

Dissertation

Contributions to Medical Image Registration

Darko Zikic

TECHNISCHE UNIVERSITÄT MÜNCHEN

Computer Aided Medical Procedures & Augmented Reality / I16

TECHNISCHE UNIVERSITÄT MÜNCHEN
Computer Aided Medical Procedures & Augmented Reality / I16

Contributions to Medical Image Registration

Darko Zikic

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Cremers
Prüfer der Dissertation: 1. Univ.-Prof. Dr. N. Navab
2. Prof. D. Rueckert, Ph.D.
Imperial College London, UK

Die Dissertation wurde am 14.04.2011 bei der Technischen Universität München
eingereicht und durch die Fakultät für Informatik am 02.12.2011 angenommen.

Abstract

Image registration, also known as image alignment, spatial normalization, or motion estimation, is the process of computing the spatial transformation between corresponding structures of objects depicted in two or more images. In medical scenarios, the spatial transformation between corresponding points can have various causes, such as natural patient movement, different patient positioning, or anatomical changes. Sometimes, the apparent misalignment is caused by the task of comparing images of different subjects. Thus, image registration is often the foundation for further analysis and applications, such as fusion of different modalities for diagnosis and interventions.

Image registration presents an active field of research, among other reasons also due to the complex nature of the associated optimization problem. Among the registration methods, the so called deformable registration problem, which treats high-dimensional transformation models, is particularly challenging. In the first part of this work, we present an overview of existing intensity-based registration methods, with focus on deformable approaches. Our aim is to point out the relations between the different approaches, and to highlight the common elements. In the second part of this thesis, we present several of our contributions to the field of registration of medical images, most of them affecting deformable registration methods:

1) We propose a simple and efficient preconditioning scheme for improvement of the convergence speed of gradient-based methods for arbitrary image-based difference measures in deformable registration. The proposed scheme is especially useful for deformable multi-modal registration employing statistical difference measures, since in these cases, the range of applicable efficient optimization schemes is strongly limited, due to the structure and size of the problem.

2) Alignment of angiographic 3D scans to 2D projections is an important issue for navigation during interventions. For the common single-view setting, in which only one 2D projection is available, we introduce a method for deformable 2D-3D registration of vascular structures. Prior to our work, methods employing a rigid transformation model presented the state of the art for this problem.

3) For linear registration, which is often performed prior to deformable registration, we discuss the use of Markov Random Field (MRF) modeling and discrete optimization. The approach is based on the approximation of the original energy, such that efficient discrete optimization becomes applicable. The resulting framework allows to transfer the advances from the active field of discrete optimization to linear registration.

4) For generation of statistical deformation models, we analyze the deformations resulting from deformable registration and find that they contain significant linear components. Based on this observation, we propose the use of minimal deformations, from which the linear transformation components are removed. We demonstrate that this step is non-optional for creation of accurate models representing the major deformations within a population.

Zusammenfassung

Bildregistrierung bezeichnet die Bestimmung der räumlichen Transformation zwischen korrespondierenden Strukturen eines Objektes, welches in zwei oder mehreren Bildern dargestellt wird. Die örtliche Verschiebung zwischen korrespondierenden Punkten in unterschiedlichen Aufnahmen kann in medizinischen Anwendungsszenarien vielfältige Ursachen haben, wie zum Beispiel die Bewegung, unterschiedliche Positionierung, oder anatomische Veränderungen des Patienten. In manchen Fällen kann die scheinbare Verschiebung daraus resultieren, dass ein Vergleich von korrespondierenden Strukturen in verschiedenen Patienten durchgeführt werden soll. Aufgrund der häufigen Problemstellung bildet die Bildregistrierung den Grundstein für die weitere Verarbeitung von medizinischen Bildern in zahlreichen Anwendungen. Eine Beispielanwendung ist die Überlagerung von Bildern verschiedener Modalitäten, welche für diagnostische oder interventionelle Zwecke vorteilhaft eingesetzt werden kann.

Die Bildregistrierung bleibt ein aktives Forschungsfeld, unter anderem durch die anspruchsvolle Natur der damit verbundenen Optimierungsprobleme. Dies ist insbesondere der Fall für die sogenannte deformierbare Registrierung, welche hochdimensionale Transformationsmodelle behandelt. Der erste Teil dieser Arbeit soll eine grobe Übersicht über die existierenden Methoden für deformierbare Registrierung geben. Hierbei liegt der Schwerpunkt auf den Gemeinsamkeiten und Parallelen zwischen den verschiedenen Ansätzen. Der zweite Teil der Arbeit fasst unsere methodologischen Beiträge zum Bereich der Registrierung medizinischer Daten zusammen, welche im Rahmen dieser Arbeit entstanden sind:

1) Wir stellen ein Vorkonditionierungsschema für die Optimierung beliebiger Bild-basierter Ähnlichkeitsmaße in Registrierungsproblemen vor. Basierend auf diesem Schema läßt sich eine einfache und effiziente Optimierung durchführen, welche für alle weit verbreiteten Registrierungsmethoden eingesetzt werden kann. Die vorgestellte Methode ist von besonderem Interesse für die deformierbare multi-modale Registrierung mit statistischen Ähnlichkeitsmaßen. Für diese Klasse von Registrierungsproblemen können die üblichen effizienten Optimierungsmethoden oftmals nicht eingesetzt werden, aufgrund der Größe und Struktur der entsprechenden Optimierungsprobleme.

2) Die Registrierung von angiographischen 3D Aufnahmen zu 2D Projektionsaufnahmen ist notwendig für die Navigation bei intraoperativen Anwendungen. Für das weit verbreitete "single-view" Szenario, in dem nur eine einzelne 2D Projektion vorliegt, stellen wir die erste Methode zur deformierbaren 2D-3D Registrierung von vaskulären Strukturen vor.

3) Wir stellen eine Methode zur linearen (nicht-deformierbaren) Registrierung vor, welche auf der Markov Random Field Modellierung beruht und diskrete Optimierungsmethoden verwendet. Die Methode basiert auf der Approximation der ursprünglichen Energie, die es erlaubt, dass effiziente diskrete Optimierungsmethoden eingesetzt werden können.

4) Im Kontext der Erstellung von statistischen Deformationsmodellen analysieren wir Deformationsfelder welche mittels deformierbarer Registrierung berechnet werden, und stellen fest, dass diese signifikante lineare Anteile enthalten. Daher schlagen wir den Einsatz von minimalen Deformationen vor, aus welchen die linearen Anteile entfernt wurden. Wir zeigen, dass dieser Schritt für die Erstellung von genauen Deformationsmodellen notwendig ist.

Acknowledgments

I would like to use this opportunity to express my gratitude to a number of people who helped me during my thesis.

First, I would like to thank Nassir Navab for making this work possible and for his supervision and support during this time. Also, I am very grateful to Frank Sauer and Ali Kamen at Siemens Corporate Research (SCR), for their continuous support and a great collaboration.

I would like to thank everyone at CAMP who helped in making this such a great experience. I deem myself extraordinarily lucky to have been surrounded by so many curious, enthusiastic, smart, fun and kind people. Guys, you really are too many to name - you know who you are. Martin, Ben, and Max, as my office mates and closest collaborators deserve a special mention for their stoic endurance of my whimsies. Very special thanks are due to Martina for somehow making things always work out.

Also thanks to all the folks at SCR for the productive collaboration, and the great time during my visits.

Finally, endless thanks to my parents, Miro and my very own Alma Mater, for their limitless support. And of course to Pia - Du weißt schon...

CONTENTS

Abstract	iii
Zusammenfassung	v
Table of Contents	ix
Introduction	1
I Review of Deformable Intensity-based Registration	7
1 Introduction and Overview	9
1.1 Review of Deformable Registration Methods	17
2 Elastic Variational Registration	19
2.1 Optimality Condition: Euler-Lagrange PDE	21
2.2 Estimating the Solution of the Euler-Lagrange PDE	22
2.3 Time Discretization	22
2.3.1 Explicit Time Discretization	23
2.3.2 Semi-implicit Time Discretization	24
2.3.3 Demons Time Discretization	25
2.3.4 Time Discretization by Quadratic Energy Approximation	27
2.4 Spatial Discretization	28
2.4.1 Finite Difference Method	28
2.4.2 Finite Element Method	28
2.5 Some Examples of Elastic Approaches	30
2.6 Discussion	31
3 Large Deformation Registration	33
3.1 Fluid Registration	34
3.2 Parametrization of Deformations as Flows	35
3.3 Regularization by Minimization of Geodesic Length	37
3.4 Time-Varying and Stationary Velocity Fields	37

