables the annotation software to automatically *take over* the manual matching process after a number (usually >30) of good matches. A similar process was described by [Castillo et al., 2009] in order to efficiently annotate very large numbers (≥ 3000) of landmarks in thoracic 4D CT scans. Intra-observer differences for selecting landmarks in intra-patient CT scan pairs are usually less than the image resolution (for 4D-CT scans the resolution is typically $1 \times 1 \times 2.5$ mm and intra-observer errors are ≈ 1 mm). For multi-modal scan pairs (such as the CT/MRI dataset we used) landmark detection and matching is more challenging. Therefore, normally fewer landmark pairs can be reliably selected and the observer error tends to be larger (for the MRI scans with resolution of $0.7 \times 0.7 \times 8$ mm this error was >5 mm). For the validation of registration methods using annotated landmarks the target registration error (TRE) [Fitzpatrick et al., 1998] is commonly used, which is the Euclidean distance between manually located cor-

responding point and automatically determined location (based on deformation fields).

3.2.4 Higher-level clinical measures

If the application of deformable registration is to predict a certain medical condition, e.g. differentiation between healthy and diseased subjects for neurodegenerative disorders, the outcome of a retrospective study of the prediction accuracy can be used to implicitly evaluate the registration quality. In [Bhushan et al., 2011], we compared the ability of different registration strategies to predict the treatment response of colorectal cancer patients. For this purpose, the Kolmogorov-Smirnov distance of distributions of parameter-maps estimated from dynamic contrast enhanced MRI sequences was used to predict the response of patients to radio-chemotherapy in the early stage of treatment. A better prediction or classification does not necessarily imply greater registration accuracy. However, usually the target of clinical applications is to derive clinically useful information and not necessarily to achieve a perfect registration, thus higher-level measures are often the best choice.

3.3 Evaluation using landmark localisation

Evaluating medical deformable multi-modal image registration in a controlled manner is not trivial. As discussed above, finding and accurately marking corresponding anatomical landmarks across modalities is difficult even for a clinical expert. Random deformation experiments, as are usually performed in the literature for multi-modal registration (e.g. in [D'Agostino et al., 2003; Glocker et al.,

2008a; Mellor and Brady, 2005; Wachinger and Navab, 2012]), are not very realistic. In order to perform a simulated deformation for multi-modal data, an aligned scan pair must be available. Moreover, simulated deformations hardly ever capture the complexity and physical realism of patient motion (an exception would be a phantom based on a number of scans acquired at different times). Few thorough and principled comparisons of (multi-modal) image similarity in the medical domain have been made to-date. To overcome the aforementioned challenges, we propose a new approach to compare similarity metrics and evaluate their robustness and accuracy: *regional landmark localisation*. For this purpose, we employ the Visible Human dataset (VHD), see Fig. 3.3. Because the scans were taken post-mortem, no motion is present and different modalities are intrinsically in perfect alignment. We selected two MRI sequences, T1 and PD weighted volumes, as they offer a sufficient amount of cross-modality variations. The images are up-sampled from their original resolution of $1.875 \times 4 \times 1.875$ mm to form isotropic voxels of size 1.875 mm³.

A simple window-based similarity cost aggregation is used (which avoids the differentiation of the similarity function and setting of a regularisation weighting) to derive local similarity maps across modalities for a large number of automatically detected landmarks. This will be explained in detail below (Sec. 3.3.1). In Sec. 3.3.2, we present our experimental setting to evaluate the robustness of different similarity measures with respect to clinically relevant image distortions: Gaussian noise and locally varying multiplicative bias fields. In [Hirschmüller and Scharstein, 2009] similar experiments have been performed to compare matching costs for the application of stereo depth estimation under the influence of linear and non-linear (but monotonic, so not multi-modal) grey-value transformations.

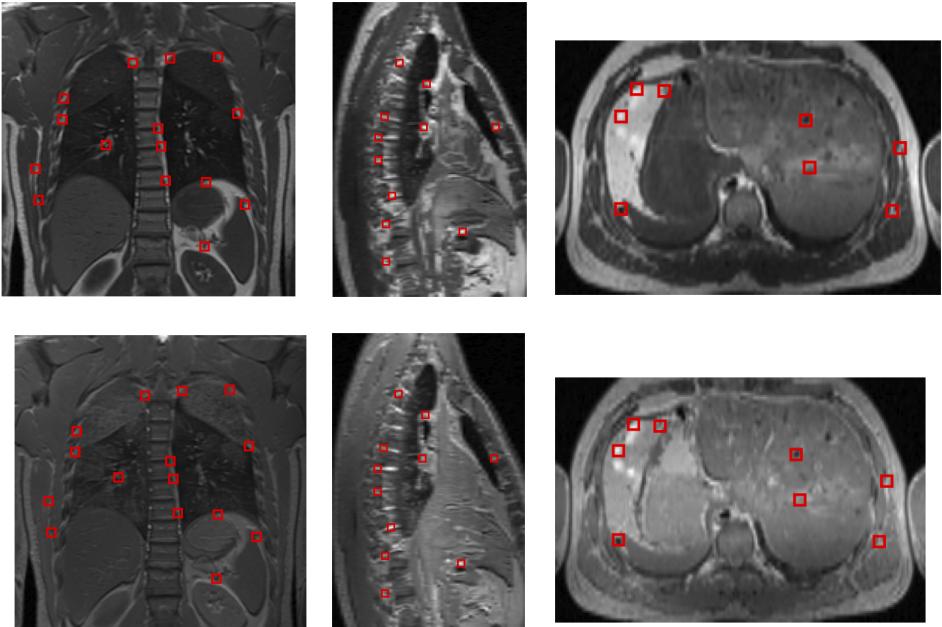


Figure 3.3: Visible Human Dataset used for landmark localisation experiment. The T1 and PD MRI scans, acquired post-mortem, are intrinsically aligned. A representative subset of the 619 landmarks, which were found using the 3D Harris corner detector, is plotted using red squares.

We have used a second dataset for evaluation: a pair of an inhale-exhale CT lung scans [Castillo et al., 2009]. Here, the challenges are the locally varying contrast, due to the change in density within the lungs during respiration, image noise because the scans are acquired with low radiation dose, artefacts due to the 4D reconstruction, and, importantly, local deformations. The localisation of landmarks using only the window-based similarity cost is therefore too challenging for simple metrics, such as SAD or SSD. For this reason, the dataset is also suitable to evaluate multi-modal similarity measures. The resolution of these scans is $0.97 \times 0.97 \times 2.5$ mm.

3.3.1 Similarity-based landmark localisation

For evaluation purposes, we propose to measuring the ability of a similarity metric (or image representation) to robustly and accurately define similarity across scans in the presence of geometric and/or intensity distortions. First, we select a number of anatomical landmarks or feature locations in one scan. In the case of the VHD dataset, this is done automatically using the 3D Harris (or Förstner) corner detector [Rohr, 1997] (see Fig. 3.3). Given the smoothed spatial image gradients ∇I_σ , we construct the matrix $C = \nabla I_\sigma \nabla I_\sigma^T$ for every voxel. A high ratio R between its determinant and trace (see Eq. 3.3) indicates a point feature (ϵ is set to 0.001).

$$R = \frac{\det C}{\text{trace } C + \epsilon} \quad (3.3)$$

Using non-maximum suppression, we automatically select 619 well distributed landmarks, with an example shown in Fig. 3.4. These landmarks are relatively easy to locate, since they have strong gradient responses in all three dimensions. For the second dataset, 300 corresponding landmarks, have been manually annotated by [Castillo et al., 2009], at inner-lung features (e.g. vessel bifurcations) with an intra-observer error of ≈ 1 mm.

After the definition of landmarks in the first scan, the similarity measure (which is being evaluated) has to be computed within a search region extracted from the second (multi-modal) scan. The size of the search region S is set to $32 \times 32 \times 32$ mm for the first experiment (T1- and PD-MRI of Visible Human Dataset) and $24 \times 24 \times 62$ mm for the second experiment (inhale and exhale phase of 4D-CT scan) in order to cover the large respiratory motion. The window W for similarity cost aggregation has an edge length of $l = 11$ voxels. For each displacement within

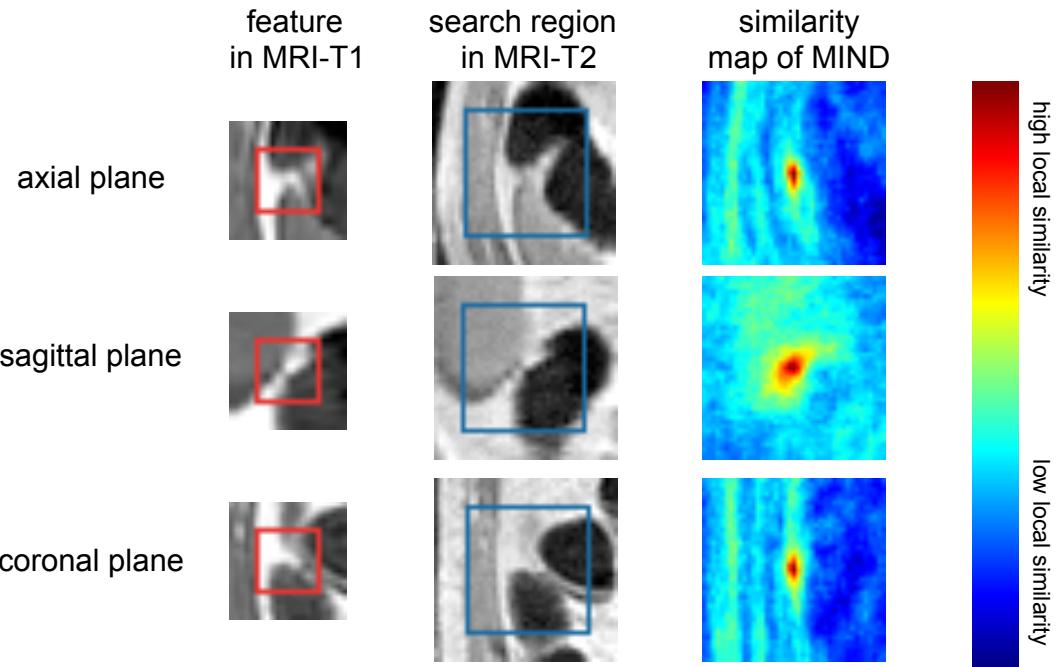


Figure 3.4: Overview of the proposed landmark evaluation experiment. Left: For each landmark a small feature cubic window (edge length $l = 11$ voxels) is extracted from the first scan (marked with red box). Centre: A larger search region (capturing the corresponding landmark location) is defined within the second scan (marked with blue box). Right: The similarity term (here SAD of MIND) is computed for each location within the search region (sliding the window for aggregation). High local similarity is displayed in red, low similarity in blue. The optimal position is then selected and compared with the gold standard location and used to evaluate the localisation accuracy.

the search region, the sum of all point-wise similarity costs for each of the l^3 voxels within the aggregation window is calculated. This forms an intrinsic regularisation of the similarity map and is based on the assumption that neighbouring voxels (within the small window) move together.

Subsequently, the location within S with the highest similarity (lowest similarity cost) is selected. A subpixel optimum is found by fitting a quadratic function to the neighbouring similarity values in each dimension. The Euclidean distance

between the gold standard landmark displacement and the obtained subvoxel optimum of the similarity search is used as landmark localisation error (LLE) (see Fig. 3.5).

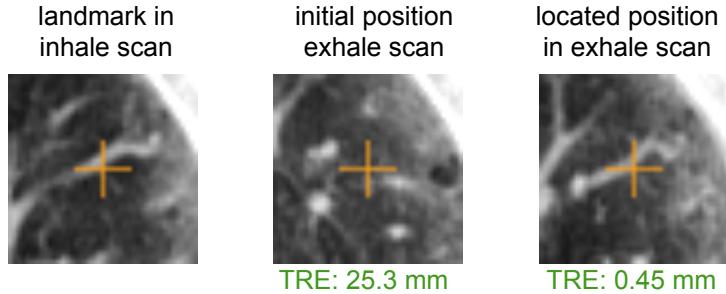


Figure 3.5: Example of landmark localisation in the inhale-exhale CT dataset. Left: The manual landmark is shown on the axial slice of the inhale volume. Centre: The initial position within the search window in the exhale volume. Right: The located position based on the optimum of the similarity map (c.f. Fig. 3.4 right). Note that the optimum lies in a different axial slice. The resulting landmark localisation error (LLE) is calculated with respect to the manually annotated position of the corresponding landmark.

3.3.2 Evaluation of robustness against intensity distortions

To test the robustness of different approaches, we simulate realistic intensity distortions (which are common in clinical scans). First, we apply a non-uniform bias field (multiplicative linear gradient from left to right) of varying strength to the MRI-T1 scan. Figure 3.6 (top row) shows a coronal slice of the T1 scan with very little, moderate, and a strong bias field. The strength is varied in 5 steps up to a maximum amplitude of the multiplicative field of 2 (so that it is in the range of $[0, 2]$). Robustness against bias fields plays an important role in image registration of MRI scans and many approaches have been made to address this challenge. [Ashburner and Friston, 2005] proposed a simultaneous image registration and

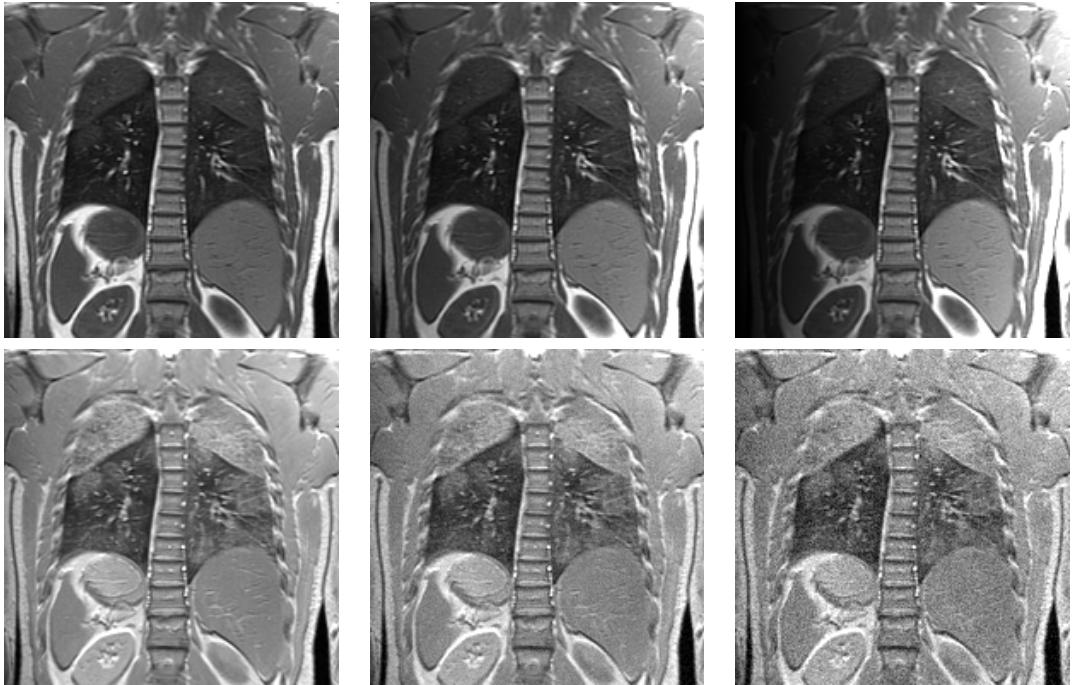


Figure 3.6: Top row: Simulated intensity distortion of MRI-T1 scan using a multiplicative bias field (linear gradient). From left to right: Coronal slice with very little, moderate (range [0.5, 1.5]) and strong bias field (range [0, 2]). Bottom row: MRI-PD scan with additive Gaussian noise (intensity $\mu = 200$). From left to right: very little, moderate ($\sigma = 15$) and strong noise ($\sigma = 30$).

intensity correction method in order to deal with non-uniform bias fields in combination with brain tissue segmentation. The coding complexity of the residual image is minimised in [Myronenko and Song, 2010] to estimate a smooth intensity correction field. Similarly, in [Modersitzki and Wirtz, 2006] a regularised correction function is employed in a variational optimisation framework to compensate for inhomogeneous intensity mappings.

Directly removing or modelling the bias field within registration is challenging, in particular for multi-modal scans, due to the often non-functional relationship between image intensity distributions. The use of a similarity metric, which is sufficiently insensitive to locally varying intensity inhomogeneities is therefore very

useful. Another advantage, of avoiding an explicit model of the bias field, is the simplicity and general applicability of such an approach.

Robustness of a method against Gaussian noise is an indicator that is commonly used to test its usefulness in a real world scenario. Clinical scans are often acquired with lower signal-to-noise (SNR) ratios than research scans, because it is desirable to reduce both scan time for MRI and radiation exposure for CT. Figure 3.6 (bottom row) shows a coronal slice of a PD scan with no, moderate, and strong additive Gaussian noise (a Gaussian noise distribution is close to the Rician noise, which reflects the physical MRI acquisition). The strength is varied in 6 steps up to a maximum variance of $\sigma = 30$ (the mean image intensity is 200).

The inhale-exhale CT dataset exhibits locally varying intensity distribution due to lung compression. A local change in lung volume is expressed as a relative difference in the corresponding Hounsfield values within the breathing cycle. Due to the low-dose radiation setting employed in clinical 4D-CT sequences it also has an intrinsically greater noise level than the VHD dataset. Therefore, no further intensity distortions are applied to this dataset.

3.4 Summary

This chapter has presented different commonly used evaluation criteria for deformable registration. Measuring the target registration error for a (preferably large and well distributed) number of manually annotated anatomical (or geometric) landmarks is the most suitable and widely used gold standard for pulmonary registration methods.

The validation of deformable multi-modal registration is particularly difficult

and little research has addressed this so far. In Sec. 3.3, we have presented a new approach to evaluate and compare different multi-modal similarity measures using landmark localisation experiments. An advantage of this procedure, which will be used in Chapters 5 and 6, is that it can be performed independently of a specific regularisation penalty, transformation model and optimisation technique and is therefore ideally suitable to distinguish between of different ways of defining multi-modal similarity.

Measures to assess the quality of the deformation fields (based on its Jacobian and inverse constancy) of a registration algorithm have been presented. The next chapter discusses in more detail the modelling of lung motion with specific regularisation terms and transformation models.

Chapter 4

Modelling of complex lung motion

The objective of this chapter is to discuss the suitability of current approaches to regularisation and parameterisation for lung deformations, with particular focus on sliding motion, and its implications for the optimisation of the resulting energy functions. We demonstrate the limitations of robust norms for the regularisation lung motion estimation. Instead, we propose the use of an image-derived minimum spanning tree to form an efficient discontinuity preserving connectivity of control points within the parametric grid. We present a simple and accurate approach to ensure symmetric and diffeomorphic transformations for both continuous and discrete optimisation methods.

A realistic mathematical model of the underlying physical motion is key to any image registration method. For this reason, there should be a good balance between how close a model is to the problem and its robustness against image noise and artefacts. Furthermore, the numerical representation of a model and its computational complexity need to be taken into consideration. The main challenge for the estimation of lung motion is the occurrence of sliding organ motion.

While the motion can be assumed to be smooth within most organs and individual lung lobes, there is sliding motion along surfaces, in particular the lung pleura. Figure 4.1 illustrates the notion of sliding motion, due to respiratory motion, on an example slice of a thoracic 4D CT scan (overlay of inhale and exhale phase). The main anatomy of thorax, including lung fissures and diaphragm motion, is outlined. The main part of the lung motion (indicated by white arrows on right side) is approximately parallel to the lung boundary, but with opposing directions for structures within and outside of the lung (the rib cage moves only marginally during this breathing cycle) causing discontinuities at the interfaces. A short overview of variational approaches to regularise deformations will be presented in Sec. 4.1. The standard approach of homogenous diffusion regularisation, which will be presented in Sec. 4.1.1, is not appropriate to model this motion. Three alternative methods, which address the sliding motion: robust norms, direction-dependent and image-adaptive regularisation will be discussed in the following.

Parameterisation, primarily using the concept of Free-form deformations (FFD) [Rueckert et al., 1999], plays an important role in medical image registration. It reduces the large number of degrees of freedom and at the same time imposes a specific regularity on the deformations. A standard FFD transformation model is usually unsuitable to deal with the complex sliding motion, as it enforces the smoothness constraint across the motion boundary. A simple solution is to use a segmentation mask and perform two separate registrations, treating the motion inside and outside of the lungs independently [Wu et al., 2008]. However, this requires very accurate segmentation, it is not trivial extendable to multi-organ sliding motions and a combination of the two deformations might cause gaps or overlap. In Sec. 4.3, we propose the use of an image-derived minimum-spanning-

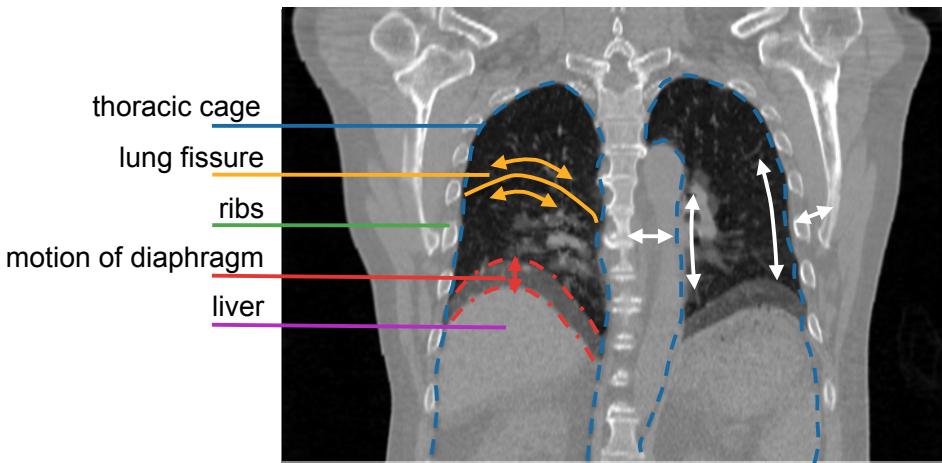


Figure 4.1: Illustration of sliding motion in coronal plane of thoracic CT scan during respiration. Overlay of inhale and exhale phase of 4D-CT scan [Castillo et al., 2009]. The anatomy, which plays a role in respiratory motion is depicted. On the right, it can be seen that the motion inside the lung is mainly parallel to the thoracic cage, while the ribs stay almost static, causing a discontinuity of the motion field. The motion of the diaphragm is roughly 20 mm in this case (larger than the thickness of vessels within the lungs).

tree (MST) for sliding-preserving parametric transformations. The MST removes edges between control-points at locations of high image gradients (which are likely to coincide with discontinuous motion).

For particular applications of image registration certain properties are desirable for the estimated deformations. Usually, deformable registration is performed by defining one scan as fixed, the other as moving image. Symmetric and inverse consistent registration approaches remove the bias of choosing one or the other image as the moving image. Diffeomorphic mappings ensure that the inverse transformation is invertible and free of singularities. We propose a novel, efficient approach to obtain both transformations, which are both diffeomorphic and symmetric in Sec. 4.4. It can be seen as a modular independent component, which is easy to incorporated into almost any existing registration algorithm.

4.1 Variational regularisation

Deformable image registration is an ill-posed, underdetermined problem and therefore requires some form of regularisation to obtain physically plausible motion estimates. Regularisation ensures a smooth deformation field in image areas with low information content (e.g. homogeneous regions) by penalising the deviation of the motion vectors of spatially close locations.

4.1.1 Linear or homogenous diffusion

Diffusion regularisation penalises the squared Euclidean norm of the deformation field and has been used widely in image registration. In a variational method, a global regularisation assumption can be incorporated using a combined energy term. A general form of a non-parametric registration approach, which estimates the transformation $\phi = \mathbf{Id} + \mathbf{u}$ (where $\mathbf{Id} : \mathbf{x} \rightarrow \mathbf{x}$ is the identity transform and \mathbf{u} a dense displacement field) between target image I_t and moving image I_m , can be formulated as the following optimisation problem:

$$\phi^* = \operatorname{argmin}_{\phi} E(\phi, I_t, I_m) \quad (4.1)$$

$$E(\phi) = \sum_{\mathbf{x} \in \Omega} \underbrace{\mathcal{S}(I_t(\mathbf{x}), I_m(\mathbf{x} + \mathbf{u}))}_{\text{dissimilarity term}} + \alpha \sum_{\mathbf{y} \in \mathcal{N}} \underbrace{\mathcal{R}(\mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{y}))}_{\text{regularisation term}} \quad (4.2)$$

where \mathcal{S} describes the image dissimilarity and \mathcal{R} is a penalty to ensure the regularity of the transformation, which depends on the deviation of \mathbf{u} within a neighbourhood \mathcal{N} .

[Horn and Schunck, 1981] introduced the use of the squared Euclidean norm of the gradient of the deformations as penalty function \mathcal{R} . For a d dimensional image the regularisation term, which is also known as *diffusion regularisation*, is defined as:

$$\mathcal{R}(\mathbf{u}(\mathbf{x})) = \sum_{i=1}^d |\nabla u_i(\mathbf{x})|^2 = |\nabla \mathbf{u}(\mathbf{x})|^2 \quad (4.3)$$

Another popular approach to diffusion regularised deformable registration is the demons framework [Thirion, 1998]. Here the optimisation problem is decoupled into two parts. First a force field is estimated aiming to minimise the sum of squared differences (SSD) of image intensities. The regularisation is then performed in a second step, where the force field is convolved with a Gaussian filter kernel G_σ yielding the deformation field $\mathbf{h} = G_{\text{sigma}} \star \mathbf{u}$. The Gaussian filtering can be seen as an approximation to the diffusion regularisation in Eq. 4.3. However, the decoupling of similarity computation and regularisation comes at the cost of slower convergence and generally requires more iterations than the variational approach discussed above.

4.1.2 Robust norms

To improve the performance of motion estimation at discontinuities, functionals $\rho(x)$ other than the squared error term in Eq. 4.3 have been proposed, which reduces the regularisation penalty for step changes in the motion field. The modulus (L1 norm) $\rho(x) = |x|$ would be an obvious choice, but because it is not differentiable at the origin it cannot be directly used for continuous optimisation. Two differentiable robust variants, namely the Charbonnier penalty $\rho(x) = \sqrt{x^2 + \epsilon^2}$ [Charbonnier et al., 1994] and the Lorentzian $\rho(x) = \log(1 + \frac{x^2}{2\sigma^2})$ are widely used

(c.f. [Scharstein and Szeliski, 1996]).

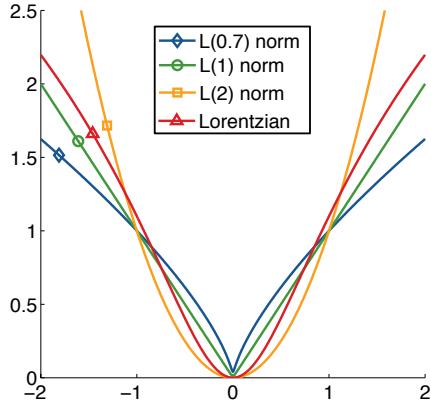
In [Heinrich et al., 2010b], we have proposed the use of the modified L^p -norm (or generalised Charbonnier penalty) $\rho(x) = (x^2 + \epsilon^2)^{\frac{p}{2}}$ for deformable lung registration, in order to preserve sliding motion. A comparison of the aforementioned robust penalties is shown in Fig. 4.2. Our experiments showed, that when using a norm with $p \leq 1$, a correctly estimated step change at the interface of lung and rib cage (see magnitude of the deformations in Fig. 4.2, bottom right). However, the remaining part of the motion field is not sufficiently smooth and unrealistic discontinuities are visible in other areas as well.

One other problem of robust norms compared to diffusion regularisation is that the penalty function is non-convex for $p \leq 1$ (or for the Lorentzian). Therefore the energy minimisation is hindered, a practical solution is to use a weighted average of squared and robust penalty with gradually increasing weight of the non-convex part (graduated non-convexity (GNC) [Blake and Zisserman, 1987]).

4.1.3 Direction-dependent diffusion regularisation

The robust norms discussed above, are entirely driven by the estimated deformations themselves. The basic intuition that motion discontinuities coincide with changes in appearance is not explicitly modelled, which might lead to unrealistic discontinuities (as seen in the example in Fig. 4.2). [Nagel and Enkelmann, 1986] proposed an anisotropic image-driven regularisation weighting and [Zimmer et al., 2011] used a joint image- and flow-driven approach.

[Schmidt-Richberg et al., 2012a] proposed an approach, which explicitly models sliding motion. Prior knowledge is used to determine the locations where discon-



Different penalty functions $\rho(x)$
with $\epsilon = 0.1$ and $\sigma = 0.5$



Coronal plane of CT lung scan
used in registration experiment



Magnitude of deformations using
quadratic (diffusion) penalty



Deformations for L^p norm
regulariser with $p = 0.7$

Figure 4.2: Influence of robust penalty functions on deformation fields using image data from [Murphy et al., 2011b]. Sliding motion at the interface between lungs at thoracic cage can be seen (see red box) using a regularisation norm with $p = 0.7$. However, the motion within the lungs is unrealistic.

tinuous motion occurs. An anisotropic (direction-dependent) smoothing is only used in close proximity to the sliding motion boundary. The diffusion regularisation of Eq. 4.3 is applied to a decoupled displacement vector field, in which the perpendicular \mathbf{u}^\perp and parallel \mathbf{u}^{\parallel} components of the motion (with respect to the motion boundary) are separated. To achieve a smooth motion of \mathbf{u}^\perp while preserving discontinuities in \mathbf{u}^{\parallel} along sliding surfaces, the image domain Ω is divided

into two parts: the inside and outside of the lungs $\Gamma \subset \Omega$.

Locations where discontinuities may occur are explicitly defined by an automatic or manual segmentation. To obtain a smooth solution of Eq. 4.3 for the decoupled vector field, a spatial weighting function $\omega(\mathbf{x})$ is introduced based on the distance $\delta(\mathbf{x})$ from the segmentation boundary. The direction-dependent regularisation (DDR) penalty \mathcal{R}^{DDR} is then defined as:

$$\begin{aligned}\mathcal{R}^{DDR} = & \sum_{i=1}^d \left(\int_{\Omega} \omega |\nabla u_i^\perp|^2 + (1 - \omega) |\nabla u_i^||^2 d\mathbf{x} \right. \\ & \left. + \int_{\Gamma} \omega |\nabla u_i^||^2 d\mathbf{x} + \int_{\Omega/\Gamma} (\omega |\nabla u_i^||^2) d\mathbf{x} \right)\end{aligned}\quad (4.4)$$

It has been shown in [Schmidt-Richberg et al., 2012a] that a direction-dependent regularisation approach yields higher registration accuracy than homogenous regularisation, in particular close to the sliding motion interface. Similar approaches have been subsequently used for elastic [Baluwala et al., 2013], demons [Risser et al., 2012b] and FFD registration models [Delmon et al., 2013].

4.1.4 Non-local regularisation

While the previously discussed approaches only consider the immediate neighbours (six in 3D) of a voxel to be part of \mathcal{N} for the regularisation energy, it may be beneficial to include a larger, non-local neighbourhood to explicitly enforce smoothness across larger objects. Current state-of-the-art optical flow estimation methods often use non-local filtering of the motion field (as a post-processing) to ensure both discontinuity preservation across object boundaries and smoothness within

an object [Sun et al., 2010]. A common approach to non-local edge-preserving filtering is the joint bilateral filter [Kopf et al., 2007]. The filtered motion $u_i^*(\mathbf{x})$ for dimension i of the vector field $\mathbf{u}(\mathbf{x}) = \sum_i^d u_i(\mathbf{x})$ is expressed as a weighted average within a large, non-local neighbourhood \mathcal{N} :

$$u_i^*(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{N}} w(\mathbf{x}, \mathbf{y}) u_i(\mathbf{y}) \quad (4.5)$$

In contrast to linear translation invariant (LTI) filters (such as Gaussian filters), the weights w are spatially varying and image-dependent. For the case of the bilateral filter the weights for an image I are given by:

$$w(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \exp \left(-\frac{|\mathbf{x} - \mathbf{y}|^2}{\sigma_s^2} \right) \exp \left(-\frac{|I(\mathbf{x}) - I(\mathbf{y})|^2}{\sigma_i^2} \right) \quad (4.6)$$

where N is a normalisation constant, σ_s a parameter to adjust the spatial weighting, and σ_i a constant to weight the intensity similarity (in relation to the image noise). Including the non-local weighting directly into the energy minimisation is more difficult, most importantly because of the high computational complexity [Krähenbühl and Koltun, 2012; Werlberger et al., 2010], which is linearly dependent on the size of \mathcal{N} . In [Heinrich et al., 2013b] (see Sec. 9.3.2), we have presented an alternative approach to non-local regularisation using multiple layers of super-voxels.

[Glocker et al., 2009] models non-local interactions by starting from a fully connected graph and learning the co-dependencies of control points with a clustering approach from training data. When applying these deformation priors for the registration of unseen images, an improvement over the conventional local reg-

ularisation was shown in particular for images with strong noise or artefacts. A very efficient approach for inference in fully-connected conditional random fields (CRF) has been recently proposed by [Krähenbühl and Koltun, 2011]. Their work uses a filtering approach to obtain a very good approximation to non-local regularisation, when restricting themselves to Gaussian pair-wise potentials. Improved multi-class segmentation results have been presented, which indicate the benefit of considering long-range connections.

4.2 Parameterisation of deformations

Instead of estimating a displacement vector for each voxel, non-rigid motion can also be parameterised with a lower dimensional model. The Free-form deformation (FFD) model [Lee et al., 1997; Rueckert et al., 1999], which is based on cubic B-spline basis function, is a popular choice, because it allows for a compact representation of complex motion using a much lower number of control points (which are uniformly spaced) than image voxels and intrinsically produces smooth and C^2 continuous transformations. In contrast to thin-plate splines [Bookstein, 1989], B-splines have a limited (local) support. The order of the employed B-spline basis determines the number of local neighbours of each control point, which affect the local transformation. Figure 4.3 shows the basis functions for B-splines of different orders. Given the displacements \mathbf{d} of all control points, the dense motion field $\mathbf{u}(\mathbf{x}) = (u, v, w)$ for a three-dimensional FFD at $\mathbf{x} = (x, y, z)$ can be found using:

$$\mathbf{u}(\mathbf{x}) = \sum_{i=0}^3 \sum_{j=0}^3 \sum_{k=0}^3 B_i(u) B_j(v) B_k(w) \mathbf{d}_{m+i,n+j,o+k}, \quad (4.7)$$

where $m = \lfloor x/\delta_x \rfloor - 1$, $n = \lfloor y/\delta_y \rfloor - 1$, $o = \lfloor z/\delta_z \rfloor - 1$, $u = x/\delta x - \lfloor x/\delta_x \rfloor$, $v = y/\delta y - \lfloor y/\delta_y \rfloor$, and $w = z/\delta z - \lfloor z/\delta_z \rfloor$. The l th basis function of the B-spline is denoted as B_l , the control points spacings by δ_x , δ_y and δ_z . The values for B_l are according to [Lee et al., 1997]: $B_0(u) = (1-u)^3/6$, $B_1(u) = (3u^3 - 6u^2 + 4)/6$, $B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6$ and $B_3(u) = u^3/6$. For an improved computational efficiency of B-spline interpolations (in order to map from image to control point domain) the coefficients can be pre-computed and stored in a lookup table.

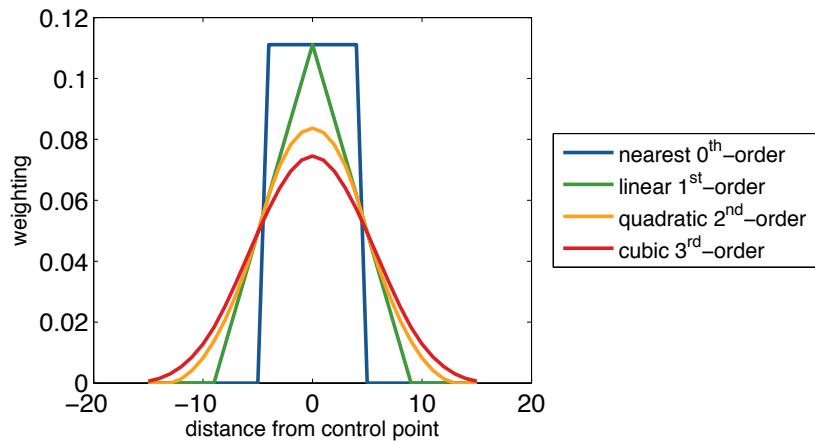


Figure 4.3: Different orders b of (1D) B-spline basis functions

Regularisation of B-spline transformations can be performed by penalising their bending energy [Rueckert et al., 1999]:

$$\mathcal{R} = \left(\frac{\partial^2 \mathbf{u}}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial z^2} \right)^2 + 2 \left(\left(\frac{\partial^2 \mathbf{u}}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial x \partial z} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial y \partial z} \right)^2 \right) \quad (4.8)$$

This is equivalent to a curvature regularisation and imposes no penalty for affine transformations. Penalising the squared Euclidean norm of the first order derivatives results in a diffusion regularisation similar to Eq. 4.3.

4.3 Image-adaptive regularisation using trees

The previously described approaches for regularisation of the complex lung motion, have primarily focussed on decoupling regularisation along sliding surfaces according to segmentation cues. Based on a parametric transformation model (as described above), we propose a more elegant and simpler concept. We use a global regularisation model (e.g. Eq. 4.3) with an image-adaptive weighting of the contribution of each of the neighbouring control points (or voxels) within \mathcal{N} . This idea, which has been widely used in computer vision [Boykov et al., 2001], encourages discontinuities only at locations with high intensity gradients without requiring an explicit segmentation. We propose to extract a minimum spanning tree (MST), based on intensity differences of neighbouring voxels. A weighting of 1 is given to any neighbour whose edge is part of the MST, otherwise it is set to 0.

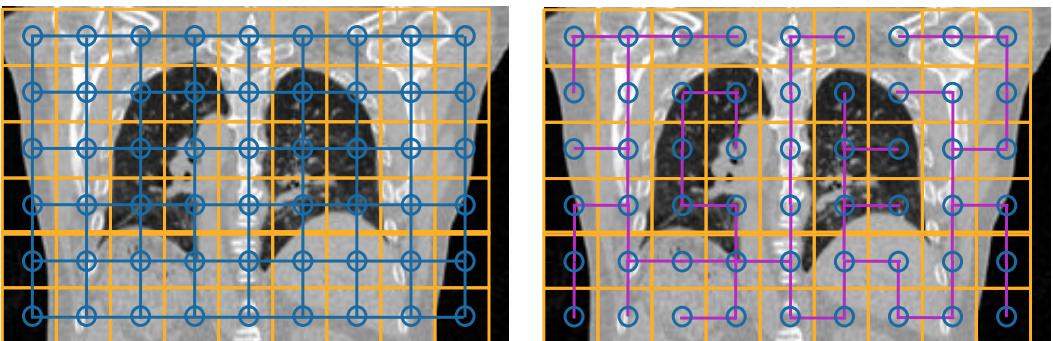


Figure 4.4: Concept of image-adaptive regularisation, demonstrated on coronal CT slice of chest. Left: Equally spaced control points $p \in \mathcal{P}$ are shown with blue circles. Influence regions η are limited by yellow lines. Edges in four-connected graph are shown with blue lines. Right: Edges in minimum-spanning-tree (edge weights are based on SAD of intensities in η) are shown with purple lines. There are no edges and thus no regularisation across the sliding boundary between lungs and rib cage. Note that in practice the spacing of control points is much smaller.

We adapt a parametric registration approach, because it forms an additional in-

trinsic regularisation and is computationally more efficient. A number n of equally spaced control points $p \in \mathcal{P}$ are defined with a grid-spacing of g forming a regular grid with 6-connectivity (see Fig. 7.6 left). Each control point can be seen as a vertex in an undirected graph, where its six neighbours $q \in \mathcal{N}_p$ are connected by potential edges e with corresponding edge weights $w(p, q)$. In our approach, we define the edge weights to be the sum of absolute differences (SAD) between the intensities $I(\mathbf{x})$ of all voxels within the influence region $\eta = \{-g/2+1, -g/2+2, \dots, g/2\}^3$ around a control point p (centred at \mathbf{x}_p) and the respective voxels for a neighbouring control point q (centred at \mathbf{x}_q):

$$w(p, q) = \sum_{\Delta\mathbf{x} \in \eta} |I(\mathbf{x}_p + \Delta\mathbf{x}) - I(\mathbf{x}_q + \Delta\mathbf{x})| \quad (4.9)$$

To find the spanning-tree, which connects all the vertices (without loops) with minimal total edge weight, we employ the greedy algorithm by [Prim, 1957] which has $\mathcal{O} = n \log(n)$ complexity. Any vertex can be chosen as the root node (this has no effect on the MST formation). Starting from the root node, the tree is grown by repeating the following four steps at each new vertex q^* in the tree (see Fig. 4.5):

1. add new candidate edges ($e \subset \mathcal{N}_{q^*}$) to priority queue
2. remove edges connecting two vertices, which are already both part of the MST (i.e. forming a loop), from queue
3. select the candidate edge with lowest weight from priority queue
4. add the newly connected vertex to the tree

Using a priority queue (a heap data-structure) to keep all candidate edges in order is computationally more efficient than using an adjacency matrix.

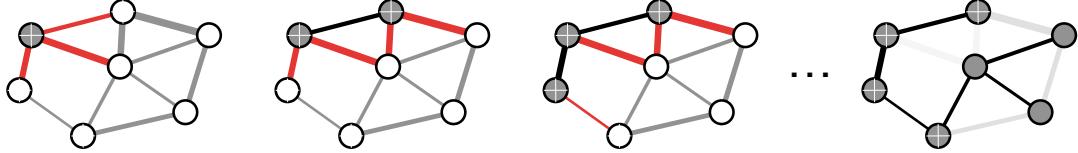


Figure 4.5: Example of Prim’s algorithm to find the minimum-spanning-tree (MST). Vertices, which are already part of the MST are shaded in grey (others in white). The candidate edges, which are sorted in a priority queue are shown in red. The edge weight is depicted by the line width. An edge, which has been selected as part of the MST is drawn in black. The images from left to right show the evaluation of the tree until completion.

The resulting tree is well balanced, and, as a consequence, the maximum width is approximately $|\mathcal{P}|/\log|\mathcal{P}|$. The output of Prim’s algorithm consists of a sorted list of all nodes (with increasing tree depth) and the index of each node’s parent. A similar approach has been used as image representation for finding stereo correspondence by [Veksler, 2005], however in that case each node only represented a single pixel. This makes the approach almost unfeasible for 3D image registration. [Lei and Yang, 2009] used an MST in conjunction with a mean-shift over-segmentation for optical flow estimation. An example of an image-derived MST is shown in Fig. 7.6 (right). It is visible that areas of large image gradients (where sliding motion could be expected) are not connected, and therefore less regularisation will be applied between these nodes.

4.4 Symmetric and diffeomorphic transforms

As discussed in Sec. 3.2.2, it is desirable for a number of applications to impose certain restrictions on the estimated deformations. Diffeomorphic transformations avoid singularities in the deformation field, are invertible and ensure a one-to-one mapping. Symmetric registration approaches reduce the inverse consistency error (ICE, see Eq. 3.1 and [Christensen and Johnson, 2001]) and remove the bias, which of two scans is chosen as the target image. There is a large number of approaches, which use either soft constraints (as part of the registration cost) or hard constraints to ensure a certain deformation quality. We present a novel approach, in Sec. 4.4.1, to obtain symmetric and diffeomorphic transformations for almost any registration using an efficient modular approach. Our approach improves the flexibility and accuracy of symmetric-diffeomorphic approaches for the large deformations that are common for lung motion.