3.4.1	Non-stationary Velocity Fields	38
3.4.2	Stationary Velocity Fields	38
3.5	Discussion	39
4	Demons Registration Approaches	41
4.1	Interpretation as Alternating Optimization	42
4.2	Interpretation as Time Discretization Scheme	42
4.3	Generalization of Difference Measures	43
4.4	Fluid Demons	43
4.5	Generalization of Regularization	44
4.6	Update Schemes and Diffeomorphic Demons	45
4.7	General Demons Framework	46
4.8	Further Extensions	47
4.9	Discussion	47
5	Parametric Registration Methods	49
5.1	Characterization of Linear Transformation Models	51
5.2	Examples of Linear Transformation Models	53
5.2.1	B-Spline Free-form Deformations (FFD)	53
5.2.2	Trigonometric Functions	54
5.2.3	Radial Basis Functions (RBF)	55
5.3	Relation to Other Approaches	56
5.4	Discussion	57
II	Methodological Contributions	59
6	Egalitarian Preconditioning: A Preconditioning Scheme for Difference Measures in Deformable Registration	61
6.1	Introduction	62
6.1.1	Optimization of Statistical Difference Measures for High- dimensional Transformation Models	62
6.1.2	Local Gradient Bias	64
6.1.3	Related Work	67
6.1.4	Outline	68
6.2	The Egalitarian Preconditioning Scheme	69
6.2.1	Application to the Demons Framework	70
6.2.2	Application to Variational and Parametric Approaches	71
6.2.3	Application to Registration in Groups of Diffeomorphisms	72
6.3	Analysis of the Preconditioning Scheme	73
6.3.1	Model for Condition Analysis	73
6.3.2	Restriction of Analysis to Permissible Directions	74
6.3.3	Condition Analysis of Difference Measures	76
6.4	Evaluation	78
6.4.1	Evaluation for the Demons Method	79
6.4.2	Evaluation for Registration in Groups of Diffeomorphisms	83
6.5	Summary and Discussion	85
6.A	Appendix: Derivation of Optimal Model Energy with Lagrange Multipliers	87

7	Deformable 2D-3D Registration	89
7.1	Introduction	89
7.1.1	Relation to Prior Work	92
7.2	Method	93
7.2.1	Setting and Preprocessing	94
7.2.2	Preliminaries and Notation	94
7.2.3	The Model	96
7.2.4	Difference Measure	96
7.2.5	Length Preservation Constraint	97
7.2.6	Diffusion Regularization	97
7.2.7	Optimization Scheme	98
7.2.8	Solving for 2D-3D Correspondences	99
7.2.9	Overall Algorithm	101
7.3	Results and Evaluation	101
7.3.1	Parameter Values	103
7.3.2	Tests on Synthetic Data	104
7.3.3	Real Data with Artificial Deformation	104
7.3.4	Real Data with Natural Deformation	104
7.3.5	Test in a Clinical Setup	105
7.4	Discussion	109
7.4.1	Vessel Deformation in Clinical Routine	110
7.4.2	Segmentation	111
7.4.3	Rigid 2D-3D Registration	111
7.4.4	Runtime	112
7.4.5	Ambiguities	112
8	Linear Registration with MRFs and Discrete Optimization	115
8.1	Introduction	115
8.2	Intensity-based Estimation of Linear Transformations by Markov Random Fields	117
8.2.1	Discrete Markov Random Fields	118
8.2.2	Energy Approximation for Estimation of Linear Transfor- mations by Second-Order MRFs	120
8.2.3	Parameterization	122
8.2.4	Discretization of the Parameter Space	123
8.2.5	Optimization	125
8.2.6	Implementation Details	125
8.2.7	Method Discussion	125
8.3	Method Evaluation	126
8.3.1	Quality of Approximation	126
8.3.2	Dependency on Parameters	128
8.4	Applications	129
8.4.1	Linear Intensity-based Image Registration	129
8.4.2	Intensity-based 2D-3D Registration	134
8.5	Conclusion	137

9 Minimal Deformations	141
9.1 Introduction	141
9.1.1 Nonlinear Registration and Resulting Deformations	142
9.1.2 Statistical Shape Models	143
9.2 Methods	144
9.2.1 Notation and Basic Definitions	144
9.2.2 Computation of Minimal Deformations	144
9.2.3 Statistical Shape Models	146
9.3 Results and Evaluation	148
9.3.1 Test Settings	148
9.3.2 Quantification of Extracted Similarity Transformation Components	150
9.3.3 Effects of Proposed Method on SSMs	151
9.4 Discussion and Conclusion	153
List of Author's Publications	155
References	159

'What luck! Here's a deep, damp ditch on the other side,
which I shall now proceed to fall into.'

LORD PETER WIMSEY

INTRODUCTION

Image registration describes the task of computing the spatial transformation (i.e. motion) between corresponding structures of a certain object, which are depicted in a given set of input images. Here, we use the term *image* in a general sense, and it can represent any spatial measurement of a certain signal or quantity. The measured quantity can be for example the brightness intensity in gray-valued images, but also higher-dimensional, vector-valued measurements are possible, for example the different color channels in photographs. Examples of images are photographs, depth images, or medical images from any arbitrary modality, such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), or ultrasound (US).

In medical scenarios, there are numerous reasons which cause the misalignment of corresponding structures. These can include natural patient movement, different patient positioning, or anatomical changes in the patient over time. In certain applications, we are interested in comparing images of different subjects, and here, the apparent misalignment of corresponding structures is due to the actual differences in the subjects.

There are numerous synonyms for image registration, which highlight different aspects of the problem, such as image-based motion estimation, image alignment, or spatial normalization of images.

Image registration has been a very active research area in the last four decades. Its importance stems from the fact that the knowledge about the motion or correspondence of structures is an important building block and prerequisite for many applications. Furthermore, images present a very attractive, since non-invasive and non-obtrusive, type of measurements, from which the information about the motion and the correspondence can be extracted.

The medical domain offers many examples for applications of image registration. One wide-spread application is the fusion of complementary information from different modalities - for example the combination of anatomical information (e.g. from CT) and functional information (e.g. from PET). For the correct fusion of the information, it is crucial that the single images are aligned. To enable a meaningful fusion, the motion between the two scans has to be undone, and for that, it first has to be estimated - enter registration.

A further example are so-called longitudinal studies, in which the patient is

scanned at different points in time, in order to monitor the progress of the illness or the treatment. Again, in order to facilitate the assessment of the treatment and allow its quantification, the motion between the single scans has to be estimated. For some applications, the registration of images of different patients is required. An example is the field of *computational anatomy* which treats the creation of so called *atlases* of single organs, which describe the average anatomy and the variation of the given population (thus allowing potential early detection of diseases). To this end, large sets of images from different patients have to be registered. For some organs with high variability between individuals such as the brain, this is an especially challenging task, since the assumption of corresponding structures in the input images is no longer clearly given. In this case with no corresponding structures, it is also not possible to assume physical properties to model plausible deformations.

There are also numerous applications in non-medical domains, for which image registration is used, such as robotics, surveillance applications, reconstruction of 3D scenes, video compression, or creation of large panoramic scenes by image stitching [Szeliski, 2010]. These applications often employ photographs or single video frames as input images, and the corresponding deformable image registration problem is often referred to as the *optical flow problem*.

The registration problem can be formalized as an optimization of a certain energy model, where the energy is modeled such that it is possibly minimized by the true transformation. In this case, the registration task boils down to the definition of energy and performing an optimization. These steps are however not trivial for image registration, due to the inherent properties of the problem. Based on the setting of the registration problem, and the employed model, as well as the used optimization method, there are some widely accepted classifications, which we briefly discuss next.

The available input images and the application requirements determine the properties of the registration method to be used. One important classification of the registration problem is the one into *mono-modal* and *multi-modal* registration scenarios. Mono-modal registration describes a setting in which both input images are acquired by the same type of device. On the other hand, multi-modal registration refers to the problem of computing the transformation between images of different modalities, for example between a CT and a PET image. Multi-modal registration is of special importance in the medical domain, as it enables the fusion of the information from different modalities. Therefore, the applicability in multi-modal scenarios is an important requirement for general purpose registration methods in medical scenarios.

A further common distinction of registration methods is based on the type of the transformation to be estimated. Two broad classes are usually identified: the so called *rigid* and *deformable* registration problems. The synonyms for rigid transformations are global transformation and linear transformations, and they highlight the different aspects of this class, which describes a comparably simple motion, which is globally valid for all structures in the image domain. On the other hand, the class of deformable (or non-rigid, non-linear, or local) transformations are more complex than rigid transformations and can encode different independent motions of individual structures in the image domain. Problems arising from deformable registration result in more complex and irregular energy functions, and thus pose more and harder challenges than rigid registration problems, and are thus currently a more active field of research.

For this reason, most of the work in this thesis is performed in the context of deformable registration methods.

A further point of discrimination between the methods is the dimensionality of the input images. While many applications provide input images of same dimensionality, for example two 2D photographs, or two 3D CT scans, in some cases, the input images have different dimensions, which can lead to estimation of the transformation from partially lower-dimensional input. One such example is the case of 2D - 3D registration in medical settings, where for example the motion in a 3D volume has to be estimated from an initial 3D CT scan and one or several corresponding 2D X-ray images.

When it comes to different registration approaches, a commonly accepted high-level classification of registration methods distinguished between *feature-based* and *intensity-based* algorithms. While approaches from both groups can be formulated as optimization problems, the modeling of the energy, and the resulting optimization schemes differ considerably.

The defining characteristic of feature-based registration algorithms is that they operate based on specific, sparse and spatially localized features. The three major steps of feature-based approaches are 1) the extraction of features, 2) the subsequent matching of corresponding features, 3) computation of the transformation based on estimated correspondences. The selection of sufficiently many precisely localized features, and their correct matching are challenging tasks. While a number of automatic methods has been proposed for mono-modal scenarios, the extension to multi-modal scenarios is not straightforward.

In intensity-based registration, on the other hand, there is no explicit extraction of features. The energy model is based on a certain difference measure between the two input images, which takes into account the complete information (e.g. the intensity values) from the images, possibly after certain pre-processing steps. In contrast to feature-based approaches, intensity-based approaches are commonly applied in multi-modal scenarios. This is possible due to a number of difference measures, which are able to assess the similarity of images acquired by different devices.

For the above reasons, in this work we focus on intensity-based methods.

Outline and Overview of the Thesis

In the first part of this work, we provide a brief review of existing deformable intensity-based registration methods. We base the presentation on the initial motivation for the single approaches, and discuss the following groups of methods

- Elastic Variational Registration (Chapter 2)
- Large Deformation Registration (Chapter 3)
- Demons Registration Approaches (Chapter 4)
- Parametric Registration (Chapter 5)

We try to point out the basic principles and the main ideas behind the single lines of development. Also, since the resulting methods share numerous methodological aspects, we attempt to highlight the unifying links between the different approaches when possible.

In the second part, we discuss a number of contributions to the field of image registration, which we proposed during this work. While the methodological

contributions are generally applicable, their development was mostly motivated by problems arising from medical settings. In the following, we give a short overview of the main contributions.

Egalitarian Preconditioning: A Preconditioning Scheme for Difference Measures in Deformable Registration (Chapter 6)

We propose a preconditioning scheme for efficient optimization of arbitrary image-based difference measures in the context of deformable registration.

Our scheme is based on the observation and analysis of a fundamental negative property of steepest gradient descent (SGD) for registration problems: For image-based difference measures, SGD exhibits slower convergence speed in areas corresponding to low-gradient regions of the source image. We refer to this property as *Local Gradient Bias*, and our approach aims at avoiding the above effect in a principled way, for all difference measures. Since the key idea behind our approach is to remove the inequalities of the updates in the different image regions, we call the resulting approach *Egalitarian Preconditioning*.

We perform a theoretical analysis of the condition of difference measures in registration problems, and demonstrate that the proposed scheme improves the condition of the original problem. Because of the simplicity of the proposed scheme, its application improves the convergence speed while adding only negligible computational cost, thus resulting in shorter effective runtimes.

The proposed preconditioning is of particular interest for high-dimensional deformable registration with statistical difference measures such as mutual information (MI). In these settings, the range of applicable standard methods for efficient optimization is strongly limited, due to the structure and the size of the resulting optimization problem. Here, our approach presents a conceptually simple, yet a theoretically justified alternative, which can easily be integrated into any gradient-based registration scheme. We demonstrate the application of the proposed preconditioning for registration in the group of diffeomorphisms, and for the demons method. For the demons scheme in particular, our approach can be seen as an *efficient* generalization to arbitrary difference measures.

Single-View 2D-3D Deformable Registration of Vascular Structures (Chapter 7)

Alignment of angiographic 3D scans to 2D projections is an important issue for 3D depth perception and navigation during interventions. The common single-view setting, in which only one 2D projection is available, is particularly challenging due to the inherent ill-posedness of the problem. Prior to our work, methods employing a rigid transformation model have presented the state of the art for the single-view problem. We introduce a method for deformable registration of 3D vessel structures to a respective single projection of the scene. Our approach addresses the inherent ill-posedness of the problem by incorporating a priori knowledge about the vessel structures into the formulation in the form of regularization terms.

Linear Registration with MRFs and Discrete Optimization (Chapter 8)

We propose a framework for linear intensity-based registration, based on a discrete Markov Random Field (MRF) formulation. Here, the challenge arises from the fact that optimizing the associated energy requires a high-order MRF model. Currently, methods for optimizing such high-order models are less general, easy to use, and efficient, than methods for the popular second-order models. The main idea in this work is to perform an approximation to the original high-order energy by an MRF with tractable second-order terms.

The resulting framework allows to transfer advances from the currently very active research field of discrete optimization to linear registration problems. We demonstrate the applicability of the framework by intensity-based registration, and 2D-3D registration of medical images.

Minimal Deformations (Chapter 9)

Nonlinear registration is mostly performed after initialization by a global, linear transformation, computed by a linear registration method. For the further processing of the results, it is mostly assumed that this preregistration step completely removes the respective linear transformation. We perform an analysis which demonstrates that in deformable settings, this is not the case. As a consequence, a significant linear component is still existent in the deformation computed by the nonlinear registration algorithm. We propose a method which performs an a posteriori extraction of a similarity transformation from a given nonlinear deformation field, resulting in what we refer to as *minimal deformations*.

For certain applications, such as construction of Statistical Shape Models (SSM) for analysis of shape variability, it is an unwanted property that deformations contain linear transformation components: SSMs should be invariant to similarity transformations, since these do not capture information about shape. Actually, we demonstrate that without the proposed pre-processing, the major modes of the SSM are corrupted and mostly contain information about the linear transformation, and not about the actual deformations. This effect is significantly reduced by the application of the proposed pre-processing step.

Part I

Review of Deformable Intensity-based Registration Methods

INTRODUCTION AND OVERVIEW

As defined previously, registration is the process of estimating the spatial transformation between corresponding structures depicted in the input images. More specifically, in contrast to linear registration, *deformable registration* is the estimation of more complex, non-linear transformations. Over the last three decades, many different approaches to perform this task have been proposed, employing different intuitions, models, and techniques. Thus, the field of deformable registration is very heterogeneous. At times, this makes it hard to compare the different methods directly, and it is possible to gain the impression that the single methods have rather little in common. However this is not the case. Actually, almost all standard methods for intensity-based deformable registration share a small set of the same underlying principles and building blocks. First, these methods all can be formulated as an energy minimization problem. Second, all of these methods can be described in terms of the major building blocks, namely the *energy model*, the *transformation model*, and the *optimization method*. These blocks are not independent. The choice of the energy, and the transformation models strongly influences the range of applicable optimization schemes.

In this chapter we present a high-level and general introduction to the intensity-based registration problem, and the mentioned building blocks. In the following chapters, we then present a brief review of some of the most common approaches for deformable registration, which have been proposed over the last decades. We focus on methods developed in the context of medical image analysis, and present the methods in the following groups: *variational elastic registration*, *large deformation registration*, *demons registration methods*, and *parametric registration*. The boundaries between the single groups of methods are rather loose, and many approaches can be seen as belonging to several of these groups. Furthermore, in most cases, specific methodological developments can be readily transferred between these groups, meaning that modifications of one particular method can often be applied to other approaches as well.

The following exposition aims at generality and thus we do not distinctly distinguish between rigid and deformable methods. The presentation mostly treats the methods in a general way, such that any transformation model can be employed, such that rigid registration methods result from employing a specific low-dimensional model.

In order to keep the following exposition simple, we constrain the treatment

to two input images, namely the *source image* I_S and the *target image* I_T . Here, the source image denotes the image to be warped, such that it matches the static target image. There are numerous synonyms for the input images, for example, the source image is commonly also denoted by *moving image* or *template image*, and the target image by *fixed image*, the *study*, or *reference image*. Furthermore, we assume that the input images are of the same dimensionality d , and that they map to one-dimensional intensity values, i.e. we have $I_S, I_T : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$. Here, Ω denotes a spatial domain, on which the images are defined. The goal of the registration is now to estimate the transformation $\phi : \Omega \rightarrow \Omega$, which warps the source image to $I_S \circ \phi$, such that it corresponds to the target image I_T .

We assume that the transformation is defined with respect to the reference frame of the warped source image, i.e., that the transformation maps a point x from the warped source image, to the corresponding point $\phi(x)$ in the original source image. The above assumption in particular implies two points. The first one is that the reference frame, in which the transformation is defined is changing with the estimate of the transformation, in the course of the registration process. The second point is that for the final result, in which the warped source is supposed to correspond to the target image, the final reference frame corresponds to the reference frame of the target image. Please note that in this setting, the actual motion of particles in the source image, with respect to its original location is described by the inverse transformation ϕ^{-1} . This approach for image warping is also known as *backward warping*, and has the advantage of computational efficiency. The alternative, which is far less frequently employed, is the so called *forward warping* can be modeled in our setting by $I \circ \phi^{-1}$.

Representing Deformations There are two major takes on modeling the deformation function on the infinite-dimensional level: by employing either a linear space, or by assuming a group structure for deformations. For the actual implementation, the infinite-dimensional formulation has to be approximated by a finite-dimensional transformation model. For the choice of the finite-dimensional model, issues such as the expressiveness of the model, efficiency, and the ease of implementation play an important role.

For the infinite-dimensional setting, the first approach is to model the deformation in terms of the corresponding *displacement* field u . In this context, the transformation is often expressed as $\phi = \text{Id} + u$, in terms of the identity transformation Id , with $\text{Id} : x \mapsto x$, and the displacement $u \in \mathbb{H}$. Here, we assume the displacement $u \in \mathbb{H}$ to be an element of a certain Hilbert space \mathbb{H} . A common assumption is that \mathbb{H} is the space of square integrable functions L^2 . Another possibility is to chose \mathbb{H} as a Sobolev space with certain smoothness properties. In this case, the combination of two transformations $\phi_u = \text{Id} + u$ and $\phi_v = \text{Id} + v$ is modeled by the addition of the corresponding displacements, i.e. by ϕ_u “+” $\phi_v = \text{Id} + (u + v)$.

An alternative approach is to assume that deformations are elements of a certain group \mathcal{G} with composition as the group operation. A common choice for \mathcal{G} is to assume that it is the group of diffeomorphic transformations¹. One advantage of this model is that the composition seems to be a more appropriate model for a combination of transformations than the addition, since it describes

¹A transformation is diffeomorphic iff it is invertible and both the original transformation and the inverse are continuous.

exactly the application of one transformation followed by the other. The additive combination can be seen as an approximation which is only valid in the case of small displacements.