[Rueckert et al., 2006] demonstrates the use of a hard constraint for B-spline transformations, which can be used to guarantee diffeomorphic deformations, if the maximum displacement of each control point is limited to ≈ 0.4 times the grid spacing. This constraint was also later used by [Glocker et al., 2008a], we however found that it is too restrictive for deformable lung registration, when sliding motion is present. [Sotiras and Paragios, 2012] enforced symmetric transformations in a discrete optimisation framework by defining the transformation grid (in a common intermediate space between both images) and restricting the forward displacements to be the opposite (negative) of backward displacements. However, this constraint can only be easily enforced in a discrete optimisation framework and requires the minimum displacement quantisation (without interpolation) to be 2 voxels.

[Christensen and Johnson, 2001] presented a penalty term (soft constraint) to avoid inverse inconsistency. [Avants et al., 2008] expanded on this idea and added a symmetry constraint into the registration cost function. In [Rohlfing et al., 2003] introduced an incompressibility constraint based on the Jacobian of the deformation, which has been used in lung registration [Rühaak et al., 2011] to avoid singularities. Another popular approach to obtain diffeomorphic transformations is the scaling and squaring method [Arsigny et al., 2006]. It is based on the fact that the composition of two diffeomorphic transformations will again yield a diffeomorphic transformation. If the maximal displacement for every voxel is less than 0.5 voxels, any deformation field is diffeomorphic. Therefore the initial deformation \mathbf{u} is divided by the power of two of a number N , such that

$$\mathbf{u}^0 = 2^{-N} \cdot \mathbf{u}, \|\mathbf{u}^0\| < 0.5$$

This is called *scaling*. The diffeomorphic transformation $\hat{\mathbf{u}}$ is then obtained by N -times composing \mathbf{u}^0 with itself.

$$\mathbf{u}^n = \mathbf{u}^{n-1} \circ \mathbf{u}^{n-1} \text{ for } n = \{1, 2, \dots, N\} \quad (4.10)$$

This part, called *squaring*, yields $\hat{\mathbf{u}} = \mathbf{u}^N$ after N steps.

4.4.1 Novel approach to symmetric and inverse-consistent transformations

The disadvantages of the aforementioned approaches is that they are either too restrictive for large lung motion or not applicable independently of the employed

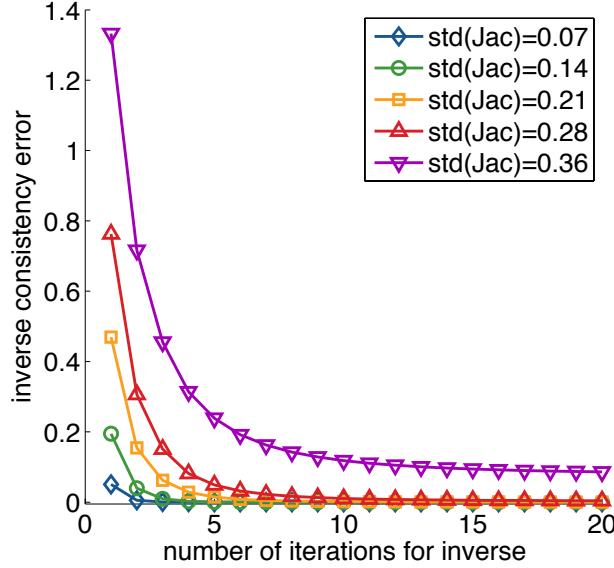


Figure 4.6: Inverse consistency error (ICE) in voxels of the described iterative inversion method, applied to random displacement fields of different complexity (without singularities, and $\mu(Jac) \approx 1$). For reasonably smooth fields (with $std(Jac) < 0.3$) the error quickly converges to 0, only for the inversion of very strong deformations a residual error remains.

transformation or optimisation technique. In contrast, our approach can be used as a modular processing step in any registration method. We first estimate both full transformations ϕ_F and ϕ_B independently. In order to calculate a valid inverse transformation the transformations have to be diffeomorphic. To avoid the restrictions of a hard constraint, arbitrarily large displacements are allowed and a diffeomorphic mapping is afterwards obtained by applying the scaling and squaring method. We then use a fast iterative inversion method, as presented in [Chen et al., 2007], to invert the half-length transformations $\phi_F(0.5)^{-1}$ and $\phi_B(0.5)^{-1}$. Let \mathbf{u} denote a given displacement field, e.g. $\phi_F(0.5)$. The displacement field of its inverse transform \mathbf{u}^{-1} will be estimated iteratively, starting from $\mathbf{u}_0^{-1} = 0$, and

updating:

$$\mathbf{u}_n^{-1}(\mathbf{x}) = -\mathbf{u}(\mathbf{x} + \mathbf{u}_{n-1}^{-1}(\mathbf{x})) \quad (4.11)$$

At convergence \mathbf{u}_n^{-1} approaches the true inverse \mathbf{u}^{-1} . Equation 4.11 can only yield the inverse if the Jacobian of \mathbf{u} is always positive. For a small number of non-positive Jacobians this procedure will yield an approximate inverse. Fig. 4.6 shows the inverse consistency error $\text{ICE} = \|\mathbf{u}(\mathbf{x}) + \mathbf{u}^{-1}(\mathbf{x} + \mathbf{u})\|^2$ obtained when the above inversion method is applied to random displacement fields (using FFDs) with increasing complexity. It can be seen that 5-10 iterations are usually sufficient to obtain a very low residual ICE.

After obtaining the inverses, the two symmetric transformations are calculated by composition: $\phi_{FS} = \phi_F(0.5) \circ \phi_B(0.5)^{-1}$ and $\phi_{BS} = \phi_B(0.5) \circ \phi_F(0.5)^{-1}$, which ensures that these transformations are inverse consistent. We use this symmetric approach in all deformable registration experiments and demonstrate its advantages in Secs. 7.4 and 7.6 for continuous and discrete optimisation, respectively.

4.5 Summary

In this chapter, different approaches for regularisation and parameterisation of deformations were presented, which aim at modelling lung motion. The techniques can be roughly divided into three groups depending on their motion model: piecewise constant, piece-wise smooth and globally smooth. Robust norms, such as the L_p -norm (with $p \leq 1$), the Lorentzian and the Pott's model [Boykov et al., 2001], fall in the first category and are widely used to estimate stereo correspondences and optical flow, where strong discontinuities, occlusions and large areas of constant

displacements are common. We have applied robust norms to lung registration and found the results unconvincing, because too many unrealistic motion discontinuities were found. Globally smooth models include: parametric spline transformations, diffusion regularisation and isotropic Gaussian smoothing (used in the demons framework). They are most widely used in medical image registration, because the majority of deformable organ motion is smooth. The second category, piece-wise smooth motion, is the most applicable one for lung motion. Direction-dependent regularisation has become very popular recently, however most of them rely on accurate manual segmentations of the lungs and are not easily extendable to the sliding motion of multiple organs. We presented a new approach to model piece-wise smooth motion for parametric graph-based transformations using image-adaptive trees. This approach has also great benefits for discrete optimisation, which will be discussed in Sec. 7.5. Non-local regularisation is another relatively new way of modelling piece-wise smooth motion, which will be discussed as an outlook in Sec. 9.3.2. A new approach using layers of supervoxels has been very recently published in [Heinrich et al., 2013b]. Additionally, a novel flexible approach to obtain invertible and symmetric transformations (which are physically more plausible) is presented that overcomes some of the limitations of previous techniques, which used hard constraints. The evaluation of our contributions will be presented in Chapter 7. The next chapters will describe an equally important aspect of deformable lung registration, the definition of image similarity across scans. Using a robust and discriminative similarity measure enables more accurate correspondences and can therefore further aid in modelling lung motion.

Chapter 5

Spatial context for statistical similarity metrics

The objective of this chapter is to improve both the robustness and accuracy of statistical image similarity metrics for challenging registration tasks. We address this problem by introducing the use of spatial contextual information into the calculation of mutual information. Our new metric, textural mutual information, incorporates spatial context by using all image intensity values within a small patch rather than single intensities only. The dimensionality of this extremely large space of potential image patches is reduced using cluster trees, which enables an efficient estimation of the statistical distribution of textures.

The alignment of multi-modal images helps to relate clinically relevant information that is often complementary across modalities. As discussed in Sec. 2.3, multi-modal registration can be used, amongst other things, to fuse scans for image-guided interventions, diagnostic tasks and the monitoring of disease treatment. Here, the registration of CT and MRI is useful, as it can combine the good

spatial resolution and dense tissue contrast of a CT with the better soft tissue contrast of MRI. In image-guided radiotherapy, multi-modal registration can be applied to compensate for a different patient positioning between a pre-treatment (planning) scan and the intra-operative scan to achieve a more accurate delivery of radiation dose. Figure 5.1 shows two example slices of an MRI and a CT scan.

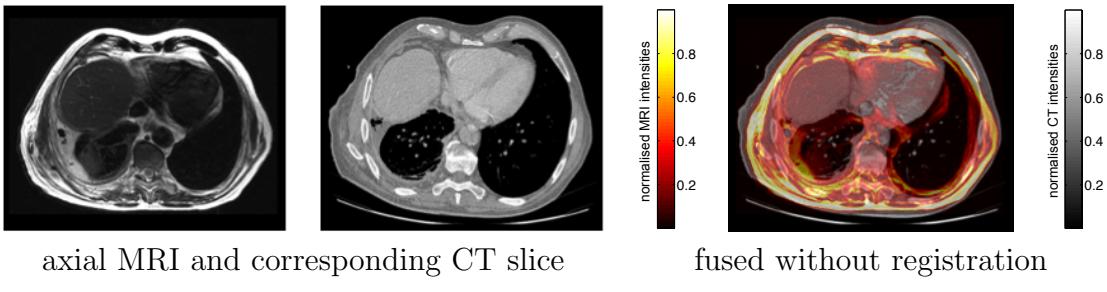


Figure 5.1: An example of two corresponding slices of CT and MRI scans from the empyema dataset. The overlay (fusion) of both modalities without registration demonstrates the misalignment and need for multi-modal registration.

The similarity between images is the main driving force for registration. There are a number of challenges for particular applications, which have to be addressed in order to design a suitable similarity metric. First of all, a metric needs to be robust against noise or intensity distortions, for example MRI bias fields. In CT lung imaging, a particular challenge is the locally changing intensity of the lung parenchyma due to compression or expansion during respiration. Motion compensation for images from dynamic sequences, such as dynamic contrast-enhanced (dce) MRI or perfusion CT, require the similarity metric to be invariant to local change in contrast, i.e. a functional intensity mapping. The relationship between the intensities of images taken using different physical phenomena (e.g. computed tomography (CT) and magnetic resonance imaging (MRI) as discussed in Sec. 2.1) cannot be expressed, in general, in a functional way (see e.g. Fig. 5.1). Therefore

statistical similarity metrics (or structural image representations) have been used for the task of deformable multi-modal image registration [Andronache et al., 2008; Loeckx et al., 2007; Rogelj et al., 2003].

Most image similarity metrics are intensity-based (iconic similarity). The other group of approaches, geometric (or feature-based) methods, use only a sparse set of salient points in each image and try to extract very rich descriptors, which can be matched robustly across images. Examples of descriptors for feature-based approaches are the *scale invariant feature transform* (SIFT) [Lowe, 1999], *gradient location and orientation histograms* (GLOH) [Mikolajczyk and Schmid, 2005] and *geometric moments invariants* (GMI) [Shen and Davatzikos, 2002]. Spatial context is usually employed to achieve high discrimination and better matching. These descriptors may be invariant to some extent to changes of intensity (or illumination), since they rely on image gradients, local orientations or tissue classifications, but they have not been successfully applied to medical multi-modal registration, where the intensity variations across modalities are more severe. One exception is the use of Gabor wavelets by [Ou et al., 2011], where a particular subset of features is selected by training for a specific registration task. Another disadvantage of feature-based registration is that a dense deformation field can only be obtained by using an interpolating transformation (e.g. thin-plate splines) between the sparse point matches. This makes the resulting deformations highly dependent on the characteristics of the underlying parametric transformation model and can be problematic for many applications, e.g. when computing statistics on deformations in population studies.

The key idea of the work presented in this and the next chapter is to incorporate spatial context in a dense (voxel-wise) manner into the calculation of image

similarity. Dense image similarity, which is used in iconic registration approaches, yields high accuracy and globally valid deformation fields. Spatial context, which has so far been mainly used in geometric registration, improves the robustness of correspondences across scans and is more discriminative for different image features than simple image intensities alone.

Sec. 5.1 discusses different approaches to statistical (or information theoretic) similarity metrics. The most commonly used metric, mutual information (MI), and its point wise and hierarchical estimation are described in Sec. 5.2. Previous work on incorporating a spatial variable into the computation of MI, namely conditional mutual information [Loeckx et al., 2007], and its advantages in the presence of non-uniform bias fields is discussed.

In Sec. 5.3 we derive a novel formulation of mutual information using *textural mutual information* (TMI), which incorporates small image patches and increases its robustness by making use of this spatial context. Our approach effectively performs a dimensionality reduction of the large space of image patches using cluster trees and generalises many recently proposed extensions to MI, e.g. regional, conditional and higher-order, adaptive bin width and normalisation.

Section 5.4 presents experimental results for all discussed similarity metrics using the landmark localisation experiment described in Sec. 3.3. First, the sensitivity of localisation accuracy with respect to the choice of parameters is studied. Then, the robustness of the metrics against common intensity distortions in medical imaging (additive Gaussian noise and non-uniform multiplicative bias field) is evaluated.

Figure 5.2 gives a visual categorisation of the presented methods in terms of their underlying principles driving similarity: statistical, structural and context-

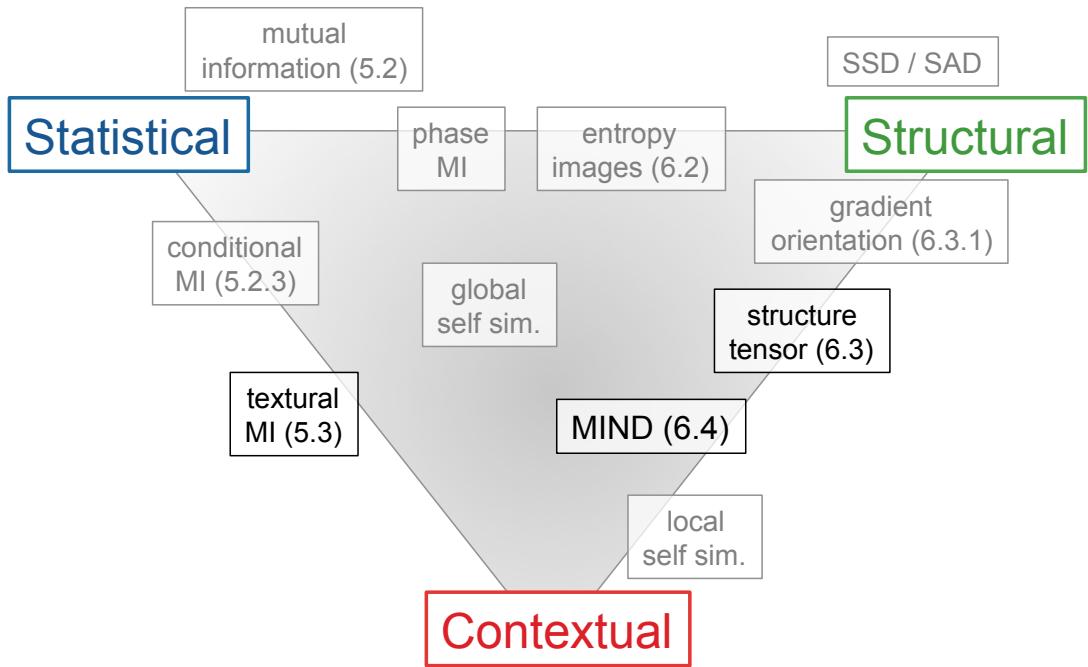


Figure 5.2: Categorisation of popular approaches for image similarity metrics. The three corners represent the underlying principles used to derive image similarity (statistical, structural and contextual information). Methods are positioned according to what extent they rely on either of these principles. New approaches proposed in this work are set in black font.

tual information. In the next chapter the use of structural image representations are discussed, including two novel contributions by us (see Secs. 6.3 and 6.4): the *structure tensor orientation measure* and the *modality independent neighbourhood descriptor* (MIND). Our contributions towards the field of similarity metric all make use of additional contextual information to improve robustness (textural mutual information for statistical metrics and MIND for structural image representations). As discussed above, the use of contextual information is closely related to feature-based methods. Our contributions therefore help bridging the gap between geometric and iconic image registration.

5.1 Introduction to statistical similarity metrics

For monotonic intensity transformations, many approaches model a linear intensity relationship. The normalised cross-correlation (NCC) or local cross-correlation (LCC) [Avants et al., 2008] are examples of such metrics, which allow for an offset (change of mean intensity) and gain (change of contrast) across image acquisition.

The NCC between target image I_t and moving image I_m , is defined as:

$$\text{NCC} = \frac{\sum_{\mathbf{x} \in \Omega} (I_t(\mathbf{x}) - \bar{I}_t)(I_m(\mathbf{x}) - \bar{I}_m)}{\sqrt{\sum_{\mathbf{x} \in \Omega} (I_t(\mathbf{x}) - \bar{I}_t)^2} \sqrt{\sum_{\mathbf{x} \in \Omega} (I_m(\mathbf{x}) - \bar{I}_m)^2}} \quad (5.1)$$

where Ω is the image domain, and \bar{I}_t and \bar{I}_m are the means of I_t and I_m respectively.

The correlation ratio [Roche et al., 1998] generalises the cross-correlation from linear to arbitrary functional relations. The dissimilarity between images I_t and I_m with intensities $i \in \mathcal{I}$ at a voxel \mathbf{x} based on the correlation ratio $\eta(I_t|I_m)$ can be written as:

$$\text{CR}(\mathbf{x}) = 1 - \eta(I_t(\mathbf{x})|I_m(\mathbf{x})) = \frac{1}{\sigma_{I_t}^2} \sum_{i \in \mathcal{I}} p_i(I_t(\mathbf{x})) \sigma_{ic}^2 \quad (5.2)$$

where $\sigma_{I_t}^2$ is the variance of I_t , $p_i(I_m) = \sum_i \delta[I_m(\mathbf{x}) - i]$ is the marginal distribution of I_m and σ_{ic}^2 is the conditional variance of I_t and I_m :

$$\sigma_{ic} = \frac{1}{p_i(I_m)} \sum_{\mathbf{x}} (I_t(\mathbf{x}) - \bar{I}_t) \delta[I_m(\mathbf{x}) - i] \quad (5.3)$$

The Kronecker delta ($\delta[I_m(\mathbf{x}) - i]$) is often replaced by a Parzen window estimator

[Viola and Wells III, 1997] with standard deviation σ_p :

$$\frac{1}{\sigma_p \sqrt{2\pi}} \exp \left(-\frac{(I_m(\mathbf{x}) - i)^2}{2\sigma_p^2} \right) \quad (5.4)$$

Figure 5.3 shows a comparison of these two statistical metrics alongside with mutual information. It can be seen that for the non-functional intensity mapping (which is common e.g. between CT and MRI scans) in this example cannot be adequately modelled by the cross-correlation or the correlation ratio.

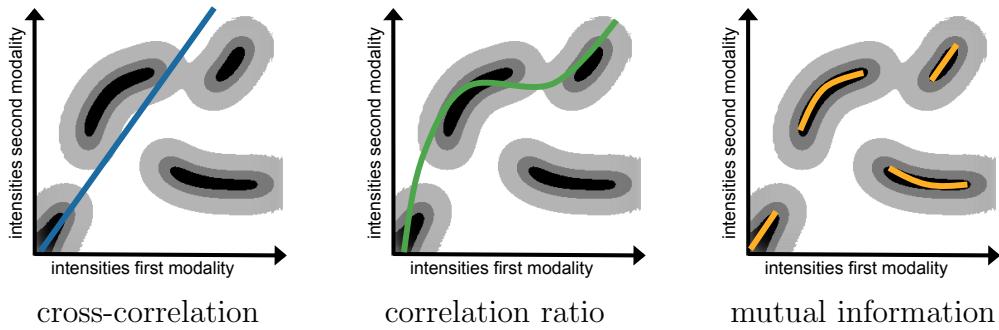


Figure 5.3: Visual comparison of statistical similarity metrics (adapted from [Hermosillo et al., 2002] Fig. 1). The joint distribution for an exemplary non-functional intensity mapping across two modalities is shown (high and low probabilities are indicated by black and white respectively). Cross-correlation assumes a linear relationship between intensities, and the correlation ratio assumes a functional relationship. They can therefore not sufficiently model this complex intensity mapping. Registration methods based on MI tend to drive the joint distribution towards cases with strong clusters (marked in yellow).

5.2 Mutual information

Mutual information (MI) is derived from information theory and measures the statistical dependency of two random variables. It was first introduced to medical image registration for the rigid alignment of multi-modal scans [Maes et al., 1997;

Viola and Wells III, 1997]], and later used successfully in a variety of applications, including deformable registration [Rueckert et al., 1999; Meyer et al., 1997]. It is based on the assumption that a lower entropy of the joint intensity distribution corresponds to a better alignment. Given the joint probability $p(I_t(\mathbf{x}), I_m(\mathbf{x}))$ of a voxel \mathbf{x} with intensity $I_t(\mathbf{x})$ in the target image and $I_m(\mathbf{x})$ in the moving image and its two marginal intensity probabilities $p(I_t(\mathbf{x}))$ and $p(I_m(\mathbf{x}))$, mutual information is defined as the difference between the sum of marginal entropies $H(I_t)$, $H(I_m)$ and the joint entropy $H(I_t, I_m)$:

$$\begin{aligned} MI(I_t, I_m) &= H(I_t) + H(I_m) - H(I_t, I_m) \\ &= -\sum_{\mathbf{x} \in \Omega} p(I_t(\mathbf{x}), I_m(\mathbf{x})) \log \frac{p(I_t(\mathbf{x}), I_m(\mathbf{x}))}{p(I_t(\mathbf{x}))p(I_m(\mathbf{x}))} \end{aligned} \quad (5.5)$$

where Ω defines the image overlap domain.

[Studholme et al., 1999] introduced normalised mutual information (NMI), which divides MI by the marginal entropy of the target image, in order to cope with the effect of changing image overlap on MI.

An example of the joint distributions before and after maximisation of NMI, using a pair of corresponding CT and MRI slices (see Fig. 5.1) is shown in Fig. 5.4. The joint distribution after alignment is less dispersive and exhibits one fewer peak and therefore has a higher NMI.

5.2.1 Pointwise normalised mutual information

In [Hermosillo et al., 2002] and [Rogelj et al., 2003], variants of mutual information to obtain a point-wise similarity metric have been proposed. For the implementation of NMI as a comparison method, the approach of [Rogelj et al., 2003] is

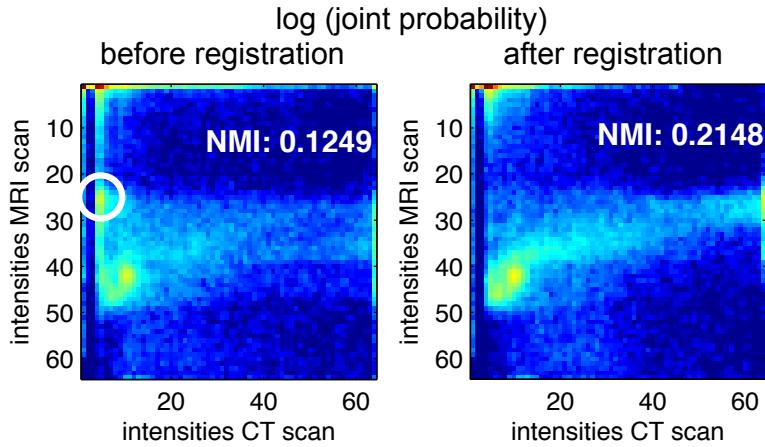


Figure 5.4: Joint intensity distributions of MRI and CT images, shown in Fig. 5.1, demonstrating the concept of mutual information. The non-aligned image pair shows one more peak (which results from a mixture of tissue classes, see white circle), which is not visible in the registered image pair. The normalised mutual information (NMI) is therefore increased.

used in this work. The joint and marginal histograms of two images I_t and I_m are obtained in a conventional manner. A Parzen window kernel is used to improve the robustness of the joint histogram estimation. The local contribution $\text{NMI}(\mathbf{x})$ for each voxel can then be obtained using:

$$\text{NMI}(\mathbf{x}) = -\log \left(\frac{p(I_t(\mathbf{x}), I_m(\mathbf{x} + \mathbf{u}(\mathbf{x})))}{p(I_t(\mathbf{x}))p(I_m(\mathbf{x} + \mathbf{u}(\mathbf{x})))} + \epsilon \right) \frac{1}{\sum_{\Omega} p(I_t(\mathbf{x})) \log p(I_t(\mathbf{x}))} \quad (5.6)$$

The second term is as a constant, because it is dependent on the (fixed) target image only. It is assumed that the images are (almost) in alignment and therefore $\mathbf{u}(\mathbf{x}) \approx \mathbf{0}$. Equation 5.6 can be efficiently represented as a look-up table to speed-up each point-wise similarity evaluation.

Another simplification has been presented by [Kim et al., 2003] for robustly finding stereo correspondences under (artificially) changing illumination. Using a Taylor series expansion, they find the following simplified equation for each point-

wise MI contribution:

$$\text{MI}(\mathbf{x}) = -\frac{1}{n} \log(p(I_t(\mathbf{x}), I_m(\mathbf{x} + \mathbf{u}(\mathbf{x}))) * \sigma_p + \epsilon) \quad (5.7)$$

where n is the number of pixels and σ_p the Parzen smoothing kernel. This formulation ignores the marginal distributions. Thus for images with highly non-uniform intensity distributions (e.g. with many background voxels) the MI calculation will be adversely affected. We therefore employ Eq. 5.6 for our experiments. As noted by [Hirschmüller and Scharstein, 2009] a small constant ϵ has to be added within the logarithm of Eqs. 5.6 and 5.7 to avoid $\log(0)$, and we empirically set $\epsilon = 0.001$.

For variational registration methods the derivative of the cost function with respect to the transformation parameters needs to be evaluated. [Klein et al., 2005] compared three different approaches to obtain these derivatives for MI: finite differences (FD) [Kiefer and Wolfowitz, 1952], simultaneous perturbation (SP) [Spall, 1992] and the analytic solution (AS) for parametric transformations [Thévenaz and Unser, 2000]. In this work, the best results are obtained with FD and AS: for non-parametric registration methods the FD approach is most suitable. The gradient of NMI for an arbitrary dimension x of the deformation field is obtained by:

$$\frac{\delta \text{NMI}}{\delta x} \approx \frac{\text{NMI}(\mathbf{x} + \Delta x) - \text{NMI}(\mathbf{x} - \Delta x)}{2\Delta x} \quad (5.8)$$

where Δx is a small spatial step (usually $|\Delta x| = 1$ voxel) in direction of x .

5.2.2 Hierarchical mutual information

Another challenge for the use of MI for deformable registration is the fact that the joint intensity distributions depend on the alignment of the images, and therefore larger deformations can make the initial estimates unreliable. [Kim et al., 2003] address this problem by alternating between updating the displacement fields and recalculating the joint histogram and thereby increasing the number of iterations needed for the registration. [Hirschmüller and Scharstein, 2009] improve on this idea by using a hierarchical MI estimation, in which the displacement vectors from a lower resolution are used to initialise the MI calculation for the finest resolution.

In our Gauss-Newton registration framework this is implicitly included by using a coarse-to-fine strategy. For the landmark localisation experiments, which are used in this chapter, a dense displacement field is first obtained for volumes of half the original resolution using the sliding window approach discussed in Sec. 3.3.1 (integral images, c.f. [Tapia, 2011], are used for an efficient implementation of the window aggregation). The obtained displacements $\mathbf{u}(\mathbf{x})$ are then used to repeat the calculation of the joint intensity distribution in Eq. 5.6, where corresponding locations are now \mathbf{x} in image I_t and $\mathbf{x} + \mathbf{u}(\mathbf{x})$ in image I_m .

Alternatively, a local joint histogram estimation could be used, which however would limit the number of samples and would require more sophisticated histogram strategies like NP-windows [Dowson et al., 2008], which are computationally extremely demanding for 3D volumes. A more efficient computation for this technique was recently presented by [Joshi et al., 2011].

5.2.3 Conditional mutual information

A number of disadvantages when using MI for deformable multi-modal registration have been analysed by [Loeckx et al., 2007], [Haber and Modersitzki, 2007], and [Studholme et al., 2006]. They have often been attributed to the fact that MI ignores the spatial neighbourhood of a particular voxel within one image and consequently it does not use the context of spatial information shared across images. In the presence of image intensity distortions, such as non-stationary bias fields in MRI scans, this can deteriorate the quality of the alignment, especially in the case of non-rigid registration where the geometric constraints of the transformation are relaxed compared to global alignment.

One approach to overcome this problem is to include spatial information into the joint and marginal histogram computation. In [Rueckert et al., 2000] a second-order mutual information measure is defined, which extends the joint entropy estimation to the spatial neighbours of a voxel and therefore uses a 4D histogram, where the third and forth dimensions define the probability of the spatial neighbours of a voxel having a certain intensity pair. A problem that arises here is the curse of dimensionality, meaning that a lot of samples are needed to populate the higher-dimensional histogram. The authors therefore limit the number of intensity bins to 16, which might again decrease the accuracy.

[Studholme et al., 2006] introduce a third channel into the joint histogram, containing a spatial or regional label. A similar approach called conditional mutual information (CMI), has been introduced by [Loeckx et al., 2007]. A third dimension is added to the joint histogram and a second dimension is added to the marginals representing the regional location of an intensity pair. The image is subdivided

into a number of overlapping regions and each intensity pair only contributes to its specific regional histograms. A number of anchor points are evenly distributed on the image grid. Each voxel in a 3D volume is then attributed to its 8 nearest anchor points, and its contribution to this regional label $r(\mathbf{x})$ is weighted by the reciprocal spatial distance between voxel and anchor point. CMI is then defined as:

$$\text{CMI}(\mathbf{x}) = -\sum_{\mathbf{x} \in \Omega} w(r(\mathbf{x})) \log \left(\frac{p(I_t(\mathbf{x}), I_m(\mathbf{x}))}{p(I_t(\mathbf{x}))p(I_m(\mathbf{x}))} \right) \quad (5.9)$$

In [Loeckx et al., 2007] it was shown that this reduces the negative influence of bias fields and yields a higher registration accuracy for a small number of realistic test cases. The drawbacks lie again in the difficulty of populating this 3D histogram, and in the fact that corresponding anatomical structures, which are spatially further apart, are not taken into account.

5.3 Textural mutual information

To address the previously discussed problems related to using mutual information for deformable multi-modal registration, we have introduced a novel way of incorporating spatial context in [Heinrich et al., 2012a]. *Textural mutual information* (TMI) is a new image similarity metric, which efficiently incorporates intensity information from local neighbourhoods into the estimation of the joint histogram. This is achieved by representing each voxel location by the most representative image patch, where the dimensionality of the patch space is reduced using a cluster tree and/or texton dictionary.

In [Russakoff et al., 2004] a regional implementation of MI based on small im-

age patches has been introduced. The authors make the simplifying assumption that the high-dimensional data are normally distributed and transform it into a space where they are uncorrelated. This approach demonstrates increased robustness, but is limited to rigid registration. In [Knops et al., 2006] non-equidistant histogram binning was introduced using k-means clustering.

[Yi and Soatto, 2011], proposed, independently of us, a method which also uses image patches to introduce spatial context into MI. They organise image patches into orbits under the action of Euclidean transformations. However, their approach is computationally complex and so far limited to rigid 3D or deformable 2D registrations.

For the following derivation and implementation of TMI, Fig. 5.5 provides an overview of the steps. Some extensions, which have been proposed for mutual information, such as optimal bin width selection [Knops et al., 2006], higher-order MI [Rueckert et al., 2000], and conditional MI [Loeckx et al., 2007] (see Sec. 5.2.3) can be generalised by the concept of textural mutual information.

5.3.1 Textons

Mutual information, as a multimodal similarity measure, is based on the assumption that voxels of corresponding anatomical structures are represented by a common intensity pair. However, due to degradations of medical images this is not fulfilled in real life. Imaging-related artefacts can cause a complex intensity distribution within the same tissue. Textural mutual information is motivated by the fact that although single intensities might provide only a limited representation of the underlying anatomical structure, a patch including several neighbouring vox-

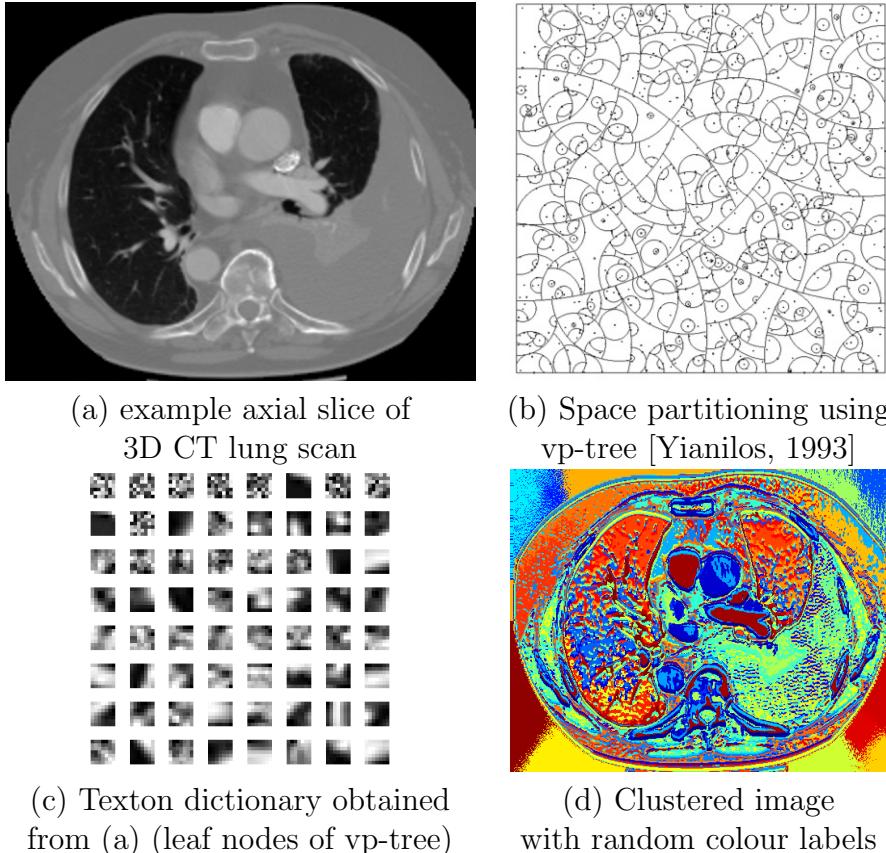


Figure 5.5: Overview of our method (shown on an example CT slice (a)). A representative texton dictionary (c) is learnt from the image using hierarchical tree clustering (c). Each pixel is then assigned to the closest texton using nearest neighbour search (pixels with the same colour belong to the same cluster (d)).

els can more effectively capture the texture and thus better describe the actual situation.

Representing the images directly by their local patches would increase the dimensionality of their joint histogram. This is because the number of samples is too low to populate such a high-dimensional histogram. Therefore, a lower dimensional representation of the space of possible patches or texture representations needs to be found.

Textural representation and classification has previously mainly employed the

responses of high-level filter banks (e.g. Gabor filters [Ou et al., 2011]). A filter bank approach has the advantages of representing a large support area while still having a low-dimensional feature vector. However, the choice of filters is crucial and has to be learnt for each specific image. In [Varma and Zisserman, 2003], it is shown that an excellent representation of texture can be directly obtained by using small image patches (as small as 3×3). In their application of texture classification, so called textons are learnt based on segmented supervised training data and clustered forming a texton dictionary. Another related approach was used in [Deselaers and Ferrari, 2010] to define global self-similarity of object detection. In the next section, we present our approach, which employs hierarchical tree clustering to obtain a representative texton dictionary. In [Brox et al., 2008b] a cluster tree is used to efficiently detect similar textures for image denoising. Because the number of possible classes is relatively low in their work, a (non-hierarchical) k-means clustering is employed. The principle of the clustering of similar patches is illustrated in Figure 5.5.

5.3.2 Cluster trees

Finding similar patches using a tree structure has been extensively studied and can be usually performed with $\mathcal{O}(n \log n)$ complexity. We use the vantage-point tree, which was introduced by Yianilos [Yianilos, 1993] and achieved the best results in a recent comparison of clustering methods [Kumar et al., 2008] in terms of computational complexity for both the clustering and the retrieval of image patches. First, a pivot is chosen (an optimised pivot is selected using the element that results in the largest spread for a random subset) and the distances to all

remaining elements are calculated. The elements are then split into two equal-sized branches based on the median distance. These steps are repeated recursively until a fixed tree depth T_D is reached. The number of leaf nodes is equivalent to the bins b in the histogram and defined by $b = 2^{T_D+1}$.

The distance measure $D_p(\mathbf{x}_1, \mathbf{x}_2)$ between two voxels \mathbf{x}_1 and \mathbf{x}_2 within the same image I is chosen to be the sum of squared differences (SSD) of all voxels between the two patches P of size $(2p + 1)^d$ (with image dimension d) centred at \mathbf{x}_1 and \mathbf{x}_2 . The spatial Euclidean distance $\|\cdot\|^2$ can be added using a weighting term λ :

$$D_p(I, \mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{\mathbf{p} \in P} (I(\mathbf{x}_1 + \mathbf{p}) - I(\mathbf{x}_2 + \mathbf{p}))^2 + \lambda \|\mathbf{x}_1 - \mathbf{x}_2\|^2} \quad (5.10)$$

The median element within a leaf node is selected as a representative texton and stored in the texton dictionary T . There are now two possibilities of finding the nearest texton element for every voxel in the images in order to obtain the texton labelling $t(\mathbf{x}) \in T$:

- *approximate labelling* using the best-bin-first method
- *exact labelling* using nearest neighbour search in the metric space

The approximate labelling is readily available after the computation of the tree. Each voxel will be labelled according to the leaf node it is located in. Since the vp-tree does not guarantee a globally optimal labelling, in which each voxel is assigned to the closest representative leaf node element (texton), there are possibly other textons with smaller distance.

To obtain the exact labelling, a nearest neighbour search (NNS) has to be

run for every patch (by backtracking through the tree). As Eq. 5.10 describes a metric, the triangle inequality can be used to accelerate the search for the closest textons. For the NNS, we first calculate an exhaustive distance matrix $D(t_i, t_j)$ between all textons (the number of textons is a lot smaller than the number of patches, therefore the computations for this step are negligible). For a given query element q , we start from an arbitrary initial texton assignment t_0 and calculate the respective distance $d(q, t_0)$ and set $d_{\min} = d(q, t_0)$. This forms an upper bound on the patch distance and can be used to exclude all textons t_i , for which the following equation holds:

$$|D(t_i, t_0) - d(q, t_0)| \geq d_{\min} \quad (5.11)$$

Then, a new texton is chosen, its distance computed and more textons may be discarded according to Eq. 5.11. If the new distance is smaller than d_{\min} , the upper bound is replaced by it and the current texton assignment is chosen as being optimal. This procedure is repeated until all textons t_i have been either compared or excluded based on Eq. 5.11. On average the number of distance comparisons can be reduced by more than an order of magnitude using this technique.

Additionally, the second closest textons may be found and included for the calculation of the joint histogram (in conventional intensity histograms, a similar approach is usually performed, using e.g. linear interpolation). It has to be noted that the histogram obtained from the texton labels does not preserve any ordering (in contrast to an intensity histogram), and therefore techniques like Parzen window estimation (smoothing of the histograms) cannot be employed.

Cluster forest

One straightforward solution to deal with the problems arising from a non-optimal tree clustering is to use a multiple cluster trees (i.e. a cluster forest) and assign multiple labels to each voxel. This also is computationally more efficient than performing the exact nearest neighbour search within one tree. A similar concept has been used by [Kleinschmidt et al., 2008] to find the closest patches for denoising. This approach is visualised in Fig. 5.6, where one element (depicted by a red circle) is assigned to three different clusters. This idea could be potentially

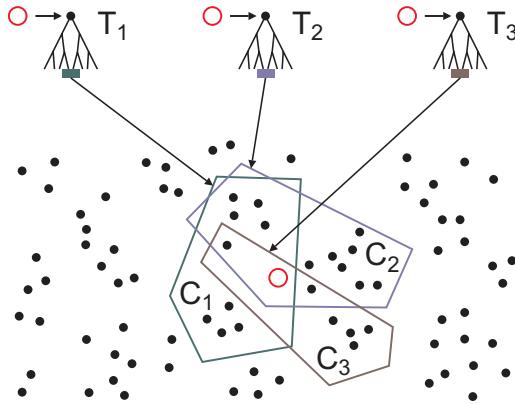


Figure 5.6: Illustration of the concept of cluster forests by [Kleinschmidt et al., 2008]. The search element \textcircled{O} is assigned to three different clusters. This helps alleviate problems caused by a non-optimal clustering and can therefore potentially provide a better representation of each element.

further extended so that trees within the cluster forest represent different features (e.g. different scales of patches and thereby enabling a simultaneous coarse and fine registration).

Local estimation of textural mutual information

For the local estimation of TMI, the local derivation of mutual information presented by [Rogelj et al., 2003] is used (see Sec. 5.2.1). The point-wise similarity $\text{TMI}(I_t, I_m, \mathbf{x})$ between target image I_t and moving image I_m can be defined by the statistical dependency of their texton representations T_t and T_m . If a cluster forest with F trees is employed, each voxel in image I_t has been assigned to F representations $T_t^1, T_t^2, \dots, T_t^F$, and the point-wise TMI is obtained with a summation over F :

$$\text{TMI}(I_t, I_m, \mathbf{x}) = \sum_{i=1}^F \log \left(\frac{p(T_t^i(\mathbf{x}), T_m^i(\mathbf{x}))}{p(T_t^i(\mathbf{x})) p(T_m^i(\mathbf{x}))} + \epsilon \right) \quad (5.12)$$

where ϵ is again a small constant to avoid $\log(0)$.

In the following section, the suitability of the textural mutual information as multi-modal image similarity and its robustness to parameter choice and intensity distortion is experimentally studied and compared to normalised and conditional mutual information.

5.4 Experiments and results

In Sec. 3.3 the similarity-based landmark localisation experiment was described, which is used here to study the ability of different metrics to measure similarity across scans. In this experiment not a full deformable registration is performed (which would make a direct comparison of similarity metrics challenging, because transformation model and optimisation might influence the performance of each metric differently). Each similarity metric is evaluated densely between locations around a landmark in the target scan using a sliding window aggregation

of point-wise similarities over a certain search region (in the moving scan). The search region is chosen so that the gold standard location (determined by landmark correspondences) is captured. The position (displacement) for each metric and landmark is determined by choosing the maximal similarity within the search region, the discrepancy between gold standard location is defined as landmark localisation error (see Sec. 3.3.1). Two datasets are employed for this experiment, one T1-weighted and one PD-weighted MRI chest scan from the Visible Human Dataset (VHD) (which were acquired post-mortem and are intrinsically aligned) and an inhale and exhale lung scan from a 4D-CT sequence.