Actually, diffeomorphisms form a Lie group, that is, a group which additionally has a structure of a differentiable manifold, with the associated Lie algebra corresponding to the tangential spaces of the manifold. The tangents spaces host the derivatives of elements from \mathcal{G} , and thus the update vector fields v . The tangent spaces are in general modeled by Sobolev spaces, as this ensures a certain smoothness of v , and in consequence results in $\phi \in \mathcal{G}$ being diffeomorphic [Trouvé, 1995, Dupuis et al., 1998].

In this context, the representation of deformations as a series of composed updates is commonly employed. It was popularized and heavily used in the context of methods for large deformation registration. A deformation ϕ can be expressed as a concatenation of small deformations $\text{Id} + v_k$, which are represented by the corresponding update displacement fields or velocities v_k by

$$\phi = (\text{Id} + v_{t_0}) \circ (\text{Id} + v_{t_1}) \circ \dots \circ (\text{Id} + v_{t_n}) . \quad (1.1)$$

Following this intuition, the transformation can be extended by a virtual temporal dimension and formalized by $\Phi : \Omega \times [0, 1] \rightarrow \Omega$, with the final transformation given for $t = 1$ by $\phi = \Phi(1)$. The structure of the temporal transformation, corresponding to (1.1) for a continuous time variable is then described by

$$\frac{d\Phi(t)}{dt} = v(t) \circ \Phi(t) . \quad (1.2)$$

The single velocities are assumed to be elements of a certain Sobolev space, i.e. $v \in \mathbb{H}$, which is the tangent space of the deformation manifold.

It is interesting to note that this representation naturally resembles the iterative way in which the transformations are estimated. This point of view emphasizes the fact that the deformation estimate evolves during the registration process². A corresponding time-dependent formulation of the transformation for linear spaces would be $\phi_t = \text{Id} + u_t$. Please note that the formulation of the transformation as $\phi = \text{Id} + u$ does not contradict the alternative formulation of the deformation. It can be used to represent any deformation, so it can also be used to encode any deformation computed by the group-based representation. However, without further modifications, the assumption of a linear space for u , for which the representation $\text{Id} + u$ was originally used, lacks the theoretical properties which are required for the derivation of diffeomorphic methods in the context of large deformations.

An example of a modification which can be used with the assumption of a linear space to derive algorithms corresponding to the ones naturally arising from the group assumption, is the treatment of the image domain as a deformable continuum, and the appropriate employment of the reference frames [Christensen, 1994]. Eulerian and Lagrangian frames are two standard reference frames for describing phenomena in deforming materials. The Eulerian frame performs the computation on a fixed spatial grid, which does not change over time and is not influenced by the deformation. For the computation of derivatives in the

²Please note that this evolution of the transformation in general does not describe an actual evaluation of the structures in the image. The “time” is a continuous analogon to the discrete concept of iterations.

Eulerian frame, the so called material derivative has to be taken into account. This has to be considered for registration, since the driving forces are derivatives of the difference measure, and the iterative updates can be interpreted as the discretization of the velocity, i.e., the derivative of the displacement field. In the Lagrangian framework, the bookkeeping is not performed with respect to a fixed reference frame, but with respect to the original position of the material particles. The effect in the context of registration is that for a particle originally located at point $x = \phi(x, t_0)$, the update at time t (that is at the corresponding iteration) is computed at the current position of the particle, that is $\phi(x, t)$. We omit further technical details, and refer to the nice descriptions in the context of registration in [Christensen, 1994, Sec. 9.3] and [Bro-Nielsen, 1996, Sec. 4.3.1]. The major point is that the assumption of the deformable continuum in the modeling changes the computation of the derivatives/updates, and this leads to an approach similar to the one resulting from the group assumption.

As previously mentioned, for the actual implementation, finite-dimensional approximation of the infinite-dimensional formulation is required. This step is performed by employing a finite-dimensional transformation model. Depending on the application, transformations with different degrees of complexity can be required. This requirement is reflected in the choice of the *transformation model* for the problem at hand, thereby constraining the space of admissible transformations, and often reducing the problem complexity by reducing the number of parameters. We review some common different types of transformation models in Section 5.2.

The Energy Model With the above prerequisites, the intensity-based registration is generally formulated as optimization of a certain energy E , that is by

$$\hat{\phi} = \arg \min_{\phi \in \mathbb{H}} E(\phi, I_T, I_S) . \quad (1.3)$$

In the actual implementation, the solution to the registration problem is estimated by the optimization of the above energy with respect to a finite set of parameters, which govern the transformation model.

Based on the application requirements and the available input, different *energy models* have to be employed. In general, the energy is a weighted sum of a difference measure E_D and a regularization term E_R

$$E(\phi) = E_D(I_T, I_S, \phi) + \lambda E_R(\phi) . \quad (1.4)$$

The difference measure encodes the degree of similarity between the target image I_T and the warped source image $I_S \circ \phi$. It is supposed to be minimized, when I_T and $I_S \circ \phi$ are perfectly aligned. Arguably the simplest example for a difference measure is the sum of squared differences (SSD)

$$E_D = \frac{1}{2} \int_{\Omega} (I_T(x) - I_S(\phi(x)))^2 dx , \quad (1.5)$$

which assumes that the correctly warped source image is identical to the target image, i.e. $I_T(x) = I_S(\phi_{\text{opt}}(x))$. For the multi-modal registration in medical settings, a variety of difference measures has been proposed. The two major classes of difference measures are the point-wise difference measures, which

compare certain quantities at each image point, independently of the values in the remaining image, and the statistical difference measures, such as Correlation Coefficient (CC), Correlation Ratio (CR), or Mutual Information (MI), which depend on the statistical distribution of image intensities. The purpose and interpretation of the regularization term E_R are more varied. It can be used to penalize unlikely transformations, model physically the behavior of the imaged structures, or in some methods, it is required to make the optimization problem well posed. In contrast to the difference measure, a regularization term is not a part of the energy model in all methods. Probably the simplest type of regularization is the diffusion regularization, which penalizes the first-order derivatives of the displacement fields, thus imposing smoothness

$$E_R = \frac{1}{2} \int_{\Omega} \sum_{i=1}^d \|\nabla u_d(x)\|^2 dx . \quad (1.6)$$

One restrictive assumption we will make in the exposition in the following chapter is that the regularization term is a quadratic form, i.e. that it can be expressed as an error term $e_R(u) = L_R u$, which is linear in u , and penalized in the squared L^2 norm

$$E_R = \frac{1}{2} \|L_R u\|^2 = \frac{1}{2} \langle L_R u, L_R u \rangle = \frac{1}{2} \langle L_R^* L_R u, u \rangle . \quad (1.7)$$

Here, L_R represents a linear operator, and L_R^* is the corresponding adjunct operator. Most regularization operators correspond to this assumption, such as linear elasticity, curvature, or bending energy regularization terms. For simplicity, we will define

$$A_R = L_R^* L_R . \quad (1.8)$$

As an example, in the case of diffusion regularization, we have $L_R = \nabla$ and $A_R = -\Delta$.

For all intensity-based registration methods, the energy in (1.3) is non-linear with respect to the unknown transformation, since the transformation acts through warping of the source image. Because of this non-linearity, in general the resulting energy is non-convex and we are dealing with numerous local minima. In consequence, the *optimization* of (1.3) is local, i.e. it does not attempt at estimating the global energy minimum, and it depends on the initialization. In order to reduce the effects of unwanted local minima, registration is mostly performed by coarse to fine focusing in a scale space setting. Also, because of the non-linearity of the energy, the optimization is performed by an iterative optimization process. The choice of the optimization process depends on the employed energy model and the transformation model. The choice of optimization method for high-dimensional transformation models is particularly dependent on the used difference measure. In contrast to point-wise measures such as SSD, statistical measures such as MI restrict the practically employable optimization methods due to their global nature.

Derivation of Registration Methods There are two principled ways of deriving registration methods: the so-called *variational* and *parametric* derivation. They are both starting from the same energy model, and are schematically

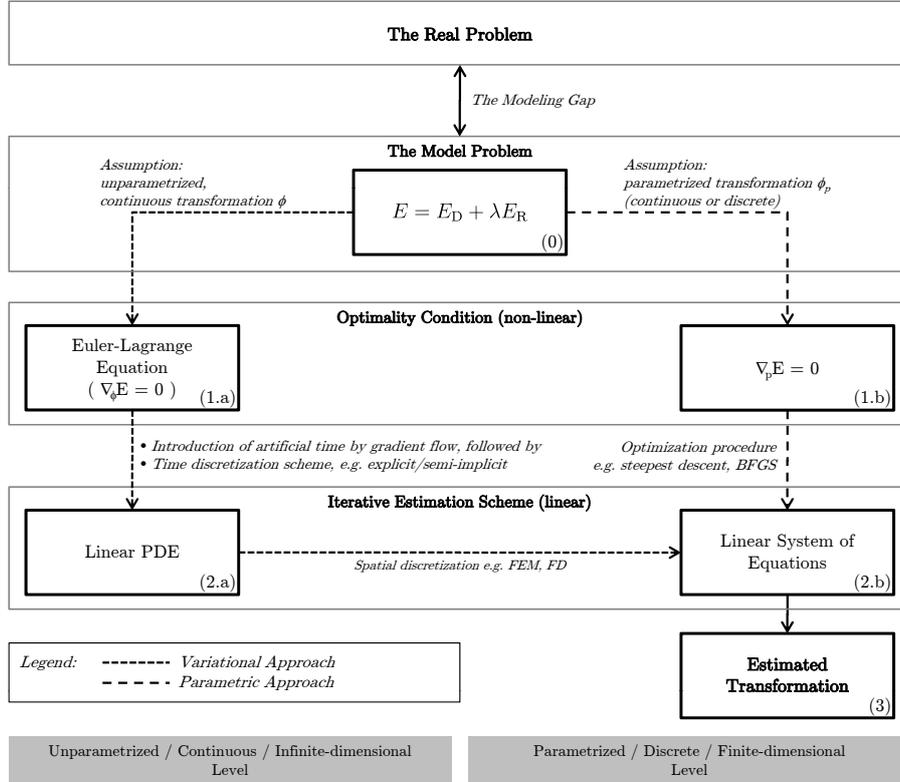


Figure 1.1: Overview of possible approaches for derivation of algorithms for estimating the solution to the deformable registration problem.

illustrated in Figure 1.1. While the variational derivation is performed in the infinite-dimensional setting, with subsequent finite-dimensional parametrization, the parametric derivation directly assumes a finite-dimensional settings by employing the parametrization of the transformation, and derives the registration method directly with respect to the finite-dimensional model.³

It is to be noted that ultimately, and independent of the derivation, for the actual implementation of the derived methods, a finite-dimensional formulation must be employed. It is interesting to note that for many of the final, finite-dimensional formulations, both variational and parametric derivations can be formulated.

Also, a very important point to keep in mind is the *Modeling Gap*, cf. Figure 1.1. Registration approaches are generally based on the energy model, which is always an approximation of the actual real problem. Quite often, due to the computational and modeling constraints, the energy model is rather simplistic compared to the infinitely complex reality. The modeling gap describes this difference between the actual problem, and the model of the problem, which we employ to estimate a reasonable solution. As an example, consider some

³These different derivations are referred to as optimize-discretize and discretize-optimize in [Modersitzki, 2009].

Algorithm 1.1 Basic Generic Intensity-based Registration

Given:

input images I_T and I_S
energy model E
transformation model \mathcal{M}
initial transformation estimate ϕ_0

Goal:

compute $\hat{\phi} = \arg \min_{\phi \in \mathcal{M}} E(\phi, I_T, I_S)$

- 1: initialize transformation: $\phi \leftarrow \phi_0$
- 2: *//perform optimization*
- 3: **repeat**
- 4: compute transformation update: $v \leftarrow \text{optimizer}(I_T, I_S, E, \phi)$
- 5: update transformation estimate: $\phi \leftarrow \phi \oplus v$
- 6: **until** convergence
- 7: $\hat{\phi} \leftarrow \phi$

Note: The symbol \oplus denotes a generic update scheme.

of the requirements for a perfect registration scheme for estimating human breathing motion. This would involve an infinitely realistic, and patient specific body model, including spatially varying regularization, describing the actual motion of the single organs, as well as their interaction. This again would require patient specific organ segmentations, realistic (and computationally expensive) non-linear models of the single organs with the corresponding exact tissue parameters. Not to forget the sliding motion of the organs, resulting in discontinuous and anisotropic displacement fields. Keeping in mind that these are just some of the requirements, it is easy to get the first idea about the depths of the complexity of the real problem. In practical applications, the body model is mostly approximated by a regularization term, requiring that the displacement is smooth in a certain way, mostly assuming homogeneous behaviour in the whole domain, such as in Equation (1.7). In this light, it is clear, that accuracy of such a model is limited, and any estimated result is only an approximation to the actually desired real solution. We believe strongly that the modeling gap should be kept in mind, when reasoning about the registration process. This is of particular interest when one considers the quality of the optimization result. It is often encountered in experiments that the optimization of the model energy beyond a certain level does not result in a decrease of the actual error. In some cases, it is even observed that the exact optimization of the model energy results in worse results in terms of the actual error, compared to an intermediate result with a higher corresponding energy.

A Generic Gradient-based Deformable Registration Boiled down, all approaches for deformable registration known to us can be seen as an iterative accumulation of updates to the initial deformation estimate. Plainly speaking, the difference between the approaches consists in the way of estimating the updates and in the way of composing them, and is a consequence of the chosen energy, transformation model, and optimization scheme. This generic framework

is represented by Algorithm 1.1.

Furthermore, most optimization methods used for deformable registration are based on the gradient of the energy model ∇E , and these approaches were the focus of our work. In the case of gradient-based registration method, the iterative updates mentioned above are essentially smoothed versions of the (potentially pre-processed) energy gradient.

So, although not strictly describing all methods, most deformable registration algorithms can be thought of in the form of the following pseudo algorithm

$$\begin{aligned} v_i &\leftarrow \text{smooth}\left(f\left(-\nabla E_D(I_T, I_S \circ \phi_i)\right)\right) \\ \phi_{i+1} &= \phi_i \oplus v_i \end{aligned} \quad (1.9)$$

The smoothing step in these approaches either results from the combination of the numerical method and the regularization term (\leftrightarrow variational elastic methods with semi-implicit time discretization, \leftrightarrow elastic-type demons method), or is a consequence of the choice of the transformation model (\leftrightarrow Group of Diffeomorphisms, \leftrightarrow Sobolev spaces, \leftrightarrow low-dimensional transformation models). The pre-processing of the gradient, denoted above by the function f , depends on the choice of the optimization method.

For gradient based methods, the computation of the energy gradient plays a central role. It is interesting to note that the energy gradient has some structural properties which hold for arbitrary energy models. Due to the linearity of the energy term, the gradient can be computed independently for the difference measure and the regularization term, so that we have

$$\nabla E = \nabla E_D + \lambda \nabla E_R . \quad (1.10)$$

Furthermore, for linear regularization terms as defined in (1.7), we have

$$\nabla E_R(u) = A_R u . \quad (1.11)$$

And finally, the structure of the gradient of the difference measure also has a particular structure cf. e.g. [Hermosillo et al., 2002, Chéfd'hotel et al., 2002]. Independently of the choice of the difference measure, the gradient can be written as

$$\nabla E_D(I_T, I_S, \phi) = \omega_D(I_T, I_S, \phi) \nabla I_S \circ \phi . \quad (1.12)$$

Here, ω_D is a scalar-valued function, and it is the only part of which depends on the actual choice of the difference measure. Also, please note that the term $\nabla I_S \circ \phi$ amounts to either $\nabla(I_S \circ \phi)$ for the assumption of deformations as elements of linear spaces or $(\nabla I_S) \circ \phi$, if a group structure is assumed, cf. e.g. [Stefanescu et al., 2004].

One further point should still be mentioned: the update scheme, which we denoted until now by \oplus . In the early works on deformable registration, transformations were mostly assumed to be elements of a linear vector spaces, and the update operator was chosen simply as the addition of the two corresponding displacement fields. In the work on large deformation registration, it was suggested that it is more appropriate to treat transformations as a group. If this view is taken, then the updates must be performed such that the resulting transformation is still a member of the group. For transformations, a natural

operator which fulfills this requirement is the composition of two transformations, so that we have

$$\phi \oplus v = \phi \circ (\text{Id} + v) . \quad (1.13)$$

Although first discussed in the context of large deformation registration, the results regarding the update schemes are also transferable to other methods.

1.1 Review of Deformable Registration Methods

While there are many overviews of rigid registration methods which agree on a consistent framework [Brown, 1992, Van den Elsen et al., 1993, Maintz and Viergever, 1998, Fitzpatrick et al., 2000, Hill et al., 2001, Hajnal et al., 2001b, Zitova and Flusser, 2003], the field of deformable registration is much less sorted.

Our goal is to provide an overview of the different approaches for deformable registration, and include recent developments in the field. We structure the presentation based on the original motivation and derivation of the single methods, and discuss the following lines of work: *variational elastic approaches*, *large deformation registration*, *demons registration approaches*, and *parametric registration*. In spite of the highly heterogeneous derivations, all the methods share the same underlying structure, and many links between the single methods have been established in the literature. For example, all discussed methods fit either to the variational or the parametric way of derivation, as illustrated in Figure 1.1. The elastic, large deformation, and demons registration methods belong to the variational methods, which are derived in the infinite-dimensional setting, with a subsequent parametrization step. Parametric methods, on the other hand, employ the finite-dimensional parametrization of the transformation directly in the energy model, as the starting point for the derivation of the algorithm. Please note that rigid approaches are implicitly included in the group of parametric methods.

In the following we give brief overviews of the single groups of the approaches and the lines of work on deformable registration, which we will discuss in more detail in the following chapters.

Elastic Variational Registration (Sec. 2) represents the oldest approaches for deformable registration. They treat the problem in a continuous infinite-dimensional setting and postpone the parametrization step to the final stage of the derivation. Many of these approaches were originally motivated by the intuition to simulate the behavior of an elastic deformable body, which responds to a set of forces derived from a difference measure between the input images. For this reason, this general group of approaches is also known as *elastic registration* in the medical domain. An equivalent and more general view to this approach is to interpret the problem as a minimization of an energy model consisting of a difference measure E_D and a regularization term E_R . One limitation of these approaches is the restriction to comparably small deformations. This behavior is caused by the regularization term, which penalizes large and complex deformations since they would cause too high regularization energies. While appropriate in some settings, this behaviour is too constraining in others, for example for inter-subject brain registration.