First, the sensitivity of particular parameter choices for all similarity metrics is tested. In Sec. 5.4.2 the robustness against intensity distortions is evaluated and the metrics are compared among another. The particular challenges for VHD are the non-functional mapping of intensity distributions across scans, and the additional intensity distortions, which were described in detail in Sec. 3.3.2. The landmark localisation in the 4D-CT dataset is difficult for three reasons: deformable deformations, initial misalignment (on average 15 mm) and spatially varying contrast due to lung compression / expansion. Additionally there is an intra-observer error for the manual landmark annotation of ≈ 1 mm [Castillo et al., 2009]. The results are evaluated using the landmark localisation error averaged over 619 and 300 landmarks for the VHD and 4D-CT dataset respectively.

5.4.1 Parameter sensitivity

The choice of parameters plays an important role for the practical application of image registration methods. Fewer parameters and less sensitivity to their setting

is favourable. In the following the number of histogram bins for all metrics and the labelling approach and number of trees for TMI are studied.

Normalised mutual information

Figure 5.7 (a) shows the landmark localisation error for NMI for both datasets used in our experiment. The standard deviation of the Parzen window kernel g_σ has been empirically chosen to $\sigma = 0.75$. The number of bins for the histogram calculations is successively doubled from 8 to 512. Based on these experiments, at least 64 histogram bins should be used (which is in agreement with published algorithms, e.g. [Rueckert et al., 1999]). The hierarchical MI estimation improves the results when the two scans are substantially misaligned (4D-CT dataset).

Conditional mutual information

Figure 5.7 (b) displays the localisation accuracy for CMI for the two datasets. Roughly 50^3 voxels are used to form one spatial histogram bin (third histogram dimension) yielding 32 bins for the VHD and 48 bins for the 4D-CT dataset. The localisation error is slightly lower than for NMI, however hierarchical MI cannot be directly employed, since the change of spatial histogram labels of displaced voxels adversely affects the local CMI estimation.

Textural mutual information

The influence of the tree depth of TMI, which is directly related to the number of histogram bins, on the landmark localisation accuracy is studied in Fig. 5.7 (c). The weighting for the spatial distance in Eq. 5.10 is set to $\lambda = 0$. In a similar way to the results obtained for NMI and CMI, a larger number of bins results in

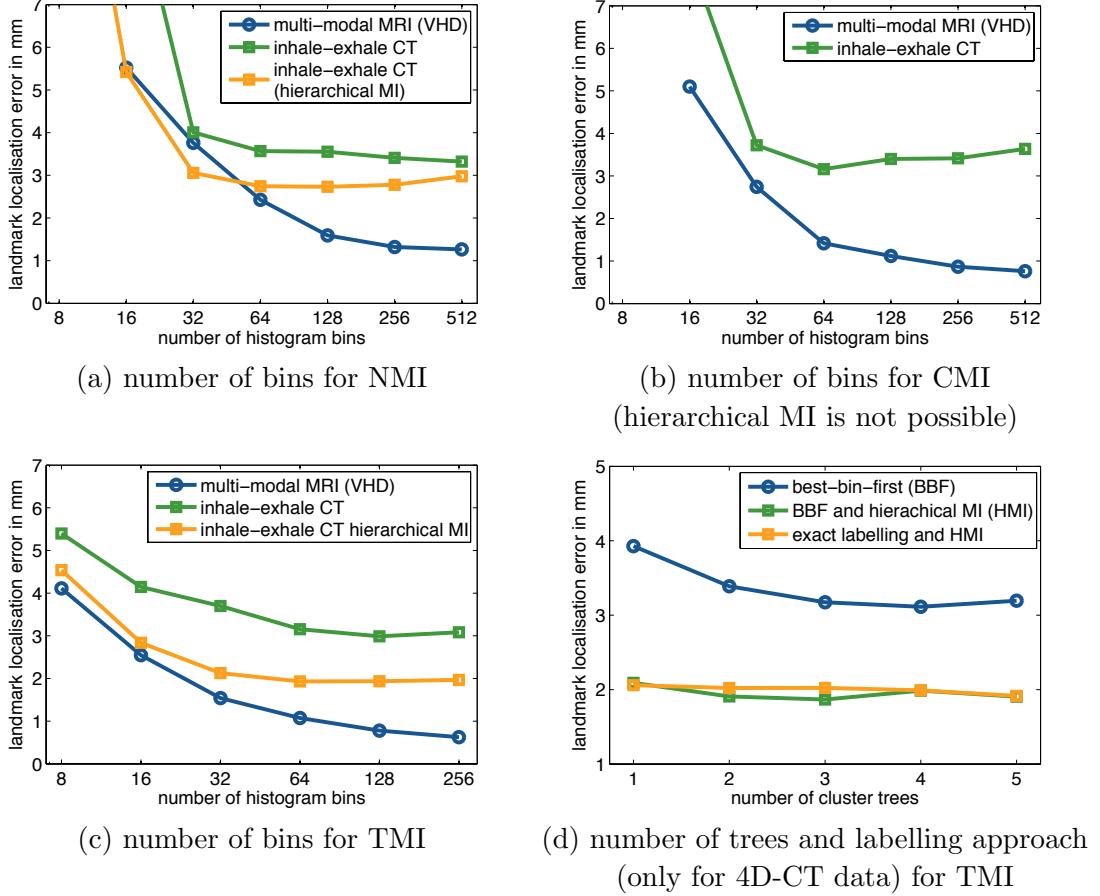


Figure 5.7: Parameter sensitivity of all presented methods. The Parzen smoothing is set to $\sigma_i = 0.75$ in all cases. The number of spatial bins in CMI is set to 32 (VHD) or 48 (4D-CT). Landmark localisation error in mm for multi-modal MRI (\ominus) and 4D-CT dataset (\square and \blacksquare with hierarchical MI (HMI) estimation). The localisation error is in general smaller when using more bins (which is equivalent to a greater tree depth for TMI). The results for TMI using fewer bins are, however, substantially better than for NMI or CMI. The use of the HMI estimation reduces the error for the inhale-exhale CT dataset for both NMI and TMI substantially. For TMI, the exact labelling strategy (see Sec. 5.3.2) and using multiple trees (see Sec. 5.3.2) is in general better, yet these improvements are negligibly when HMI estimation is used.

higher accuracy. However, for a small number of bins (8, 16, 32) the results are substantially better than for the two other state-of-the-art approaches. This can be explained by the near optimal population of each histogram bin (since the number of voxels within each leaf of the tree is the same). Therefore, the sensitivity to this parameter setting is greatly reduced.

Figure 5.7 (d) shows the landmark localisation error for the 4D-CT dataset with increasing number of trees. The trees use the same patch distance, but are based on different randomly chosen subsets (samples) of all voxels/patches. In the case where the hierarchical MI estimation is used (to reduce the influence of initial misregistration to the estimation of the joint distribution) no benefit can be seen by using multiple trees. Otherwise, a small improvement is achieved, suggesting an improved robustness of this approach for scans with large deformations. The use of approximate (best-bin-first) or exact (using Eq. 5.11) labelling has little influence on the landmark localisation accuracy. We therefore conclude that the time-consuming step of an exact nearest-neighbour search can be avoided for most registrations tasks.

5.4.2 Robustness against intensity distortions

We now test the sensitivity of the different statistical similarity metrics in the presence of intensity distortions. Figure 5.8 shows the cumulative distribution of the localisation error for the inhale-exhale CT dataset for the three presented MI formulations, SAD and the initial landmark distance. This is a particularly challenging case with an average landmark displacement of 15 mm. The complex intensity distortions due to the compression of lung tissue can not be adequately

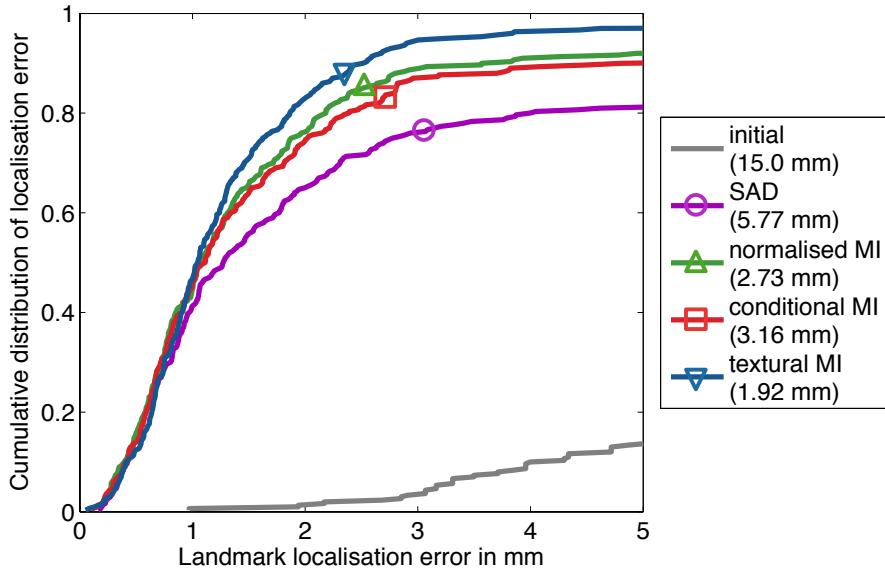


Figure 5.8: Cumulative error distribution for presented MI-based similarity metrics for the inhale-exhale CT dataset using the best individual settings for each method. Textural MI yields the smallest average error and an improved robustness (fewer large errors) against normalised and conditional MI (the improvements are statistically significant with $p = 0.029$ and $p = 0.006$). SAD is not suitable for this task, due to the intensity distortions between inhale and exhale scans.

modelled using SAD (localisation error 5.77 mm). The best settings for normalised MI (128 bins, including HMI estimation), conditional MI (64 bins), and textural MI (64 bins, including HMI estimation) have been used. While the correct location (within an error of ≤ 1.5 mm) of the majority of landmarks can be recovered by all methods, the inclusion of contextual information in TMI provides a greatly improved robustness. Using TMI resulted in only 6.7 % of landmarks being located more than one voxel ($=2.85$ mm) away from its gold standard position, while this proportion is 14 % for CMI and 12 % for NMI.

To evaluate the robustness of the metrics against intensity distortions, first a locally varying multiplicative bias field, and second, additive Gaussian noise, have been applied to the VHD dataset as described in Sec. 3.3.2. Figure 5.9 demon-

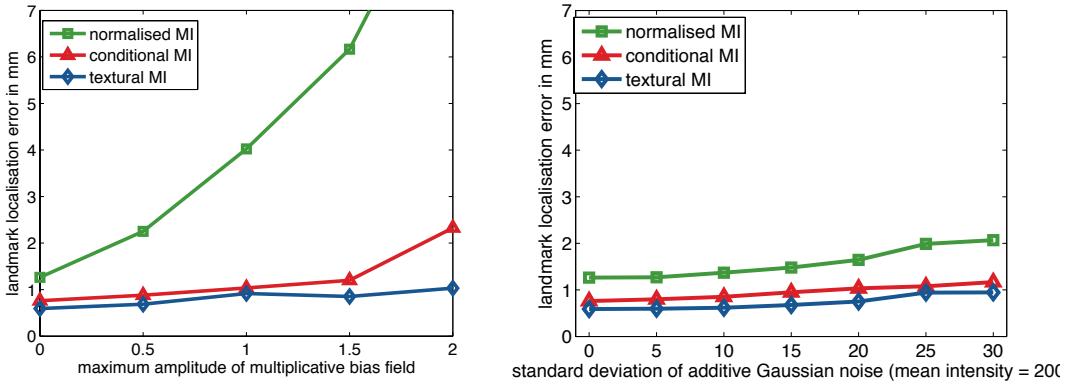


Figure 5.9: Landmark localisation error for VHD images with simulated intensity distortions. A multiplicative bias field (linearly varying from left to right) strongly affects the accuracy of NMI, while both other metrics are robust against it. All metrics deal well with additive Gaussian noise.

strates the robustness of both conditional and textural MI to the bias field. The localisation accuracy of normalised MI deteriorates substantially with increased bias field, a behaviour that has been also shown by [Loeckx et al., 2007]. The additional regional channel in CMI, and the use of contextual information (based on image patches) in TMI, show a clear advantage over the traditional MI formulation. All metrics are relatively insensitive to increased Gaussian noise (see Fig. 5.9 right). TMI yields the lowest localisation error (0.59 mm) for the original images, a two-fold improvement over NMI (1.26 mm).

5.5 Summary

In this chapter, the theoretical and practical implications of using statistical metrics, in particular mutual information (MI), to define similarity across multi-modal scans, has been studied. Important extensions for the calculation of MI for deformable registration, point-wise and hierarchical estimation, were discussed. In

Sec. 5.3, our contribution to the field: textual mutual information (TMI), was presented and discussed. TMI efficiently incorporates spatial context by using a clustering of images based on small patches of image intensities. Two state-of-the-art metrics, normalised and conditional MI are used for comparison purposes. TMI demonstrates significantly improved performance for the localisation of landmarks in two challenging datasets compared to NMI and CMI. It is also less sensitive to its internal parameter settings and more robust against intensity distortions (bias fields and noise) than NMI.

While mutual information-based similarity metrics are theoretically well suited to address non-functional intensity relations across images, they remain challenging from an optimisation point of view. For deformable registration, local evaluations of cost functions are necessary. Therefore, for a globally defined statistical metric, approximations about the independence of the global joint statistics and local deformations have to be made (see Secs. 5.2.1 and 5.2.2). This motivates the use of a structural image representation, which is independent of the underlying image acquisition and can be minimised with standard metrics such as sum of squared differences (SSD). The concept of structural representations will be explained in the next chapter and two novel formulations for multi-dimensional structural image representations are proposed.

Chapter 6

Multi-dimensional structural image representation

The objective of this chapter is the introduction of novel multi-dimensional structural image representations for the use in deformable multi-modal registration. The first contribution is the introduction of a new concept for representing local image orientation through decomposing of the gradient structure tensors. The second contribution is the modality independent neighbourhood descriptor (MIND). This novel concept enables a representation of images using multi-dimensional descriptors, which encapsulate the local image structure independently of contrast, noise and modality. MIND can be minimised using sum of squared differences (SSD), the most widely applicable similarity metric.

This chapter is closely linked with the previous one. The driving idea is again to increase the use of contextual information to define image similarity in order to improve the robustness and accuracy of correspondences across images, in particular for multi-modal scans. In contrast to the previous chapter, no global statistical

relation between intensities across scans is necessary for structural image representations. Another advantage of the concept is that structural representations can be used in deformable registration with efficient optimisation methods (see Chapter 7) using any point-wise similarity metric, such as sum of absolute or squared differences (SAD, SSD). While previously structural content had been represented by only scalar values, we introduce two novel approaches to multi-dimensional structural image representation, which efficiently capture contextual information without relying on global statistical relations across images.

In Sec. 6.1 the concept of a structural image representation is introduced and an overview of previous and novel approaches is given. Entropy images are discussed in more detail in Sec. 6.2. We present a new approach to extract local image orientation based on the structure tensor in Sec. 6.3, which successfully tackles the challenges of gradient reversal and normalisation of orientation vectors. In Sec. 6.4 the *modality independent neighbourhood descriptor* (MIND) is introduced, which is a multi-dimensional representation. We will show that, besides its robustness against intensity distortions and noise, it is highly discriminative for different image features. A further extension is proposed in Sec. 6.4.2, the *self-similarity context* (SSC), which reduces the impact of localised image noise or artefacts and increases the contextual information content.

Experimental results for the presented methods in this chapter will be evaluated using the landmark localisation experiments presented in Sec. 3.3. For this purpose, we will again employ two datasets: an inhale-exhale pair of a 4D-CT lung sequence and an MRI-T1 and MRI-PD thorax scan of the visible human (VHD).

6.1 Overview of structural image representations

Structural image representation for multi-modal registration aims to solve the problem of defining a suitable mapping of intensities across modalities by transforming the image intensities so that they are independent of the specific modality and local contrast. The respective representations can then be compared across images, using simple similarity metrics such as SAD or SSD. Ideally, the derived representation should be highly discriminative for different anatomical or geometric features and be robust against noise.

Entropy is a general measure for information content. In [Penney et al., 1998] the entropy of difference images has been applied to the registration of fluoroscopy and CT images, since it is invariant to an intensity offset and more robust to noise than direct differences. Local image entropy has been proposed by [Wachinger and Navab, 2012] as a structural (scalar) image representation for multi-modal registration (see Sec. 6.2), which can be minimised using SAD or SSD. Here, each voxel is represented by its local entropy value, which is estimated based on the intensity histogram within a small (weighted) neighbourhood. Higher entropy can be found for a location at which there are intensity changes, and the value should in principle be independent of the intensity differences and therefore image contrast.

Assuming constant image gradients across acquisitions has been successfully used to match images with slight changes in brightness [Brox et al., 2004]. For multi-modal images, however, the strength of image gradients of corresponding features is not constant and can point in opposite directions. [Mellor and Brady, 2005] used the local phase, which can be extracted from the images using the monogenic

signal, as an alternative image representation. An improvement of the robustness of the registration has been found for MRI and ultrasound scans. However, in their work mutual information was used between local phase images, which implies that there was still no direct dependency between the representations across modalities. The use of local orientation (assuming it can be reliably estimated) is also possible and has the attractive property of being very discriminative for the location along image edges. [Pluim et al., 2000] successfully used the local orientation in addition to mutual information for rigid registration, and showed that the number of local minima during optimisation could be substantially reduced. [Haber and Modersitzki, 2007] maximised the inner product of normalised image gradients for deformable multi-modal registration. A problem arises when the orientations of two corresponding locations are opposed to one another, and thus the normalised gradients are reversed. We have addressed this issue in [Heinrich et al., 2011c] by using an eigenvector analysis of the image structure tensor, which is invariant to gradient reversal: called the *structure tensor-based orientation measure* (STORM), which will be further described in Sec. 6.3. The STORM representations of the images can then be compared across modalities using SAD or SSD.

[Zabih and Woodfill, 1994] presented two approaches (the *rank filter* and *census transform*) to structural image representation, which are widely used for stereo or optical flow computation of images with radiometric differences [Hirschmüller and Scharstein, 2009]. The rank filter represents each voxel in the image by its rank (ordering) within a local neighbourhood \mathcal{N} and can be defined by $I(\mathbf{x})_{rank} = \sum_{\mathbf{y} \in \mathcal{N}} H(I(\mathbf{x}) - I(\mathbf{y}))$, where $H(\cdot)$ is the heaviside step function. Similarly, the census transform defines a binary variable based on whether or not the intensity of a voxel in \mathcal{N} is greater than the central voxel's intensity. Here, the spatial

structure of the local neighbourhood is also stored, resulting in a bitwise vector of length $|\mathcal{N}|$ for each voxel. The similarity across images is then defined as the Hamming distance [Hamming, 1950] of all bitwise entries. It can be easily shown that both transforms are invariant to order preserving changes in offset (change of mean intensity) and gain (change of contrast) across images. The census transform, which keeps the contextual information of the local neighbourhood, has been shown in [Zabih and Woodfill, 1994] and [Hirschmüller and Scharstein, 2009] to outperform the rank filter and several other metrics, which are invariant to monotonic greyscale mapping. Since the local ordering of intensities is not preserved across modalities, it is however not applicable to define representations for multi-modal images.

[Ojala et al., 2002] proposed a similar concept called, local binary patterns (LBP), which also uses a pixel-wise comparator to extract a binary descriptor. LBPs differ from the census transform in that only pixel locations on a circle around the centre pixel are considered. Additionally, rotational invariance is achieved by mapping all patterns, which are equal up to a (quantised) rotation, to a common descriptor string. The descriptors are then often pooled spatially into histograms. Further extensions to this concept utilise the uniformity of these circular (or spherical in 3D [Liao and Chung, 2009]) patterns as a texture descriptor.

BRIEF [Calonder et al., 2010] is another descriptor related to the census transform, which deals well with monotonic greyscale mappings. Its main difference is the spatial arrangement of the binary tests, which are not limited to centre-to-neighbour comparisons, but randomly sampled (from e.g. a Gaussian distribution) within the neighbourhood. Two approaches have been presented by [Calonder et al., 2010] to obtain rotationally invariance. Binary Robust Invariant Scalable

Keypoints (BRISK) [Leutenegger et al., 2011] further improve on the concept of BRIEF by pre-scaling and pre-rotating a deterministic sampling pattern using local gradient information and reducing the number of sampling locations.

In Sec. 6.4, we introduce a new concept based on the distance of small image patches within a local neighbourhood. The *modality independent neighbourhood descriptor* (MIND) is invariant to non-functional intensity mappings across images, local changes in contrast, image noise and robust against geometric distortions. It is based on the principle of image self-similarity [Shechtman and Irani, 2007; Buades et al., 2005]. Similarly to the previously introduced descriptors, the spatial configuration of the local image structure is stored in a multidimensional vector, making use of the contextual information. Instead of using only a binary representation, the distances between patches are represented by continuous (or quantised) values. This makes MIND more sensitive to local orientation, and using patches increases its robustness. By using a patch distance function, which depends on squared differences MIND is not limited to monotonic greyscale mappings. The MIND representation can be compared across modalities using SAD or SSD. While MIND is based on local self-similarity, and therefore very much dependent on the image patch around the voxel of interest. In Sec. 6.4.2, we introduce *self-similarity context* (SSC), an extension to MIND. Similar to the concept of BRIEF [Calonder et al., 2010], all pair-wise patch distances within a local neighbourhood are employed for SSC, thus reducing the impact of the central patch and increasing the contextual information content.

6.2 Entropy images

Local patch-based entropy images have been proposed by [Wachinger and Navab, 2012], which were minimised using SSD across modalities. In their experiments, a similar registration accuracy, compared to mutual information for rigid multimodal registration and some synthetic non-rigid experiments, is demonstrated. The basic assumption that drives the registration, based on entropy images, is that intensity changes occur at the same locations in different modalities. The Shannon entropy E for a voxel \mathbf{x} within a local neighbourhood \mathcal{N} in an image I containing intensities of the range $i \in \mathcal{I}$ is defined as:

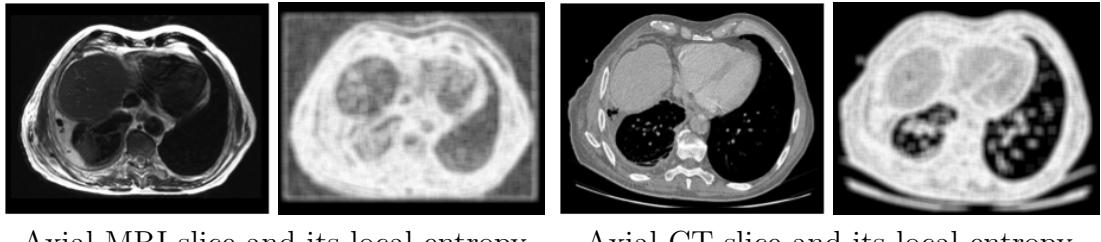
$$E(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{N}} \sum_{i \in \mathcal{I}} p(I(\mathbf{x}) = i) \log p(I(\mathbf{x}) = i) \quad (6.1)$$

where $p(I(\mathbf{x}) = i)$ is the probability of the occurrence of a certain intensity $i \in \mathcal{I}$ based on a local histogram estimate. The histogram calculation can be weighted based on the spatial distance of \mathbf{x} within \mathcal{N} . A patch histogram using a spatial Gaussian weighting σ_s is defined as:

$$p(\mathbf{x}, i) = \frac{1}{\sigma_s \sqrt{2\pi}} \sum_{\mathbf{y} \in \mathcal{N}} \exp\left(-\frac{|\mathbf{x} - \mathbf{y}|^2}{2\sigma_s^2}\right) \delta[I(\mathbf{y}) - i] \quad (6.2)$$

where $\delta[\cdot]$ defines the Kronecker delta function. The well-known Parzen window estimation with a standard deviation of σ_i can be obtained by replacing the delta function in Eq. 6.2 with $\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(I(\mathbf{y}) - i)^2}{2\sigma_i^2}\right)$.

Examples of the entropy image representations for an example MRI and CT slice are shown in Fig. 6.1, using a uniformly weighted patch of 7×7 voxels. The sensitivity to locally different noise can be clearly seen from this example.



Axial MRI slice and its local entropy Axial CT slice and its local entropy

Figure 6.1: Entropy images for MRI and CT slices from Fig. 5.1, calculated using a uniformly weighted patch of 7×7 voxels. This example demonstrates the concept’s ability to transform image into a structural representation, but also its susceptibility to noise (e.g. in the background of the MRI).

The intensity range \mathcal{I} can either be set by a global or local normalisation. According to [Wachinger and Navab, 2012], the number of intensity bins should be sufficiently small to ensure a well populated local histogram, for the application of entropy images for deformable registration. However, this reduces the sensitivity to small intensity changes. A local intensity normalisation has been suggested. Since the straightforward implementation of Eqs. 6.1 and 6.2 requires $|\mathcal{N}|$ computations per voxel, this approach would be impractical for larger neighbourhoods. Integral images can be used instead to reduce the computational complexity to $6|\mathcal{I}|$ for a uniform weighting within patches [Porikli, 2005].

A challenge for entropy images is a changing level of noise within and across images, which influences the entropy calculation and might have a negative impact on the registration.

6.3 Structure tensor gradient orientation

Boundaries between neighbouring tissues carry significant information in medical images. The gradient of tissue boundaries might not have the same magnitude

for images of different modalities, but the orientation of the gradient should be the same or flipped by 180° . Here, we will present an approach for structural image representation based on the principal orientation of the structure tensor of image gradients, which is called *structure tensor orientation measure* (STORM). STORM represents each voxel in the images with a vector of length d (image dimension) and can be efficiently compared across images using SSD and SAD. It can also intrinsically deal with the problem of gradient reversal for corresponding structures in multi-modal images.

6.3.1 Gradient orientation

In [Pluim et al., 2000] gradient orientation (GO) was used to improve the MI measurement for rigid image registration. It was shown that the use of this measure can be formulated to obtain a convex optimisation problem. Their results showed that GO, in contrast to MI, was able to find the same unique minima for a range of random misalignments. The local gradient orientation can be directly estimated from the spatial image gradients $\nabla I(\mathbf{x})$, where the normal direction $\mathbf{n}(I, \mathbf{x})$ is then given by:

$$\mathbf{n}(I, \mathbf{x}) := \frac{\nabla I(\mathbf{x})}{\|\nabla I(\mathbf{x})\|} \quad (6.3)$$

As this measure is not well defined in homogeneous image regions and would be too much affected by noise, [Haber and Modersitzki, 2007] addressed this problem by substituting the denominator in Equation (6.3) with a robust norm:

$$\|\nabla I(\mathbf{x})\|_\epsilon := \sqrt{\nabla I(\mathbf{x})^T \nabla I(\mathbf{x}) + \epsilon^2} \quad (6.4)$$

using an automatic choice of the small constant ϵ based on the image noise level.

Defining the similarity \mathcal{S} between target image I_t and moving image I_m based on their gradient orientation can be defined by the inner product of their corresponding normals $\mathbf{n}(I_t)$ and $\mathbf{n}(I_m)$:

$$\mathcal{S}(I_t, I_m, \mathbf{x}) = \frac{|\nabla I_t(\mathbf{x}) \cdot \nabla I_m(\mathbf{x})|}{\|\nabla I_t(\mathbf{x})\| \cdot \|\nabla I_m(\mathbf{x})\|} = \mathbf{n}(I_t) \cdot \mathbf{n}(I_m) \quad (6.5)$$

This will yield 1 for perfectly parallel normals (or same orientations) and -1 for reversed gradients (orientations, which are flipped by π). To fully benefit from a structural representation, it should be possible to use similarity metrics like SAD and SSD. It can be easily shown that the SSD of the d -valued representation gives an equivalent optimisation objective:

$$\begin{aligned} \text{SSD}(\mathbf{n}(I_t), \mathbf{n}(I_m), \mathbf{x}) &= \sum_{i=1}^d (n_i(I_t(\mathbf{x})) - n_i(I_m(\mathbf{x})))^2 \\ &= \underbrace{\sum_{i=1}^d n_i(I_t(\mathbf{x}))^2}_{=1} - 2 \sum_{i=1}^d n_i(I_t(\mathbf{x})) n_i(I_m(\mathbf{x})) + \underbrace{\sum_{i=1}^d n_i(I_m(\mathbf{x}))^2}_{=1} \\ &= 2 - 2 (\mathbf{n}(I_t) \cdot \mathbf{n}(I_m)) \end{aligned} \quad (6.6)$$

To address the problem of gradient reversal and in order to formulate a cost function, which is well defined for small deviations around the minimum, [Pluim et al., 2000] use a cosine function of the difference in orientation δ to define the *absolute gradient orientation (AGO)* measure:

$$\text{AGO}(\delta) = \cos(2\delta + 1)/2 = \cos^2(\delta) \quad (6.7)$$

If the orientation is defined by its normal vector in both images $\mathbf{n}(I, \mathbf{x})$ and $\mathbf{n}(J, \mathbf{x})$, this measure is simply given by the squared dot product of the unit normals. The disadvantage of this metric is that it cannot be directly used in SSD based optimisation methods. In the following section we will show how the principal eigenvector of the structure tensor of image gradients can be used to address this problem and find a structural image representation, which is invariant to gradient reversal and can be minimised using SSD or SAD.

6.3.2 Orientation based on structure tensors

The structure tensor \mathbf{T} , which is also called second-moment matrix [Koenderink and Pont, 2003], is a matrix derived from the image gradients $\nabla I(\mathbf{x})$: $\mathbf{T}(\mathbf{x}) = \nabla I(\mathbf{x})\nabla I(\mathbf{x})^T$. Gradients of discrete images are usually obtained by using finite differences, which can be efficiently done by a convolution with a five-point stencil:

$$\frac{\partial I(\mathbf{x})}{\partial x} = \frac{-I(\mathbf{x} + 2e_x) + 8I(\mathbf{x} + e_x) - 8I(\mathbf{x} - e_x) + I(\mathbf{x} - 2e_x)}{12} \quad (6.8)$$

where e_x defines a unit step in the direction of x . To compensate for the influence of noise on the gradient computation, the image gradients are then spatially smoothed using Gaussian kernel σ_g .

We now show how the structure tensor can be used to find the orientation of an image $I(\mathbf{x})$ within a small region \mathcal{N} around \mathbf{x} . The image gradient vector $\nabla I(\mathbf{x})$ is defined to be of unit length. The unit normal vector $\mathbf{n}(\mathbf{x})$ of the image intensity orientation can then be found where:

$$\mathbf{n}(\mathbf{x})^T \nabla I(\mathbf{x}) = 1 \quad (6.9)$$

Equation 6.9 can be solved by the least-squares optimisation of $E(\mathbf{n})$ [Derpanis, 2005] (using the notation I_x as an abbreviation for $\frac{\partial I}{\partial x}$ etc.):

$$\begin{aligned} E(\mathbf{n}) &= \max_{\|\mathbf{n}\|=1} \sum_{\mathbf{x} \in \mathcal{N}} (\mathbf{n}^T \nabla I(\mathbf{x}))^2 = \max_{\|\mathbf{n}\|=1} \mathbf{n}^T \left(\sum_{\mathbf{x} \in \mathcal{N}} \nabla I(\mathbf{x}) \nabla I(\mathbf{x})^T \right) \mathbf{n} \\ &= \max_{\|\mathbf{n}\|=1} \mathbf{n}^T \mathbf{T} \mathbf{n} \text{ with } \mathbf{T} = \begin{vmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{vmatrix} \end{aligned} \quad (6.10)$$

where the matrix \mathbf{T} is the structure tensor of spatial image gradients obtained using a smoothing kernel with σ_g to obtain the summation within \mathcal{N} . Finding the normal orientation in this energy maximisation is equivalent to the following eigenvalue problem:

$$\mathbf{T} \mathbf{n} = \lambda \mathbf{n} \quad (6.11)$$

The solution of the maximisation problem, Equation (6.10), is the eigenvector which corresponds to the largest eigenvalue. The obtained eigenvalues contain additional information about the local image structure. A single dominant orientation is present if $\lambda_1 \gg \lambda_2, \lambda_3$, homogenous regions are characterised by $\lambda_1 \approx \lambda_2 \approx \lambda_3 \approx 0$. Using λ_1 to weight the local influence of the similarity term (i.e. higher weighting for more pronounced local orientation) could potentially be useful.

The eigenvector $\mathbf{v} = (v_1, v_2, v_3)^T$, which corresponds to the largest eigenvalue λ_1 is used to extract the three-dimensional orientation described by the two angles θ and φ :

$$r = \sqrt{v_1^2 + v_2^2 + v_3^2}, \theta = \cos^{-1} \left(\frac{v_3}{r} \right), \varphi = \tan^{-1} \left(\frac{v_1}{v_2} \right) \quad (6.12)$$

The obtained orientation is defined up to a rotation of 180° , because a reversal of

the gradient between modalities is possible.

Substituting $\mathbf{v}(I_t, \mathbf{x})$ and $\mathbf{v}(I_m, \mathbf{x})$ into Equation 6.6 yields the proposed similarity term, STORM, for two images I_t and I_m . The main contribution here lies in the fact that we estimate the local orientation based on the structure tensor, and not directly from the smoothed image intensity gradients as done e.g. in [De Nigris et al., 2010] or [Pluim et al., 2000].

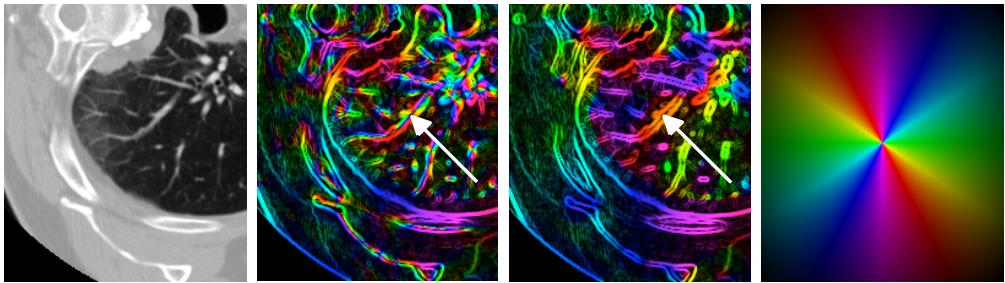


Figure 6.2: Estimation of local image orientation. Left: grayscale image of axial slice of lung CT scan. Centre-left: Orientation based on image gradients. Centre-right: Orientation based structure tensor. Right: Colour-coding for angles of the orientation. The structure tensor estimate avoids the reversal of orientation for small thin features (see white arrow).

Visual example of STORM

One advantage of using STORM compared to image gradient orientation is illustrated in Figure 6.2. Here, the estimation of the local orientation of an axial slice of a lung CT scan is demonstrated. It can be seen that the eigenvector decomposition of the structure tensor leads to more consistent orientation estimation around small structures, such as the pulmonary vessels. The reversal of orientation estimates between positive and negative image gradients is avoided. Therefore STORM can be directly used as a structural representation for multi-modal image registration, using SAD or SSD as metric. STORM is comparatively simple to

calculate and yields a vector of size d (image dimension) for each voxel.

While STORM offers theoretically a good structural representation for scans independent of modality, the spatial contextual information is only weakly covered by the orientation of image gradients. In the following, we will present a different approach to multi-dimensional image representations based on the calculation of patch-based self-similarities in a larger spatial region, which should be beneficial (based on the findings of Chapter 5) for challenging multi-modal registration tasks.

6.4 Modality independent neighbourhood descriptor (MIND)

In this section the *modality independent neighbourhood descriptor* (MIND) [Heinrich et al., 2012b] is presented and its use to define the similarity between two images based on the SSD of their descriptors is demonstrated. First we motivate the use of image self-similarity for the construction of an image descriptor. We will then propose the definition of self-similarity by using a Gaussian-weighted patch-distance and explain the spatial capture range of the descriptor.

Self-similarity

Our approach uses the principle of self-similarity, a concept which has first been introduced in the domain of image denoising by [Buades et al., 2005]. The authors make use of similar image patches across a noisy image to obtain a noise-free pixel, which is computed as a weighted average of all other pixels in the image. The weights $w(i, j)$ used for the averaging are based on the sum of squared differences

between the patch, which surrounds the pixel of interest, and all other patches in the image I . The denoised pixels $NL(i, I)$ are then calculated using the following equation:

$$NL(i, I) = \sum_{j \in \Omega} w(i, j)I(j) \quad (6.13)$$

where Ω is the image domain. The approach demonstrated a very good performance for image denoising.

The use of patches to measure similarity based on the weights $w(i, j)$ within the same image can easily capture a variety of image features, because it treats regions, edges, corners and textures in a unified way and is thus much more meaningful than using single intensities. In subsequent work, this approach was simplified to search for similar patches only within a local search region \mathcal{N} (replacing Ω in Eq. 6.13 by \mathcal{N} defining the local neighbourhood, see [Coupé et al., 2006]). Figure 6.3 gives an example of how well the self-similarity pattern can describe the local structure around an image location. Mainly because of this property, the concept has subsequently been used in a variety of applications. One example is its application to object localisation by [Shechtman and Irani, 2007]. Here, a correlation surface is extracted using colour patch distances and then stored in a log-polar histogram, which can be matched across images using the L_1 norm.

Motivation and Concept

Our aim is to find an image representation, which is independent of the modality, contrast and noise level of images from different acquisitions and at the same time sensitive to different types of image features. Our approach is based on the assumption that a local representation of image structure, which can be estimated

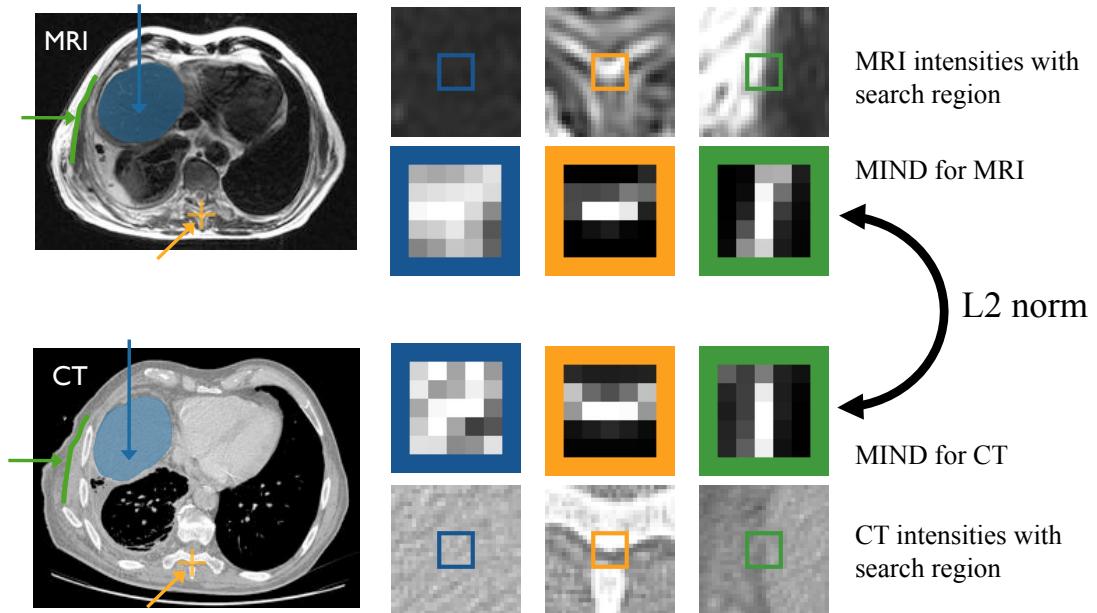


Figure 6.3: Concept for the use of MIND as multi-modal similarity metric. MIND is calculated in a dense manner in CT and MRI. Three example locations with different image features: ■ homogenous intensities (liver), □ corner points at one vertebra and ▨ image gradients at the boundary between fat and non-fat tissue. The corresponding descriptors (in coloured boxes, high intensities correspond to small patch distances) are independent of the respective modality and can be easily compared using the L_2 norm.

through the similarity of small image patches within one modality, is shared across modalities. Many different features may be used to derive a similarity cost function for image registration, such as corner points, edges, gradients, textures or intensity values. Figure 6.3 shows some examples on two slices of a CT and MRI volume. Using the SSD of the corresponding intensities between small image patches as similarity metric is sensitive to all these different types of image features. However, the use of these patch distances is limited to single-modal images. In our approach, a multi-dimensional image descriptor, which represents the distinctive image structure in a local neighbourhood, is extracted based on patch distances

for both modalities separately and afterwards compared using simple single-modal similarity measures.

6.4.1 Derivation of MIND representations

MIND can be specified by a distance D_p , a variance estimate V and a spatial search region R :

$$\text{MIND}(I, \mathbf{x}, \mathbf{r}) = \frac{1}{n} \exp \left(-\frac{D_p(I, \mathbf{x}, \mathbf{x} + \mathbf{r})}{V(I, \mathbf{x})} \right) \quad \mathbf{r} \in R \quad (6.14)$$

where n is a normalisation constant (so that the maximum value is 1) and $\mathbf{r} \in R$ defines the search region. By using MIND, an image will be represented by a vector of size $|R|$ at each location \mathbf{x} .

Patch-based distance

To evaluate Eq. 6.14 we need to define a distance measure between two voxels within the same image. As discussed before, a straightforward choice of a distance measure $D_p(\mathbf{x}_1, \mathbf{x}_2)$ between two voxels \mathbf{x}_1 and \mathbf{x}_2 is the sum of squared differences (SSD) of all voxels between the two patches P of size $(2p + 1)^d$ (with image dimension d) centred at \mathbf{x}_1 and \mathbf{x}_2 .

$$D_p(I, \mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{p} \in P} (I(\mathbf{x}_1 + \mathbf{p}) - I(\mathbf{x}_2 + \mathbf{p}))^2 \quad (6.15)$$

The distance value defined in Eq. 6.15 has to be calculated for all voxels \mathbf{x} in the image I and all search positions $\mathbf{r} \in R$. The naïve solution (which is e.g. used in [Coupé et al., 2006]) would require $3(2p + 1)^d$ operations per voxel and is therefore

computationally very expensive.

We propose an alternative solution to calculate the exact patch-distance very efficiently using a cubic convolution filter kernel of size $(2p+1)^d$. First a copy of the image I' is translated by \mathbf{r} yielding $I'(\mathbf{r})$. Then the point-wise squared difference between I and $I'(\mathbf{r})$ is calculated. Finally, these intermediate values are convolved with the kernel C , which effectively substitutes the SSD summation in Eq. 6.15:

$$D_p(I, \mathbf{x}, \mathbf{x} + \mathbf{r}) = C \star (I - I'(\mathbf{r}))^2 \quad (6.16)$$

This procedure is now repeated for all search positions $\mathbf{r} \in R$. The solution of Eq. 6.16 is equivalent to the one obtained using Eq. 6.15. Using this method it is also easily possible to include a Gaussian weighting within the patches by using a Gaussian kernel C_σ of size $(2p+1)^d$. The computational complexity per patch distance calculation is therefore reduced from $(2p+1)^d$ to $d(2p+1)$ for an arbitrary separable kernel and $3d$ for a uniform patch weighting (and integral images for fast filtering c.f. [Tapia, 2011]). A similar procedure has been proposed in the context of windowed SSD aggregation by [Scharstein and Szeliski, 1996].