Large Deformation Registration (Sec. 3) was introduced to allow the estimation of large deformations, in scenarios where elastic methods proved too restrictive. The first approaches in this direction were based on the idea of replacing the simulation of an elastic object by the simulation of a viscous fluid. While preserving certain regularity properties of the resulting transformation, this approach does not limit the magnitude of the deformations. Because of this original motivation, this group of approaches is sometimes also referred to as *fluid registration*. Further work in this area extended and generalized this approach by demonstrating that the fluid simulation corresponds to a minimization of a difference measure, without a regularization term, in the group of diffeomorphisms. A further set of methods in this group are the so called Large Deformation Diffeomorphic Metric Mapping (LDDMM) approaches, which use as the regularization term the geodesic length of the deformation in the space of diffeomorphisms. Methods for large deformation registration are usually also classified as belonging to the group of variational methods, as they in mostly perform the derivation of the algorithms in infinite-dimensional spaces.

Demons Registration Approaches (Sec. 4) were originally introduced as an efficient and heuristic alternative to the computation-heavy fluid registration approaches, which were based on the physical simulation of elastic bodies or fluids. Originally, the demons approach was assuming a mono-modal scenario, and was implementing an elastic-type strategy. In the meanwhile, the method has been extended to fluid-type registration problems, and it has been generalized to arbitrary difference measures and regularization terms. Through several interpretations, the method is now theoretically well founded, and can be interpreted as an efficient numerical approach for solving the optimization problems arising from the elastic or fluid-type registration approaches.

Parametric Registration (Sec. 5) are characterized by employing a finite and mostly low-dimensional transformation model directly in the definition of the energy model, and by performing the derivation of the actual algorithm in the finite-dimensional setting. This is in contrast to the previously considered variational approaches. The derivation of the registration algorithm is based on the optimization of the energy with respect to the transformation parameters. Please note that this group naturally includes the rigid registration approaches. A motivation for the development of parametric methods was the increase of robustness and reduction of computation time by employing a comparably low number of parameters, resulting in low-dimensional problems. Although the parametric approaches take a different route in deriving the actual algorithms (cf. Figure 1.1), in general, with appropriate transformation models, same methods can be derived in this way as with variational derivations. A further advantage of parametric approaches is their conceptual simplicity. For example, theoretical questions which occur in infinite-dimensional derivations, such as the existence of the solution, do not arise in the finite-dimensional context.

ELASTIC VARIATIONAL REGISTRATION

The variational method for optimizing an energy consisting of a difference measure and a regularization term has been a very popular approach for deformable registration since its introduction in [Broit, 1981] and [Horn and Schunck, 1980, Horn and Schunck, 1981]. Actually, these works are generally considered to be the first significant approaches for estimation of non-linear transformations. Originally, a common motivation for elastic variational approaches in the medical domain was to model the imaged body as an elastic material, which reacts to forces derived from the difference measure, cf. e.g. [Broit, 1981]. This process can be modeled as a partial differential equation (PDE), and following the variational method¹ an equivalent energy formulation is derived, in which the elastic behavior of the body is represented by the regularization energy, which in this context is also often called internal energy. The point of view of energy minimization is equivalent to the simulation intuition and the resulting PDE-based modeling for the relevant cases, and it has been generally adopted in most of the subsequent research.

In [Horn and Schunck, 1981] the motivation for the use of a regularization term is different. Here, the regularization term was employed as a remedy to the observation that the minimization of the difference measure alone would result in an under-constrained and thus ill-posed problem (\leftrightarrow aperture problem). Please note that it was later observed that while the optimization of the SSD is under-constrained in L^2 , it can become a well posed problem if appropriate Sobolev spaces are chosen [Trouvé, 1995, Dupuis et al., 1998].

So, in summary, the basic idea of the elastic variational approach is to formulate a global energy model, such as in (1.4), consisting of a difference

¹A short note on my understanding of the term *Variational Methods*: Originally, the term *variational method* denotes an approach for solving a certain class of PDEs. Starting with a given PDE (e.g. (2.2)), the idea is to find a corresponding energy E (e.g. (1.4)), such that the solutions of PDE are critical points of E . As summarized in [Evans, 1998, Ch. 8]: “The point is that whereas it is usually extremely difficult to solve [the PDE] directly, it may be much easier to discover critical points of the functional.” In the context, the energy formulation is referred to as the *variational formulation* of the problem. Since most modern registration approaches start by formulating directly the energy, and not the PDE itself, they do not strictly correspond to the original variational approach, but take the variational route in the opposite direction. In context of registration, the term *variational* is thus nowadays mostly used in a more general sense, to denote methods which perform the derivation in the continuous setting.

measure and a regularization term, that is

$$E(u) = E_{\mathcal{D}}(u) + \lambda E_{\mathcal{R}}(u) , \quad (2.1)$$

and minimize it with respect to the displacement field u , representing the non-linear transformation $\phi = \text{Id} + u$. The derivation is performed in the infinite-dimensional continuous setting, and the displacement u is assumed to belong to a certain Hilbert space \mathbb{H} - in most works L^2 is assumed. Correspondingly, the necessary condition for a local optimum of the energy is derived in the form of an Euler-Lagrange partial differential equation (PDE) system

$$\lambda A_{\mathcal{R}}u + f_{\mathcal{D}}(u) = 0 . \quad (2.2)$$

The Euler-Lagrange equation corresponds to the derivative of the energy (1.4) with respect to the displacement [Evans, 1998], that is

$$\nabla E(u) = \lambda A_{\mathcal{R}}u + f_{\mathcal{D}}(u) , \quad (2.3)$$

so that (2.2) can also be read as $\nabla E = 0$ and presents a necessary condition for an extremal point of (1.4). In (2.2), $A_{\mathcal{R}}u$ corresponds to $\nabla E_{\mathcal{R}}(u)$, and the so called *force* $f_{\mathcal{D}}(u)$ corresponds to $\nabla E_{\mathcal{D}}(u)$. The Euler-Lagrange equation in (2.2) is non-linear due to the non-linearity of $E_{\mathcal{D}}(u)$, and $f_{\mathcal{D}}(u)$, in which the transformation acts through the warping of the source image, which introduces the non-linearity. Please compare Section 2.1 for more details.

Please note that because of its non-linearity, (2.2) cannot be solved directly. Thus, while it characterizes the extremal point, it does not directly provide a way of computing the solution. In order to estimate the transformation satisfying the Euler-Lagrange equation, a time-dependent process is introduced, which features an artificial time and corresponds to a gradient descent strategy

$$\frac{\partial u(t)}{\partial t} = -\nabla E(u(t)) = -[\lambda A_{\mathcal{R}}u(t) + f_{\mathcal{D}}(u(t))] , \quad (2.4)$$

with a given $u(t_0)$. This step transforms the non-linear Euler-Lagrange PDE from (2.2) into a dynamic (time dependent) PDE system.

In order to arrive at a formulation which can be practically implemented, a discretization of the dynamic non-linear PDE system in (2.4) is required. This discretization is twofold. The first step of the discretization process is the time discretization, and yields a series of parabolic linear PDE systems of the form

$$Au(t + \tau) = -f . \quad (2.5)$$

The above system is linear, since at time t , the unknown is $u(t + \tau)$, and $u(t)$ is given, so that $f = f_{\mathcal{D}}(u(t))$ is assumed fixed. So, the result of the time discretization is an iterative process, which corresponds to the iteration loop in Algorithm 1.1. Common approaches for time discretization are the *explicit* and the *semi-implicit* discretization schemes. The temporal discretization in variational approaches loosely corresponds to the choice of the optimization method in other approaches. For details on the time discretization, please compare Section 2.3.

The second discretization step is the spatial discretization, which transforms the linear PDEs from (2.5) into finite dimensional linear equation systems, the

solution of which can be estimated numerically. The common approaches for the spatial discretization are the *finite difference* (FD) and the *finite element* (FE) method, please compare Section 2.4 for more details.

For a schematic outline of variational approaches in general, please compare also Figure 1.1. We now discuss the single components of the elastic variational approach in some more detail in the following sections.

2.1 Optimality Condition: Euler-Lagrange PDE

Starting with the energy formulation E from (1.4) and assuming that the unknown displacement $u \in \mathbb{H}$ is not parametrized, and an element of an appropriate function Hilbert space \mathbb{H} , the condition for the extremal point u' of E is characterized by the vanishing of the first variation of u' with respect to all possible test functions v , that is

$$0 = \delta E(u', v) , \forall v \in \mathbb{H} . \quad (2.6)$$

The first variation of E , at $u \in \mathbb{H}$, in the direction $v \in \mathbb{H}$ is defined as

$$\delta E(u, v) = \lim_{\epsilon \rightarrow 0} \frac{E(u + \epsilon v) - E(u)}{\epsilon} \Big|_{\epsilon=0} = \frac{dE(u + \epsilon v)}{d\epsilon} \Big|_{\epsilon=0} . \quad (2.7)$$

The first variation is also known as the *Gâteaux derivative*.

The gradient $\nabla_{\mathbb{H}} E(u)$ of E at u in \mathbb{H} is defined by

$$\delta E(u, v) = \langle \nabla_{\mathbb{H}} E(u), v \rangle , \forall v \in \mathbb{H} . \quad (2.8)$$

With the definition of the gradient, one can re-write the optimality condition (2.6) as

$$0 = \langle \nabla_{\mathbb{H}} E(u'), v \rangle , \forall v \in \mathbb{H} , \quad (2.9)$$

and derive an alternative formulation of the optimality condition

$$0 = \nabla_{\mathbb{H}} E(u') . \quad (2.10)$$

The above equation system is called the *Euler-Lagrange equation*, associated with the energy E . For convenience, in the following we will simply write ∇E for $\nabla_{\mathbb{H}} E$, where the context is clear.