To obtain a normalisation in Eq. 6.14, one could simply divide each descriptor by its maximum value. Here, another approach is used, where the minimal patch distance is subtracted from D_p before inserting it into the exponential:

$$D_p^*(I, \mathbf{x}_1, \mathbf{x}_2) = D_p(I, \mathbf{x}_1) - \min_{\mathbf{x}_2 \in \mathcal{N}} (D_p(I, \mathbf{x}_2)) \quad (6.17)$$

Based on the assumption that at least one patch within \mathcal{N} resembles the same underlying structure (which is a fairly safe assumption, since anatomical structures

are rarely confined to a single point in 3D medical images) the minimum patch-distance in Eq. 6.17 is an estimate of the SSD of the additive image noise.

Variance measure for the Gaussian function

We want to obtain a high response for MIND for patches that are similar to the patch around the voxel of interest, and a low response for everything that is dissimilar. A Gaussian function (see Eq. 6.14) is used for this purpose. The denominator $V(I, \mathbf{x})$ in Eq. 6.14 is an estimation of the local variance. A smaller value for V yields a sharply decaying function, and higher values indicate a broader response. The parameter should be related to the amount of noise in the image. The variance of the image noise can be estimated via pseudo-residuals ϵ calculated using a six-neighbourhood \mathcal{N} (see [Coupé et al., 2008]):

$$\epsilon_i = \sqrt{\frac{7}{6}} \left(I(\mathbf{x}_i) - \frac{1}{6} \sum_{\mathbf{x}_j \in \mathcal{N}} I(\mathbf{x}_j) \right) \quad (6.18)$$

ϵ is averaged over the whole image domain Ω to obtain a constant variance measure $V(I, \mathbf{x}) = \frac{1}{|\Omega|} \sum_{i \in \Omega} \epsilon_i^2$. This, however, increases the sensitivity of the image descriptors to spatially varying noise. Consequently, a locally varying function is beneficial. Therefore, $V(I, \mathbf{x})$ is determined by using the mean of the patch distances themselves within the local neighbourhood $\mathbf{n} \in \mathcal{N}$:

$$V(I, \mathbf{x}) = \frac{1}{|\mathcal{N}|} \sum_{\mathbf{n} \in \mathcal{N}} D_p(I, \mathbf{x}, \mathbf{x} + \mathbf{n}) \quad (6.19)$$

Using this approach (Eq. 6.19), MIND can be automatically calculated without the need for any additional parameters.

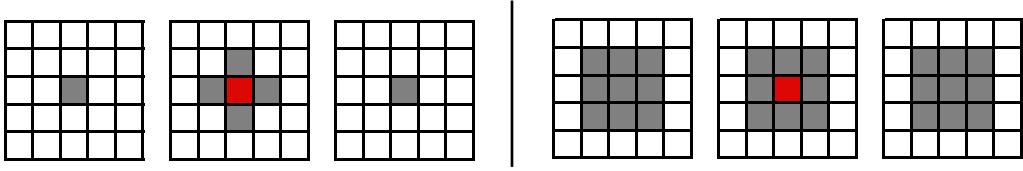


Figure 6.4: Different sampling strategies of the search region: Left: six-neighbourhood (6-NH). Right: dense sampling (26-NH). Red voxel is the voxel of interest, grey voxels are being sampled within $\mathbf{r} \in R$. The left and right column show the 3D slice above and below the central voxel.

Example responses of the obtained descriptors for three different image features for both CT and MRI are shown in Fig. 6.3 (second and third row on the right), where a high intensity corresponds to a small patch distance. Fig. 6.3 demonstrates how well descriptors represent these features, independent of modality.

Spatial search region

An important issue using MIND is the spatial extent of the search region (see R in Eq. 6.14) over which the descriptor is calculated. In the original work of [Buades et al., 2005], self-similarity was defined across the whole image domain, thus coining the term: “non-local filtering”. In object detection, [Shechtman and Irani, 2007] used a sparse ensemble of self-similarity descriptors calculated with a search radius of 40 pixels, which was stored in a log-polar histogram. For the use of MIND in image registration, however, a smaller search region was found to be sufficient. We have defined two different types of spatial sampling for the spatial search region R : dense sampling and a six-neighbourhood. Figure 6.4 illustrates these configurations, where the red voxel in the centre is the voxel of interest, and all the grey voxels define R . The computational complexity is directly proportional to the number of sampled displacements, and therefore the

six-neighbourhood clearly offers the best time efficiency. The radius of the search range can be increased to capture a larger neighbourhood. However, if the chosen neighbourhood is too large, the resulting descriptor might be affected by non-rigid deformations.

MIND is not invariant to rotations. While this would be a disadvantage for feature-based registration approaches, for dense deformable registration the local orientation helps to drive the alignment of edges. Note, that in practice the MIND image representation can be efficiently re-computed at each iteration of the registration.

6.4.2 Extension to the self-similarity context

The concept of self-similarity can be further extended to be based on spatial contextual information [Heinrich et al., 2013c]. Spatial context has been successfully applied to object detection [Tu, 2008], [Heitz and Koller, 2008] and is also the main driving force of pictorial structures [Felzenszwalb and Huttenlocher, 2005]. Here our aim is not to find a good structural representation of the underlying shape, but rather the context within its neighbourhood. An example of a patch configuration for the proposed self-similarity context (SSC) is shown in Fig. 6.5. We make use of the same images patches within the six-neighbourhood as before. However, now the patch distances are not defined with respect to the central patch, but all other patches within the neighbourhood. The advantage of this is that a noisy central patch does not adversely affect the descriptor calculation. Since the number of distance comparisons obviously increases quickly by including more patches, we therefore restrict the number of patches involved to the six-neighbourhood and

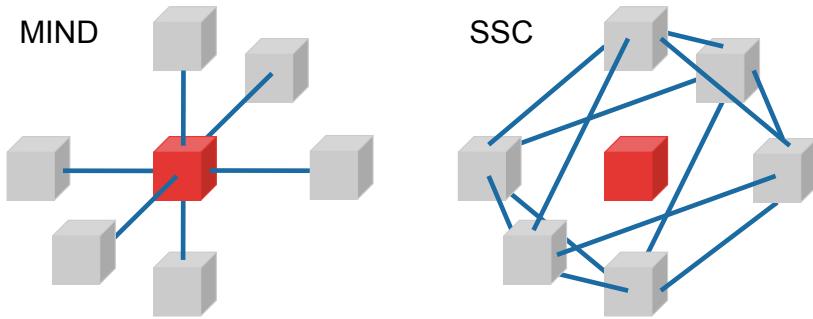


Figure 6.5: Concept of self-similarity context (SSC) compared to MIND with six-neighbourhood (6-NH). The patch around the voxel of interest is shown in red, all patches within its immediate 6-NH in grey. Left: All patch distances (shown with blue lines) used for MIND within 6-NH take the centre patch into account. Right: Geometrical and/or structural context can be better described by SSC using all patch to patch distances, of which none is dependent on the central patch.

neglect the three largest spatial connections (left-most to right-most, etc.) so that there are 12 distances (of which only half have to be calculated, since each distance is part of two descriptors within the whole image). All other steps described in Secs. 6.4.1 and 6.4.1 to obtain the result of Eq. 6.14 are performed equivalently.

6.4.3 Multi-modal similarity metric using MIND

One motivation for the use of MIND is that it enables multi-modal images to be aligned using a simple similarity metric across modalities. Once the descriptors are extracted for both images, yielding a vector for each voxel, the similarity metric between the two images can be chosen to be SAD or SSD between their corresponding descriptors. Therefore efficient optimisation algorithms, which converge rapidly, can be used without further modification. Using SAD as an example, the similarity term $\mathcal{S}(\mathbf{x})$ of two images I and J at voxel \mathbf{x} can be defined as the

SAD between the descriptors:

$$\mathcal{S}(\mathbf{x}) = \frac{1}{|R|} \sum_{\mathbf{r} \in R} |\text{MIND}(I, \mathbf{x}, \mathbf{r}) - \text{MIND}(J, \mathbf{x}, \mathbf{r})| \quad (6.20)$$

For the Gauss-Newton optimisation (see Chapter 7) used here, which minimises the linearised error term in a least-square sense ([Madsen et al., 2004]), we simply treat the MIND representation as a multi-channel image. The derivatives with respect to deformations in the spatial directions are calculated for each element of the descriptor individually. We will discuss this in more detail in the context of the continuous optimisation for deformable registration used in Sec. 7.3.3.

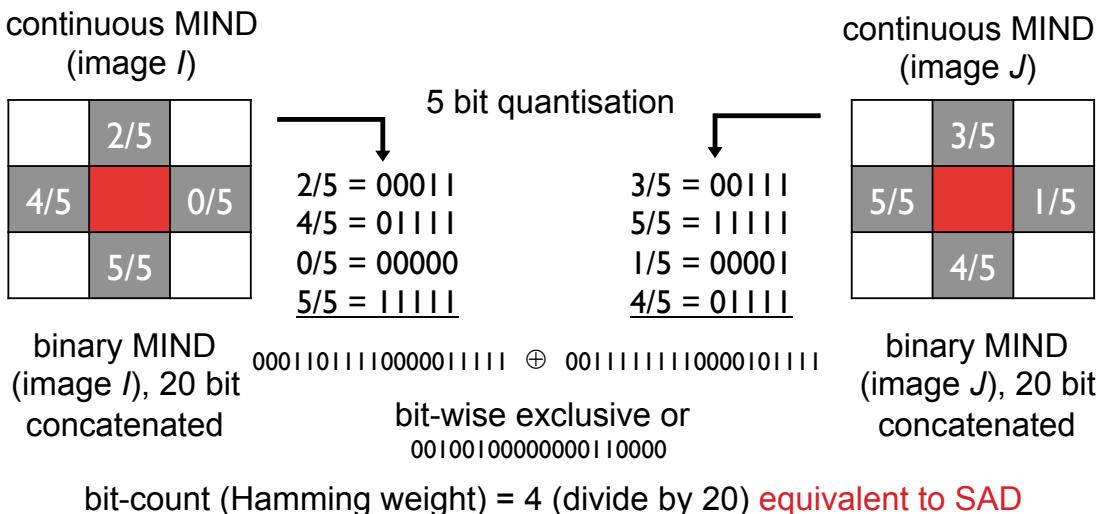


Figure 6.6: Concept of using the Hamming distance to speed up similarity evaluations. Continuous valued descriptor entries (here: $|R| = 4$) are quantised to a fixed number of bits and concatenated. The similarity of Eq. 6.20 can now be evaluated using only one bitwise XOR and bit-count operation.

Eq. 6.20 requires $|R|$ computations to evaluate the similarity at one voxel. Some algorithms, especially the discrete optimisation framework presented in Chapter 7 (or [Glocker et al., 2008a] and [Shekhovtsov et al., 2008]), as well as the land-

mark localisation experiments, use many cost function evaluations per voxel. In order to speed up these computations the descriptor can be quantised to a vector of only 64 bit, without significant loss of accuracy. The exact similarity evaluation of Eq. 6.20 can then be obtained using the Hamming distance between two descriptors using only one operation per voxel (a speed-up of $|R|$). A descriptor using self-similarity context as the neighbourhood definition consists of 12 elements, for which we use 5 bits per element, which translates into 6 different possible values (note that we cannot use a quantisation of 2^5 because the Hamming weight only counts the number of bits, which differ). Figure 6.6 illustrates the concept.

In the following section, the suitability of the two novel multi-dimensional structural image representations STORM and MIND for multi-modal image similarity and their robustness to parameter choice and intensity distortion is experimentally studied and compared to the state-of-the-art: entropy images [Wachinger and Navab, 2012].

6.5 Experiments and results

The performance of the presented structural image representations is compared in the following with the same experiments used in the previous chapter. Details of this similarity-based landmark localisation are described in Sec. 3.3. As before, two datasets as used: T1- and PD-MRI scans of Visible Human Dataset (VHD) and an inhale-exhale lung CT scan pair. The image resolutions are $1.875 \times 1.875 \times 1.875$ mm and $0.97 \times 0.97 \times 2.5$ mm respectively. The initial average displacement for the CT dataset is 15.0 mm (the intra-observer error for manual annotations, which is the lower bound on localisation accuracy is ≈ 1 mm) [Castillo et al., 2009].

6.5.1 Parameter sensitivity

In the following the sensitivity of all metrics with respect to the spatial Gaussian smoothing σ_s is tested. Additionally, the influence of the number histogram bins on the calculation of entropy images, and the neighbourhood layout for MIND are evaluated.

Entropy images

Figure 6.7 (a,b) shows the results for using SAD of the entropy images for the two experiments (landmark localisation in multi-modal MRI scans and inhale-exhale CT). The number of bins and the spatial Gaussian smoothing with σ_s are varied over a range of values. The size of the smoothing kernel is $\rho_s = 2\lceil \frac{3}{2}\sigma_s \rceil + 1$ and the Parzen smoothing is set empirically to $\sigma_i = 0.75$. A local normalisation is performed within a neighbourhood \mathcal{N} of size ρ_s^3 . It can be seen that the Gaussian smoothing kernel has a great influence on the results (values in the range of $\sigma_s = [0.5, 0.875]$ appear to be best), whereas the entropy calculation is fairly insensitive to the number of bins.

Structure tensor orientation measure (STORM)

Figure 6.7 (c) shows the results for using SAD of STORM for the two experiments (landmark localisation in multi-modal MRI scans and inhale-exhale CT). Only a single parameter the Gaussian smoothing of gradients σ_g has to be chosen. The size of the smoothing kernel is $\rho_s = 2\lceil \frac{3}{2}\sigma_g \rceil + 1$. It can be seen that the results are fairly sensitive to σ_g (values in the range of $\sigma_g = [0.625, 1.5]$ appear to be best) and the absolute landmark localisation accuracy is substantially better than using

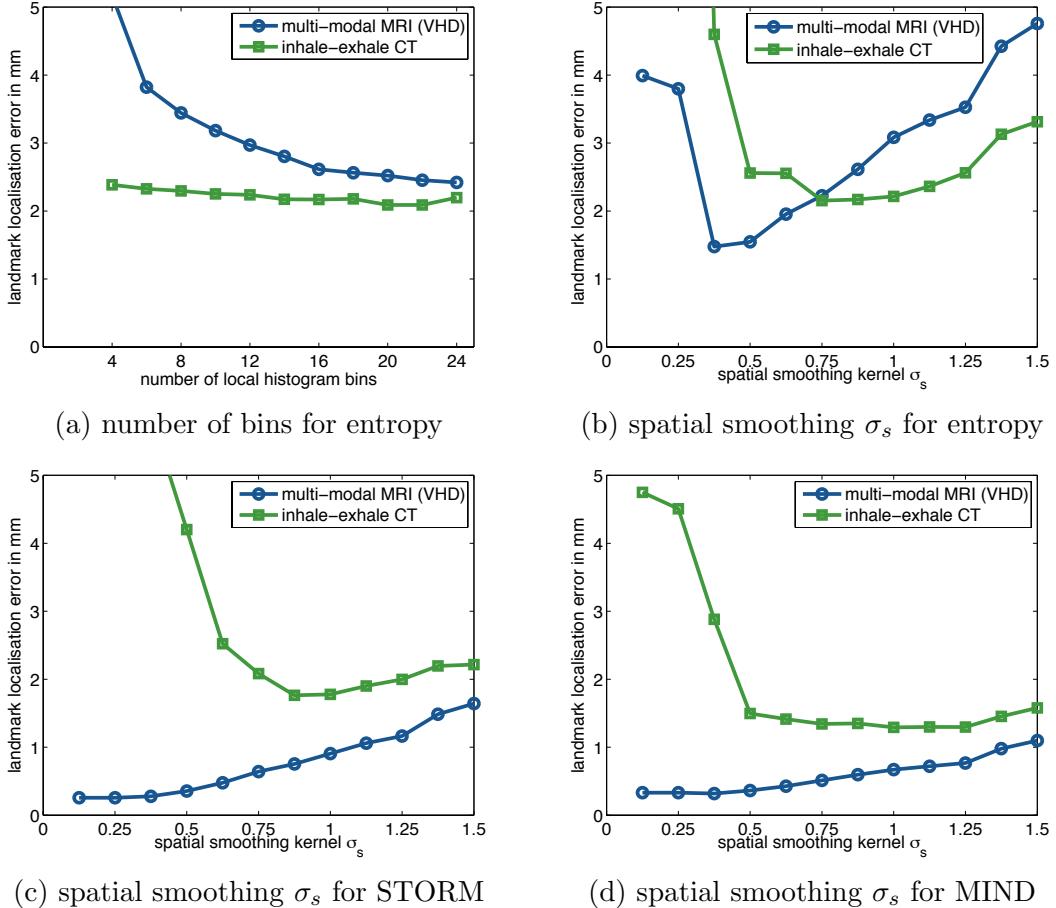


Figure 6.7: Landmark localisation error in mm for multi-modal MRI (\ominus) and 4D-CT dataset (\square). The size of smoothing kernel and neighbourhood for local normalisation is $\rho_s = 2\lceil \frac{3}{2}\sigma_s \rceil + 1$. The Parzen smoothing is set to $\sigma_i = 0.75$ in all cases. (a) The number of histogram bins has little influence on localisation accuracy (using $\sigma_s = 0.875$). (b) The variation of the spatial Gaussian smoothing, σ_s , shows a high sensitivity of entropy images to this parameter (using 16 bins). The sensitivity to σ_s is low for STORM (c) and MIND (d). MIND achieves the best results for the inhale-exhale CT dataset.

entropy images (see Fig. 6.7 (b)).

Modality independent neighbourhood descriptor (MIND)

The influence of the Gaussian weighting parameter σ_s for the patch distance Eq. 6.15 for the two landmark localisation experiments is shown in Fig. 6.7 (d). It can be seen that similarly good results are obtained for a large range of values for $\sigma_s > 0.5$. Figure 6.8 (left) gives an overview of the landmark localisation error for the inhale-exhale CT dataset for different neighbourhood configurations of MIND compared to SSC. It can be seen that SSC is consistently better than using either a six or 26 neighbourhood (6-NH, 26-NH). The influence of the Gaussian weighting σ_s for the patch distance computation is also lower. SSC is also more robust against additional Gaussian noise applied to the inhale-exhale CT scans as shown in Fig. 6.8 (this experiment was not performed for the other methods). We conclude that avoiding the central patch for self-similarity computations greatly reduces the adverse affect of image noise. Therefore, SSC is better suited (than MIND) for challenging modalities with strong noise. This is further demonstrated for the registration of ultrasound to MRI in Sec. 9.3.1.

6.5.2 Robustness against intensity distortions

The presented structural image representations have been compared in terms of landmark localisation error for the multi-modal VHD and inhale-exhale CT datasets. As similarity metric, sum of absolute differences (SAD) is used throughout the tests. Entropy images are simply represented by scalar values, STORM with a vector of length 3, and MIND with a 64 bit integer (obtained from the

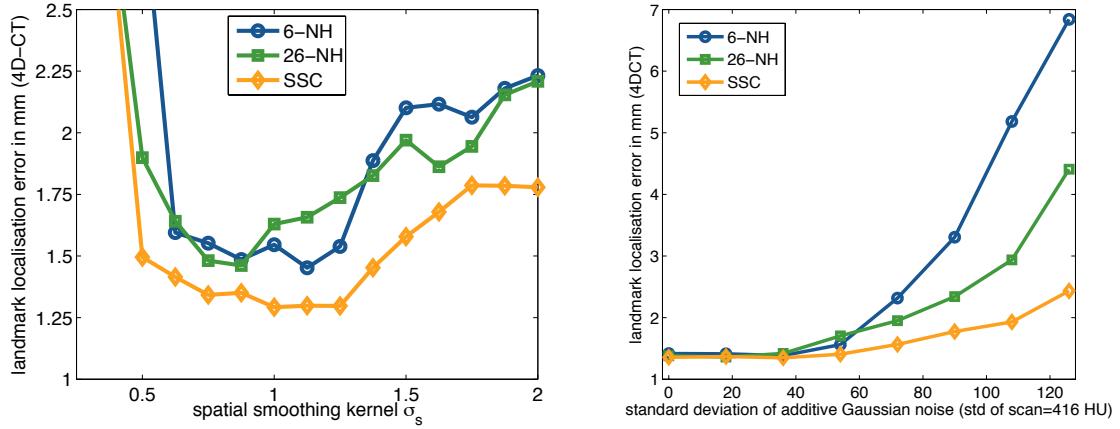


Figure 6.8: Left: Results for the landmark localisation experiments for MIND with different spatial search regions, using the inhale-exhale CT dataset. The consistently best accuracy (over varying Gaussian weightings σ_s of the patch distances) is obtained using SSC. Right: SSC is also more robust against additional Gaussian noise applied to the inhale-exhale CT scans.

quantisation described above and compared by Hamming distances). Figure 6.9 shows the cumulative distribution of the localisation error for the inhale-exhale CT dataset for SAD of the three representations, SAD of image intensities and the initial landmark distance (average initial displacements are 15 mm).

All structural representations achieve better results than using SAD on the original intensities, which yields an localisation error of 5.77 mm. MIND, using the self-similarity context (SSC) neighbourhood, achieves the best results (1.29 mm), followed by STORM (1.76 mm) and entropy images (2.15 mm). We can therefore conclude that the scalar entropy representation only, based on the local information content (Shannon entropy), is not sufficiently discriminative to match complex geometric features. Both MIND (implicitly) and STORM (explicitly) make use of the local orientation, which indicates that this is one of the main driving forces for multi-modal image matching.

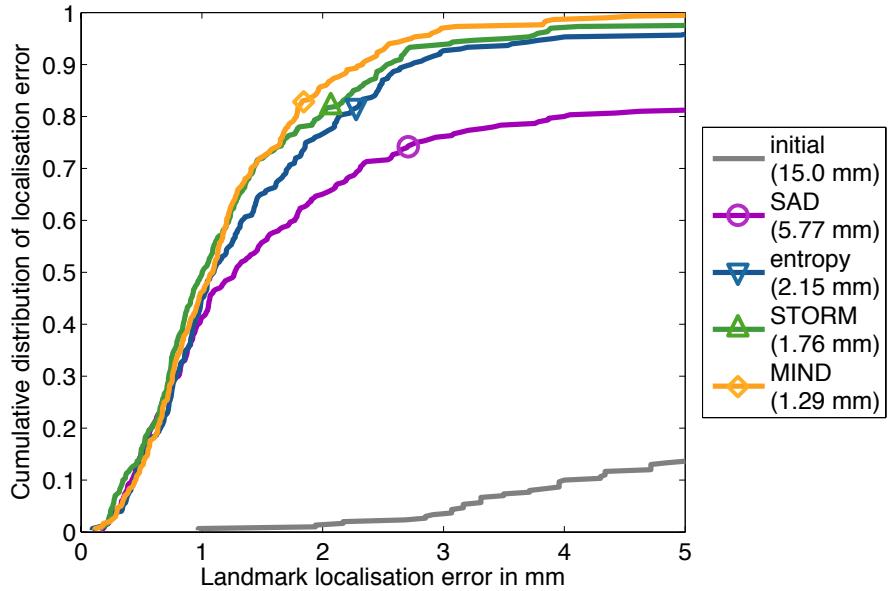


Figure 6.9: Cumulative error distribution for SAD of the various image representations for the inhale-exhale CT dataset. MIND (with SSC neighbourhood) achieves the best results (with an accuracy close to the ground truth inter-observer error) and an improved robustness (fewer large errors) compared to STORM and entropy images (the improvements are statistically significant with $p = 0.002$ and $p = 0.033$). STORM gets closest to the true landmark location for many landmarks, but is less robust than MIND. SAD of intensities shows a significant error for about a third of the landmarks.

MIND achieves the highest robustness and a localisation distance of more than one voxel ($=2.85$ mm) from the ground truth position for only 4.7 % of landmarks, while this ratio is 6.7 % for STORM and 9 % for entropy images. We conclude that this improvement over STORM is mainly due to additional use of contextual information (especially using SSC). The robustness of the representations is further evaluated on the multi-modal VHD images against intensity distortions: a locally varying multiplicative bias field and additive Gaussian noise (see Sec. 3.3.2). Figure 6.10 demonstrates the robustness of both MIND and STORM to the additive noise. The localisation accuracy of entropy images deteriorates sub-

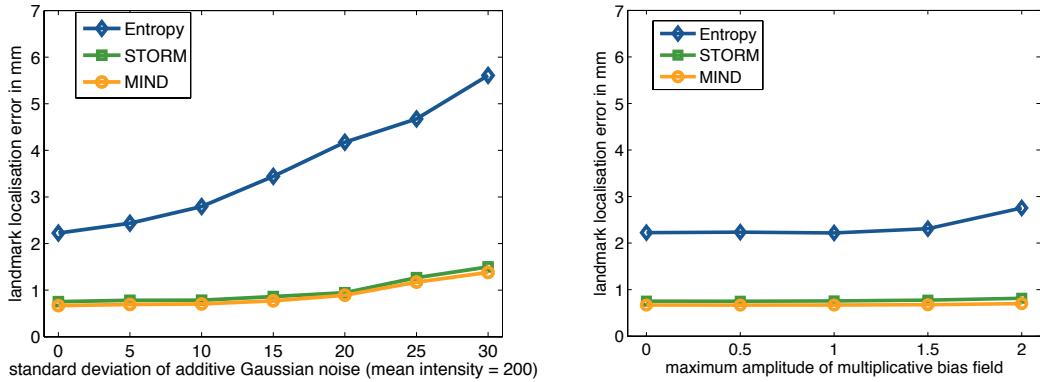


Figure 6.10: Landmark localisation error for VHD images with simulated intensity distortions. Entropy images are strongly affected by additive Gaussian noise, while both STORM and MIND are very insensitive to image distortions (and achieve substantially better results than the entropy representation for the original images). All metrics are nearly invariant to the multiplicative bias field (linearly varying from left to right).

stantially with increased noise, confirming the intuitive concern that entropy is, by definition highly dependent on the image SNR (in fact in [Tsai et al., 2008] entropy was used to assess the quality of radiographs). A denoising of the images prior to the entropy calculation might help, but would be less principled than STORM or MIND, which can intrinsically deal with noise. All representations are nearly unaffected by locally varying bias fields, which is a strong argument in favour of using structural image representations, since it removes the need for an additional treatment of intensity variations as discussed in Sec. 3.3.2.

6.6 Discussion of multi-modal similarity metrics and representations

In this and the previous chapter, we have presented a thorough analysis and experimental evaluation of three state-of-the-art multi-modal similarity metrics and three novel contributions, which have been made during this thesis. The methods are divided into two different groups: statistical similarity metrics and structural image representations (optimised with SAD). Based on our experiments, the latter approach has advantages both in terms of generalisation to different optimisation strategies and in achieved accuracy, which has been evaluated using regional landmark localisation on two challenging datasets (multi-modal MRI scans and inhale-exhale CT scans) and with respect to additional intensity distortions.

Figure 6.11 shows a comparison of the best performing approaches from both groups: conditional and textural mutual information (CMI, TMI), structure tensor based orientation (STORM) and modality independent neighbourhood descriptors (MIND) in terms of localisation accuracy. The two proposed methods, which make additional use of contextual information (TMI and MIND) show improved robustness and a reduction in the number of landmarks with large localisation errors. Entropy images and normalised mutual information, despite achieving an improvement compared to SAD of the original intensities (for the CT dataset), have been shown to be inferior to the other methods. Especially the former with respect to additive noise and the latter with respect to multiplicative bias fields. Textural mutual information achieves the highest accuracy for the VHD dataset (without additional intensity distortions) of 0.59 mm, closely followed by MIND with 0.67 mm.

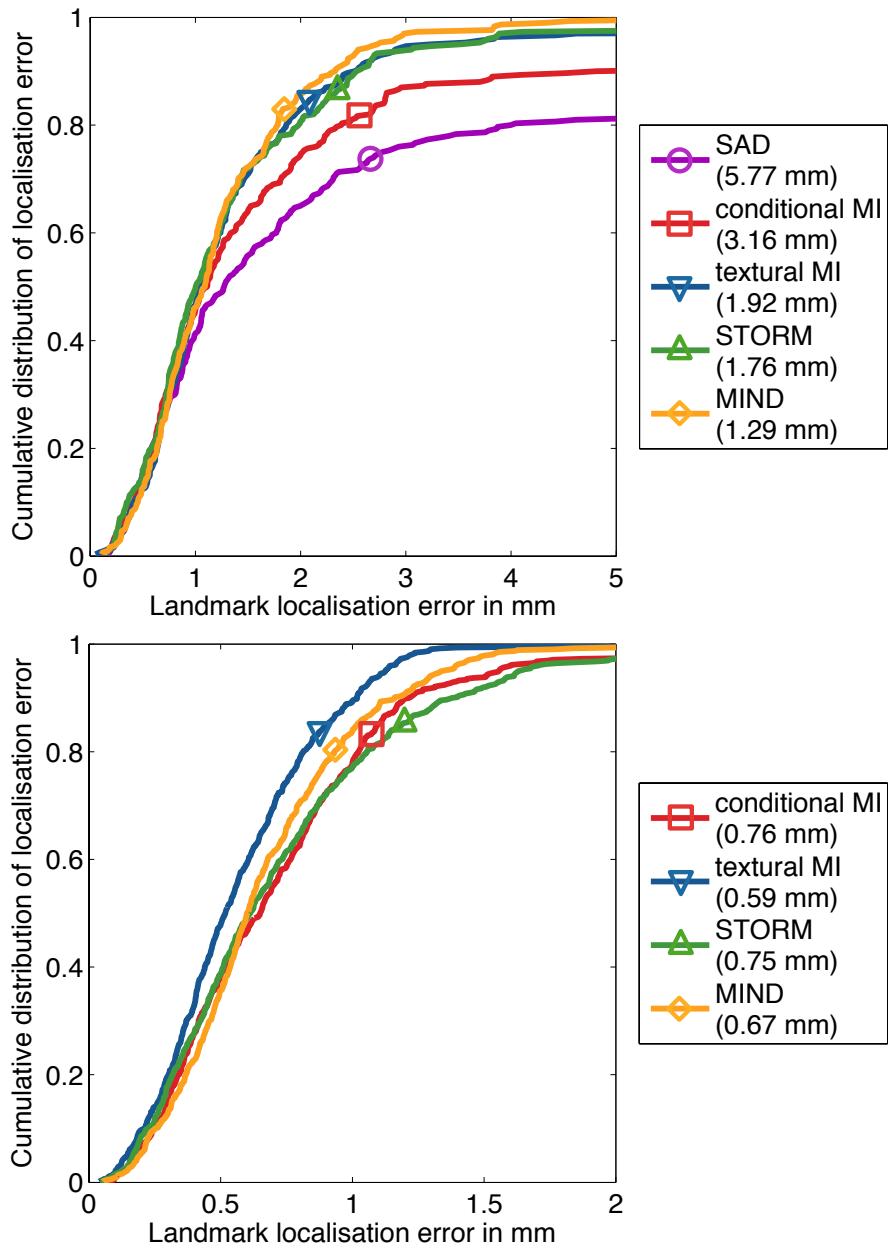


Figure 6.11: Cumulative error distribution for best performing methods. Top: MIND (with SSC neighbourhood) achieves the best results (1.29 mm) for the inhale-exhale CT dataset. Bottom: TMI achieves a slightly higher accuracy (0.59 mm) for the multi-modal VHD images (without additional distortions).

Efforts have been made to make the newly proposed methods both computationally very efficient, applicable to different optimisation methods, less dependent on parameter settings and therefore user friendly. Using the self-similarity context (SSC) neighbourhood for the MIND representation has been shown to be the best approach based on these experiments, because it makes the most use of contextual information and achieves both the highest robustness and accuracy. It has only one free parameter to set (a wide range of values gives similarly good results c.f. Fig. 6.8) and can be efficiently calculated with linear complexity (in ≈ 5 seconds for the highest resolution of one of our test images). The use of quantised descriptors and the Hamming distance (see Fig. 6.6) require no more operations (and no additional memory) for the similarity cost evaluations than SAD or SSD (of scalar intensities with double precision).

6.6.1 Outlook: Metric learning

In the previous two chapters, we have presented a number of current and novel approaches to define a model of similarity across scans, with particular focus on multi-modal data. Another area, which has currently gained a lot of interest in research, is to learn a metric or model of similarity from training data. The concept of textural mutual information (see Sec. 5.3) uses unsupervised learning to cluster data points based on their relationships (pair-wise distances of some chosen metric) without training data.

Choosing the right metric and the weighting of elements of feature vectors to find correspondences across images is not a trivial task. Therefore, hand-crafting these models might not be ideal for all applications. [Ou et al., 2011] trained the

optimal weighting of a large number of Gabor wavelet features using aligned multi-modal images. An improvement of the registration accuracy was found using only a subset of elements of the original feature vectors. For more general clustering applications a number of distance metric learning approaches have been presented in the past. [Xing et al., 2002] use a supervised learning, where point pairs (x, y) have been labelled as similar or dissimilar. The aim is to learn a distance metric $d(x, y)$ (of the family of Mahalanobis distances), which is defined by a positive semi-definite matrix A :

$$d(x, y) = \|x - y\|_A = \sqrt{(x - y)^T A (x - y)} \quad (6.21)$$

A constraint is added to avoid the trivial solution of $A = 0$. When restricting A to be diagonal, this problem is easy to solve and results in a similar weighting scheme as used by [Ou et al., 2011]. Using a full matrix A also incorporates cross-element distances and has been optimised by [Xing et al., 2002] using gradient descent together with iterative projections. Note, that the this method does not explicitly deal with multi-modal data since all data points are organised in the same space.

[Weinberger et al., 2005] notice that for many tasks a tight clustering of all similar data points is less important than the discrimination between similar and dissimilar ones. Their approach: large margin nearest neighbour (LMNN), which is inspired by support vector machines (SVM) directly optimises the kNN classification performances. LMNN finds a matrix A , which maximises the margin (also called hinge loss) of dissimilar data point clusters. This approach can be implemented more efficiently than the one of [Xing et al., 2002] and an extension has been proposed by [Weinberger and Saul, 2008], which can also adapt multiple,

locally different metrics. A similar idea was used by [Tang et al., 2012], who learn a locally varying weighting of different similarity metrics in order to best reflect the complementary challenges of aligning different anatomies in medical scans. In their experiments, a small improvement was found for the non-rigid registration of lung and brain scans, when using a weighted combination of the metrics: cross-correlation, local mutual information, intensity differences and differences of image features (here vesselness filters).

While the previously discussed approaches, still rely on a pre-defined metric between different elements of the feature vectors (usually L1 or L2 norm), [Lee et al., 2009] propose to directly learn a linear mapping over a certain feature representation of image patches using max margin structured prediction [Tschantaridis et al., 2005]. They applied this technique successfully to linear registration of multi-modal brain scans with improved results compared to several mutual information variants. [Bronstein et al., 2010] learn a dimensionality reduction based on multi-modal training data, which embeds input data from two arbitrary spaces into a Hamming space, where a simple metric (L1 norm) can correctly represent the data similarity. Their approach, which can be also seen as a structural image representation, extends the work of [Shakhnarovich et al., 2003] on parameter (or similarity) sensitive hashing by learning two separate mappings (one for each modality). The disadvantages of the former two approaches are that they requiring well-aligned images from all modalities and a new similarity measure has to be learnt for each possible combination of modalities. [Wachinger and Navab, 2012] proposed an unsupervised learning of a low-dimensional embedding of the data using Laplacian eigenmaps to find scalar structural representations, which show promising results for 2D multi-modal images. This approach, however, comes at a

very high computational cost.

When using local image descriptors, such as MIND (see Sec. 6.4) for registration, not only the weighting of different elements of a descriptor vector, but also the spatial scale and location for calculating each entry can be learnt. [Simonyan et al., 2012] learn the spatial pooling region of gradient orientation features by a discriminative dimensionality reduction, which includes a sparsity constraint and can be formulated as a convex optimisation problem. A similar idea has been used by [Trzcinski et al., 2013] to learn pooling configurations for a binary descriptor using a boosting framework.

After the discussion and evaluation of different regularisation and similarity cost terms for deformable lung registration, the next chapter deals with the optimisation of these terms. The focus will lie on methods, which find good approximations to a global optimum of the cost function in tractable computation time. A gradient-based Gauss-Newton optimisation framework will be presented, which enables the minimisation of multi-dimensional image representations (such as MIND). This is followed by a more flexible gradient-free approach using discrete optimisation. A number of techniques are introduced to make this approach computational very efficient, while preserving excellent accuracy.

Chapter 7

Towards globally optimal energy minimisation

The objective of this chapter is to develop new concepts in optimisation of deformable image registration that address a number of shortcomings of the existing state-of-the-art. A focus lies on methods that are not only locally driven but able to robustly find an approximately global optimum. The main contribution is a new discrete optimisation framework that employs a dense sampling of displacements, and an efficient inference of regularisation using a minimum-spanning-tree.

Mathematical and numerical optimisation are at the core of every deformable image registration method. Registration methods consist in general of a transformation model, a registration term, a similarity term and an optimisation method. In chapters 4-6, we have developed new concepts for similarity and regularisation. This chapter deals with new approaches for more efficient and accurate optimisation. Optimisation for deformable registration is often ill-posed (or under-

determined), non-convex and non-linear. In the past, most research in the domain of medical image registration has focussed on the development of novel similarity metrics and transformation models, while reverting to general-purpose optimisation methods. The choice of the right optimisation framework is very much dependent on the registration problem. While there are a number of more universally applicable optimisation methods, these are rarely the best choice in terms of accuracy and computational complexity for a given medical application.

Iconic registration methods, employ the image intensity information densely, and have a number of important advantages compared to geometric – feature based approaches (they are less dependent on the employed transformation model). We will therefore focus on optimisation for intensity-based registration between target and moving image I_t and I_m , which aims to estimate the best permissible transformation ϕ^* :

$$\phi^* = \operatorname{argmin}_{\phi} E(\phi, I_t, I_m) \quad (7.1)$$

The energy term to be minimised, which consists of a similarity term \mathcal{S} and a regularisation penalty \mathcal{R} , can be written as:

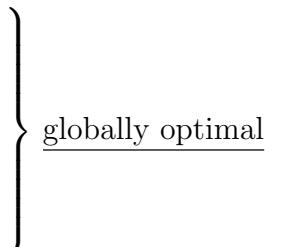
$$E(\phi) = \mathcal{S}(I_t, I_m, \phi) + \alpha \mathcal{R}(\phi) \quad (7.2)$$

where α is a scalar weighting between the two terms. The local transformation can then be described by a dense displacement field $\mathbf{u} = (u, v, w)^T$ and the identity transform: $\phi = \text{Id} + \mathbf{u}$. In order to obtain smooth deformations and a well-posed optimisation problem, parametric models (which include linear transformations) restrict ϕ to follow a specified type of transformation (see Sec. 4.2). Non-parametric registration methods mainly rely on penalising deviations in \mathcal{R} , but may also

impose further restrictions to obtain e.g. only diffeomorphic or symmetric transformations (see Sec. 4.4). Finding the global optimum of Eq. 4.1.1 is usually NP-hard and certain approximations and trade-offs have to be made to obtain a tractable computational complexity.

7.1 Optimisation for lung registration

The main objective of this chapter is to propose new strategies to improve the efficiency of the optimisation framework, and at the same time estimates a solution that is as close as possible to the registration cost term defined by Eq. 7.2. In the following, three major aspects of optimisation for deformable registration will be discussed:

- Generalisation to different types of similarity metrics (e.g. statistical, point-wise and multi-dimensional representations)
 - Robustness to local minima in energy function
and avoidance of bias of initialisation
 - Enforcing both local and global regularisation
and obtaining smooth, plausible deformations
- 

globally optimal

The importance of these aspects is motivated by the challenges that are present for the application of deformable lung registration in this thesis. The three major challenges are: complex intensity mappings across scans (different modalities, see discussion in Chapter 5) and within scans due to changing density, large motions of small features leading to local minima and large initial misalignments, and discontinuous sliding motion at the interface between lungs and rib cage (and

other organs). A detailed discussion of these three challenges was presented in Sec. 2.4.

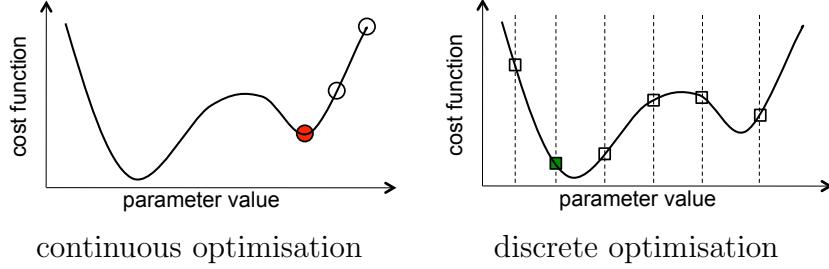


Figure 7.1: Comparison of continuous and discrete optimisation principles for a 1D cost function. Continuous optimisation usually finds a local minimum with high-accuracy within few cost function evaluations (\circlearrowright). Discrete optimisation requires the parameter space to be finitely quantised (\square), and it finds the approximate global optimum (up to a quantisation error) and is therefore more robust against a non-convex cost function and initialisation.

A demonstrative comparison between continuous and discrete optimisation and their respective advantages and shortcomings is shown in Fig. 7.1. Gradient-based methods are only guaranteed to find the global optimum of a convex cost function. A typical registration cost function is highly non-convex and therefore only a local minimum is guaranteed to be found, which might be far away from the global optimum in some instances. Multi-resolution or multi-scale approaches can help avoiding some of the local minima, but more complex motion consisting of both coarse and fine scale deformations remains problematic. Both the similarity and the regularisation term need to be differentiable. This most importantly limits the choice of possible similarity metrics or requires approximations to the derivatives. An overview of continuous optimisation methods for non-linear cost functions is given in [Madsen et al., 2004].

Simulated annealing [Kirkpatrick et al., 1983], is a combinatorial optimisation technique, which can theoretically find the global optimum to many problems by permitting the current parameter vector to escape out of local minima by accepting worse intermediate solutions that have a higher energy with a certain probability (determined by the annealing temperature). However, the high computational complexity limits its use in practical applications.

Graph-based discrete optimisation overcomes some aspects of the two main problems of continuous optimisation and therefore offers attractive properties for deformable medical image registration. However, this comes at the cost of highly increased computational complexity, because the space of deformations needs to be discretised and therefore the number of degrees of freedom is drastically increased. Consequently, suitable approximations have to be made so that the discrete optimisation problem becomes computationally tractable, while retaining its superior robustness against local minima, and in order to make full use of the higher flexibility of the search space and similarity metric.

The remainder of the chapter is organised as follows. In Sec. 7.2, we give a short overview of standard continuous optimisation methods for deformable registration. For a more comprehensive recent survey see [Sotiras et al., 2013]. A discussion of the popular and widely used demons framework and its extensions for diffeomorphic mappings is presented in Sec. 7.2.1. We highlight some of their shortcomings and motivate the use of a Gauss-Newton optimisation framework (see Sec. 7.3), which has been employed in parts of this thesis. Extensions to deal with multi-modal registration by using a multi-channel optimisation or preconditioning (see Sec. 7.3.3), the use of Gauss-Newton for linear registration (see Sec. 7.3.4), and an improved multi-resolution scheme (see Sec. 7.3.2) will be pre-

sented. In Sec. 7.4, the convergence of these continuous optimisation methods is experimentally compared for the registration of challenging inhale-exhale CT lung scans (in terms of registration accuracy and deformation complexity).

In Sec. 7.5 a discrete optimisation framework for deformable registration using a graph-based model [Bishop et al., 2006, Chapter 8] is presented as a very promising alternative to continuous optimisation. Less research has been directed at the use of MRF-based deformable registration for medical applications so far, mainly because of the large number of degrees of freedom involved. A range of new approaches is developed in Secs. 7.5.2–7.5.4, which reduce the complexity of the discrete optimisation problem with little sacrifice of robustness and accuracy of the estimated deformations, including:

1. a stochastic approximation for the similarity cost term (similar to [Klein et al., 2007] for continuous optimisation) with a dense spatial sampling
2. a simplified graph structure using an image-derived minimum-spanning-tree (see also Sec. 4.3), for improved efficiency of inference of regularisation
3. a multi-scale refinement with subvoxel accuracy for efficient belief propagation based on lower envelope computations
4. and an additional processing step, which ensures diffeomorphic and symmetric transformations (as discussed in Sec. 4.4.1)

The first two steps make use of a discrete, quantised displacement space $\mathcal{L} \Rightarrow \mathbb{Q}$, while the third and fourth step allow us to estimate continuous valued deformations $\mathbf{u} \Rightarrow \mathbb{R}$. The benefits of each individual part of our approach are evaluated in terms of registration accuracy for respiratory motion estimation in CT scans. The most

important differences are compared to the discrete optimisation framework **drop** [Glocker et al., 2008a].

A more comprehensive validation of the presented approaches for challenging clinical registration tasks (including a direct comparison to other state-of-the-art algorithms) will be left for Chapter 8.

7.2 Continuous optimisation for registration

Solving Eq. 7.2 requires a mathematical or numerical optimisation procedure. If both similarity \mathcal{S} and regularisation term \mathcal{R} are analytically differentiable or a derivative can be approximated numerically, the optimal transformation parameters can be found by searching for the extremal point, where $\partial E / \partial \mathbf{u} = 0$. Since the problem is non-linear, a linearisation has to be performed around the initial transformation (assuming reasonably small deformations) and the solution is found iteratively, following small steps in a search direction determined by the gradient of the cost function. Methods such as gradient descent (or steepest descent), conjugate gradient descent and Newton-like approaches differ in the way the step size is determined: explicit, semi-explicit or implicit (see [Zikic, 2011] for a more detailed discussion). Setting the right step size is very difficult and either too low convergence rates or unstable results are potentially problematic for the former two approaches [Zikic et al., 2010b]. Newton-like approaches make a quadratic approximation of the error term, which provides an automatic choice of the step size and much faster convergence. The disadvantage is they can only be applied if the energies can be stated as a least square problem (this is the case for SSD as similarity metric and diffusion regularisation). [Klein et al., 2007] demonstrated

the advantages of Newton-like methods over standard descent approaches (but also found that for B-spline registration with mutual information, stochastic gradient descent can be even more efficient). In Sec. 7.3 the choice of the Gauss-Newton method is motivated, which has also been widely used in optical flow estimation [Horn and Schunck, 1981], [Bruhn et al., 2005] and for spatial normalisation (inter-patient brain registration) [Avants et al., 2008]. While Newton’s method is applicable for all twice-differentiable functions, many image registration approaches can be formulated as non-linear least square problems. A comprehensive overview of optimisation methods to solve these problems, including Levenberg-Marquart [Ashburner et al., 2007] and Quasi-Newton can be found in [Madsen et al., 2004] or [Boyd and Vandenberghe, 2004]. An alternative approach, which decouples the optimisation of \mathcal{S} and \mathcal{R} , is the demons framework, which will be presented in the next section.

7.2.1 Demons framework and diffemorphism

The demons framework was first presented by [Thirion, 1998] and later popularised by its diffeomorphic [Vercauteren et al., 2009] and large-deformation [Beg et al., 2005] extensions. It avoids the difficulties of directly optimising the regularised cost function in Eq. 7.2. Instead, the optimisation is coarsely approximated and decoupled by alternatively minimising the similarity metric and application of Gaussian smoothing (as a substitute for the regularisation term). The so-called demons forces are related to the Gauss-Newton optimisation of the un-regularised registration cost, using a normalised SSD criterion [Pennec et al., 1999]. The

update force field \mathbf{f} given a previous or initial motion estimate \mathbf{u} , is obtained by:

$$\mathbf{f} = \frac{(I_t - I_m(\mathbf{u}))\nabla I_m(\mathbf{u})}{||\nabla I_m(\mathbf{u})||^2 + (I_t - I_m(\mathbf{x}))^2} \quad (7.3)$$

A Gaussian smoothing is applied twice: to the update field (approximating fluid regularisation) and the composition of the previous and the update field (similar to diffusion regularisation). Unfortunately, the setting of the parameters (standard deviation) of these two smoothing operations is far from trivial and the results are very sensitive to this choice.

An alternative formulation that restricts the transformations to be diffeomorphic was presented by [Vercauteren et al., 2009]. In practice a very simple approximation, the scaling-and-squaring approach (see Sec. 4.4, [Arsigny et al., 2006]), is used to obtain a diffeomorphic mapping for almost any displacement field.

While being reasonably efficient and easy to implement, this framework lacks most of the aforementioned desirable aspects of an optimisation framework. For example, it cannot be directly generalised to similarity metrics other than SSD. Although recently, efforts towards using other metrics have been made by [Zikic et al., 2010b], [Modat et al., 2010] and [Risser et al., 2011b]. Relying solely on local gradient information, it is susceptible to local minima and initial misalignments. Global regularisation can only be approximated using Gaussian smoothing, therefore it is in practice very difficult to obtain both globally and locally smooth transformations (some improvements may be achieved using simultaneous multi-scale smoothing c.f. [Risser et al., 2011a]).

To overcome these problems, we used a Gauss-Newton optimisation approach, which includes a regularisation penalty and thus avoids the decoupling of the

energy terms. It will be demonstrated in the Sec. 7.4 that the higher complexity involved with solving the coupled registration cost (which results in a 3×3 matrix inversion for each pixel) is offset by a much faster convergence, especially in areas with little intensity information. Additionally, there are fewer parameters to set and the sensitivity to their choice is in general lower.

7.3 Gauss-Newton optimisation

In order to explain the use of Gauss-Newton optimisation for registration, we first consider the general case of a non-linear vector function \mathbf{f} with $\mathbb{R} \Rightarrow \mathbb{R}^d$ [Madsen et al., 2004] (in our case $d = 3$ is the image dimension). The aim is to find the optimal coordinates \mathbf{x}^* , which minimise the least square error $\|\mathbf{f}(\mathbf{x})\|^2$:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{f}(\mathbf{x})\|^2 = \frac{1}{2} \mathbf{f}(\mathbf{x})^T \mathbf{f}(\mathbf{x}) \quad (7.4)$$

In order to efficiently solve the non-linear least square problem, we first need to linearise $\mathbf{f}(\mathbf{x})$ using the first-order Taylor series approximation: $\mathbf{f}(\mathbf{x} + \mathbf{h}) \approx \mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{h}$, where $\mathbf{J} \in \mathbb{R}^d$ is the Jacobian, a matrix containing first-order spatial derivatives of \mathbf{f} : $J_i = \frac{\partial \mathbf{f}}{\partial x_i}$ for each voxel. We can now estimate the optimal $\mathbf{x}^* = \mathbf{x} + \mathbf{u}$ iteratively by solving for the Gauss-Newton update step \mathbf{h}_{gn} (see [Madsen et al., 2004] for a full derivation):

$$(\mathbf{J}^T \mathbf{J}) \mathbf{h}_{gn} = -\mathbf{J}^T \mathbf{f} \quad (7.5)$$

The relation to Newton's method (see e.g. [Boyd and Vandenberghe, 2004]) is that the matrix of second derivatives (Hessian) is approximated by $\mathbf{J}^T \mathbf{J}$ using first-order

derivatives. The linearisation is accurate for small $|\mathbf{h}|$. Therefore, the convergence rate is quadratic when close to the final solution. This results in performance gains of Newton-like methods (as shown e.g. in [Klein et al., 2007]) compared to first-order descent methods, which have only a linear convergence rate. Using simple descent methods might also lead to inaccurate registration results, if stopping criteria cause the optimisation to finish before reaching the true minimum [Zikic et al., 2010b]. It is worth noting that the efficient second order minimisation (ESM), which claims to have quadratic convergence [Benhimane and Malis, 2004], and which has been employed for deformable registration in [Wachinger and Navab, 2012], is closely related to the Gauss-Newton optimisation. It only differs in using a symmetric formulation of the Jacobian: $\mathbf{J}_{\text{ESM}} = (\mathbf{J}(\mathbf{x}) + \mathbf{J}(\mathbf{x} + \mathbf{h}))/2$, and relates to the symmetric formulation of the demons framework [Vercauteren et al., 2008].

In the following, the registration framework based on Gauss-Newton optimisation, which has been further developed in this work will be presented.

7.3.1 Diffusion-regularised deformable registration

For deformable registration, we aim to minimise the following cost function with respect to the deformation field $\mathbf{u} = (u, v, w)^T$, consisting of a non-linear similarity term \mathcal{S} (dependent on \mathbf{u}) and a diffusion regularisation term $\mathcal{R}(\mathbf{u}) = ||\nabla \mathbf{u}(\mathbf{x})||^2$:

$$\operatorname{argmin}_{\mathbf{u}} \left(\sum_{\mathbf{x} \in \Omega} \mathcal{S}(I_t(\mathbf{x}), I_m(\mathbf{x} + \mathbf{u})) + \alpha \sum_{\mathbf{y} \in \mathcal{N}} \mathcal{R}(\mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{y})) \right) \quad (7.6)$$

where Ω defines the image domain and \mathcal{N} the chosen neighbourhood (here: von Neumann neighbourhood). When using a quadratic similarity term, the objective function to be minimised is of the form $\sum_i \mathbf{f}_i^2$ and we can apply the Gauss-Newton

optimisation method. We simplify the notation to $\mathcal{S} = \mathcal{S}(I_t(\mathbf{x}), I_m(\mathbf{x}))$ and $\nabla\mathcal{S} = (\frac{\partial\mathcal{S}}{\partial u}, \frac{\partial\mathcal{S}}{\partial v}, \frac{\partial\mathcal{S}}{\partial w})^T$.

The linearisation of $\mathcal{S}(\mathbf{x})$ for the SSD as similarity criterion $\|I_t(\mathbf{x}) - I_m(\mathbf{x} + \mathbf{u})\|^2$ is given by: $\mathcal{S}(\mathbf{x} + \mathbf{u}) \approx (I_t(\mathbf{x}) - I_m(\mathbf{x} + \mathbf{u})) + \nabla\mathcal{S}^T \mathbf{u}$. This is equivalent to the brightness constancy assumption utilised in optical flow estimation [Horn and Schunck, 1981], which has been first noted by [Zikic et al., 2010a].

The diffusion regularisation term $\mathcal{R}(\mathbf{x}) = \|\nabla\mathbf{u}(\mathbf{x})\|^2$ is quadratic with respect to \mathbf{u} since the differential operator is linear. The Laplacian of \mathbf{u} is defined as $\Delta\mathbf{u} = \nabla \cdot (\nabla\mathbf{u})$. The resulting update step given an initial or previous deformation field \mathbf{u}_{prev} becomes:

$$(\nabla\mathcal{S}^T \nabla\mathcal{S} + \alpha\Delta) \mathbf{u}_{gn} = -(\nabla\mathcal{S}^T \mathcal{S} + \alpha\Delta\mathbf{u}_{prev}) \quad (7.7)$$

Equation 7.7 can be rewritten as a system of linear equations of the form $\mathbf{A}\mathbf{u} = \mathbf{b}$. It can be solved using the successive overrelaxation (SOR) method [Young, 1954], which is a modification of the iterative Gauss-Seidel algorithm. It is computed by sequentially updating the values of \mathbf{u} with the iteration step:

$$u_i^{(k+1)} = (1 - \omega)u_i^{(k)} + \frac{\omega}{A_{ii}}(b_i - \sum_{j>i} A_{ij}u_j^{(k)} - \sum_{j<i} A_{ij}u_j^{(k+1)}) , \quad i = 1, 2, 3. \quad (7.8)$$

where k is the current number of iterations, A_{ij} and b_i the elements of \mathbf{A} and \mathbf{b} , respectively, and $\omega \in]0, 2[$ is a relaxation parameter. First-order derivatives of the point-wise similarity terms are calculated using finite differences [Kiefer and Wolfowitz, 1952]. The partial derivative of \mathcal{S} for each dimension x_i of the

deformation vector field is obtained by:

$$\frac{\partial \mathcal{S}}{\partial x_i} \approx \frac{\mathcal{S}(\mathbf{x} + \mathbf{e}_i) - \mathcal{S}(\mathbf{x} - \mathbf{e}_i)}{2} \quad (7.9)$$

where \mathbf{e}_i is a unit step in direction of x_i . For the case of SSD as similarity metric, this simplifies to: $\frac{\partial \mathcal{S}}{\partial x_i} = (J(\mathbf{x} - \mathbf{e}_i) - J(\mathbf{x} + \mathbf{e}_i))/2$.

The Laplacian of the deformation field, $\Delta \mathbf{u}$, is approximated by the six direct neighbours \mathcal{N} of a voxel $\Delta u_i = (\sum_{j \in \mathcal{N}} u_j) - 6u_i$. Note, that because $\Delta \mathbf{u}$ depends on the current values of u_i^k in Eq. 7.8, it has to be re-calculated for each SOR iteration. The relaxation parameter ω was set to 1.95 to achieve fast convergence. In our experiments we found that a fixed number of 25 inner fixed-point iterations was sufficient.

The final deformation field is calculated by the addition of the update steps \mathbf{u}_{gn} . The parameter α balances the similarity term with the regulariser, which is necessary to overcome the under-determination of the problem and ensures a globally smooth motion estimation. The value of α has to be found empirically.

7.3.2 Coarse-to-fine image registration

As mentioned in the introduction and shown in Fig. 7.1, continuous optimisation can be susceptible to local minima (for non-convex cost functions), and in particular it relies on a good initialisation. A common approach to improve the performance of deformable registration is to use a coarse-to-fine image representation [Unser et al., 1993]. This can be achieved by using a multi-resolution strategy and/or a parametric transformation model (such as B-spline transformations).

In a multi-resolution framework, image derivatives and deformation fields are

initially estimated for downsampled versions of the images. The resulting deformations are also applied to downsampled images for successive iterations. After convergence of the Gauss-Newton optimisation at a low resolution, the deformations are upsampled to initialise the estimation of the deformation fields at the next finer resolution. This can cause problems for small but often important anatomical structures (e.g. vessels), which undergo a large deformation (or motion), because the registration algorithm does not take them into account in lower resolutions.

A parametric transformation model (as discussed in Sec. 4.2) may be beneficial, it however substantially increases the complexity of the similarity gradient computation, because they have to be performed at the highest resolution (see [Klein et al., 2007] for a thorough discussion of this problem). In the following, it will be explained how the computation of the derivative of the similarity term can be simplified for parametric models with similar accuracy and much lower computational cost.

Assume, we are using a B-spline transformation model with a uniform control-point spacing of g . Using a multi-resolution strategy with the same downsampling factor of g , low resolution image representations are obtained by a filtering of the original images with the B-spline function, followed by a subsampling with a spacing of g . Afterwards, derivates are obtained by a finite differences approximation of Eq. 7.9 with a unit step size $|\mathbf{e}| = 1$. For a parametric registration, similarity derivatives are first calculated using the original image resolution but only evaluated for a subset of control points (here the number of control points is equivalent to the low-resolution image dimensions). Both, B-spline filtering and derivative estimation are linear operators. Thus, for the specific choice of a fixed step size of $|\mathbf{e}| = g$ for derivative approximations, the parametric scheme

yields the exact same coarse scale derivatives for the similarity term as the multi-resolution approach (which has a much lower computational cost, especially when using multi-dimensional image representations). In contrast to a standard multi-resolution scheme, the obtained deformations after each Gauss-Newton update step (see Eq. 7.7) have to be upsampled and applied to the images at their original resolution (which is slightly more complex). We refer to the this modified approach as *parametric* in the experimental section 7.4.

7.3.3 Gauss-Newton for multi-modal registration

In general Eq. 7.9 can be applied to arbitrary quadratic point-wise similarity metrics, if we assume their Jacobian is block-sparse (the similarity criterion is only dependent on its immediate neighbours). Mutual information, unlike SSD or NCC, is usually not minimised in a least-square sense, nevertheless [Wachinger and Navab, 2009] show similarly good results for MI² as a similarity metric. It might however be useful in some cases to use a pre-conditioner to improve the stability of the optimisation and avoid a diverging solution, especially in cases of complex deformations and many local minima in the cost function. [Zikic et al., 2010b] propose to use an approximate normalisation of the point-wise gradient vectors, thus relying more on the direction of the derivatives than their amplitude. For SSD-based similarity metrics this is not necessary, since here vectors are already approximately normalised as discussed in Sec. 7.2.1.

Another possibility to use Gauss-Newton in multi-modal deformable registration is to use structural image representations as discussed in Chapter 6. If a scalar representation such as entropy images is used, Gauss-Newton optimisation can be

directly used with an SSD similarity of these structural representations. For multi-dimensional image representations (e.g. MIND, see Sec. 6.4) with r channels, the Jacobian \mathbf{J} has to be represented as a matrix ($\mathbb{R}^{r \times 3}$) and not as a vector ($\mathbb{R}^{1 \times 3}$) for each voxel. The dot products of partial derivatives (and function values) to obtain the terms $\mathbf{A} = \mathbf{J}^T \mathbf{J}$ and $\mathbf{b} = -\mathbf{J}^T \mathbf{f}$ in Eq. 7.3 therefore become products of matrices (instead of vectors), involving a summation of partial derivatives over all channels. The partial derivatives are obtained for each channel individually by applying Eq. 7.9. In order to save memory the partial derivatives and function values are computed on-the-fly and added to the matrix \mathbf{A} and the vector \mathbf{b} for each voxel.

7.3.4 Rigid registration with Gauss-Newton optimisation

Rigid-body image registration aims to find the best transformation to align two images while constraining the deformation to be parameterised by a rigid-body (translation and rotation, 6 parameters). Extending this model to the more general affine transformation, the transformed location $\mathbf{x}' = (x', y', z')^T$ of a voxel $\mathbf{x} = (x, y, z)^T$ can be parameterised by $\mathbf{q} = (q_1, \dots, q_{12})$:

$$\begin{aligned} u &= x' - x = q_1 x + q_2 y + q_3 z + q_{10} - x \\ v &= y' - y = q_4 x + q_5 y + q_6 z + q_{11} - y \\ w &= z' - z = q_7 x + q_8 y + q_9 z + q_{12} - z \end{aligned} \tag{7.10}$$

where $\mathbf{u} = (u, v, w)^T$ is the displacement of \mathbf{x} . For a quadratic image similarity function \mathbf{f}^2 , the Gauss-Newton method can also be applied here. We again use a

linear approximation of the error term:

$$\mathbf{f}(\mathbf{x}') \approx \mathbf{f}(\mathbf{x}) + \mathbf{J}^T(\mathbf{x})\mathbf{u} \quad (7.11)$$

where Jacobian $\mathbf{J}(\mathbf{x})$ is the derivative of the error term with respect to the transformation. The iterative update step \mathbf{u}_{gn} is given by Eq. 7.3. We insert Eqs. 7.10 into Eq. 7.11 and differentiate with respect to \mathbf{q} to calculate $\mathbf{J}(\mathbf{x})$. The advantage of this method is that we can directly use the point-wise spatial derivatives of the similarity term to obtain an affine transformation, so that structural image representations have to be computed only once per iteration.

Parameterising a rigid-body transformation directly is not possible in a linear way. Therefore, at each iteration the best affine matrix is first estimated and then the best rigid-body transformation is found using the solution presented in [Arun et al., 1987].

7.4 Evaluation of continuous optimisation

In order to evaluate the performance of the aforementioned continuous optimisation strategies, a number of registration experiments are performed with inhale-exhale CT scan pairs. In this section, we used the two extreme breathing phases of the most challenging cases (#7 and #8) of the DIR-Lab 4DCT dataset [Castillo et al., 2009]. For each scan, 300 anatomical landmarks have been carefully annotated by thoracic imaging experts. The average landmark error before registration is 13.4 mm for these two scan pairs, the expert landmark annotation error ≈ 1 mm, and the scan resolution is $0.97 \times 0.97 \times 2.5$ mm. An example of coronal and

axial slices at the two extreme breathing phases is shown in Fig. 7.2 (top row).

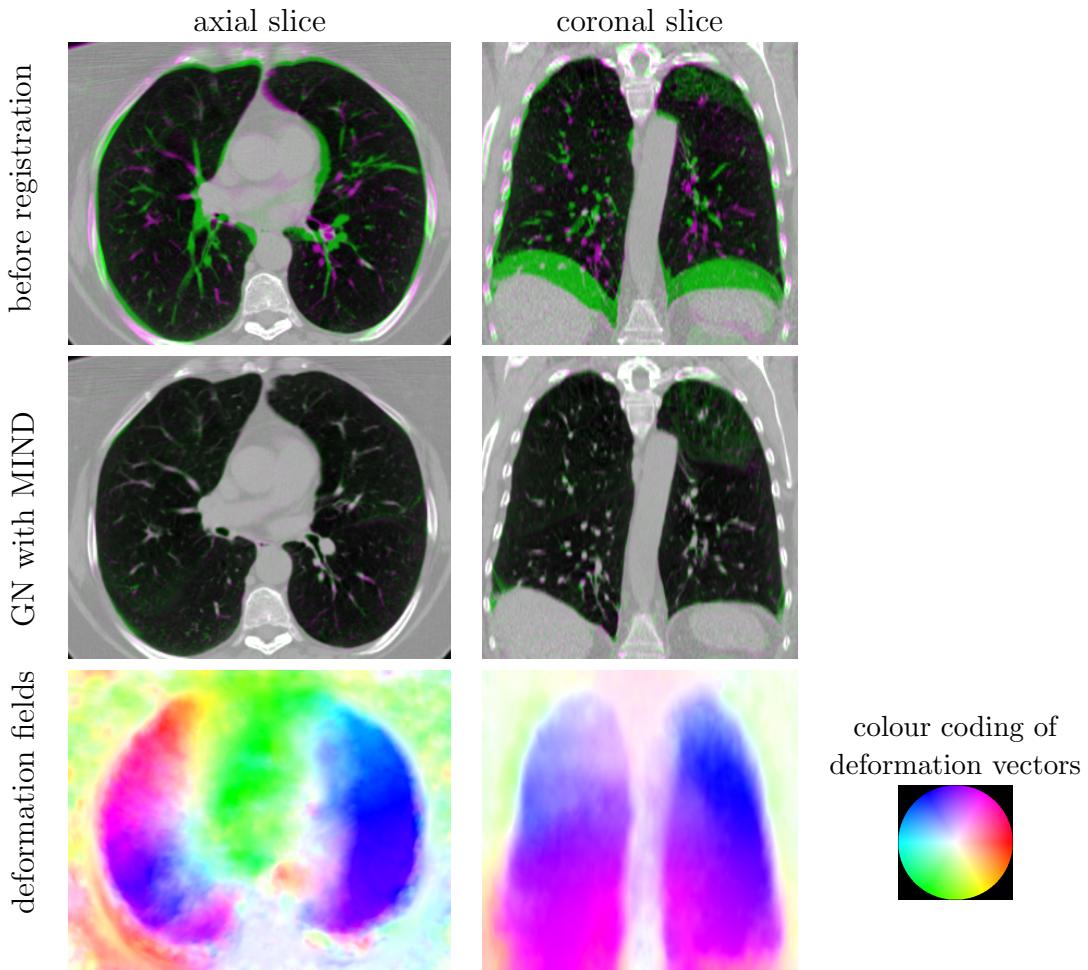


Figure 7.2: Visual outcome of deformable registration for case #8 of DIR-lab dataset. Top row shows overlay of inhale (green) and exhale (magenta) scan before registration. Centre row demonstrates the accurate alignment after registration using Gauss-Newton optimisation and MIND representations. Bottom row shows the corresponding deformation fields.

SSD (of image intensities or MIND representations) is used as similarity metric, and the smoothness of deformations is enforced by diffusion regularisation (see Sec. 4.1.1), which is approximated using a Gaussian smoothing for the demons framework. We compare five different approaches:

1. symmetric diffeomorphic demons [Vercauteren et al., 2009] (see Sec. 7.2.1),
2. Gauss-Newton (GN) without symmetric constraints (see Sec. 7.3.1),
3. GN using a multi-resolution scheme (see Sec. 7.3.1),
4. GN using a parametric multi-scale approach (see Sec. 7.3.2), and
5. multi-channel parametric GN optimisation using MIND (see Sec. 7.3.3).

For all but the second method, the symmetric and inverse-consistent approach, detailed in Sec. 4.4.1, is used. The resulting transformations show on average a negligible inverse consistency error of less than 0.2 voxels, and the number of negative Jacobians (singularities in the deformation field) is close to 0 (a fraction of $<10^{-5}$).

An iterative optimisation is performed using 10 iterations for each of four resolution levels (with subsampling factors of 8, 4, 2 and 1). We found that diffeomorphic demons had a substantially slower convergence, most likely because the de-coupling of similarity minimisation and regularisation results in a suboptimal optimisation overall. We therefore use twice as many iterations for the demons framework. The regularisation parameters were chosen experimentally to obtain the highest registration accuracy. For diffeomorphic demons a fluid-like Gaussian smoothing with $\sigma_{\text{fluid}} = 0.25$ voxels and a diffusion filtering with $\sigma_{\text{diffusion}} = 1.0$ showed the best performance. A step size of $\alpha_{\text{step}} = 5$ (see details in [Vercauteren et al., 2009]) helped improve the speed of convergence. An advantage of the presented Gauss-Newton optimisation is that only a single parameter α (which weights the regularisation) has to be set. We found that $\alpha = 100$ was best for SSD directly

applied to intensities, and $\alpha = 0.3$ for the MIND representations (using the self-similarity context as spatial search region see Sec. 6.4.2, and a patch-distance parameter $\sigma = 0.75$).

Figure 7.3 shows the evaluation of the different optimisation strategies in terms of registration accuracy over the number of iterations and the distribution of the final registration error. It can be seen that diffeomorphic demons (using twice as many iterations) achieves worse results than our symmetric Gauss-Newton approach. A significant improvement can be found when using the symmetric compared to the asymmetric approach. The parametric coarse-to-fine scheme achieves slightly better results than a standard multi-resolution approach, because at each iteration the deformation field is applied to the high resolution original images - preserving the motion of smaller details. Using the multi-channel optimisation of MIND representations results in a further significant improvement and the best overall performance (at the cost of slightly higher computation time, see Fig. 7.4 (right)). Figure 7.4 shows the evaluation of the complexity of deformations in terms of standard deviation of Jacobians during optimisation. Demons yield slightly less complex transformations.

7.5 Efficient MRF-based discrete optimisation

As discussed in the introduction, continuous optimisation has a number of drawbacks in particular when dealing with complex motion (i.e. respiratory lung motion), since it intrinsically relies on a convex cost function. This can be alleviated using a multi-resolution or multi-scale (parametric) transformation model, but it is still prone to local minima and there is no explicit control over the space of possi-

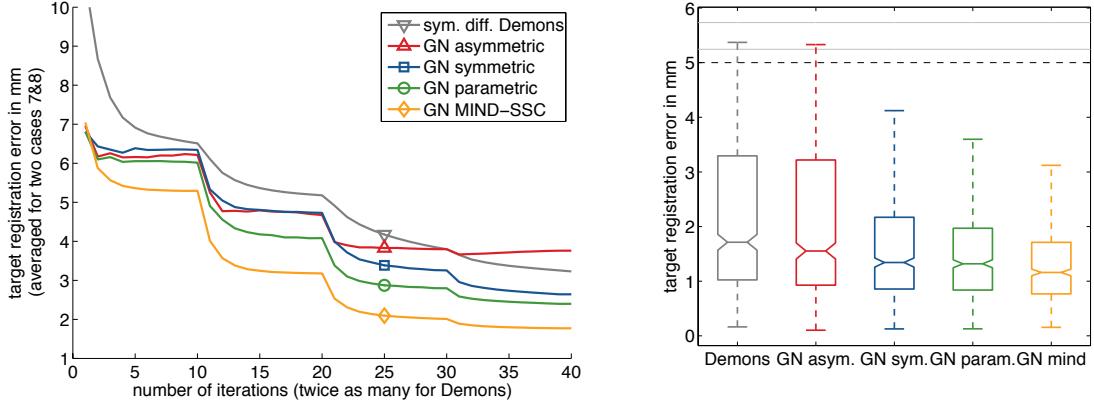


Figure 7.3: Left: Evaluation of various continuous optimisation methods applied to the deformable registration of inhale-exhale CT. Target registration error in mm (averaged for cases 7 and 8) is plotted with respect to the iteration count (twice as many for demons). Best results are obtained using a parametric coarse-to-fine approach and a multi-channel optimisation of MIND representations. Right: The distribution of the final registration error demonstrates improved robustness when using a symmetric formulation.

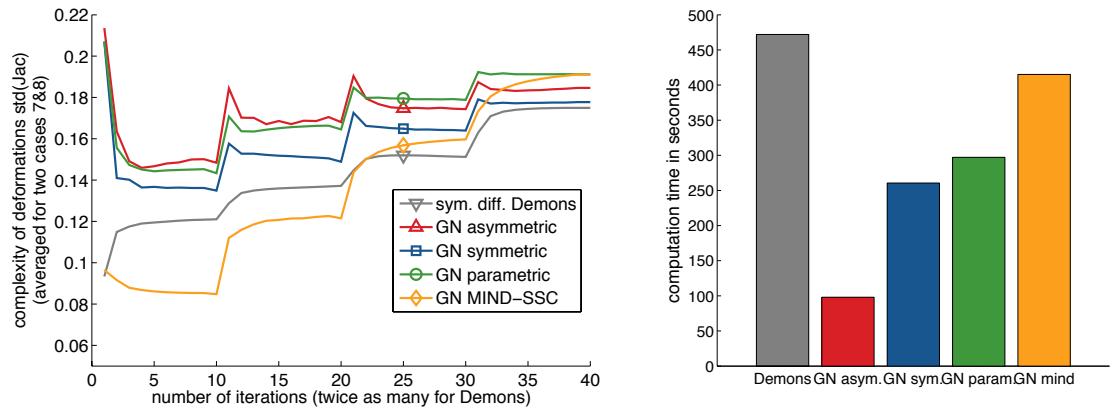


Figure 7.4: Left: Complexity of deformation for different optimisation strategies (averaged for cases 7 and 8) in terms of standard deviation of Jacobians with respect to the iteration count. Demons yield the smoothest transformations, but cannot completely capture the complex motion, possibly due to the decoupling of the regularisation. Right: Gauss-Newton has in general a lower computation time than demons.

ble displacements. A solution to this problem is to discretise (quantise) the search space. Therefore, very large deformations and potential discontinuities in the motion field can be captured without applying a multi-resolution scheme, which may result in loss of anatomical detail. Deformable registration using discrete optimisation can be formulated as Markov Random Field (MRF) labelling. To model a parametric image registration problem, a graph is defined, in which the nodes $p \in \mathcal{P}$ correspond to control points in a transformation grid with a spatial location of $\mathbf{x}_p = \{x_p, y_p, z_p\}$ (for non-parametric transformations, each node corresponds to only one voxel). For each node, we define a set of labels $f_p \in \mathcal{L}$, which correspond to a discrete displacement $f_p = \mathbf{u}_p = \{u_p, v_p, z_p\}$. The energy function to be optimised consists of two terms: the data (also called unary) cost D and the pair-wise regularisation cost $R(f_p, f_q)$ for any node q , which is directly connected ($\in \mathcal{N}$) with p :

$$E(f) = \underbrace{\sum_{p \in \mathcal{P}} D(f_p)}_{\text{data term}} + \alpha \underbrace{\sum_{(p,q) \in \mathcal{N}} R(f_p, f_q)}_{\text{regularisation term}} \quad (7.12)$$

The unary cost measures the similarity of the voxels around a control point p in one image and the set of voxels in the second image around the control point location, which is displaced by f_p . It is independent of the displacements of its neighbours. The pair-wise term enforces a globally smooth transformation by penalising deviations of the displacements of neighbouring nodes. The weighting parameter α sets the influence of the regularisation. For single-modal registration, sum of absolute differences (SAD) of the intensities in the target image I_t and the moving image I_m can be used as similarity metric (note that SAD cannot directly be used in continuous optimisation, because it is not differentiable). The

deformation field is regularised using the squared differences of the displacements of neighbouring control points, which approximates the diffusion regularisation that we used before. The energy term for a labelling f then becomes:

$$E(f) = \sum_{p \in \mathcal{P}} |I_t(\mathbf{x}_p) - I_m(\mathbf{x}_p + \mathbf{u}_p)| + \alpha \sum_{(p,q) \in \mathcal{N}} \frac{\|\mathbf{u}_p - \mathbf{u}_q\|^2}{\|\mathbf{x}_p - \mathbf{x}_q\|} \quad (7.13)$$

The complexity of the optimisation problem is in the best case linearly dependent on the number of nodes n times all possible displacements of the label space \mathcal{L} . Considering a typical 3D registration problem with $n = 10^7$ voxels and a maximum motion of 15 voxels in each direction, so that $|\mathcal{L}| \approx 3 \times 10^4$, this problem will be intractable both in terms of computation time and memory requirements.

In order to obtain both high registration accuracy and low computational complexity, we use the following five steps in our MRF-based registration framework [Heinrich et al., 2012d], which is called **deeds** (**d**ense **e**displacement **s**ampling). First, the transformations are parameterised using multiple levels of B-spline grids with coarse-to-fine control point spacings. In contrast to most commonly used approaches (with the exception of e.g. [Schnabel et al., 2001]), a constant complexity is used for each level, in order to avoid errors at coarse scales, which cannot be recovered at finer scales (see Sec. 7.5.1). In Sec. 7.5.2 a dense stochastic sampling is introduced for the computation of the similarity term. This enables the use of a very large search space $|\mathcal{L}| > 5 \times 10^3$, which deals well with complex motion, while retaining good computational efficiency. Sec. 7.5.3 describes a different graph representation using an intensity-based minimum-spanning-tree. It enables the modelling of sliding motion and offers attractive computational advantages for the inference of regularisation (using belief propagation) compared to

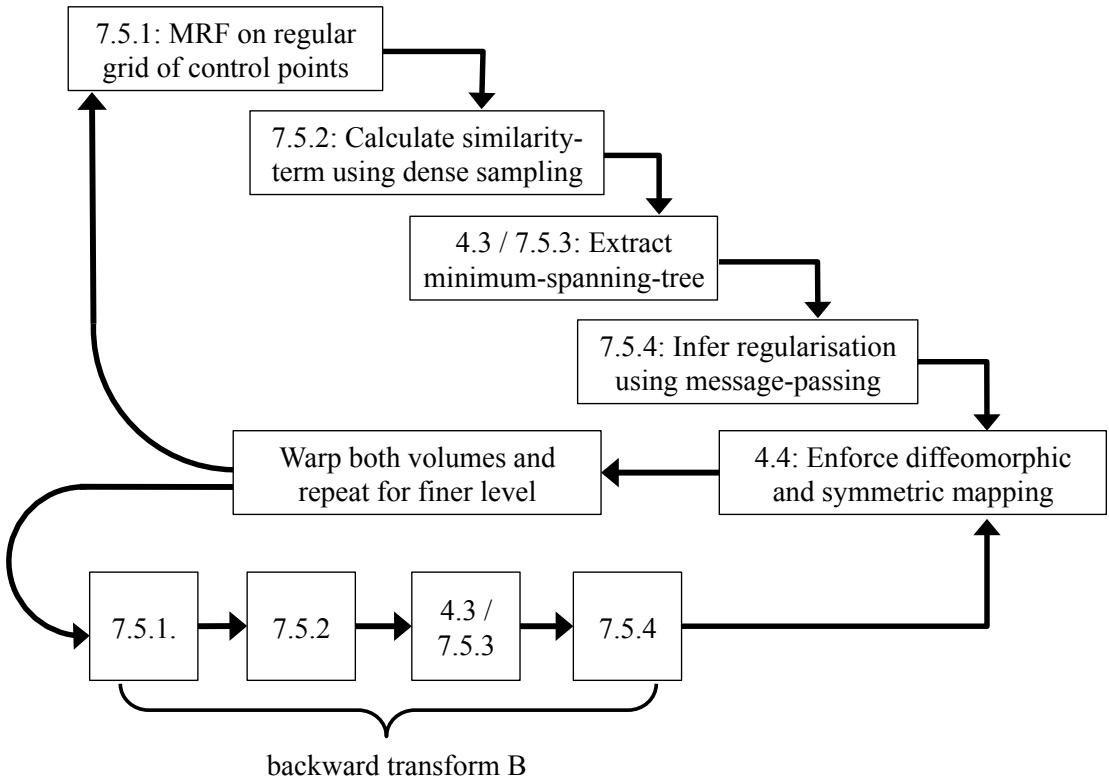


Figure 7.5: Overview of the MRF-based deformable registration approach **deeds**. The flow-chart displays the algorithmic order of individual steps presented in this section.

e.g. graph cuts (see Sec. 7.5.3). Sec. 7.5.4 presents an efficient computation of pair-wise potentials for an incremental diffusion regularisation penalty (with sub-pixel accuracy). The symmetric, diffeomorphic transformation model, described in Sec. 4.4 is used again, to obtain deformations that are inverse consistent and free from singularities . An overview of the resulting algorithm is shown in Fig. 7.5. An extension to our framework, which includes the simultaneous estimation of lung ventilation is described in Sec. 7.7 and [Heinrich et al., 2013a].

7.5.1 Parameterisation of grid and displacements

A hierarchical subdivision of the image domain into groups of voxels has been proposed for discrete optimisation by [Felzenszwalb and Huttenlocher, 2006]. The solution from a coarser scale is used to initialise the energy of nodes at subsequent levels. However, the label-space is kept the same for all levels, resulting in too many degrees of freedom (for a 3D registration) at the finest scale.

[Yang et al., 2010] proposed to discard labels with high cost (and thereby reducing the label space) at each level so that the complexity stays constant. We have adopted this *constant-space* approach for 3D medical registration in [Heinrich et al., 2011a] and achieved high registration accuracy. There are nevertheless some drawbacks of the method. Selecting only a small number of potential displacements in the coarse level can lead to a local optimum. We found that a simple SAD criterion is not sufficient, even for single-modal problems, and more complex metrics (NCC in these experiments) had to be employed. Furthermore, the initial computation of the data cost is still linearly dependent on n and the largest label space, which outweighs the performance improvements. The non-uniform reduction of label sets makes the optimisation problem harder and efficient optimisation strategies (see Sec. 7.5.4) cannot be employed.

Another approach to reduce the complexity is a parameterisation of the transformation. We avoid using a multi-resolution scheme (which was used for the continuous optimisation framework and may degrade the quality of the registration) and adopt a multi-level B-spline scheme [Schnabel et al., 2001], with the difference that we always employ the highest image resolution. For a given level, the image is subdivided according to a 0th order B-spline grid into non-overlapping cubic groups

of voxels (preserving the independence of the unary term for all nodes). The similarity cost, which is incurred when translating a cube of voxels (or equivalently move a control point in the B-spline grid) is aggregated voxel-wise as explained in the next section (only a subsample of K voxels is used). Subsequently, the regularisation term is calculated only for each control point.

Several subsequent levels with decreasing grid-spacing g are used, increasing the number of nodes for each finer level. The maximum number of displacement steps in the label space l_{\max} is decreased correspondingly. The search space is defined as $\mathcal{L} = d \cdot \{0, \pm 1, \dots, l_{\max}\}^3$ voxels, where d is a discretisation step. Using this refinement approach, both high spatial accuracy and low computational complexity are achieved. By using same numbers for g and l_{\max} for each level, the complexity is kept constant for all levels, because the number of similarity computations is $\sim K \cdot n/g^3 \cdot (2l_{\max} + 1)^3 \approx 8Kn$, thus is linear with the number of voxels.

For a finer level, the previous deformation field is obtained using the B-spline interpolation for voxels between control points and used as the prior deformation. We found that for estimating lung motion, a first order (linear) B-spline function provided best results.

7.5.2 Dense stochastic sampling

In the previously proposed **drop** method [Glocker et al., 2008a] the complexity (especially the memory requirements of the employed FastPD optimisation) of the registration had to be reduced by using a sparse sampling of the deformation space. Instead of densely sampling the deformations in all three dimensions, only displacements along the three axes are considered. This may lead to a non-optimal

registration and could severely reduce its accuracy. In [So et al., 2011] it has been shown that this leads to similar problems, which are common for gradient-based optimisation (local minima, bias of initialisation). Using a more efficient optimisation strategy (as introduced in Sec. 7.5.3), the search space can be substantially increased, so that a dense displacement sampling and small discretisation steps can still be used. While the complexity of the smoothness calculations is reduced using the parametric grid model, the number of similarity cost evaluations dependents only on the number of voxels (and is independent of the number of grid nodes). In the context of gradient-based image registration, the concept of stochastic optimisation has been introduced by [Klein et al., 2007]. When using a parametric model, the similarity term for a certain control point displacement is aggregated (summed) for all voxels within the influence range of the control point. Similar to [Robbins and Monro, 1951], we can make a stochastic approximation to this summation. Only a random subset K of all voxels is used for each control point. This makes the similarity term computation many times more efficient with very little sacrifice of registration accuracy.

7.5.3 Minimum-spanning-tree

Following the paradigm of pictorial structures [Felzenszwalb and Huttenlocher, 2005], we propose that medical images can be more efficiently treated using a relaxed graph structure. Instead of connecting each node to its six immediate neighbours, only the most relevant edges are considered, leading to a minimum spanning tree (MST). The construction of the MST is described in detail in Sec. 4.3. It can sufficiently reflect the underlying anatomical connectivity in a medical image

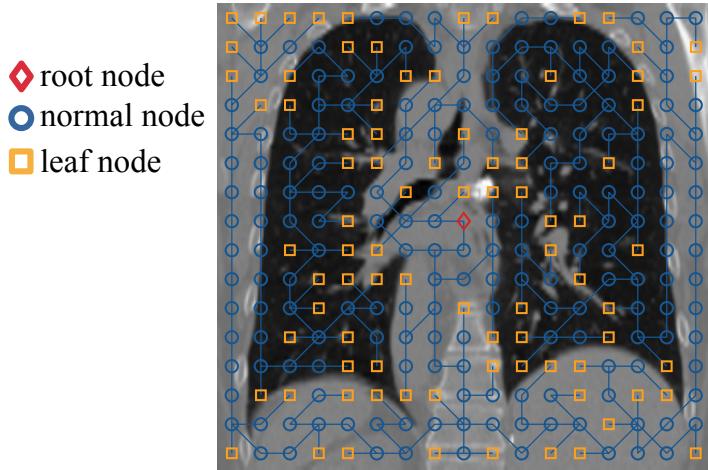


Figure 7.6: Example of minimum-spanning-tree (MST) of a 2D coronal slice of a lung CT based on 8-connectedness (\diamond root, \circ normal, and \square leaf nodes). In practice the tree is constructed in 3D (using 6-connectedness).

(see Fig. 7.6). Belief propagation on a tree (BP-T) enables us to find the global optimum, without iterations, in only two passes [Felzenszwalb and Huttenlocher, 2005].