For the model energy from (1.4), the Euler-Lagrange equation has the general form

$$0 = \nabla E(u) = \nabla E_D(u) + \lambda \nabla E_R(u) . \quad (2.11)$$

Since the corresponding energy E is based on input images, and thus non-linear, the optimality condition will in general characterize only local critical points.

Based on the linear model of the regularization term in (1.7), we set

$$A_R := \nabla E_R = L_R^* L_R . \quad (2.12)$$

Furthermore, we use

$$f_D := \nabla E_D , \quad (2.13)$$

for the non-linear derivative of the difference measure $\nabla E_D(u)$, which is commonly referred to as the *force* term. We will freely switch between the gradient notation and the use of A and f . This denomination of ∇E_D as the force is due to the interpretation of PDEs resulting from simulation of mechanical models, which are commonly written as

$$Au = -f . \tag{2.14}$$

Here, the right hand side represents the external forces f acting on a certain object, while the response of the object is modeled by the operator A . The resulting displacement in each step $u(t + \tau) = -A^{-1}f(u(t))$ can thus be seen as the result of the simulation of the interaction of the object model, and the set of external forces. Intuitively, the equation (2.14) is satisfied, when the *inner forces* Au exercised by the object exactly oppose the external forces f . This simulation point of view presented a motivation for many approaches developed for deformable registration, cf. e.g. [Broit, 1981].

Please note that while the Euler-Lagrange equation from (2.11) characterizes the optimality condition, because of its non-linearity, it cannot be directly solved, and thus it does not directly provide the means to computing the solution. The following section gives an overview of a standard approach for iterative estimation of the solution of the Euler-Lagrange equation.

2.2 Estimating the Solution of the Euler-Lagrange PDE

The Euler-Lagrange equation arising from the deformable registration problem is non-linear. A standard way of estimating the solution of (2.11), is by introducing an artificial time, thus transforming the static (time-independent) PDE from (2.11) to a dynamic (time-dependent) PDE, the steady state of which should correspond to the equilibrium of the static problem.

To this end, the unknown displacement is augmented by an artificial time to $u : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^d$, and the so called *gradient flow* is defined by

$$\frac{\partial u(t)}{\partial t} = -\nabla E(u(t)) , \tag{2.15}$$

with the initial displacement $u(0)$. In the following, we employ the more compact notation ∂_t for the temporal derivative, where convenient. Please note that (2.15) corresponds to a gradient descent strategy [Hermosillo, 2002, p. 39].

It is also interesting to note that the steady state of (2.15), that is, when the estimate of u does not change any more, i.e. $\partial_t u \rightarrow 0$, corresponds to the static PDE from (2.2). In order to compute the steady state of (2.15), a linearization of the PDE is required. This step is performed by discretization of the time, which is discussed in the next section.

2.3 Time Discretization

The time discretization transforms the dynamic PDE from (2.15) to an iterative process with linear PDEs in the single iterations. The idea of time discretization

is to approximate the continuous temporal variable by discrete time steps in an iterative process. For simplicity, we assume here that a fixed step size τ is employed, so that the time t corresponding to the iteration i can be represented by $t_n = n\tau$ (otherwise, we have $t_n = \sum_{i=0}^n \tau_i$).

The time discretization step corresponds to the choice of the optimization method in other approaches. For example, we will see that the explicit time discretization is equivalent to steepest gradient descent, the semi-implicit scheme can be seen as steepest descent in varying Sobolev space, and the time discretization based on quadratic approximation of the energy in each iteration corresponds to Newton-type optimization. Thus the time discretization has a major influence on the behavior of the overall process, for example the convergence speed, and the quality of the estimated result.

Numerous time discretization schemes are available in the literature. We discuss the common *explicit* and *semi-implicit* time discretization schemes in the following two sections.

Although usually not presented in the context of classical elastic variational methods, the smoothing strategy of the elastic-type demons method can be interpreted as a specific efficient time discretization technique. We discuss this scheme in Section 2.3.3.

Finally, an important group of time discretization techniques operates by a quadratic approximation of the difference measure and the corresponding linearization of the force term, cf. Section 2.3.4. It is important to note that this strategy is not applicable to all difference measures (e.g. not applicable to MI). When applicable, however, these approaches result in efficient Newton-type optimization, which considerably improves the convergence speed compared to steepest descent strategies.

2.3.1 Explicit Time Discretization

The first step in the time discretization by the *explicit time discretization scheme* (also known as (*forward Euler*)) is the discretization of the temporal derivative of the displacement, which can be given by

$$\frac{\partial u(t)}{\partial t} = \lim_{\tau \rightarrow 0} \frac{u(t + \tau) - u(t)}{\tau} . \quad (2.16)$$

The second step is the choice of the point in time at which the energy gradient is computed. For the explicit time discretization the gradient is computed at the current time point, that is, we have $\nabla E(u(t))$.

These two choices result in the explicit time discretization scheme

$$\frac{u(t + \tau) - u(t)}{\tau} = -\nabla E(u(t)) \quad (2.17)$$

$$= -[\nabla E_D(u(t)) + \lambda \nabla E_R(u(t))] , \quad (2.18)$$

which can be re-written to

$$u(t + \tau) - u(t) = -\tau \nabla E(u(t)) \quad (2.19)$$

$$v = -\tau [\nabla E_D(u(t)) + \lambda \nabla E_R(u(t))] \quad (2.20)$$

$$v = -\tau [f_D(u(t)) + \lambda A_R u(t)] . \quad (2.21)$$

Here we denote the displacement update by v , i.e. the difference in the displacement between the current and the next iteration $v := u(t + \tau) - u(t)$, which is the discretized approximation to $v = \partial_t u$.

The first observation to be made in regard to (2.20), is that this iterative rule for the computation of the update is equivalent to the standard steepest descent optimization method for non-linear optimization [Nocedal and Wright, 2000]. The update is simply a multiple of the negative energy gradient.

The second striking property of (2.18) is the simplicity of its computation, and the low computational cost per iteration - only the evaluation of the energy gradients is required. This simplicity however does come at a price: The explicit scheme is numerically stable only for small values of τ . In consequence, a large number of iterations can be required to achieve convergence. Based on the choice of the regularization term, a maximal value of τ for which stable behavior is achieved can be given analytically.

2.3.2 Semi-implicit Time Discretization

The second standard time discretization approach is the so called semi-implicit scheme

$$\frac{u(t + \tau) - u(t)}{\tau} = - [\nabla E_D(u(t)) + \lambda \nabla E_R(u(t + \tau))] . \quad (2.22)$$

Same as the explicit method, it assumes the discretization of the time derivative as in (2.16). The difference between (2.22) and (2.18), is that the semi-implicit scheme assumes that the gradient of the regularization term is evaluated at the next point in time. The advantage of the semi-implicit method is that it is numerically stable for all possible time steps τ , however, this stability comes at the cost of increased computational complexity.

Intuitively, the difference between the explicit and the semi-implicit scheme can be explained as follows. Since the explicit scheme computes the gradient of the regularization term at the previous point in time t , the regularization only reacts to the irregularities which were present in $u(t)$, and cannot take into account the irregularities introduced in the current update by $\tau \nabla E_D(u(t))$. Large irregularities (for large τ) can thus contaminate the estimate beyond repair. In contrast, the regularization in the semi-implicit approach acts at the future time point $t + \tau$, and therefore can cope with the irregularities from $\nabla E_D(u(t))$. This makes the semi-implicit scheme stable: No matter how irregular the update might be, it is taken into account at the future time step. On the other hand, computing the gradient “at the future point in time” is a more complex operation, which does not come for free.

It is instructive to re-write (2.22) to the update form as

$$v = -\tau (\text{Id} + \tau \lambda A_R)^{-1} [\lambda A_R u(t) + f_D(u(t))] \quad (2.23)$$

$$= -\tau (\text{Id} + \tau \lambda A_R)^{-1} \nabla E(u(t)) . \quad (2.24)$$

Alternatively, in term of the complete displacement u and not the update v , we have

$$u(t + \tau) = (\text{Id} + \tau \lambda A_R)^{-1} [u(t) - \tau f_D(u(t))] . \quad (2.25)$$

The obvious difference between the explicit scheme update in (2.20) and (2.24) is that for the semi-implicit scheme, a solution of the linear system, with the linear operator $(\text{Id} + \tau\lambda\nabla E_{\text{R}})$ is required. It is this step which increases the computational complexity. Please note that the linear operator depends on the regularization term E_{R} , or more precisely, its derivative ∇E_{R} . Also, the resulting system is only linear with the underlying assumption of E_{R} being a quadratic form. For all standard definitions of E_{R} , the application of $(\text{Id} + \tau\lambda\nabla E_{\text{R}})^{-1}$ corresponds to a smoothing operation.

On the positive side, there is no numerical restriction on the size of the time step τ for the solution of the linear system in (2.24).² Consequently, less iterations are required, which results in faster convergence speed. Usually, the gain in convergence outweighs the increased cost of single iterations, so that the resulting overall runtime is also improved when a semi-implicit scheme is used.