At each node p , the cost C_p of the best displacement can be found, given the displacement f_q of its parent q :

$$C_p(f_q) = \min_{f_p} \left(D(f_p) + \alpha R(f_p, f_q) + \sum_c C_c(f_p) \right) \quad (7.14)$$

where c are the children of p . The best displacement can be found by replacing min with argmin in Eq. 7.14. For any leaf node, Eq. 7.14 can be evaluated directly (since it has no children). Thereafter, the tree is traversed from its leaves to the root node. It is worth noting that only costs C_p for the next tree level have to be stored (only the argmin is needed to select the best displacement). Once the root node is reached, the best labelling for all nodes can be selected in another

pass through the tree (from root to leaves). Another advantage of using belief propagation on a MST is that the exact marginal distributions for each node can be directly obtained, which can be used to quantify the local uncertainty of the registration.

Inference of regularisation

Apart from BP-T, a number of other methods can be used to solve the MRF labelling: e.g. loopy belief propagation (LBP) [Felzenszwalb and Huttenlocher, 2006]; sequential tree-reweighted message passing (TRW-S) [Kolmogorov, 2006], α -expansion moves graph cuts (α -GC) [Boykov and Kolmogorov, 2004]; and the fast primal-dual strategy (FastPD) [Komodakis et al., 2008]. Graph cuts can solve binary energy minimisation problems exactly by finding the minimum cut, which separates a graph, in which each node is connected to its neighbours and two additional nodes (source and sink). α -expansion moves are an extension of graph cuts to solve multi-labelling problems. Even though they are guaranteed to converge, they do not find the global optimum in most applications. Since α -GC relies on the pair-wise potential to be a metric, the most commonly used regularisation term, squared differences of displacements, cannot be used (only the square root of this, the L2 norm, is a metric). FastPD shows an improved performance compared to α -GC and relaxes the metric-requirement. However, this comes at the cost of substantially increased memory requirements (FastPD requires roughly 1000 bytes memory per degree of freedom, compared to 6 bytes for α -GC or message passing approaches – e.g. BP-T).

7.5.4 Incremental diffusion regularisation

Finding the minimum naïvely requires $|\mathcal{L}|^2$ calculations for the regularisation cost per pair of nodes. In [Felzenszwalb and Huttenlocher, 2006] the min-convolution technique is introduced, which reduces the complexity to $|\mathcal{L}|$ by employing a lower envelope computation. For most commonly used (pair-wise) regularisation terms, such as diffusion (squared difference of displacements) and total variation (absolute difference) regularisation, this simplification is possible. Each label f_p can be represented by an upward facing parabola rooted at $(f_p, D^*(f_p))$, where $D^*(f_p) = D(f_p) + \sum_c C_c(f_p)$. The minimisation in Eq. 7.14 is defined by the lower envelope of these parabolas. In order to find this lower envelope in a single pass over all labels, the intersection between the parabola of the current label and the right-most parabola of the lower envelope needs to be calculated (see Fig. 7.7 for a visualisation). This technique requires the label spaces \mathcal{L} of both nodes to be equivalent. We make an extension to this method, which enables the use of an incremental regularisation in a multi-level scheme. If a previous deformation field is known, we first warp the moving image towards the fixed image. Since the regularisation cost depends only on the difference between two labelings, only the (subpixel) offset $\Delta = x_p - x_q, (p, q) \in \mathcal{N}$ for each dimension between displacements has to be considered. The lower envelope can be found in a similar way as in [Felzenszwalb and Huttenlocher, 2006], however the coordinates of the intersections s of the parabolas now depend on the offset Δ :

$$s = \frac{(D^*(f_q) + (f_q + \Delta)^2) - (D^*(f_p) + (f_p + \Delta)^2)}{2 \cdot (f_q - f_p)} \quad (7.15)$$

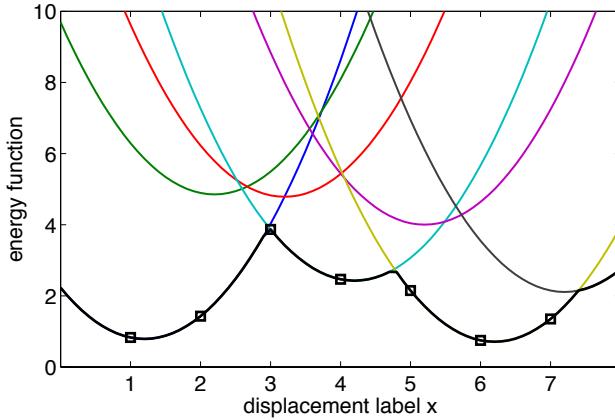


Figure 7.7: Lower envelope computation for efficient belief propagation. Each label is drawn with a coloured parabola with a minimum height of $D^*(f_p)$. Note that the minima have a uniform offset (here $\Delta=0.35$) from the integer position, determined by the index of a label. The lower envelope is depicted by a bold, black line and the sampling positions for Eq. 7.14 are indicated with squares.

7.6 Experiments using discrete optimisation

We use the same experiment as presented in Sec. 7.4, to evaluate the influence of each of our contributions (presented in the previous sections) within our discrete optimisation framework and quantify the registration accuracy. However, now all 10 cases of inhale-exhale CT scans are employed. The first five cases have been cropped (to include the thoracic cage) and resampled to form an in-plane dimension of 256×256 by [Castillo et al., 2009]. We apply a similar cropping to the remaining five datasets, but apply no resampling. Figure 7.10 (top left) shows one scan pair of the dataset, with the inhale image shown in green and the exhale image overlaid in magenta.

The challenges of the dataset require us to optimise over a large number of degrees of freedom. Only the highest image resolution is used, but the number of nodes is reduced by using a uniform grid of control points. Three levels with

decreasing grid-spacing of $g = \{8, 6, 4\}$ voxels are employed. This means that the number of nodes is increased for each subsequent level. The number of labels is correspondingly decreased, and the maximum search radius is set to $l_{\max} = \{8, 6, 4\}$ steps. The search space is defined as $\mathcal{L} = d \cdot \{0, \pm 1, \dots, \pm l_{\max}\}^3$ voxels, where d is a discretisation step, which is defined as $\{2.0, 1.0, 0.5\}$ voxels for the three levels. Sub-pixel displacements are achieved by upsampling the moving image using trilinear interpolation.

For the similarity term, $K = 64$ random samples are used for each control point p within the cubic region $R_p = \{-g/2+1, -g/2+2, \dots, g/2\}^3$. The sampling locations are uniformly distributed within the local support region R_p of a node. The same locations are used for all displacements \mathbf{x}_p for one node, but the sampling is updated for every new node, to avoid a bias. Given a grid point spacing of 8 voxels, the standard deterministic similarity cost computation would require 512 calculations per node, the stochastic approach therefore yields an 8-fold improvement in computation time. By using the same numbers for g and l_{\max} for each level the complexity is kept constant for each level, the number of similarity computations is $\sim K \cdot n/g^3 \cdot (2l_{\max} + 1)^3 \approx 8Kn$, thus linear with the number of voxels. By using the lower envelope technique discussed in Sec. 7.5.4 the complexity of the regularisation term is also linear with n .

7.6.1 Influence of regularisation weighting α

In addition to the implicit regularisation of the parametric B-spline grid, the regularisation terms $R(f_p, f_q)$ in Eq. 7.13 is of importance. A higher value of α results in a smoother deformation field. Figure 7.8 shows the resulting registration accuracy

and the mean squared error (MSE) between the deformed moving and fixed image with varying strength of the regularisation α for our proposed method. Interestingly, the lowest MSE coincides with the lowest target registration (TRE), which helps in selecting a suitable value for α . In order to compare different approaches, they should generally have a similar complexity of the resulting transformations (measured by $\text{std}(J)$ as before).

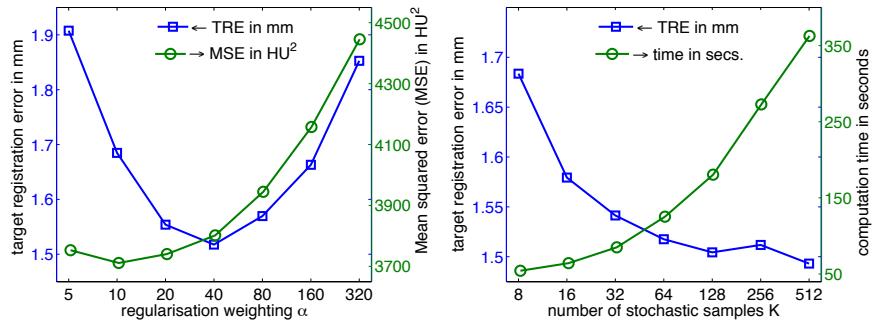


Figure 7.8: Influence of α on registration accuracy (averaged for all cases) and mean squared error (MSE) is shown in the left plot. The right plot shows the accuracy with increasing number K of stochastic samples. The TRE does not improve substantially with more than 64 samples per control point.

7.6.2 Evaluation of the influence of our contributions

The deformable registrations for all ten cases between maximum inhalation and maximum exhalation are first performed using a symmetric transformation model, the dense stochastic sampling approach and using **deeds**. This forms the **baseline** for the subsequent experiments. An average registration accuracy of **1.52 mm**, with a smoothness of $\text{std}(Jac)=0.109$ is obtained for the baseline of our proposed algorithm with a computation time of 2.04 minutes per case. Next, each individual contribution of this work is tested separately and compared to the baseline, thereby

Table 7.1: Results for individual contributions assessed using average target registration error (TRE in mm) for 3000 expert selected landmarks. A significant ($p < 0.05$) improvement (lower median) compared to the baseline is indicated by +, a higher error by - (non-significant results are in brackets). The computation time (for optimisation and complete registration) per scan pair is given in minutes. The optimal weighting for the regularisation α and the resulting complexity of the deformations, evaluated using the standard deviation of the Jacobians, are given.

	time (min)		α	std(Jac)	TRE in mm	p -value
	optim.	total				
initial					8.46	
axial	0.12	2.37	40	0.103	2.25	$9 \cdot 10^{-19}$ -
random tree	0.62	2.11	40	0.116	2.05	$7 \cdot 10^{-9}$ -
asymmetric	0.48	1.95	40	0.147	1.83	$9 \cdot 10^{-5}$ -
graph cut	17.8	19.3	5	0.118	1.73	0.152 (-)
non-stochastic	0.61	6.04	40	0.109	1.49	0.588 (+)
baseline	0.62	2.09	40	0.109	1.52	
hyper-labels	4.57	7.97	40	0.099	1.43	0.017 +

only one element is changed each time. The optimal regularisation weighting α is found for each method separately (as shown for the baseline in Fig. 7.8 left). All deformations are free from singularities. The results are summarised in Table 7.1 and Fig. 7.9.

Axial sampling: First, the dense displacement sampling is replaced by the sparse approach used in **drop** [Glocker et al., 2008a] (and similarly in [Lee et al., 2008] and [Shekhovtsov et al., 2008]). Here only displacements along each of the three axes are considered. This substantially reduces the size of the label space from $|\mathcal{L}| = (2l_{\max} + 1)^3$ to $|\mathcal{L}| = 3 \cdot 2l_{\max} + 1$. Following [Glocker et al., 2008a] we introduce an iterative loop to partially compensate for the reduced degrees of freedom. The accuracy of the registration experiments significantly deteriorates (to 2.25 mm) compared to the baseline, confirming our previous assumption that an axial sampling of the displacements results in many false local minima of the

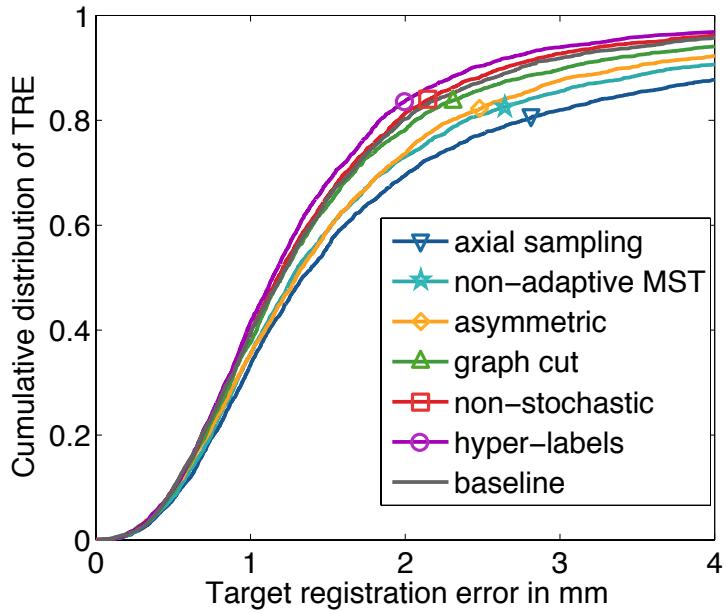


Figure 7.9: Cumulative distribution of target registration error in mm of each method using 3000 manual expert landmarks. A significant improvement, using the Wilcoxon rank-sum test, can be seen for the baseline approach compared to asymmetric registration ($p = 9 \cdot 10^{-5}$), without using an image-adaptive MST ($p = 7 \cdot 10^{-9}$) and axial sampling ($p = 9 \cdot 10^{-19}$). Hyper-labels are discussed in Sec. 7.7

registration problem.

Random tree: As described in Sec. 7.5.3, we employ an image-derived MST, which removes edges across locations with large intensity differences, which are likely to coincide with sliding motion. To evaluate the suitability of this approach, the alternative of choosing a random spanning tree is included for comparison. Using this graph significantly reduces the accuracy to 2.05 mm. Figure 7.10 demonstrates that this is mainly due to the better preservation of sliding motion of our image-adaptive approach.

Graph Cut: The BP-T optimisation is now replaced by α -expansion graph cuts [Boykov et al., 2001]. FastPD is not suitable because of its larger memory

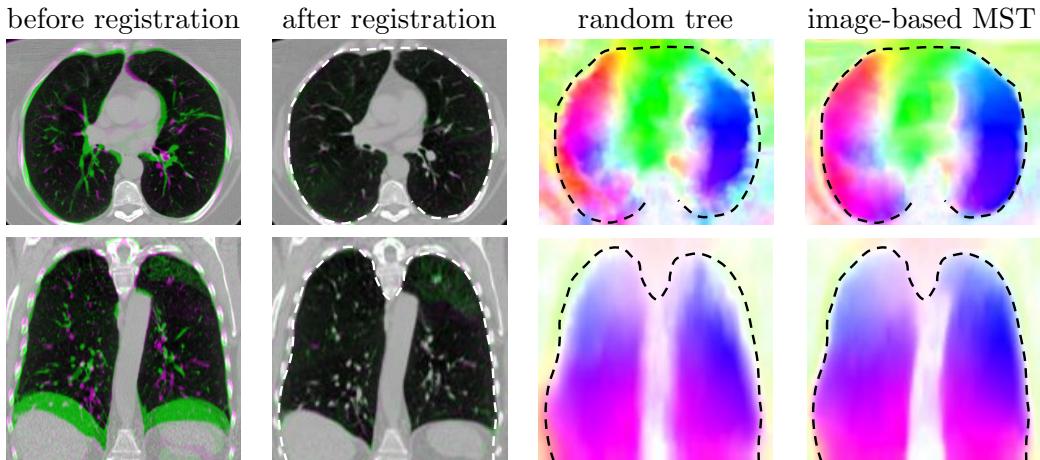


Figure 7.10: Example registration result of case 8. Top row shows the axial plane, the bottom row the coronal plane. In columns from left to right is displayed: Overlay of inhale (green) and exhale (magenta) phase before and after registration, colour-coded deformation field using a random spanning tree and image-adaptive MST respectively (the same colour coding as in Fig. 7.2 was used). The deformation fields demonstrate: the sliding of the lungs is better preserved using the MST. The outline of the thoracic cage is shown for visual guidance (dashed white or black line).

demand. Since our preferred diffusion regularisation term (squared differences) is not a metric and cannot not be optimised with α -GC, it is substituted with the L2 norm (Euclidean distance) (see also discussion in Sec. 7.5.3). During the multi-level registration, the regularisation can only be applied to the update of the deformation field when using α -GC, and the computation time for the optimisation is increased by a factor of over 30, demonstrating the improvements achieved by the use of our MST-optimisation. The accuracy is also slightly decreased to a TRE of 1.73 mm.

Asymmetric: In general, registration problems should be treated symmetrically to remove bias from the choice of fixed and moving image. The direct comparison of an asymmetric registration to the baseline also shows a significant

improvement in terms of registration accuracy for the symmetric approach. Due to the fewer constraints, the smoothness of the obtained deformations are lower compared to the symmetric approach.

Non-stochastic: As discussed in Sec. 7.5.2 the stochastic sampling approach for the similarity terms yields a greatly reduced computation time. Figure 7.8 shows that the registration accuracy improves with an increasing number of samples, but converges at $K \approx 64$. The TRE of the deterministic approach is only slightly lower than the stochastic one and this change is not statistically significant.

7.7 Physiologically motivated image registration

In many applications of medical image registration, finding the transformation parameters is only an intermediate target. One may be interested in propagating segmentation labels, estimating contrast uptake in dynamic sequences or generating a patient-specific (predictive) motion model. Parameters other than geometric displacements are of interest. We introduce hyper-labels into our registration framework, for which a fourth (and possible higher) dimension is added to the label space. This additional parameter may correspond to a non-uniform multiplicative intensity variation (to estimate bias fields or pharmacokinetic parameters), a segmentation label (c.f. [Mahapatra and Sun, 2012]) or a motion parameter (e.g. the time phase-shift for a sinusoidal motion in 4D data). In this work, hyper-labels are used to directly estimate the density change of lung tissue and therefore the local lung ventilation.

Simultaneous image registration and intensity correction has been proposed to deal with non-uniform bias fields in combination with brain tissue segmentation

in [Ashburner and Friston, 2005]. The coding complexity of the residual image is minimised in [Myronenko and Song, 2010] to estimate a smooth intensity correction field. Similarly, in [Modersitzki and Wirtz, 2006] a regularised correction function is employed in a variational optimisation framework to compensate for inhomogeneous intensity mappings. The disadvantage of such approaches is that they all rely on continuous optimisation schemes and are readily trapped in local minima. In contrast, our proposed hyper-label approach allows us to freely define the range of possible intensity correction values. It can also make use of the globally optimal regularisation, avoid local minima and is independent of the initialisation.

Two different metrics have primarily been used so far to derive ventilation images from dynamic CT [Yamamoto et al., 2011a]: Hounsfield unit (HU) change and Jacobian determinant of the deformations. Both methods were studied for ventilation estimation in [Guerrero et al., 2005], [Castillo et al., 2010], and [Kabus et al., 2008]. In [Yamamoto et al., 2011b] it was found that only the HU-based ventilation estimate had a significant correlation with lung functionality of emphysema patients. In most cases the ventilation estimation was performed in a post-processing step after deformable registration. In [Yin et al., 2009] and [Gorbunova et al., 2012] the intensity change based on the Jacobian determinant is included in the similarity metric during the registration (so called mass-preserving registration). For MRF-based optimisation this would introduce an unwanted dependency of the unary and pair-wise potentials, so we adapt the HU-based approach. According to [Yamamoto et al., 2011b] the fractional local change of lung

volume ΔV_{HU} is defined as:

$$\Delta V_{HU}(\mathbf{x}_p) = \frac{I_t(\mathbf{x}_p) - I_m(\mathbf{x}_p + \mathbf{u}_p)}{I_m(\mathbf{x}_p + \mathbf{u}_p) + 1000} \cdot V_I(\mathbf{x}_p)^{vox}, \quad (7.16)$$

where I_t represents the exhale and I_m the inhale scan, both measured in Hounsfield units, and $V_I(p)^{vox} := 1$ the exhale voxel volume, which is constant across the image. To include this locally varying density correction function into our registration framework, we introduce a fourth variable ν to the label space \mathcal{L} , so that $f_p = \{u_p, v_p, w_p, \nu_p\}$. For each node, ν_p takes quantised values of local volume change, and the smoothness of ν is ensured in the regularisation term. We simplify Eq. 7.16 by adding 1000 HU (so that air has a value of 0) to both images. This yields:

$$\begin{aligned} \nu_p &= \Delta V_{HU}(\mathbf{x}_p) = \frac{I_t(\mathbf{x}_p)}{I_m(\mathbf{x}_p + \mathbf{u}_p)} - 1 \\ &\Rightarrow (1 + \nu_p) \cdot I_m(\mathbf{x}_p + \mathbf{u}_p) = I_t(\mathbf{x}_p) \end{aligned} \quad (7.17)$$

The unified energy term to be minimised is then:

$$\begin{aligned} E(f) &= \sum_{p \in \mathcal{P}} |I_t(\mathbf{x}_p) - (1 + q_p) \cdot I_m(\mathbf{x}_p + \mathbf{u}_p)| + \\ &\quad \alpha \sum_{(p,q) \in \mathcal{N}} \frac{\|\mathbf{u}_p - \mathbf{u}_q\|^2 + \beta(\nu_p - \nu_q)^2}{\|\mathbf{x}_p - \mathbf{x}_q\|} \end{aligned} \quad (7.18)$$

where β can be used to weight the influence of density change on the regularisation. The BP-T optimisation can be applied as before, as well as the lower-envelope computation, except this is now performed over a four-dimensional array of smoothness terms. The ventilation image V_{HU} is directly given by ν_p extracted from the given

labelling f_p^* after optimising Eq. 7.18.

7.7.1 Experiments using hyper-labels

We studied the influence of the proposed hyper-labels for simultaneous registration and intensity correction using the same experimental setting as in Sec. 7.6. Five quantised intensity correction labels are added to each geometric label, therefore this approach has a 5-fold increased complexity. The average TRE is further (statistical significantly) reduced to 1.44 mm. Additionally, the obtained intensity correction field can be directly used to quantify lung ventilation, an example for which is shown in Fig. 7.11. This has several advantages over calculating the ventilation as a post-processing step (c.f. [Castillo et al., 2010], [Kabus et al., 2008]), as it is not directly affected by misalignments, avoids local minima during registration and does not need a specific smoothing of the difference images (c.f. [Yamamoto et al., 2011b]). The results in terms of registration accuracy can be found in 7.9 and Table 7.1, a small improvement compared to the proposed baseline approach can be seen. Yet, the main contribution of this approach is its ability to simultaneously and directly estimate the ventilation of the lungs (and therefore directly measure their physiological properties) as shown e.g. in Fig. 7.11. To further validate the ventilation estimation, an additional registration was performed for the POPI-model [Vandemeulebroucke et al., 2007] (equivalent to case 17 of [Murphy et al., 2011b], which provides manual segmentations) and the total lung volume change was calculated by integrating all local ventilation values inside the lungs. A good agreement between the Jacobian method (246.6 ml), our hyper-label approach (246.1 ml), and manual segmentations (247.2 ml) was found.

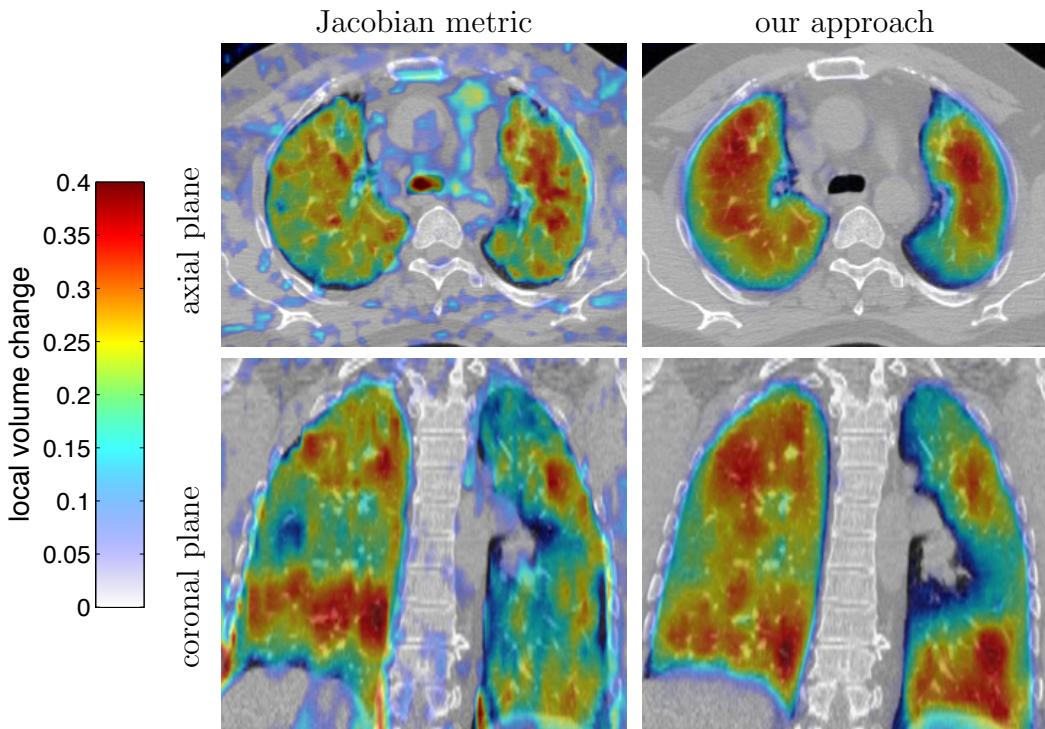


Figure 7.11: Comparison of lung ventilation for case #6 [Castillo et al., 2009]. The local lung volume change is displayed as colour overlay (transparent for low values) over greyscale intensities of the inhale scan, without using segmentation masks. Our approach using simultaneous ventilation estimation with hyper-labels demonstrates a much smoother map of lung functionality, than the two-step approach using the Jacobian metric ($Jac - 1$ as used e.g. in [Yamamoto et al., 2011a]). It can be seen that in this case the ventilation in the region around the tumour in the left lower lobe is substantially reduced.

7.8 Summary

This chapter addressed the challenges of optimisation for deformable lung registration. The Gauss-Newton method, which is an efficient second-order continuous optimisation technique, has been extended and specifically adapted for the use in medical image registration. A new rigid-body registration method has been presented based on an affine parameterisation, which can be solved with the Gauss-Newton using only the three spatial derivatives of the similarity term, and is

followed by a least-square estimation of the best rigid-body transformation. A deformable registration framework with diffusion regularisation has been presented, which includes an improved multi-scale approach and has been extended for the use of multidimensional image representations (e.g. MIND as presented in Chapter 6). This enables multi-modal deformable registration with fast convergence and improves significantly improves the accuracy of intra-modal lung registration, because it is more robust to locally varying contrast and more discriminative for geometric features. Experimental results have shown the advantages of this globally regularised approach in terms of registration accuracy and computation time compared to the demons framework, which relies on only local regularisation.

Continuous optimisation has some inherent limitations, especially with respect to the optimality of the obtained solution, the choice of similarity metrics, and its dependence on the initialisation. A comprehensive discrete optimisation framework has been developed, which improves on the current state-of-the-art with a number of technical contributions. A dense spatial sampling (**deeds**) of the similarity metric is employed to avoid local minima and a bias to initial misalignment. The high computational complexity of this step is offset by a stochastic sampling of point-wise similarity terms within a parametric B-spline transformation model. An image-derived minimum-spanning-tree (see also Sec. 4.3) is used as a relaxed graph structure, which not only enables to find a global optimum in a single iteration of belief propagation, but has also been shown to deal intrinsically well with the respiratory sliding motion of the lungs. The lower-envelope technique of [Felzenszwalb and Huttenlocher, 2006] has been extended to enable a multi-scale refinement with subvoxel accuracy. Finally, a simultaneous ventilation estimation has been introduced for lung CT, which improves the registration accuracy and

has immediate clinical value.

The use of our novel approach to obtain diffeomorphic and symmetric transformations (see Sec. 4.4.1) has been shown to give significantly improved registration accuracy for both continuous and discrete optimisation, lower complexity of deformations and much reduced inverse inconsistency.

The next chapter presents a more thorough evaluation of our registration framework for the deformable intra- and multi-modal registration of a larger number of clinical thoracic scans and a comparison to other state-of-the-art methods.

Chapter 8

Experimental evaluation on clinical scans

The objective of this chapter is to experimentally evaluate the methodological concepts introduced in this thesis. Deformable registration results are presented for two challenging clinical pulmonary imaging data sets: inhale and exhale phases of 4D-CT scans of ten cancer patients intended for radiotherapy planning; and eleven diagnostic 3D MRI and CT chest scan pairs of patients with empyema. The quantitative evaluation, for both cases, is based on manually selected anatomical landmarks.

The preceding chapters have mainly focused on the methodological development and analysis of new concepts for image similarity, modelling of deformations for lung motion and numerical optimisation. In this chapter, a thorough experimental evaluation is presented in order to demonstrate the suitability of the proposed methods for practical clinical registration tasks and to compare them to

state-of-the-art techniques. Evaluation on real clinical data is an important step for the development of new medical image analysis techniques, because simulated motion scenarios can hardly ever capture the true challenges of a real world problem. As discussed in Chapter 3, quantitative evaluation is difficult because the same evaluation metrics are not comparable across different applications. This is due to different magnitudes of motion, scan resolution and observer variability for of manual ground truth annotations. Therefore, different registration algorithms can only be compared when applied to the same dataset and using the same ground truth annotations.

In the following sections, the comprehensive deformable registration framework, which combines the various scientific contributions of this thesis is applied to two clinical applications, quantitatively evaluated, and compared to the state-of-the-art methods. In Section 8.1 the deformable motion between inhale and exhale phase of 4D-CT scans is estimated. Intensity-based similarity metrics (SAD and SSD) and the use of MIND representations (see Chapter 6) are compared. Furthermore, the performances of the presented Gauss-Newton and discrete optimisation framework (see Chapter 7) are evaluated. Additionally, published results from other state-of-the-art registration algorithms on the same dataset are included for a direct comparison. Section 8.2 presents experiments for the deformable multi-modal registration of CT and MRI chest scans of patients with empyema, a lung disease. The registration results using statistical similarity metrics (presented in Chapter 5) or MIND representations are shown.

8.1 Deformable registration of inhale-exhale CT scans

In the previous chapter, a number of experiments were already presented for respiratory motion estimation of lung CT. Its potential uses include the following. Atlas-based segmentation of the lungs and its lobes [Zhang et al., 2006], where expert segmentations from a database are propagated to a new subject. Longitudinal CT scans from the same patient can be used to monitor treatment or disease progression, e.g. for lung nodules [Staring et al., 2009b]. 4DCT scans are now widely used for motion estimation in radiotherapy planning to increase the accuracy of radiation dose delivery [Weiss et al., 2007]. Deformable registration of dynamic CT scans can also enable direct estimation of lung ventilation [Castillo et al., 2010] to assess patients with breathing disorders or spare well-functioning lung tissue from radiotherapy.

To evaluate and compare the findings of the previous chapter, we performed deformable registration on ten cases of the DIR-Lab 4DCT dataset acquired at inhale and exhale phase [Castillo et al., 2009]¹. The patients were treated for esophageal or lung cancer, and a breathing cycle CT scan of thorax and upper abdomen was obtained, with slice thickness of 2.5 mm, and an in-plane resolution ranging from 0.97 to 1.16 mm. Since these 4DCT scans are already in rigid alignment, no linear registration step was required. Particular challenges for the deformable registration task are the changing contrast between tissue and air (because the gas density changes due to compression), discontinuous sliding motion between lung lobes and the lung/rib cage interface, and large displacements of

¹This dataset is freely available at <http://www.dir-lab.com>

small features (lung vessels, airways). For each scan 300 anatomical landmarks have been carefully annotated by thoracic imaging experts, with an intra-observer error of ≈ 1 mm. For the most challenging case (#8), the average landmark error before registration is 15 mm. The first five cases have been cropped (to include the full thoracic cage) and resampled to form an in-plane dimension of 256×256 by [Castillo et al., 2009] (yielding an in-plane resolution of ≈ 1.1 mm). We apply a similar cropping to the remaining five datasets, but apply no resampling.

8.1.1 Comparison of optimisation strategies and image representations

We compare the use of the continuous Gauss-Newton optimisation (with improved the multi-resolution strategy as described in Sec. 7.3.2) and the discrete optimisation using an MST (called **deeds**, see Sec. 7.5.3). Both strategies employ a symmetric registration strategy as presented in Sec. 4.4. Additionally, we study the influence of defining the similarity metric based on image intensities or MIND representations. This leads to four different settings of our deformable registration framework.

Discrete optimisation using MIND

Using MIND within a discrete optimisation framework enables the use of the more efficient distance calculation using the Hamming distance, which was described in Sec. 6.4.3. The spatial search region for MIND is defined using the self-similarity context 6.4.2 (resulting in 12 values per voxel) for all following experiments (the Gaussian weighting parameter for patch distances is set to $\sigma = 0.75$). Storing all

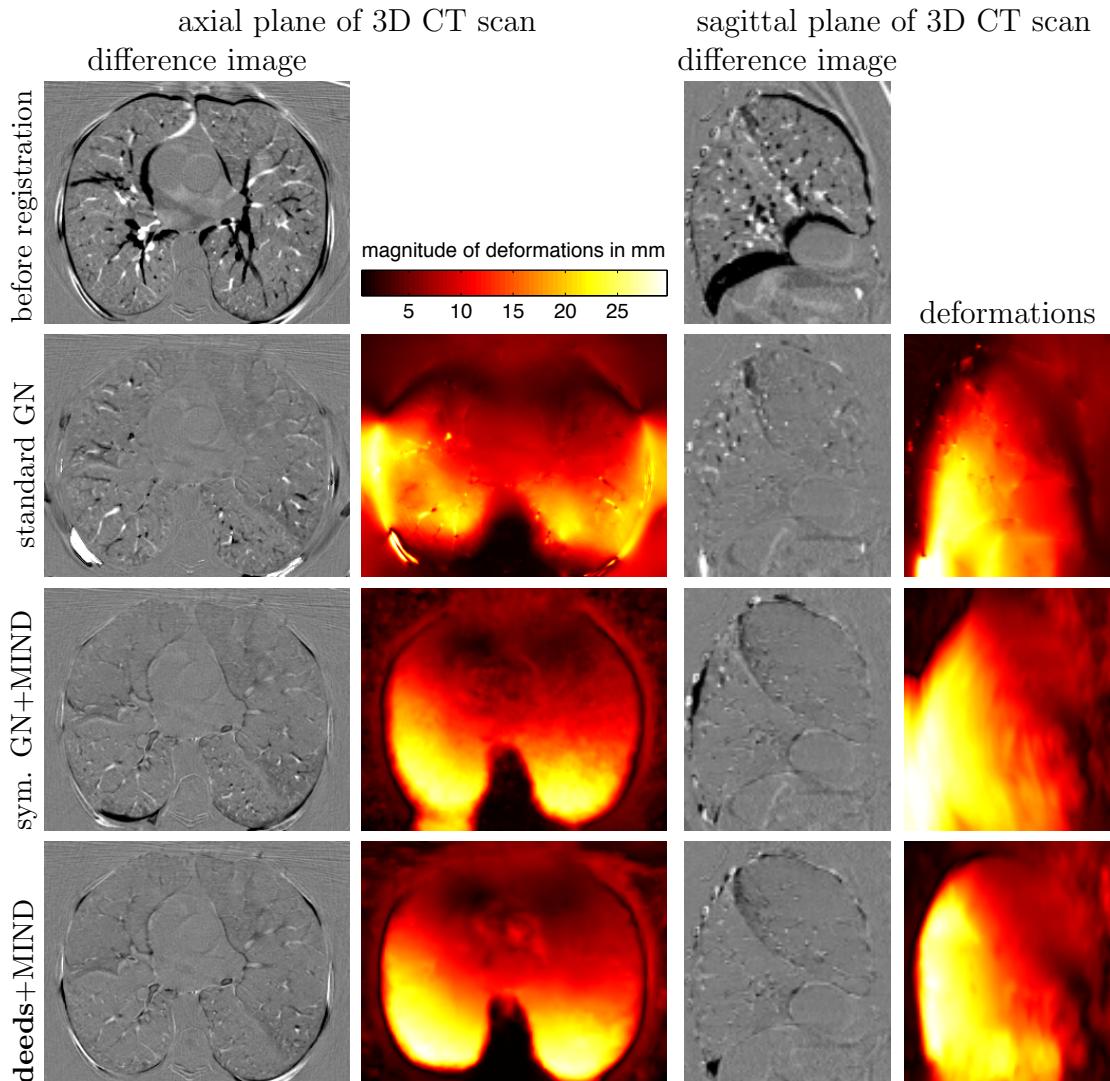


Figure 8.1: Registration results for case #8 of CT dataset. The first row shows the difference images between inhale and exhale scan before registration. A standard Gauss-Newton (GN) optimisation without symmetry constraint and SSD as similarity metric was used for the results in the second row. In particular the axial plane reveals many misaligned vessels and folding of the singularities in the deformations. The third and fourth rows show our approach using MIND with symmetric GN and **deeds** optimisation, respectively. Both methods demonstrate an improved alignment of inner and outer lung structure. The sliding motion in the sagittal plane is much better captured when using **deeds** compared to Gauss-Newton optimisation.

12 quantised descriptor values in a single 64-bit integer and calculating their L_1 norm using the Hamming weight is faster than calculating the SAD of two (single) intensity values in float precision.

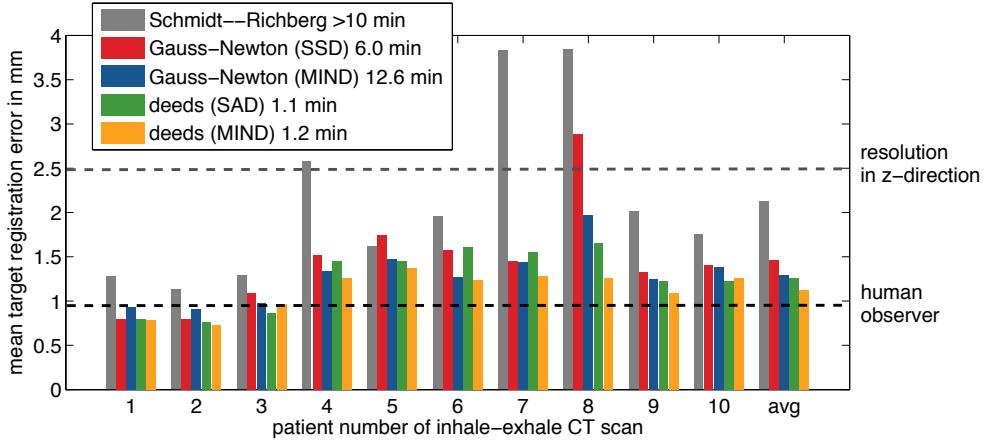


Figure 8.2: Comparison of target registration error (TRE) in mm (per case, initial TRE is 8.5 mm) for lung CT registration between [Schmidt-Richberg et al., 2012a], which achieve an accuracy of 2.13 mm, and the four different settings of our framework. The average TRE is improved from 1.46 to 1.29 mm when using MIND in the continuous optimisation framework. Our discrete optimisation method **deeds** reduces the TRE to 1.26 mm for SAD and to 1.12 mm using MIND. The residual error is lower than the image resolution and very close to the human observer error. **deeds** is also substantially faster (about 10 \times) than all other methods.

State-of-the-art methods

To put the performance of our presented registration framework into context, a comparison to state-of-the-art algorithms is performed on the same dataset. [Schmidt-Richberg et al., 2012a] presented a variational lung registration method with direction-dependent regularisation to cope with the effects of discontinuous sliding motion (see Sec. 4.1.3), which has been evaluated by the authors on the same 4DCT dataset. Figure 8.2 compares the average target registration error (TRE) in mm (for each of the ten cases) between [Schmidt-Richberg et al., 2012a]

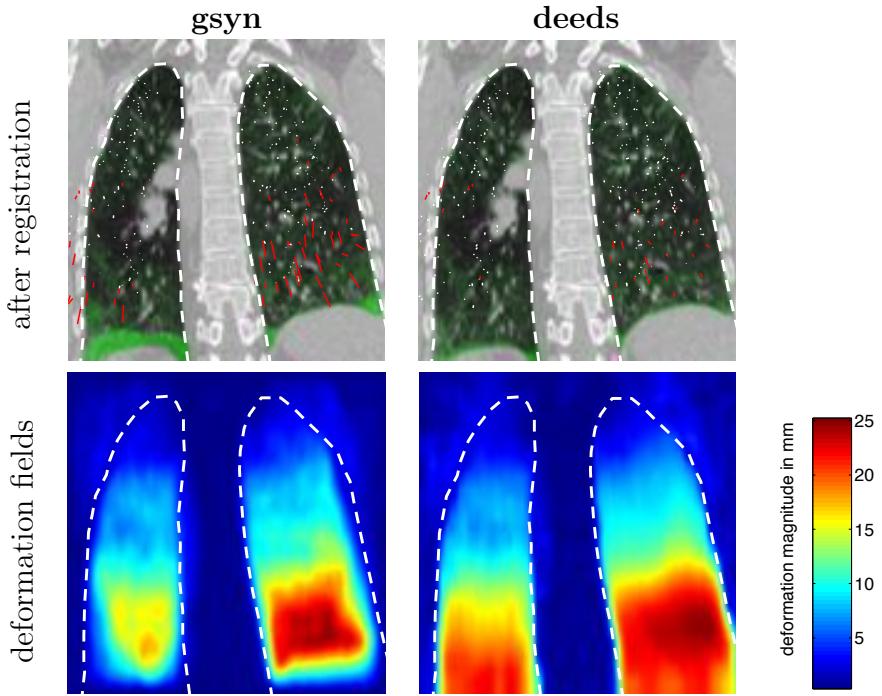


Figure 8.3: Registration result for case #6 of CT dataset. The coronal plane is shown along with vectors indicating the registration error (errors larger than the voxel size are marked in red). The magnitude of the deformation fields is shown in the second row. The sliding motion of the lungs is better preserved using **deeds**, while the deformations estimated by **gsyn** are too smooth close to the thoracic cage (segmentation shown for visual guidance with dashed white line).

and the four different settings of our framework. The same parameter settings as described in the previous chapter were used. The best results are obtained using the discrete optimisation framework with MIND representations, which has a very low computation time of only 1.2 minutes for a full registration. The most pronounced differences can be seen for case #8, which is the most challenging with a large amount of sliding motion, where the discrete framework outperforms the continuous optimisation. In general using MIND always improves the registration accuracy compared to using the directly image intensities.

We performed additional registration experiments using the **gsyn** and **drop**

methods. **gsyn**² was presented by [Avants et al., 2008] and is a symmetric, diffeomorphic, demons-like algorithm, and was chosen, because it performed best in a recent comparison study of pulmonary CT registration algorithms [Murphy et al., 2011b]. The following parameters were carefully chosen to obtain the best registration accuracy with similarly complex transformations as the proposed approach: 4 resolution levels; similarity metric: normalised cross correlation (NCC) (radius of 2 voxels); Gaussian smoothing of $\sigma=3$ and 1 voxels for gradient and deformation fields, respectively. In contrast to the results presented in [Murphy et al., 2011b] the full volumetric scans were considered and no lung masks were used to remove the outer lung information (and thereby invalidate the registration of locations outside of the mask). In a second experiment, we generate lung segmentations for all cases using thresholding and simple morphological operations and set the intensities of the outside voxels to the maximum of the ones inside the masks. We again optimised the smoothing parameters of **gsyn** for best registration accuracy, now yielding a more aggressive setting of $\sigma=3.5$ and 0.375 voxels respectively. **drop**³ developed by [Glocker et al., 2008a], which is a discrete optimisation method using a B-spline deformation grid. For **drop** the memory requirements (using 3.5 GB of RAM) limit us to use a sparse sampling of $\mathcal{L} = 3 \times \{0, \pm 1, \dots, \pm 10\}$ ($|\mathcal{L}| = 61$), with a maximum range of 24 mm. As similarity metric, SAD is used. The regularisation parameter $\lambda = 5$ was empirically chosen (when intensities are in the range [0, 255]). These settings are consistent with [Glocker et al., 2011], and no lung masks where used for **drop**. For both methods, we used four resolution levels. Our discrete optimisation framework **deeds** uses the parameters determined in

²**gsyn** is publicly available in the ANTS package <http://www.picsl.upenn.edu/ANTS/>

³**drop** is available for download at www.mrf-registration.net.

Table 8.1: Results for deformable registration of inhale and exhale CT. Average target registration error (TRE) for all 3000 expert selected landmarks (for all 10 cases) is given in mm. The complexity of the deformation fields ($\text{std}(J)$), (maximum) degrees of freedom (d.o.f.), average computation time, and the p -values of a Wilcoxon rank sum test compared to **deeds** with MIND are given where available.

	std(J)	d.o.f.	time	TRE (in mm)	p -value
initial				8.46±6.6	0
drop	15.4×10^{-2}	3.7×10^6	8 min	2.85±4.0	7.5×10^{-37}
gsyn smooth	12.9×10^{-2}	2.2×10^7	29 min	2.43±4.1	2.2×10^{-22}
gsyn masked	20.9×10^{-2}	2.2×10^7	21 min	1.57±2.1	4.2×10^{-10}
FED*	—	—	15 min	1.55±1.1	—
GN+MIND	17.7×10^{-2}	2.2×10^7	13 min	1.29±1.7	1.2×10^{-4}
deeds+MIND	15.6×10^{-2}	9.4×10^7	1.2 min	1.12±1.3	

*FED (fast explicit diffusion) [Schmidt-Richberg, 2012b] is an improved version of the direction-dependent regularisation method with better convergence.

the previous chapter: a maximum search range of $l_{\max} = 16$ voxels (using a dense stochastic sampling with $K = 64$, and a regularisation weighting of $\alpha = 50$ for SAD and $\alpha = 0.1$ for MIND. In contrast, to the previous chapter and the results reported in [Heinrich et al., 2013a], a fourth multi-scale level with a grid spacing of 2 voxels is added, which further improved the accuracy.

Results

An average TRE of 1.12 mm was achieved using **deeds+MIND**, which is a significant improvement over **drop** (TRE=2.85 mm), **gsyn** (TRE=2.43 mm) and **gsyn** with lung masks (TRE=1.56 mm). Table 8.1 summarises the results and additionally presents the complexity of deformations and computation times. **deeds+MIND** achieves the a good deformation quality, a significantly higher registration accuracy and is several times faster than all other methods. To facilitate further comparisons

to the current state-of-the-art in lung image registration, we applied our approach **deeds+MIND** to the EMPIRE10 challenge database Murphy et al. [2011b]. The exact settings of these experiments can be found in [Heinrich et al., 2013d]. Our algorithm, which is one of the very few that do not require a lung segmentation, currently ranks seventh out of 33, with the second best average TRE of 0.627 mm and the second lowest average fissure overlap error of 0.08%.

8.2 Deformable multi-modal registration of CT and MRI scans

Deformable multi-modal registration is important for a range of clinical applications. In image-guided radiotherapy it can be used to automatically propagate a manual segmentation of a tumour in a high-quality pre-treatment scan to a lower quality intra-operative scan. A second application is the use of deformable multi-modal registration to improve diagnostics tasks. Registered pairs of MRI and CT scans provide a more comprehensive view of the pathology. Additionally MRI scans can be aligned to a CT volume with a simultaneously acquired PET scan. Deformable multi-modal registration poses a more challenging task, both for registration algorithms and for validation. Finding corresponding anatomical landmarks is more difficult for the human observer, and usually much fewer locations can be found for manual annotation.

8.2.1 Chest CT and MRI of patients with empyema

We applied our proposed technique to a clinical dataset of eleven patients, which were scanned with both CT and MRI. All patients suffered from empyema, a lung disease characterised by an infection of the pleura and excess fluid within the pleural space. The extra fluid may progress into an abscess and additionally, cause the adjacent lung to collapse and/or consolidate. Both modalities are useful for detecting this pathology, but because the patients are scanned in two different sessions and at different levels of breath-hold, there are non-linear deformations, which makes it difficult for the clinician to relate the scans. The quality of the MRI scans is comparatively poor, due to motion artefacts, bias fields and a slice thickness of around 8 mm.

We asked a clinical expert to select manual landmarks for all eleven cases. 12 corresponding landmarks were selected in all image pairs, containing both normal anatomical locations and disease-specific places. It must be noted that some of the landmarks are very challenging to locate, both due to low scan quality and changes of the pathology in the diseased areas between scans. The intra-observer error has been estimated to be 5.8 mm within the MRI and 3.0 mm within a CT scan. Prior to registration, the volumes were resampled to an isotropic voxel-size of 2 mm and manually cropped to contain roughly the same field of view (and identical image dimensions).

Rigid-body registration

First, a rigid-body registration is performed for all scan pairs using the proposed Gauss-Newton framework (see Sec. 7.3.4) with MIND. The resulting landmark er-

ror is on average 9.3 mm, which is close to the error (7.1 mm) calculated for the optimal rigid-body transformation using a least square fit to the manually annotated landmark locations (this residual error reflects both non-rigid deformations and the landmark selection error). The rigid-body transformations obtained using MIND are used as initialisation of the subsequent deformable registration for all experiments, except the discrete optimisation.

Deformable multi-modal registration

Two statistical similarity metrics: normalised mutual information (NMI) and textural MI (TMI) and MIND are studied for this deformable multi-model registration experiment. The same settings (4 resolution levels, 10 iterations per level) are used for the Gauss-Newton optimisation framework (see Sec. 4.2). According to the findings from Chapter 5 we use 256 and 512 histogram bins for NMI and TMI, respectively. For TMI, using multiple trees showed a substantial improvement in our tests (we chose $F = 4$ throughout the experiments). The regularisation weighting was empirically chosen for best registration accuracy to $\alpha = 0.1$ for NMI and TMI, and $\alpha = 0.5$ for MIND. The Gaussian smoothing parameter for MIND was set to $\sigma = 1.0$. Additionally, the use of MIND in **deeds** was evaluated. The same settings as above were applied. An advantage when using the discrete framework is that no initial rigid registration is required, because of its globally optimal energy minimisation.

Figure 8.4 shows the resulting TRE in mm for all 11 cases for the four different experiments (the initial registration error is 13.5 mm). The obtained average TRE is 8.9 mm for NMI, 8.3mm for TMI, 6.5 mm for MIND using Gauss-Newton, and 6.5 mm using discrete optimisation. **deeds** with MIND is the most robust setting, with

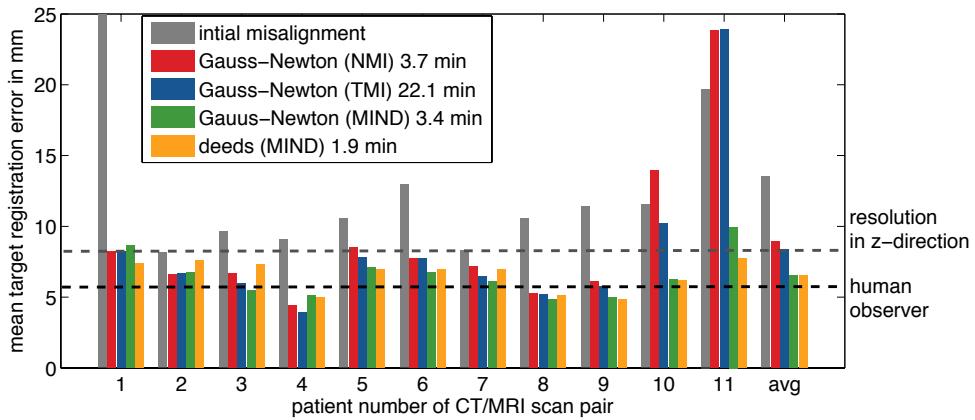
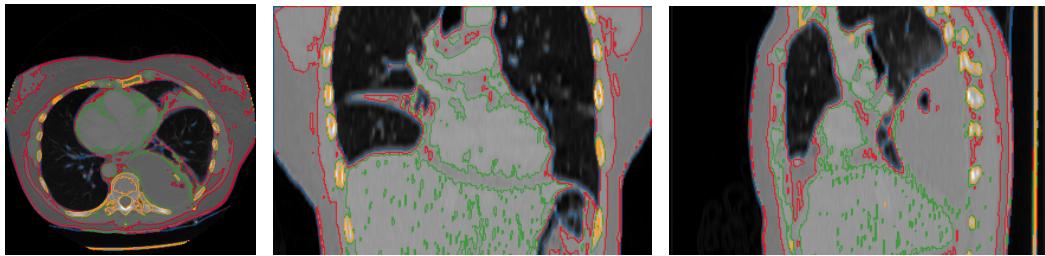
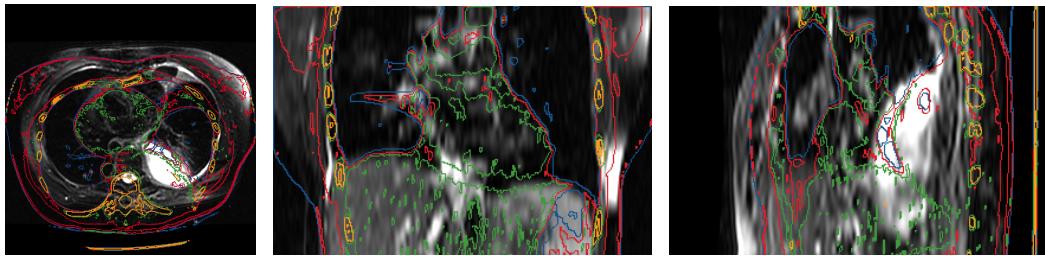


Figure 8.4: Deformable multi-modal registration of 11 cases of CT/MRI scans of empyema patients, evaluated with 12 expert landmarks per case. MIND with discrete optimisation achieves a statistically significant ($p < 0.0336$) better result than normalised mutual information (NMI). The comparatively high residual error is due to both low scan quality, (slice resolution is 8 mm) and the challenging landmark selection for the clinical expert (intraobserver error is 5.8 mm)

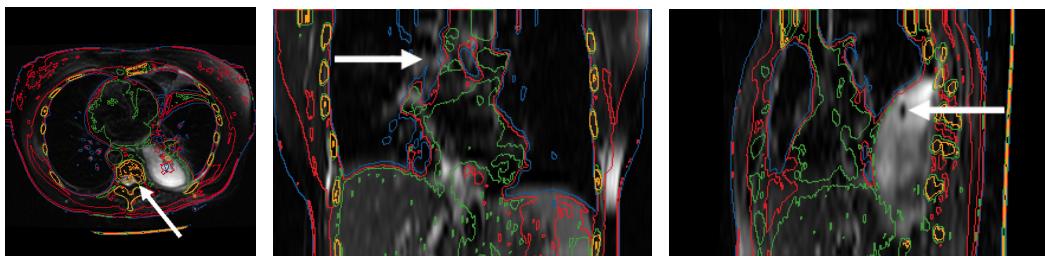
no single failure case (a failure being an increased error after registration). The average computation times per registration are 4.0 min for NMI, 24.4 min for TMI, 3.7 min for MIND and 2.1 min for **deeds**+MIND. Even though the error for MIND is much higher than for the previous CT-to-CT registration, it is lower than the spatial resolution of the MRI scans and close to the intra-observer error. A case-by-case comparison of registration errors is shown in Fig. 8.4. Using a Wilcoxon rank test, a statistically significant improvement of **deeds**+MIND compared to NMI ($p < 0.0336$) was found. The improvements from using TMI compared to NMI are almost negligible when considering the much higher complexity. The Jacobian values are all positive, thus no transformations contained any singularities. An example registration outcome for MIND and NMI is shown in Figure 8.5.



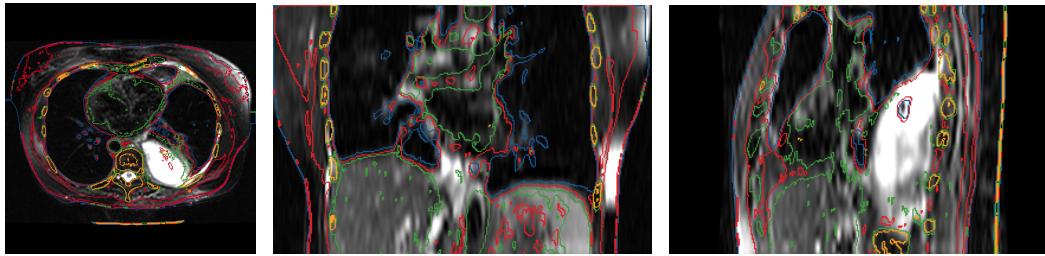
(a) CT scan of empyema patient with 4 relevant contour plots to guide the registration results.



(b) MRI scan with identical CT contour plots before registration.



(c) Identical MRI scan with CT contour plots deformed according to deformable registration using NMI. The white arrows depict inaccurate registration close to one vertebrae, the inner lung boundary and gas pocket in empyema.



(d) Identical MRI scan with CT contour plots deformed according to deformable registration using MIND and Gauss-Newton optimisation. A visually better alignment could be achieved.

Figure 8.5: Deformable CT/MRI registration results for Case 11 of empyema dataset. Left: axial, middle: sagittal and right: coronal plane. The third row shows the registration outcome using NMI. A better alignment is obtained when using MIND (forth row), here with Gauss-Newton optimisation.

8.3 Summary

In this chapter a thorough experimental validation has been presented for the methodologies developed in this thesis. The multidimensional structural image representation MIND has been shown to outperform using intensities directly in the context of single-modal respiratory CT registration. This confirms the findings of Chapter 6, where MIND performed best for landmark localisation, because of its ability to discriminate salient features and its invariance to locally (and globally) different intensity mappings across scans (which also occur for intra-modal lung CT registration due to lung compression). For deformable multi-modal registration textural MI performs slightly better than normalised MI, yet at the expense of much increased computation time. MIND achieves a further improvement on this and can be easily used with discrete optimisation, because of its point-wise representation.

The use of a minimum-spanning-tree (MST) and a dense stochastic sampling in a discrete optimisation framework offers a number of advantages over the continuous Gauss-Newton optimisation, in particular: more flexibility on the choice of a similarity metric, an intuitively definable search space (both its range and quantisation), and the avoidance of local minima and bias to initial misalignments, which removes the need for multi-resolution schemes and in the presented experiments the need for a rigid pre-registration step. Compared to other state-of-the-art registration frameworks [Avants et al., 2008; Glocker et al., 2008a; Schmidt-Richberg et al., 2012a] it offers greatly reduced computational complexity (up to $20\times$ faster registration) while being more accurate. Especially the combination of MIND and **deeds** offers a very good compromise between registration quality and computa-

tional complexity and achieves a lower registration error than the scan resolution close to intra-observer error for lung CT registration (without the need for manual lung segmentations).

The next chapter concludes this thesis and gives an outlook to further research directives and one additional application of our presented approaches: the deformable registration of intra-operative ultrasound to pre-treatment MRI brain scans with a potential application to image-guided neurosurgery.

Chapter 9

Conclusion and outlook

This thesis has presented a comprehensive framework for deformable registration of pulmonary lung images. This framework combines a number of novel mathematical formulations and computational solutions to deformable multi-modal registration, registration-based ventilation estimation and improves on state-of-the-art single-modal registration methods. The main contributions of this thesis are summarised in the following two sections: multi-modal similarity through spatial context (Sec. 9.1) and a simplified graph-based optimisation for registration (Sec. 9.2). In the following a number of promising future research directions are identified, which are building up on the work in this thesis. First, the methods developed here can be generalised for the use of other modalities and anatomical regions. In Sec. 9.3.1 a prototype for ultrasound to MR registration with application to image-guided neurosurgery will be presented with preliminary results. Second, the use of a graph-based optimisation framework offers the potential of representing the images in a more adaptive and efficient way. In Sec. 9.3.2, we propose the use of a parcellation of the image space into a number of overlapping layers of non-uniform

supervoxels, which adhere to image boundaries. We show some initial results for the use of supervoxels to estimate piece-wise smooth motion. Furthermore, the concept would also be very applicable for e.g. registration-based segmentation. Third, deformable registration is always bound to estimate displacements with a limited certainty. The presented discrete optimisation framework can be directly used to obtain an estimate of the probabilities of a dense space of potential displacements. These can also be used to reduce the adverse effect of the quantisation of deformations and improve the propagation of segmentation labels as shown in Sec. 9.3.3.

9.1 Spatial context for robust similarity

Defining similarity for deformable registration of scans from different modalities is a very challenging yet unsolved problem with many potential applications in medical image analysis, foremost computer-assisted diagnosis and image-guided interventions. Chapter 5 introduced *textural mutual information* (TMI), a new concept for the incorporation of spatial context for statistical similarity metrics. While TMI has been shown to be robust against certain intensity distortions, it is computational demanding, has several free parameters to choose and the optimisation of mutual information is challenging and has to be adapted specifically for a given registration framework.

To overcome these problems, a new multi-dimensional image representation the *modality independent neighbourhood descriptor* (MIND) was introduced in Chapter 6. MIND not only provides increased robustness against locally varying contrast, noise and is very discriminative for important image features, it also enables an

efficient solution to the estimation of dense deformation fields across scans for both continuous, gradient-based and discrete, graph-based optimisation.

Chapter 8 demonstrated the suitability of this methodological developments for clinical relevant challenging deformable registration problems, and a significant improvement in terms of accuracy compared to conventional similarity metrics for both multi-modal registration (of CT and MRI scans) and single-modal motion estimation. Being able to obtain dense similarity maps across scans of different modalities also offers a potential to improve the multi-modal fusion of the aligned images. In [Heinrich et al., 2012c], some initial experiments were presented that use a high-resolution CT scan as prior to guide the superresolution reconstruction of a lower resolution MRI scan, which is being aligned simultaneously.

Some of the remaining challenges of the use of MIND is that the scale for the self-similarity distances has to be chosen globally for each image, yet anatomical structures are varying in size. Automatically detecting the best scale for each location or employing multiple scales at once might therefore be beneficial. For the use in linear or feature-based registration the dependency of MIND on the local orientation could be a limitation and calculating descriptors for many different (three-dimensional) rotation is computationally demanding. Thus an efficient interpolation scheme for MIND representations would be useful in practice.

9.2 Efficient graph-based optimisation

Mathematical optimisation for deformable registration of medical volumetric scans is a very challenging problem mainly due to the immense computational complexity. Using "off the shelf" optimisation methods can severely restrict the employed

model of image similarity and geometric transformations and also lead to sub-optimal solutions. Graph-based optimisation techniques, which use a discretised space of permissible deformations, offer much more flexibility in the choice of the registration cost function and provide certain guarantees on the optimality of the solution. However, only few medical registration methods make full use of these possibilities.

In Chapter 7.5 a number of improvements over the state-of-the-art of discrete optimisation for deformable registration were introduced. Using an image-adaptive minimum-spanning-tree (MST) as relaxed graph structure was shown to greatly reduce the computational complexity. This enables the minimisation of larger spaces of possible geometric displacements and the estimation of hyper-parameters such as regional lung ventilation. Furthermore, it also provides for a much simpler modelling of the complex respiratory motion, which contains both smooth and discontinuous deformations, than previous approaches. This has been demonstrated in Secs. 7.6 and 8.1 for the registration of inhale-exhale CT scan pairs. The registration accuracy has been shown to be significantly improved over state-of-the-art methods with a reduction of computation time by more than one order of magnitude. The MST representation may have limitations for some particular registration applications (where the locally reduced regularisation would have adverse effects), however, the other concepts presented in Sec. 7.5 are generally applicable to other MRF inference methods, such as sequential tree-reweighted message passing (TRW-S) [Kolmogorov, 2006] and graph cut optimisation [Kolmogorov and Rother, 2007; Komodakis et al., 2008].

The flexibility of graph-based optimisation can remove some of the restrictions for the underlying registration model, and e.g. enable the use of sparse, non-

uniform image representations and spatially varying interactions between nodes. It also offers the potential of estimating the full marginal distribution over a dense space of potential deformations, which opens new directions for methodological developments and applications of medical image analysis as outlined in Sec. 9.3.3.

9.3 Outlook

Three key research directions have been identified, which could follow-up on the methodological contributions of this thesis and will be described below:

- The generalisation of the concepts of MIND and **deeds** to other imaging modalities and body organs, is proposed in Sec. 9.3.1 with initial experiments for the registration of intra-operative ultrasound to MRI scans of the brain.
- A further development of the discrete optimisation framework to use sparse supervoxel representations of images, which are edge- and detail-preserving. Initial results for the estimation of piece-wise smooth motion (with sliding boundaries) will be presented in Sec. 9.3.2 and [Heinrich et al., 2013b].
- Using the marginal distributions of the dense displacements space of **deeds** to obtain local uncertainty estimates. This can be potentially used to improve registration accuracy and to perform segmentation propagation, which will be proposed in Sec. 9.3.3.

9.3.1 US-MRI registration for neurosurgery

Registration-based image-guidance is very useful in neurosurgery, where the brain tissue exhibits non-linear deformations after opening the skull, which must be com-

pensated to relate the intra-operative ultrasound to the pre-operative MRI scan (which has higher scan quality and can give a better guidance). Image intensity distortions across modalities and time constraints make this a very challenging registration problem. We propose to apply quantised MIND or SSC image representations (as described in Sec. 6.4) together with the **deeds** optimisation of Sec. 7.5 to a set of 13 pairs of pre-operative MRI and pre-resection 3D ultrasound (US) images of the Brain Images of Tumours for Evaluation (BITE) database [Mercier et al., 2012]. [Rivaz and Collins, 2012] applied a multi-feature α mutual information (α -MI) metric as presented by [Staring et al., 2009a] with a stochastic gradient descent optimisation [Klein et al., 2007], and extend this framework using a self-similarity weighting within the feature space, calling the new metric SeSaMI. They achieve good registration results with an average TRE of 2.34 mm (see Fig. 9.2). However, this comes at the expense of very high computation times of 120 min (average per case).

Preliminary registration experiments have been carried out with our registration framework, a visual example of the registration outcome is shown in Fig. 9.1. All resulting transformations are free from singularities with an average complexity measured as standard deviation of the Jacobian of 0.08. The average target registration error (TRE) before registration in our experiments is 6.82 mm. SSC achieves the best overall registration accuracy of 2.12 mm (see Fig. 9.2) and the lowest computation time (20 sec). The TRE is lower than using MIND with six neighbours (2.45 mm), and mutual information (3.07 mm).

These experiments [Heinrich et al., 2013c] show promising results for generalisation the presented methods to different multi-modal registration applications. For the use in image-guided interventions, which could involve the tracking of surgical

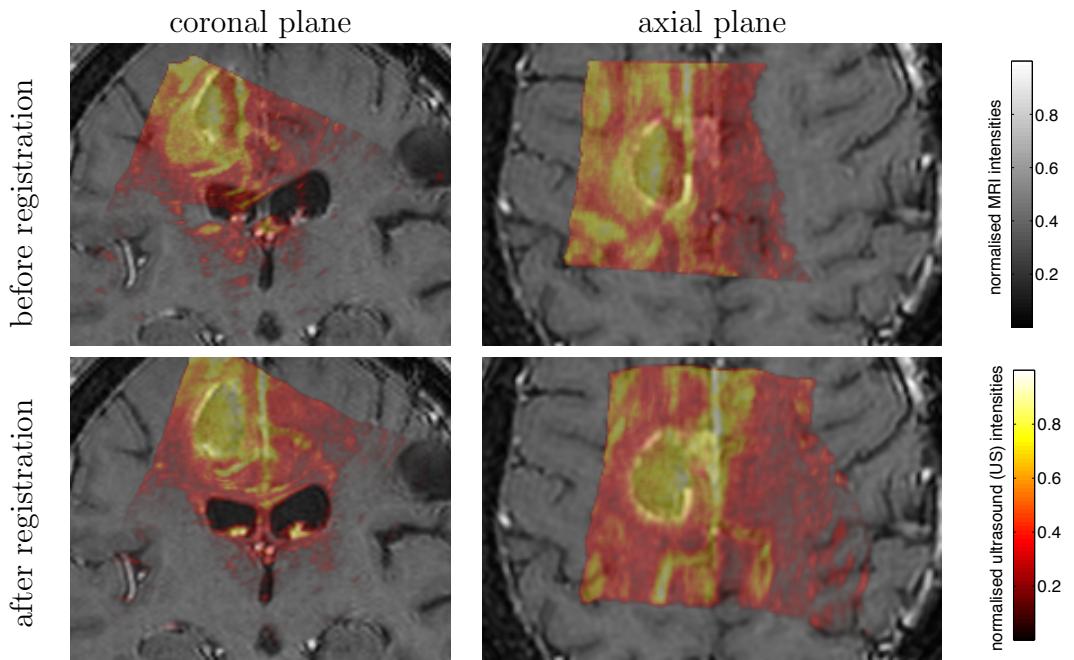


Figure 9.1: Deformable MRI-US registration results of BITE dataset using SSC and **deeds**. The intra-operative ultrasound scan is shown as false colour overlay over the grayscale MRI intensities. An clearly improved alignment of ventricles and solid tumour is visible.

instruments in intra-operative ultrasound, further improvements in computation time would be necessary. A potential way of achieving real-time performance would be a parallel implementation (on graphical processing units) of the registration framework. This is straightforward for the similarity term computations. For the belief propagation, the messages of all nodes having the same depth in the MST can be computed in parallel, which could decrease the computation time by two orders of magnitude. Using sparse image representation will be discussed in the next section as another way of improving computational efficiency.

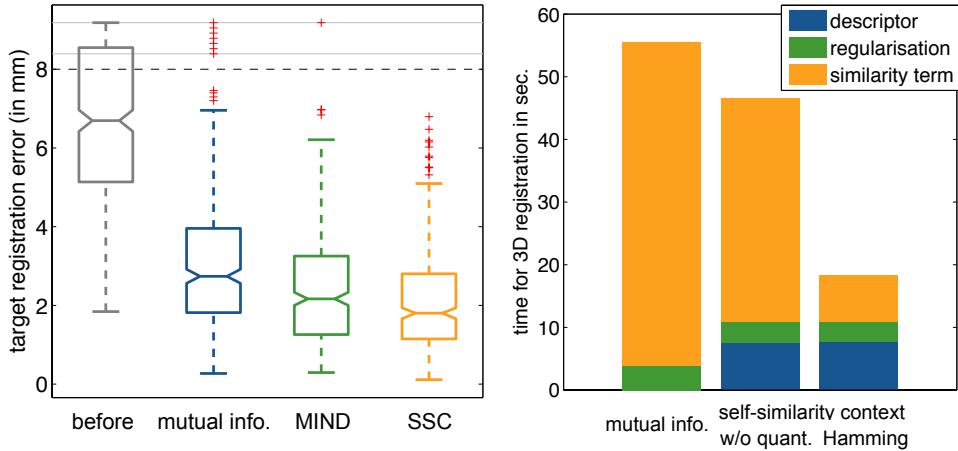


Figure 9.2: Deformable registration of 13 cases of MRI-US brain scans, evaluated with ≈ 27 expert landmarks per case. The registration error of SSC (2.12 ± 1.29 mm) is lower than that of MIND (2.45 ± 1.44 mm) and mutual information (3.07 ± 1.74 mm) within the same discrete optimisation framework. The computation time per registration using SSC and the Hamming distance (≈ 20 sec per 3D pair) is more than twice as fast compared to MI and SSC without quantisation.

9.3.2 Detail-preserving sparse image representations

Estimating a dense deformational field between three-dimensional images of high-resolution requires the optimisation of a function with a large number of degrees of freedom. Commonly used approaches, such as multi-resolution (discussed in Sec. 7.3.2) or B-spline transformation models (see Sec. 7.5.1) reduce the computational complexity, but also reduce the registration accuracy, due to the loss of detail when using homogeneous translation-invariant smoothing. Sparse image representations can potentially offer a better trade-off between efficiency and accuracy. Graph-based models intrinsically offer a representation for sparsely distributed control points. The control points could be easily selected in an adaptive manner using e.g. salient point detectors (see Sec. 3.3) or image-adaptive clusterings. Supervoxels have been shown to yield excellent low-parametric image representations with little

loss of image details and good recall of image boundaries (see Fig. 9.3 as an example). We propose to use the SLIC supervoxel algorithm [Achanta et al., 2012], to create approximately equally-sized supervoxels , which are non-uniformly shaped and adaptive to intensity boundaries.

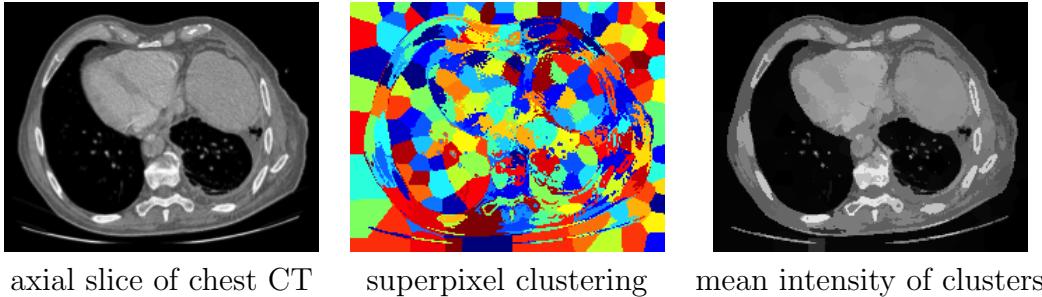


Figure 9.3: Example of sparse image representation using only 250 superpixels is able to preserve image edges and most small-scale structures.

In a graph-based registration framework not only the parameterisation of the transformation model can be represented more efficiently using sparse control points, but also the discretised search space can be reduced. A disadvantage of supervoxels is the inconsistent clustering in homogeneous or gradually changing image regions. In the context of motion estimation [Zitnick and Kang, 2007] this is a major limitation, because it is important which correspondence within a homogeneous region is selected. We propose to use multiple layers of supervoxels to obtain a piecewise smooth motion model for accurate deformable registration. SLIC is run several times with slightly different initialisations. Image regions with sufficient structural content (e.g. edges) are not affected by this disturbance and the clustering is therefore very similar for all layers. Homogeneous or gradually changing areas do not provide sufficient guidance for the supervoxel, resulting in arbitrarily different clusters for each layer. This means that when a separate op-

timisation is performed for each layer of supervoxels. The combination of these transformation is a smooth average in homogeneous regions, but adheres to discontinuities at image boundaries, which enables a non-local weighting of spatial image content, similar to approaches based on the joint bilateral filter [Kopf et al., 2007] as discussed in Sec. 4.1.4.

Experiments for using supervoxel representations for CT lung motion estimation as published in [Heinrich et al., 2013b] are presented below. The more challenging datasets #6-#10 of [Castillo et al., 2009] were used (see Sec. 8.1 for a description of the data). Four different settings of our proposed supervoxel registration were tested. First, our method was applied using only a single layer of supervoxels, similar to previous work on supervoxel matching [Lei and Yang, 2009; Zitnick and Kang, 2007], which yielded unsatisfactory results, as only the outer lung boundaries are aligned. Second, multiple layers were used, but almost no image-adaptivity of the supervoxels was used resulting in a uniform clustering and a target registration error (TRE) of 4.72 ± 3.5 mm. Figure 9.4 demonstrates the main problem of this approach, which is similar to a traditional coarse-scale image representation. The motion field is smooth across the interface at which discontinuous sliding motion occurs (see black arrows). Additionally small details (lung vessels) are lost due to uniform smoothing, resulting in an inaccurate alignment of them (see white circle). Third, multiple layers of image-adaptive supervoxels and an image similarity, based on the absolute difference of the mean cluster intensity, was used. This approach achieved a good alignment and a TRE of 2.87 ± 1.9 mm. Finally, the shape of the supervoxels (stored in a binary vector) was introduced as additional similarity metric. Adding this structural image information, which is closely related to the concept of textural MI (see Sec. 5.3), has clear advantages

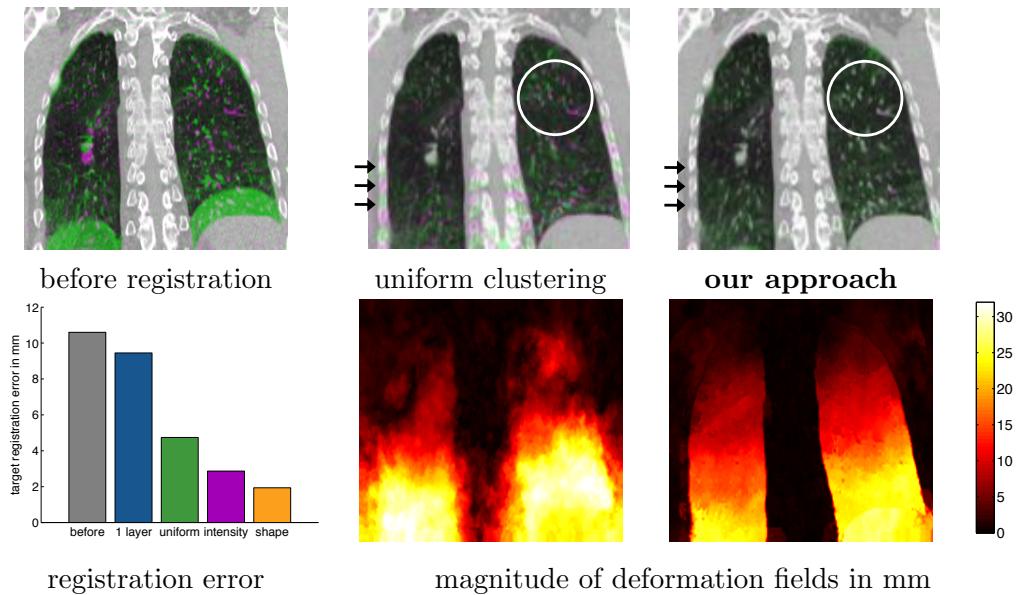


Figure 9.4: Example of deformable registration of an inhale-exhale CT scan pair. Overlay before and after registration is shown in green (inhale phase) and magenta (exhale phase). Uniform clustering approximates the traditional coarse-scale representation. Our approach using layers of supervoxels outperforms this for the matching of small vessels (see white circle) and the preservation of sliding motion (see black arrows).

to match fine image details and further reduces the registration error to 1.94 ± 1.3 mm. Figure 9.4 demonstrates the accurate alignment of our approach and the well-preserved sliding motion at the surface of the thoracic cage.

The same concept can also be used for multi-modal registration, when only the shape-based similarity metric is used. While being well-suited for motion estimation, which exhibits discontinuities (such as the sliding motion of the lungs) the use of supervoxels could have even more potential benefits for the use in registration-based propagation, which is discussed in a different context in the next section.

9.3.3 Marginal distributions for segmentation propagation

Most registration algorithms only estimate the most likely transformation between images. This is sufficient for applications where motion should be estimated in order to align scans or to detect motion abnormalities (e.g. in respiratory or cardiac motion). In other applications of deformable registration, the dense deformation field is only used as a means to propagate a manual segmentation label or estimate local volume shrinkage/expansion. Employing only the most probable transformation to propagate segmentations is unfavourable because it ignores the locally different probabilities for each individual label propagation. In [Glocker et al., 2008b] the local covariance of the uncertainty of a graph-cut solution of the registration is used to refine the displacement label space in subsequent iterations, and improve the accuracy of the alignment (an approach to obtain the full marginal distribution for graph-cut optimisation has been presented by [Kohli and Torr, 2008]). [Simpson et al., 2011] used a covariance estimate of uncertainty of a continuous-optimisation Bayesian registration framework to improve the propagation of segmentation labels.

The graph-based registration method presented in Sec. 7.5, which uses a minimum-spanning-tree and belief propagation for inference, can be directly used (with almost no additional computational cost) to estimate the exact min-marginal energies of all possible displacements. The probability $p(\mathbf{x}_i, \mathbf{u}_i)$ for each voxel $i \in \Omega$ in the image and each displacement $\mathbf{u}_i \in \mathcal{L}$ can be obtained from the min-marginal energies $E(\mathbf{x}, \mathbf{u})$ as described in [Kohli and Torr, 2008]:

$$p(\mathbf{x}_i, \mathbf{u}_i) = \exp\left(-\frac{\beta \cdot E(\mathbf{x}_i, \mathbf{u}_i)}{\text{std}(E(\mathbf{x}_j, \mathbf{u}_j))}\right) \quad \mathbf{u}_j \in \mathcal{L} \text{ and } \mathbf{x}_j \in \Omega \quad (9.1)$$

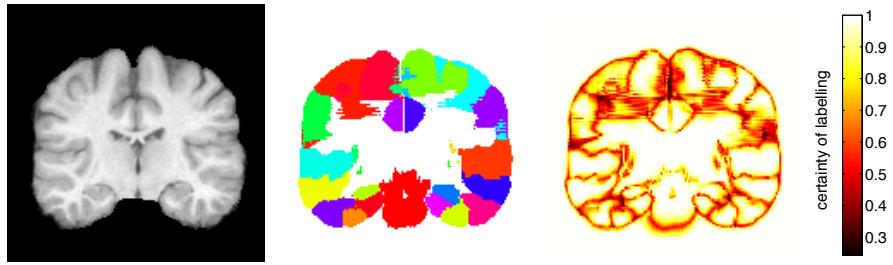


Figure 9.5: Left: Coronal slice of target MRI scan. Centre: Propagated segmentation labels from atlas scan using proposed method. Right: Local certainty of label propagation (high uncertainty indicated by dark colours). At interfaces between segmentations the uncertainty is highest. Here, the weights could be used to estimate partial tissue volume.

Dividing each marginal energy by the standard deviation over all marginals compensates for a global offset/scaling of E . β varies the spread of the probability estimates, low values result in smoother distributions and high values in more narrow peaks. Finding the optimal segmentation label is now possible by summing the probabilities for each possible segmentation label (from the atlas) and choosing the arg max of all labels.

Initial results for the inter-patient propagation of manually segmented brain structures using MRI scans are presented below. The LBPA40 dataset [Shattuck et al., 2008] was used here, which includes MRI scans of 40 normal adults and manual segmentations of 56 anatomical structures. Figure 9.5 shows an example of the proposed segmentation propagation, together with the obtained uncertainty map. Higher uncertainty is located at the boundaries of segmentation labels, which is partly due to the partial volume effect. Segmentation accuracy is measured with the DICE coefficient D (see Sec. 3.2.3). Figure 9.6 (left) shows the influence of the β parameter of Eq. 9.1. It can be seen that too small values of β cause an over-smoothing of the probability distributions and larger values are similar to the

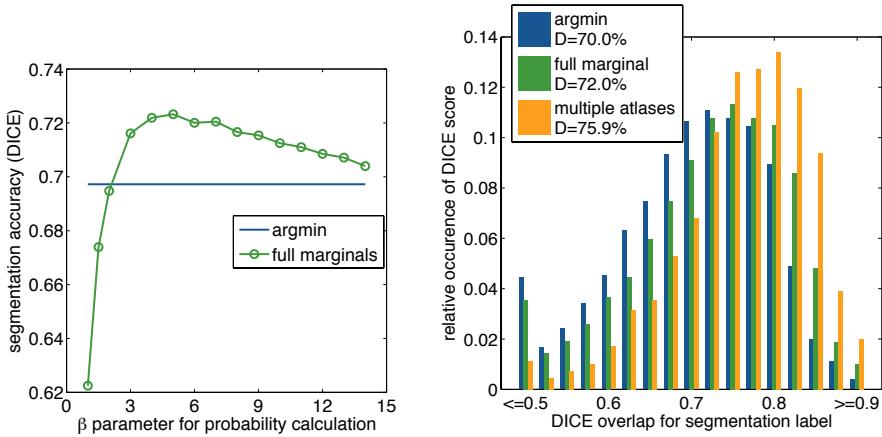


Figure 9.6: Left: Influence of β for the conversion of min-marginals to probabilities in Eq. 9.1 ($\beta \rightarrow \infty$ is equivalent to taking the argmin). Right: Distribution of DICE scores for registration experiments. Improvements of 2% in DICE overlap are obtained when using the full marginal distribution instead of selecting the argmin. Using three atlases further improves the overlap by 4%.

classical approach. Using the proposed probabilistic label weighting scheme the segmentation accuracy could be consistently improved by 2% to 72.0% (see Fig. 9.6 right), compared to the standard argmin approach (significant for 53/56 labels ($p < 0.05$)). Further improvements can be achieved when using multiple atlases for each label propagation ($D = 75.9\%$, for three atlases) or by using a different optimisation strategy, e.g. TRW-S [Kolmogorov, 2006].

9.4 Summary

This thesis has addressed some of the major challenges involved with deformable image registration in the context of lung imaging. A novel multi-dimensional image representation to define image similarity across scans of different modalities or with locally changing contrast has been developed and thoroughly tested. Methodologi-

cal and computational advances have been made to the optimisation of deformable registration, which also enables an effective way of dealing with complex lung motion, which includes sliding motion and large displacements of small anatomical structures.

For future work, it would be important to evaluate and validate the clinical impact of this comprehensive registration framework for pulmonary image analysis. The development of a prototype of an interactive software tool for the assessment of multi-modal scans has been started for this purpose. A side-by-side view of different scans gives clinicians the ability to view and perform different assessment tasks (measuring pathology or propagating manual segmentations) in both modalities simultaneously. The dense deformable motion estimation, which can be performed using the presented framework with little computation time, could then be used to establish anatomical correspondences and automatically select the correct slice and location for a point of interest for the other scan.

Another opportunity to validate the clinical relevance of this work will be possible within an ongoing clinical trial of our collaborators in the Churchill Hospital in Oxford. A very new imaging modality using hyper-polarised Xenon gas MRI (Xe-MRI) scans has been introduced, which can directly measure the spatial distribution of lung ventilation. This could become an important diagnostic modality for chronic obstructive pulmonary disease (COPD) or asthma patients. To assess the correlation between the measurements of regional lung ventilation using Xe-MRI and anatomical CT scans, multi-modal deformable registration is necessary. Additionally, the CT-based ventilation estimation, which was developed in Sec. 7.7 could be clinically validated with the help of MRI-based ventilation maps.

Appendix A

List of publications

In the following a list of all publications stemming from this thesis is given:

Heinrich, M., Schnabel, J., Gleeson, F., Brady, M. and Jenkinson, M. [2010a], Non-Rigid Multimodal Medical Image Registration using Optical Flow and Gradient Orientation, Proc. Medical Image Analysis and Understanding pp. 141–145.

Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2010b], Discontinuity preserving regularisation for variational optical-flow registration using the modified L^p norm, Med. Image Anal. Clinic: A Grand Challenge pp. 185–194.

Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2011a], 'Non-rigid image registration through efficient discrete optimization', in Proc. Medical Image Analysis and Understanding.

Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, J. and Schnabel, J. [2011b], 'Non-local shape descriptor: A new similarity metric for deformable multi-modal registration', in G. Fichtinger, A. Martel and T. Peters, eds, Medical Image Computing and Computer-Assisted Intervention MICCAI 2011, Vol. 6892 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 541–548.

Heinrich, M. P., Jenkinson, M., Gleeson, F. V., Brady, S. M. and Schnabel, J. A. [2011c], Deformable multimodal registration with gradient orientation based on structure tensors, Annals of the BMVA 2011(2), 1 – 11.

Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2012a], 'Textural mutual information based on cluster trees for multimodal deformable registration', in IEEE International Symposium on Biomedical Imaging, ISBI 2012 , pp. 1–4.

Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, M. and Schnabel, J. A. [2012b], MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration, *Medical Image Analysis* **16**(7), 1423 – 1435.

Heinrich, M. P., Jenkinson, M., Brady, S. M. and Schnabel, J. A. [2012c], Robust super-resolution reconstruction with multi-modal registration and guidance, in Proc. Medical Imaging and Understanding, pp. 81–87.

Heinrich, M. P., Jenkinson, M., Brady, S. M. and Schnabel, J. A. [2012d], Globally optimal registration on a minimum spanning tree using dense displacement sampling, in H. Delingette, P. Golland, M. Kensaku and N. Ayache, eds, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, Lecture Notes in Computer Science, Springer Berlin / Heidelberg.

Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2013a], MRF-based deformable registration and ventilation estimation of lung CT, *IEEE Transactions on Medical Imaging* **32**(7) pp. 1239–1248.

Heinrich, M., Jenkinson, M., Gleeson, F. V., Brady, M. and Schnabel, J. [2013b], Edge- and detail-preserving sparse image representations for deformable registration of chest MRI and CT volumes, in J.C. Gee, S. Joshi, K.M. Pohl, W.M. Wells and L. Zoellei, eds, *Information Processing in Medical Imaging, IPMI 2013*, Lecture Notes in Computer Science pp. 463–474.

Heinrich, M. P., Jenkinson, M., Papiez, B.W., Brady, S. M. and Schnabel, J. A. [2013c], Towards Real-time Multimodal Fusion for Image-Guided Interventions using Self-Similarities, In: *Medical Image Computing and Computer-Assisted Intervention MICCAI' 2013*, Lecture Notes in Computer Science, Springer Berlin / Heidelberg.

A.1 Co-authored publications

Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Pluim, J., Heinrich, M. P., Jenkinson, M., Schnabel, J. A. and others [2011].. Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge, *IEEE Transactions on Medical Imaging* **30**(11), 1901–1920.

Bhushan, M., Schnabel, J., Risser, L., Heinrich, M., Brady, J. and Jenkinson, M. [2011], Motion correction and parameter estimation in dceMRI sequences: application to colorectal cancer, in G. Fichtinger, A. Martel and T. Peters, eds, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011*,

Vol. 6892 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg,
pp. 476483.

Risser, L., Heinrich, M., Rueckert, D. and Schnabel, J. [2011b], Multi-modal diffeomorphic registration using mutual information: Application to the registration of CT and MR pulmonary images, in Proc. of the Fourth international workshop on pulmonary image analysis, MICCAI.

Risser, L., Heinrich, M., Matin, T. and Schnabel, J. [2012], Piecewise-diffeomorphic registration of 3D CT/MR pulmonary images with sliding conditions, in IEEE International Symposium on Biomedical Imaging, ISBI 2012 , pp. 1–4.

A.2 Awards and prizes

MICCAI Young Scientist Award (best paper/presentation) 2011 for: 'Non-local shape descriptor: A new similarity metric for deformable multi-modal registration'

Philips Healthcare Prize for best oral presentation at MIUA 2011 for: 'Non-rigid image registration through efficient discrete optimization'

Bibliography

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and Süsstrunk, S. [2012], ‘SLIC superpixels compared to state-of-the-art superpixel methods’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **34**(11), 2274 –2282.
- Ackerman, M. [1998], ‘The visible human project’, *Proceedings of the IEEE* **86**(3), 504 –511.
- Andronache, A., von Siebenthal, M., Székely, G. and Cattin, P. [2008], ‘Non-rigid registration of multi-modal images using both mutual information and cross-correlation’, *Medical image analysis* **12**(1), 3.
- Arsigny, V., Commowick, O., Pennec, X. and Ayache, N. [2006], A log-Euclidean framework for statistics on diffeomorphisms, in R. Larsen, M. Nielsen and J. Sporring, eds, ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 2006’, Vol. 4190 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 924–931.
- Arun, K. S., Huang, T. S. and Blostein, S. D. [1987], ‘Least-squares fitting of two 3-d point sets’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **9**(5), 698 –700.
- Ashburner, J. and Friston, K. J. [2005], ‘Unified segmentation’, *NeuroImage* **26**(3), 839 – 851.
- Ashburner, J. et al. [2007], ‘A fast diffeomorphic image registration algorithm’, *Neuroimage* **38**(1), 95–113.
- Avants, B., Epstein, C., Grossman, M. and Gee, J. [2008], ‘Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain’, *Medical Image Analysis* **12**(1), 26 – 41.
- Baluwala, H., Risser, L., Schnabel, J. and Saddi, K. [2013], ‘Toward physiologically motivated registration of diagnostic CT and PET/CT of lung volumes’, *Medical physics* **40**.

- Beg, M., Miller, M., Trouve, A. and Younes, L. [2005], ‘Computing large deformation metric mappings via geodesic flows of diffeomorphisms’, *International Journal of Computer Vision* **61**(2), 139–157.
- Benhimane, S. and Malis, E. [2004], Real-time image-based tracking of planes using efficient second-order minimization, in ‘Intelligent Robots and Systems, IROS 2004’, Vol. 1, pp. 943–948.
- Bhushan, M., Schnabel, J., Risser, L., Heinrich, M., Brady, J. and Jenkinson, M. [2011], Motion correction and parameter estimation in dceMRI sequences: application to colorectal cancer, in ‘Medical Image Computing and Computer-Assisted Intervention, MICCAI 2011’, Springer, pp. 476–483.
- Bishop, C. M. et al. [2006], *Pattern recognition and machine learning*, Vol. 4, Springer New York.
- Blake, A. and Zisserman, A. [1987], *Visual Reconstruction*, MIT Press.
- Bookstein, F. L. [1989], ‘Principal warps: Thin-plate splines and the decomposition of deformations’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **11**(6), 567–585.
- Boyd, S. P. and Vandenberghe, L. [2004], *Convex optimization*, Cambridge university press.
- Boykov, Y. and Kolmogorov, V. [2004], ‘An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **26**(9), 1124 –1137.
- Boykov, Y., Veksler, O. and Zabih, R. [2001], ‘Fast approximate energy minimization via graph cuts’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **23**(11), 1222 –1239.
- Brenner, D. and Hall, E. [2007], ‘Computed tomographyan increasing source of radiation exposure’, *New England Journal of Medicine* **357**(22), 2277–2284.
- Bronstein, M. M., Bronstein, A. M., Michel, F. and Paragios, N. [2010], Data fusion through cross-modality metric learning using similarity-sensitive hashing, in ‘Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on’, IEEE, pp. 3594–3601.
- Brox, T., Bruhn, A., Papenberg, N. and Weickert, J. [2004], High accuracy optical flow estimation based on a theory for warping, in T. Pajdla and J. Matas, eds, ‘European Conference on Computer Vision, ECCV 2004’, Vol. 3024 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 25–36.

- Brox, T., Kleinschmidt, O. and Cremers, D. [2008a], ‘Efficient nonlocal means for denoising of textural patterns’, *Image Processing, IEEE Transactions on* **17**(7), 1083 –1092.
- Brox, T., Kleinschmidt, O. and Cremers, D. [2008b], ‘Efficient nonlocal means for denoising of textural patterns’, *Image Processing, IEEE Transactions on* **17**(7), 1083 –1092.
- Bruhn, A., Weickert, J. and Schnörr, C. [2005], ‘Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods’, *International Journal of Computer Vision* **61**(3), 211–231.
- Buades, A., Coll, B. and Morel, J.-M. [2005], A non-local algorithm for image denoising, in ‘Computer Vision and Pattern Recognition, CVPR 2005’, pp. 60–65.
- Calonder, M., Lepetit, V., Strecha, C. and Fua, P. [2010], BRIEF: Binary robust independent elementary features, in ‘Computer Vision–ECCV 2010’, Springer, pp. 778–792.
- Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A. and Guerrero, T. [2009], ‘A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets’, *Physics in Medicine and Biology* **54**(7), 1849.
- Castillo, R., Castillo, E., Martinez, J. and Guerrero, T. [2010], ‘Ventilation from four dimensional computed tomography: Density versus Jacobian methods’, *Physics in Medicine and Biology* **55**, 4661–4685.
- Charbonnier, P., Blanc-Fueraud, L., Aubert, G. and Barlaud, M. [1994], Two deterministic half-quadratic regularization algorithms for computed imaging, in ‘IEEE International Conference on Image Processing’, Vol. 2, pp. 168–172.
- Chen, M., Lu, W., Chen, Q., Ruchala, K. J. and Olivera, G. H. [2007], ‘A simple fixed-point approach to invert a deformation field’, *Medical Physics* **35**(1), 81.
- Christensen, G., Geng, X., Kuhl, J., Bruss, J., Grabowski, T., Pirwani, I., Vanier, M., Allen, J. and Damasio, H. [2006], Introduction to the non-rigid image registration evaluation project (NIREP), in J. Pluim, B. Likar and F. Gerritsen, eds, ‘Biomedical Image Registration’, Vol. 4057 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 128–135.
- Christensen, G. and Johnson, H. [2001], ‘Consistent image registration’, *IEEE Transactions on Medical Imaging* **20**(7), 568 –582.

- Coupé, P., Yger, P. and Barillot, C. [2006], Fast non local means denoising for 3D MR images, in ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 2006’, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 33–40.
- Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C. and Barillot, C. [2008], ‘An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images’, *Medical Imaging, IEEE Transactions on* **27**(4), 425 –441.
- D’Agostino, E., Maes, F., Vandermeulen, D. and Suetens, P. [2003], ‘A viscous fluid model for multimodal non-rigid image registration using mutual information.’, *Medical Image Analysis* **7**(4), 565–575.
- De Nigris, D., Mercier, L., Del Maestro, R., Collins, D. L. and Arbel, T. [2010], Hierarchical multimodal image registration based on adaptive local mutual information, in T. Jiang, N. Navab, J. Pluim and M. Viergever, eds, ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 2010’, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 643–651.
- Delmon, V., Rit, S., Pinho, R. and Sarrut, D. [2013], ‘Registration of sliding objects using direction dependent B-splines decomposition’, *Physics in medicine and biology* **58**(5), 1303.
- Derpanis, K. G. [2005], ‘Single 2D orientation estimation via the structure tensor’, Technical Report, York University, Canada pp. 1–3.
- Deselaers, T. and Ferrari, V. [2010], Global and efficient self-similarity for object classification and detection, in ‘Computer Vision and Pattern Recognition, CVPR 2010’, pp. 1633 –1640.
- Dowson, N., Kadir, T. and Bowden, R. [2008], ‘Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **30**(10), 1841 –1857.
- Felzenszwalb, P. F. and Huttenlocher, D. P. [2005], ‘Pictorial structures for object recognition’, *International Journal of Computer Vision* **61**, 55–79.
- Felzenszwalb, P. and Huttenlocher, D. [2006], ‘Efficient belief propagation for early vision’, *International Journal of Computer Vision* **70**, 41–54.
- Fitzpatrick, J. M., West, J. B. and Maurer Jr, C. R. [1998], ‘Predicting error in rigid-body point-based registration’, *Medical Imaging, IEEE Transactions on* **17**(5), 694–702.

- Fox, N. C., Freeborough, P. A. and Rossor, M. N. [1996], ‘Visualisation and quantification of rates of atrophy in alzheimer’s disease’, *The Lancet* **348**(9020), 94–97.
- Glocker, B., Komodakis, N., Navab, N., Tziritas, G. and Paragios, N. [2009], Dense registration with deformation priors, in ‘Information processing in medical imaging’, Springer, pp. 540–551.
- Glocker, B., Komodakis, N., Tziritas, G., Navab, N. and Paragios, N. [2008a], ‘Dense image registration through MRFs and efficient linear programming’, *Medical Image Analysis* **12**(6), 731 – 741.
- Glocker, B., Paragios, N., Komodakis, N., Tziritas, G. and Navab, N. [2008b], Optical flow estimation with uncertainties through dynamic MRFs, in ‘Computer Vision and Pattern Recognition, CVPR 2008n’, IEEE, pp. 1–8.
- Glocker, B., Sotiras, A., Komodakis, N. and Paragios, N. [2011], ‘Deformable medical image registration: Setting the state of the art with discrete methods’, *Annual Review of Biomedical Engineering* **13**(1), 219–244.
- Gorbunova, V., Sporring, J., Lo, P., Loeve, M., Tiddens, H. A., Nielsen, M., Dirksen, A. and de Bruijne, M. [2012], ‘Mass preserving image registration for lung CT’, *Medical Image Analysis* **16**(4), 786 – 795.
- Guerrero, T., Sanders, K., Noyola-Martinez, J., Castillo, E., Zhang, Y., Tapia, R., Guerra, R., Borghero, Y. and Komaki, R. [2005], ‘Quantification of regional ventilation from treatment planning CT’, *International Journal of Radiation Oncology, Biology, Physics* **62**, 630–634.
- Haber, E. and Modersitzki, J. [2007], ‘Intensity gradient based registration and fusion of multi-modal images’, *Methods of information in medicine* **46**(3), 292–299.
- Hajnal, J. V., Hill, D. L. and Hawkes, D. J. [2001], *Medical image registration*, CRC.
- Hamming, R. W. [1950], ‘Error detecting and error correcting codes’, *Bell System technical journal* **29**(2), 147–160.
- Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2010b], ‘Discontinuity preserving regularisation for variational optical-flow registration using the modified l_p norm’, *Med. Image Anal. Clinic: A Grand Challenge* pp. 185–194.

- Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2011a], Non-rigid image registration through efficient discrete optimization, in ‘Proc. Medical Image Analysis and Understanding’.
- Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2012a], Textural mutual information based on cluster trees for multimodal deformable registration, in ‘International Symposium on Biomedical Imaging, ISBI 2012’, pp. 1–4.
- Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2013a], ‘MRF-based deformable registration and ventilation estimation of lung CT’, *IEEE Transactions on Medical Imaging* **32**(7), 1239–1248.
- Heinrich, M., Jenkinson, M., Brady, M. and Schnabel, J. [2013d], ‘Dense displacement sampling and self- similarities applied to the empire10 dataset’, *EMPIRE challenge website empire10.isi.uu.nl* pp. 1–7.
- Heinrich, M., Jenkinson, M., B.W., P., Brady, M. and Schnabel, J. [2013c], Towards real-time multimodal fusion for image-guided interventions using self-similarities, in ‘Medical Image Computing and Computer-Assisted Intervention, MICCAI 2011’, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 187–194.
- Heinrich, M., Jenkinson, M., Papiez, B., Brady, M. and Schnabel, J. [2013b], Edge-and detail-preserving sparse image representations for deformable registration of chest MRI and CT volumes, in J. Gee, S. Joshi, K. Pohl, W. Wells and Z. L., eds, ‘Information Processing in Medical Imaging, IPMI 2013’, Lecture Notes in Computer Science, Springer, pp. 463–474.
- Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, S. M. and Schnabel, J. A. [2012b], ‘MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration’, *Medical Image Analysis* **16**(7), 1423 – 1435.
- Heinrich, M. P., Jenkinson, M., Brady, S. M. and Schnabel, J. A. [2012c], Robust super-resolution reconstruction with multi-modal registration and guidance, in ‘Proc. Medical Imaging and Understanding’, pp. 81–87.
- Heinrich, M. P., Jenkinson, M., Brady, S. M. and Schnabel, J. A. [2012d], Globally optimal registration on a minimum spanning tree using dense displacement sampling, in H. Delingette, P. Golland, M. Kensaku and N. Ayache, eds, ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 2012’, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 115–122.

- Heinrich, M. P., Jenkinson, M., Gleeson, F. V., Brady, S. M. and Schnabel, J. A. [2011c], ‘Deformable multimodal registration with gradient orientation based on structure tensors’, *Annals of the BMVA* **2011**(2), 1 – 11.
- Heitz, G. and Koller, D. [2008], Learning spatial context: Using stuff to find things, in D. A. Forsyth, P. H. S. Torr and A. Zisserman, eds, ‘European Conference on Computer Vision, ECCV 2008’, Vol. 5302 of *Lecture Notes in Computer Science*, Springer, pp. 30–43.
- Hermosillo, G., Chefd’hotel, C. and Faugeras, O. [2002], ‘Variational methods for multimodal image matching’, *Int. J. Comput. Vision* **50**(3), 329–343.
- Hirschmüller, H. and Scharstein, D. [2009], ‘Evaluation of stereo matching costs on images with radiometric differences’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **31**(9), 1582 –1599.
- Holden, M. [2008], ‘A review of geometric transformations for nonrigid body registration’, *Medical Imaging, IEEE Transactions on* **27**(1), 111 –128.
- Horn, B. and Schunck, B. [1981], ‘Determining optical flow’, *Artificial Intelligence* **17**, 185–203.
- Hua, X., Leow, A. D., Parikshak, N., Lee, S., Chiang, M.-C., Toga, A. W., Jr, C. R. J., Weiner, M. W. and Thompson, P. M. [2008], ‘Tensor-based morphometry as a neuroimaging biomarker for alzheimer’s disease: An MRI study of 676 AD, MCI, and normal subjects’, *NeuroImage* **43**(3), 458 – 469.
- Joshi, N., Kadir, T. and Brady, S. [2011], ‘Simplified computation for nonparametric windows method of probability density function estimation’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **33**(8), 1673 –1680.
- Kabus, S., von Berg, J., Yamamoto, T., Opfer, R. and Keall, P. [2008], Lung ventilation estimation based on 4D-CT imaging., in ‘Proc. of the first international workshop on pulmonary image analysis, MICCAI’, pp. 73–82.
- Kiefer, J. and Wolfowitz, J. [1952], ‘Stochastic estimation of the maximum of a regression function’, *The Annals of Mathematical Statistics* **23**(3), 462–466.
- Kim, J., Kolmogorov, V. and Zabih, R. [2003], Visual correspondence using energy minimization and mutual information, in ‘International Conference on Computer Vision, ICCV 2003’, pp. 1033 –1040 vol.2.
- Kirkpatrick, S., Vecchi, M. et al. [1983], ‘Optimization by sammulated annealing’, *Science* **220**(4598), 671–680.

- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P. et al. [2009], ‘Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration’, *Neuroimage* **46**(3), 786.
- Klein, S., Staring, M. and Pluim, J. [2007], ‘Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-Splines’, *Image Processing, IEEE Transactions on* **16**(12), 2879–2890.
- Klein, S., Staring, M. and Pluim, J. P. W. [2005], Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration, in ‘Proc. of SPIE Vol’, Vol. 5747, pp. 192–203.
- Kleinschmidt, O., Brox, T. and Cremers, D. [2008], Nonlocal texture filtering with efficient tree structures and invariant patch similarity measures, in ‘Int. Workshop on Local and Non-Local Approximation in Image Processing, LNLA’.
- Knops, Z., Maintz, J., Viergever, M. and Pluim, J. [2006], ‘Normalized mutual information based registration using k-means clustering and shading correction.’, *Medical image analysis* **10**(3), 432–439.
- Koenderink, J. and Pont, S. [2003], ‘Irradiation direction from texture’, *Journal of the Optical Society of America A* **20**(10), 1875–1882.
- Kohli, P. and Torr, P. [2008], ‘Measuring uncertainty in graph cut solutions’, *Computer Vision and Image Understanding* **112**(1), 30–38.
- Kolmogorov, V. [2006], ‘Convergent tree-reweighted message passing for energy minimization’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(10), 1568–1583.
- Kolmogorov, V. and Rother, C. [2007], ‘Minimizing nonsubmodular functions with graph cuts-a review’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **29**(7), 1274–1279.
- Komodakis, N., Tziritas, G. and Paragios, N. [2008], ‘Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies’, *Computer Vision and Image Understanding* **112**(1), 14–29.
- Kopf, J., Cohen, M., Lischinski, D. and Uyttendaele, M. [2007], Joint bilateral upsampling, in ‘SIGGRAPH’, Proceedings of ACM Conference.

- Krähenbühl, P. and Koltun, V. [2011], Efficient inference in fully connected CRFs with gaussian edge potentials, in ‘Advances in Neural Information Processing Systems’, pp. 109–117.
- Krähenbühl, P. and Koltun, V. [2012], ‘Efficient nonlocal regularization for optical flow’, *ACM Transactions on Graphics* **31**, 4.
- Kumar, N., Zhang, L. and Nayar, S. [2008], What is a good nearest neighbors algorithm for finding similar patches in images?, in ‘European Conference on Computer Vision, ECCV 2008’, pp. 364–378.
- Lee, D., Hofmann, M., Steinke, F., Altun, Y., Cahill, N. D. and Scholkopf, B. [2009], Learning similarity measure for multi-modal 3D image registration, in ‘Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on’, IEEE, pp. 186–193.
- Lee, K. J., Kwon, D., Yun, I. D. and Lee, S. U. [2008], Deformable 3D volume registration using efficient MRFs model with decomposed nodes, in M. Everingham, C. J. Needham and R. Fraile, eds, ‘British Machine Vision Conference, BMVC 2008’.
- Lee, S., Wolberg, G. and Shin, S. Y. [1997], ‘Scattered data interpolation with multilevel b-splines’, *Visualization and Computer Graphics, IEEE Transactions on* **3**(3), 228–244.
- Lei, C. and Yang, Y.-H. [2009], Optical flow estimation on coarse-to-fine region-trees using discrete optimization, in ‘International Conference on Computer Vision, ICCV 2009’, pp. 1562 –1569.
- Leow, A., Yanovsky, I., Chiang, M.-C., Lee, A., Klunder, A., Lu, A., Becker, J., Davis, S., Toga, A. and Thompson, P. [2007], ‘Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration’, *Medical Imaging, IEEE Transactions on* **26**(6), 822 –832.
- Leutenegger, S., Chli, M. and Siegwart, R. Y. [2011], BRISK: Binary robust invariant scalable keypoints, in ‘Computer Vision (ICCV), 2011 IEEE International Conference on’, IEEE, pp. 2548–2555.
- Liao, S. and Chung, A. C. [2009], Non-rigid image registration with uniform spherical structure patterns, in ‘Information Processing in Medical Imaging’, Springer, pp. 163–175.

- Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D. and Suetens, P. [2007], Nonrigid image registration using conditional mutual information, in ‘Information Processing in Medical Imaging, IPMI 2007’, Lecture Notes in Computer Science, pp. 725–737.
- Lowe, D. [1999], Object recognition from local scale-invariant features, in ‘International Conference Computer Vision, ICCV 1999’, Vol. 2, pp. 1150–1157.
- Madsen, K., Bruun, H. and Tingleff, O. [2004], Methods for non-linear least squares problems, Informatics and Mathematical Modelling, Technical University of Denmark, DTU.
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. and Suetens, P. [1997], ‘Multimodality image registration by maximization of mutual information’, IEEE Transactions on Medical Imaging **16**(2), 187–198.
- Mahapatra, D. and Sun, Y. [2012], ‘Integrating segmentation information for improved MRF-based elastic image registration’, IEEE Transactions on Image Processing **21**(1), 170–183.
- Mellor, M. and Brady, M. [2005], ‘Phase mutual information as a similarity measure for registration’, Medical Image Analysis **9**(4), 330 – 343.
- Mercier, L., Del Maestro, R., Petrecca, K., Araujo, D., Haegelen, C. and Collins, D. [2012], ‘Online database of clinical MR and ultrasound images of brain tumors’, Medical Physics **39**, 3253.
- Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A. and Wahl, R. L. [1997], ‘Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations’, Medical Image Analysis **1**(3), 195 – 206.
- Mikolajczyk, K. and Schmid, C. [2005], ‘A performance evaluation of local descriptors’, Pattern Analysis and Machine Intelligence, IEEE Transactions on **27**(10), 1615 –1630.
- Modat, M., Vercauteren, T., Ridgway, G., Hawkes, D., N.C., F. and S., O. [2010], Diffeomorphic demons using normalized mutual information, evaluation on multimodal brain MR images, in ‘SPIE Medical Imaging’, Proceedings of.
- Modersitzki, J. and Wirtz, S. [2006], Combining homogenization and registration, in J. P. W. Pluim, B. Likar and F. A. Gerritsen, eds, ‘Biomedical Image Registration, Third International Workshop, WBIR 2006’, Vol. 4057 of Lecture Notes in Computer Science, Springer, pp. 257–263.

Murphy, K., van Ginneken, B., Klein, S., Staring, M., de Hoop, B., Viergever, M. and Pluim, J. [2011a], ‘Semi-automatic construction of reference standards for evaluation of image registration’, *Medical Image Analysis* **15**(1), 71.

Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G., Garcia, V., Vercauteren, T., Ayache, N., Commowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporrung, J., De Bruijne, M., Han, X., Heinrich, M., Schnabel, J., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J., Ourselin, S., Muenzing, S., Viergever, M., De Nigris, D., Collins, D., Arbel, T., Peroni, M., Li, R., Sharp, G., Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Smeets, D., Loeckx, D., Song, G., Tustison, N., Avants, B., Gee, J., Staring, M., Klein, S., Stoel, B., Urschler, M., Werlberger, M., Vandemeulebroucke, J., Rit, S., Sarrut, D. and Pluim, J. P. W. [2011b], ‘Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge’, *Medical Imaging, IEEE Transactions on* **30**(11), 1901 – 1920.

Myronenko, A. and Song, X. [2010], ‘Intensity-based image registration by minimizing residual complexity’, *Medical Imaging, IEEE Transactions on* **29**(11), 1882 –1891.

Nagel, H. and Enkelmann, W. [1986], ‘An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **5**, 565–593.

Ojala, T., Pietikainen, M. and Maenpaa, T. [2002], ‘Multiresolution gray-scale and rotation invariant texture classification with local binary patterns’, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(7), 971–987.

Ou, Y., Sotiras, A., Paragios, N. and Davatzikos, C. [2011], ‘DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting’, *Medical Image Analysis* **15**(4), 622 – 639.

Ou, Y., Ye, D., Pohl, K. and Davatzikos, C. [2012], Validation of DRAMMS among 12 popular methods in cross-subject cardiac MRI registration, in ‘Biomedical Image Registration’, Vol. 7359 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 209–219.

OConnor, J., Tofts, P., Miles, K., Parkes, L., Thompson, G. and Jackson, A. [2011], ‘Dynamic contrast-enhanced imaging techniques: CT and MRI’, *British Journal of Radiology* **84**, 112–120.

Pennec, X., Cachier, P. and N., A. [1999], Understanding the demons algorithm: 3D non-rigid registration by gradient descent, in ‘Conference on Medical Image

Computing and Computer Assisted Intervention, MICCAI 1999’, Vol. 1679 of Lecture Notes in Computer Science, Springer, Heidelberg, pp. 597–605.

Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D. and hawkes, D. [1998], ‘A comparison of similarity measures for use in 2-D-3-D medical image registration’, *Medical Imaging, IEEE Transactions on* **17**(4), 586 –595.

Pluim, J., Maintz, J. and Viergever, M. [2000], ‘Image registration by maximization of combined mutual information and gradient information.’, *IEEE transactions on medical imaging* **19**(8), 809.

Porikli, F. [2005], Integral histogram: A fast way to extract histograms in cartesian spaces, in ‘Computer Vision and Pattern Recognition, CVPR 2005’, Vol. 1, IEEE, pp. 829–836.

Prim, R. C. [1957], ‘Shortest connection networks and some generalizations’, *Bell System Technical Journal* **36**, 1389–1401–244.

Rey, D., Subsol, G., Delingette, H. and Ayache, N. [2002], ‘Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis’, *Medical Image Analysis* **6**(2), 163–179.

Risser, L., Heinrich, M., Rueckert, D. and Schnabel, J. [2011b], Multi-modal diffeomorphic registration using mutual information: Application to the registration of CT and MR pulmonary images, in ‘Proc. of the Fourth international workshop on pulmonary image analysis, MICCAI’.

Risser, L., Vialard, F., Baluwala, H. and Schnabel, J. [2012b], ‘Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions’, *Medical Image Analysis* .

Risser, L., Vialard, F., Wolz, R., Murgasova, M., Holm, D. and Rueckert, D. [2011a], ‘Simultaneous multiscale registration using large deformation diffeomorphic metric mapping’, *IEEE Transaction on Medical Imaging* **30**, 1746–59.

Rivaz, H. and Collins, D. [2012], Self-similarity weighted mutual information: A new nonrigid image registration metric, Lecture Notes in Computer Science, Springer, pp. 91–98.

Robbins, H. and Monro, S. [1951], ‘A stochastic approximation method’, *Annals of Mathematical Statistics* **22**(3), 400–407.

Roche, A., Malandain, G., Pennec, X. and Ayache, N. [1998], The correlation ratio as a new similarity measure for multimodal image registration, in W. Wells,

- A. Colchester and S. Delp, eds, ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 98’, Vol. 1496 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 1115–1124.
- Rogelj, P., Kovacic, S. and Gee, J. C. [2003], ‘Point similarity measures for non-rigid registration of multi-modal data’, *Comput. Vis. Image Und.* **92**(1), 112 – 140.
- Rohlfing, T. [2012], ‘Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable’, *Medical Imaging, IEEE Transactions on* **31**(2), 153–163.
- Rohlfing, T., Maurer Jr, C. R., Bluemke, D. A. and Jacobs, M. A. [2003], ‘Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint’, *Medical Imaging, IEEE Transactions on* **22**(6), 730–741.
- Rohr, K. [1997], ‘On 3D differential operators for detecting point landmarks’, *Image and Vision Computing* **15**(3), 219 – 233.
- Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V. and Hammers, A. [2006], Diffeomorphic registration using b-splines, in R. Larsen, M. Nielsen and J. Sporring, eds, ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 2006’, Vol. 4190 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 702–709.
- Rueckert, D., Clarkson, M. J., Hill, D. L. G. and Hawkes, D. J. [2000], Non-rigid registration using higher-order mutual information, in K. M. Hanson, ed., ‘SPIE Medical Imaging’, Vol. 3979, pp. 438–447.
- Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M. and Hawkes, D. [1999], ‘Nonrigid registration using free-form deformations: application to breast MR images’, *Medical Imaging, IEEE Transactions on* **18**(8), 712 –721.
- Rühaak, J., Heldmann, S. and Fischer, B. [2011], Improving lung registration by incorporating anatomical knowledge: A variational approach, in ‘Fourth International Workshop on Pulmonary Image Analysis, MICCAI’.
- Russakoff, D., Tomasi, C., Rohlfing, T. and C.R., M. [2004], Image similarity using mutual information of regions, in ‘European Conference on Computer Vision, ECCV 2004’, pp. 596–607.
- Scharstein, D. and Szeliski, R. [1996], Stereo matching with non-linear diffusion, in ‘Computer Vision and Pattern Recognition, CVPR 1996’, pp. 343 –350.

- Schmidt-Richberg, A. [2012b], ‘Fast explicit diffusion for registration with direction-dependent regularization’, *Biomedical Image Registration* pp. 220–228.
- Ehrhardt, J. and Werner, R. and Handels, H.
- Schmidt-Richberg, A., Werner, R., Handels, H. and J., E. [2012a], ‘Estimation of slipping organ motion by registration with direction-dependent regularization’, *Medical Image Analysis* **16**(1), 150–159.
- Schnabel, J., Rueckert, D., Quist, M., Blackall, J., Castellano-Smith, A., Hartkens, T., Penney, G., Hall, W., Liu, H., Truwit, C., Gerritsen, F., Hill, D. and Hawkes, D. [2001], A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations, in W. Niessen and M. Viergever, eds, ‘Medical Image Computing and Computer-Assisted Intervention, MICCAI 2001’, Vol. 2208 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 573–581.
- Shakhnarovich, G., Viola, P. and Darrell, T. [2003], Fast pose estimation with parameter-sensitive hashing, in ‘Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on’, IEEE, pp. 750–757.
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M. and Toga, A. W. [2008], ‘Construction of a 3D probabilistic atlas of human cortical structures’, *Neuroimage* **39**(3), 1064.
- Shechtman, E. and Irani, M. [2007], Matching local self-similarities across images and videos, in ‘Computer Vision and Pattern Recognition, CVPR 2007’, pp. 1 –8.
- Shekhovtsov, A., Kovtun, I. and Hlaváć, V. [2008], ‘Efficient MRF deformation model for non-rigid image matching’, *Computer Vision and Image Understanding* **112**(1), 91–99.
- Shen, D. and Davatzikos, C. [2002], ‘Hammer: hierarchical attribute matching mechanism for elastic registration’, *Medical Imaging, IEEE Transactions on* **21**(11), 1421 –1439.
- Simonyan, K., Vedaldi, A. and Zisserman, A. [2012], Descriptor learning using convex optimisation, in ‘Computer Vision–ECCV 2012’, Springer, pp. 243–256.
- Simpson, I., Woolrich, M. and Schnabel, J. [2011], Probabilistic segmentation propagation from uncertainty in registration, in ‘Proc. Medical Image Analysis and Understanding’, pp. 331–335.

- Smith, N. B. and Webb, A. [2010], Introduction to Medical Imaging: Physics, Engineering and Clinical Applications, Cambridge University Press.
- So, R. W., Tang, T. W. and Chung, A. C. [2011], ‘Non-rigid image registration of brain magnetic resonance images using graph-cuts’, Pattern Recognition **44**(10), 2450 – 2467.
- Sotiras, A., Davatzikos, C. and N., P. [2013], ‘Deformable medical image registration: A survey’, IEEE Transactions on Medical Imaging **32**(7), 1153–1190.
- Sotiras, A. and Paragios, N. [2012], Discrete symmetric image registration, in ‘International Symposium on Biomedical Imaging, ISBI 2012’, pp. 342–345.
- Spall, J. C. [1992], ‘Multivariate stochastic approximation using a simultaneous perturbation gradient approximation’, IEEE Transactions on Automatic Control **37**(3), 332–341.
- Staring, M., Pluim, J., de Hoop, B., S., K., van Ginneken, B., Gietema, H., Nossent, G., Schaefer-Prokop, C., van de Vorst, S. and Prokop, M. [2009b], ‘Image Subtraction Facilitates Assessment of Volume and Density Change in Ground-Glass Opacities in Chest CT’, Investigative Radiology **44**(2), 61 – 66.
- Staring, M., van der Heide, U., Klein, S., Viergever, M. and Pluim, J. [2009a], ‘Registration of cervical MRI using multifeature mutual information’, Medical Imaging, IEEE Transactions on **28**(9), 1412–1421.
- Studholme, C., Drapaca, C., Iordanova, B. and Cardenas, V. [2006], ‘Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change’, Medical Imaging, IEEE Transactions on **25**(5), 626 –639.
- Studholme, C., Hill, D. and Hawkes, D. [1999], ‘An overlap invariant entropy measure of 3D medical image alignment.’, Pattern Recognition **32**(1), 71–86.
- Sun, D., Roth, S. and M.J., B. [2010], Secrets of optical flow estimation and their principles, in ‘Computer Vision and Pattern Recognition’, Proceedings of IEEE Int. Conference.
- Tang, L., Hero, A. and Hamarneh, G. [2012], Locally-adaptive similarity metric for deformable medical image registration, in ‘Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on’, IEEE, pp. 728–731.
- Tapia, E. [2011], ‘A note on the computation of high-dimensional integral images’, Pattern Recognition Letters **32**(2), 197 – 201.

- Thévenaz, P. and Unser, M. [2000], ‘Optimization of mutual information for multiresolution image registration’, *IEEE Transactions on Image Processing* **9**(12), 2083–2099.
- Thirion, J. [1998], ‘Image matching as a diffusion process: An analogy with maxwells demons’, *Medical Image Analysis* **2**(3), 243–260.
- Trzcinski, T., Christoudias, C. M., Fua, P. and Lepetit, V. [2013], Boosting binary keypoint descriptors, in ‘Computer Vision and Pattern Recognition, 2013. CVPR 2013. IEEE Conference on’, pp. 1–8.
- Tsai, D., Y., L. and Matsuyama, E. [2008], ‘Information entropy measure for evaluation of image quality’, *Journal of Digital Imaging* **21**(3), 338–347.
- Tsochantaridis, I., Joachims, T., Hofmann, T. and Altun, Y. [2005], Large margin methods for structured and interdependent output variables, in ‘Journal of Machine Learning Research’, pp. 1453–1484.
- Tu, Z. [2008], Auto-context and its application to high-level vision tasks, in ‘Computer Vision and Pattern Recognition, CVPR 2008’, pp. 1 –8.
- Unser, M., Aldroubi, A. and Gerfen, C. [1993], Multiresolution image registration procedure using spline pyramids, in ‘SPIE’s 1993 International Symposium on Optics, Imaging, and Instrumentation’, International Society for Optics and Photonics, pp. 160–170.
- Van Rikxoort, E. M., de Hoop, B., van de Vorst, S., Prokop, M. and van Ginneken, B. [2009], ‘Automatic segmentation of pulmonary segments from volumetric chest CT scans’, *Medical Imaging, IEEE Transactions on* **28**(4), 621–630.
- Vandemeulebroucke, J., Bernard, O., Rit, S., Kybic, J., Clarysse, P. and Sarrut, D. [2012], ‘Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT’, *Medical Physics* **39**, 1006.
- Vandemeulebroucke, J., Sarrut, D. and Clarysse, P. [2007], The POPI-model, a point validated pixel-based breathing thorax model, in ‘Conference on the Use of Computers in Radiation Therapy’.
- Varma, M. and Zisserman, A. [2003], Texture classification: are filter banks necessary?, in ‘Computer Vision and Pattern Recognition, CVPR 2003’, Vol. 2, pp. 691–698.
- Veksler, O. [2005], Stereo correspondence by dynamic programming on a tree, in ‘Computer Vision and Pattern Recognition, CVPR 2005’, Vol. 2, pp. 384–390.

- Vercauteren, T., Pennec, X., Perchant, A. and Ayache, N. [2008], Symmetric log-domain diffeomorphic registration: A demons-based approach, in ‘Medical Image Computing and Computer Assisted Intervention, MICCAI 2008’, Lecture Notes in Computer Science, pp. 754–761.
- Vercauteren, T., Pennec, X., Perchant, A. and Ayache, N. [2009], ‘Diffeomorphic demons: Efficient non-parametric image registration’, *NeuroImage* **45**, 561–572.
- Viola, P. and Wells III, W. [1997], ‘Alignment by maximization of mutual information’, *Int. J. Comput. Vision* **24**(2), 137–154.
- Wachinger, C. and Navab, N. [2009], Similarity metrics and efficient optimization for simultaneous registration, in ‘Computer Vision and Pattern Recognition, CVPR 2009’, pp. 779–786.
- Wachinger, C. and Navab, N. [2012], ‘Entropy and laplacian images: Structural representations for multi-modal registration’, *Medical Image Analysis* **16**(1), 1 – 17.
- Weinberger, K. Q., Blitzer, J. and Saul, L. K. [2005], Distance metric learning for large margin nearest neighbor classification, in ‘Advances in neural information processing systems’, pp. 1473–1480.
- Weinberger, K. Q. and Saul, L. K. [2008], Fast solvers and efficient implementations for distance metric learning, in ‘Proceedings of the 25th international conference on Machine learning’, ACM, pp. 1160–1167.
- Weiss, E., Wijesooriya, K., Dill, S. V. and Keall, P. J. [2007], ‘Tumor and normal tissue motion in the thorax during respiration: Analysis of volumetric and positional variations using 4D CT’, *International Journal of Radiation Oncology Biology Physics* **67**(1), 296 – 307.
- Werlberger, M., T., P. and Bischof, H. [2010], Motion estimation with non-local total variation regularization, in ‘Computer Vision and Pattern Recognition, CVPR 2010’, pp. 1–8.
- West, J., Fitzpatrick, J., Wang, M., Dawant, B., Maurer Jr, C., Kessler, R., Maciunas, R., Barillot, C., Lemoine, D., Collignon, A. et al. [1997], ‘Comparison and evaluation of retrospective intermodality brain image registration techniques’, *Journal of Computer Assisted Tomography* **21**(4), 554–568.
- Wu, Z., Rietzel, E., Boldea, V., Sarrut, D. and Sharp, G. C. [2008], ‘Evaluation of deformable registration of patient lung 4dct with subanatomical region segmentations’, *Medical physics* **35**, 775.

- Xing, E. P., Jordan, M. I., Russell, S. and Ng, A. [2002], Distance metric learning with application to clustering with side-information, in ‘Advances in neural information processing systems’, pp. 505–512.
- Yamamoto, T., Kabus, S., Klinder, T., Lorenz, C., von Berg, J., Blaffert, T., Jr, B. W. L. and Keall, P. J. [2011b], ‘Investigation of four-dimensional computed tomography-based pulmonary ventilation imaging in patients with emphysematous lung regions’, *Physics in Medicine and Biology* **56**(7), 2279.
- Yamamoto, T., Kabus, S., Klinder, T., von Berg, J., Lorenz, C., Loo, B. and Keall, P. [2011a], ‘Four-dimensional computed tomography pulmonary ventilation images vary with deformable image registration algorithms and metrics.’, *Medical Physics* **38**, 1348–58.
- Yang, Q., Wang, L. and Ahuja, N. [2010], A constant-space belief propagation algorithm for stereo matching, in ‘Computer Vision and Pattern Recognition, CVPR 2010’, IEEE, pp. 1458–1465.
- Yeo, B., Sabuncu, M., Desikan, R., Fischl, B. and Golland, P. [2008], ‘Effects of registration regularization and atlas sharpness on segmentation accuracy’, *Medical image analysis* **12**(5), 603.
- Yi, Z. and Soatto, S. [2011], Multimodal registration via spatial-context mutual information, in ‘Information Processing in Medical Imaging, IPMI 2011’, Vol. 6801, Springer Berlin / Heidelberg, pp. 424–435.
- Yianilos, P. N. [1993], Data structures and algorithms for nearest neighbor search in general metric spaces, in ‘Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms’, Society for Industrial and Applied Mathematics, pp. 311–321.
- Yin, Y., Hoffman, E. and Lin, C. [2009], ‘Mass preserving nonrigid registration of CT lung images using cubic B-spline.’, *Medical Physics* **26**, 4213–22.
- Young, D. [1954], ‘Iterative methods for solving partial difference equations of elliptic type’, *Trans. Amer. Math. Soc.* **76**, 92–111.
- Zabih, R. and Woodfill, J. [1994], Non-parametric local transforms for computing visual correspondence, in J.-O. Eklundh, ed., ‘European Conference on Computer Vision, ECCV 1994’, Vol. 801 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 151–158.
- Zhang, L., Hoffman, E. and Reinhardt, J. [2006], ‘Atlas-driven lung lobe segmentation in volumetric X-ray CT images’, *Medical Imaging, IEEE Transactions on* **25**(1), 1 –16.

Zikic, D. [2011], ‘Contributions to medical image registration’. Dissertation, Technische Universität München.

Zikic, D., Baust, M., Kamen, A. and Navab, N. [2010b], Generalization of deformable registration in Riemannian Sobolev spaces, in T. Jiang, N. Navab, J. Pluim and M. Viergever, eds, ‘Medical Image Computing and Computer-Assisted Intervention MICCAI 2010’, Vol. 6362 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 586–593.

Zikic, D., Kamen, A. and Navab, N. [2010a], Revisiting horn and schunck: Interpretation as gauss-newton optimisation, in ‘British Machine Vision Conference, BMVC 2010’, pp. 1–12.

Zimmer, H., Bruhn, A. and Weickert, J. [2011], ‘Optical flow in harmony’, Intl J. Computer Vision **93**(3), 368–388.

Zitnick, C. L. and Kang, S. B. [2007], ‘Stereo for image-based rendering using image over-segmentation’, International Journal of Computer Vision **75**(1), 49–65.