Based on the observation that applying $(\text{Id} + \tau\lambda\nabla E_{\text{R}})^{-1}$ in (2.24) corresponds to a smoothing of the energy gradient, in [Zikic et al., 2010f, Zikic et al., 2010a] we provide an alternative interpretation of the algorithm resulting from semi-implicit time discretization as a steepest descent in Sobolev spaces. The argument for this is rather simple, since for the family of τ -dependent Sobolev spaces \mathbb{H}_{τ} , based on the scalar product $\langle L^*L(\tau)u, u \rangle_{\mathbb{L}^2}$, with $L^*L(\tau) = (\text{Id} + \tau\lambda\nabla E_{\text{R}})$, the gradient of an energy E corresponds to $(L^*L(\tau))^{-1}\nabla E = (\text{Id} + \tau\lambda\nabla E_{\text{R}})^{-1}\nabla E$, which is exactly the formulation from (2.24).

2.3.3 Demons Time Discretization

In this section, we briefly sum up a strategy for time discretization, which yields an algorithm which is equivalent to the *elastic* demons regularization scheme as introduced in [Thirion, 1998]. The argument is based on the derivation in [Cahill, 2009]. Please note that in this section, we refer to the *elastic* demons approach, in which the smoothing is applied to the complete displacement in each iteration, and not the *fluid* demons version, in which the smoothing is performed only on the displacement updates. We will discuss the difference between elastic and fluid demons in more detail in Section 4.

While this scheme is not commonly used in traditional variational approaches, it presents an important concept, which links “classical” variational methods to the elastic demons approach, which will be discussed in Section 4.

For the *demons time discretization scheme* for the gradient flow from (2.15), we assume that in each iteration, the force term is evaluated at the current point in time t_0 , i.e. we have $f_{\text{D}} := f_{\text{D}}(u(t_0))$, while all the other entities are evaluated at a future time point t , and get

$$\partial_t u(t) = -\lambda A_{\text{R}} u(t) - f_{\text{D}}(u(t_0)) , \quad (2.26)$$

with the initial value $u_0 = u(t_0)$. Assuming $f_{\text{D}} = f_{\text{D}}(u_0)$ fixed, the above can be re-written in terms of the operator $(\partial_t + \lambda A_{\text{R}})$ as

$$(\partial_t + \lambda A_{\text{R}}) u(t) = -f_{\text{D}} , \quad (2.27)$$

²Please note, that this does not mean, that all values of τ are meaningful. Too large values of τ will result in a bad approximation of the temporal evolution of the result, and may result in jumping over the desired solution estimate. Also, an intuitive setting-dependent step size limits exist. The maximum reasonable updates in most applications depends on the resolution of the input data, since the length of the maximal point-wise update $\|v(x)\|$ should be in the order of a pixel spacing.

where the nature of (2.27) as a linear parabolic PDE becomes more clearly visible. Please note that the distribution of current and future time points in (2.27) corresponds to the semi-implicit assumption in (2.22). The major difference to the semi-implicit scheme is that in (2.27), no numerical approximation of the temporal derivative ∂_t is performed at this point.

We can compute the solution to (2.27) by superposing the solutions of the homogeneous equation with the original source term $u_0 = u(t_0)$:

$$(\partial_t + \lambda A_R) u(t) = 0 \quad \text{with} \quad u_0 = u(t_0) , \quad (2.28)$$

and the solution of the inhomogeneous equation with a zero source term $u_0 = 0$:

$$(\partial_t + \lambda A_R) u(t) = -f_D(t) \quad \text{with} \quad u_0 = 0 . \quad (2.29)$$

Please compare [Evans, 1998, Sec. 2.3.1, p. 51] for details of the above approach. The solution to (2.28) is provided by the fundamental solution for the operator $(\partial_t + \lambda A_R)$, and (2.29) can be solved by the so called Duhamel's principle.

The resulting scheme can be written as

$$u(t) = \underbrace{\Psi_R(\lambda, t - t_0) * u(t_0)}_{\text{solution to (2.28)}} - \underbrace{\int_{t_0}^t \Psi_R(\lambda, t - s) * f_D(s) ds}_{\text{solution to (2.29)}} \quad (2.30)$$

Here $\Psi_R(\lambda, t)$ is the Green's function to $(\partial_t + \lambda A)$, i.e. it fulfills $(\partial_t + \lambda A) \Psi_R(\lambda, t) = \delta$. As an example, for diffusion regularization, Ψ_R corresponds to a Gaussian with the standard deviation of $\sigma = \sqrt{2t\lambda}$, that is, we have $\Psi_R(\lambda, t) = G_{\sqrt{2t\lambda}}$.

We can approximate the time integral in the solution of the inhomogeneous part from (2.30), based on the knowledge of $f_D(t_0)$, and get

$$\int_{t_0}^t \Psi_R(\lambda, t - s) * f_D(s) ds \approx (t - t_0) \Psi_R(\lambda, t - t_0) * f_D(t_0) . \quad (2.31)$$

By substituting (2.31) into (2.30), we obtain

$$u(t) \approx \Psi_R(\lambda, t - t_0) * u(t_0) - (t - t_0) \Psi_R(\lambda, t - t_0) * f_D(t_0) \quad (2.32)$$

$$= \Psi_R(\lambda, t - t_0) * [u(t_0) - (t - t_0) f_D(t_0)] . \quad (2.33)$$

Finally, the formal connection to the introduced registration framework is established simply by renaming the future time point from t to $t + \tau$ and the initial/current time point from t_0 to t . This yields the *demons time discretization scheme*

$$u(t + \tau) = \Psi_R(\lambda, \tau) * [u(t) - \tau f_D(u(t))] . \quad (2.34)$$

As already mentioned, in the case of diffusion regularization, we have $\Psi_R = G_{\sqrt{2\tau\lambda}}$, and the equation from (2.34) is identical to the original elastic demons scheme from [Thirion, 1998]. The scheme in (2.34) demonstrates how the elastic demons approach (generalized to arbitrary linear regularization operators) can be seen in the context of classical variational methods as a particular and efficient numerical scheme for time discretization. The above derivation for all regularization terms, for which an appropriate Green's function to $(\partial_t + \lambda A_R)$

Explicit scheme	$u(t + \tau) = u(t) - \tau [\lambda A_R u(t) + f_D(u(t))]$
Semi-implicit scheme	$u(t + \tau) = (\text{Id} + \tau \lambda A_R)^{-1} [u(t) - \tau f_D(u(t))]$
Demons scheme (generalized)	$u(t + \tau) = (\partial_t + \tau \lambda A_R)^{-1} [u(t) - \tau f_D(u(t))]$ $= \Psi_R(\lambda, \tau) * [u(t) - \tau f_D(u(t))]$

Table 2.1: Overview of standard time-discretization methods for deformable registration.

exists. This includes all standard homogeneous regularization terms, i.e. terms with constant regularization strength in the spatial domain.

It is interesting to note that the scheme from (2.34) corresponds to

$$u(t + \tau) = (\partial_t + \lambda A_R)^{-1} [u(t) - \tau f_D(u(t))] , \quad (2.35)$$

which allows a direct comparison to the semi-implicit scheme from (2.22). Please compare also Table 2.1 for an overview.

2.3.4 Time Discretization based on Quadratic Energy Approximation

The last scheme for time discretization which we discuss is based on the quadratic approximation of the non-linear energy in every iteration. Since the regularization term is assumed to be quadratic, this approximation only affects the difference measure, and with the assumption of the L^2 norm, the approximation results in the linearization of the error term of the difference measure. So, with the current estimate u' , and $u = u' + v$, we have

$$E(u) = E_D(u) + \lambda E_R(u) = \frac{1}{2} \|e_D(u)\|^2 + \lambda \frac{1}{2} \|L_R u\|^2 \quad (2.36)$$

$$\approx \frac{1}{2} \|e_D(u') + J_{e_D}(u')v\|^2 + \lambda \frac{1}{2} \|L_R u\|^2 . \quad (2.37)$$

Such an approximation has been performed in a series of works, with early examples in [Broit, 1981] and [Horn and Schunck, 1981]. Similar time discretization strategies have been also considered in [Brox et al., 2004, Papenberg et al., 2006]. Also, this scheme applied to SSD corresponds to the computation of the forces in the original demons approaches [Pennec et al., 1999, Vercauteren, 2008, Vercauteren et al., 2009].

It is interesting to note that the quadratic approximation of the energy by linearization of the error terms corresponds to the Gauss-Newton optimization method [Madsen et al., 2004, Sec. 3.1]. In scenarios where it is applicable, this scheme results in much faster convergence speed than strategies based on steepest gradient descent [Zikic et al., 2010e, Zikic et al., 2011]. However, one disadvantage of this scheme is that it is not practically applicable to arbitrary difference measures, especially statistical difference measures, such as MI [Modat et al., 2010, Zikic et al., 2010a, Zikic et al., 2011], when high-dimensional parametrizations are employed. The reason is that for statistical measures such as MI, the Jacobian of

the error term J_{e_D} , which is needed for the linearization, does not have a sparse form, in contrast to point-wise measures such as SSD. For high-dimensional parametrizations and 3D problems, the computation and storage of this Jacobian become impracticable.

2.4 Spatial Discretization

The result of any of the above time discretization schemes is a linear PDE. The last step which is required to develop an actual implementation is the parametrization of the problem, which approximates the infinite dimensional continuous problem by a finite dimensional version. This step is performed by a spatial discretization of the said PDE. The most popular strategies for the discretization are the *finite difference* (FD) scheme, and the *finite element* (FE) scheme.

This section is heavily based on [Strang, 1986].

2.4.1 Finite Difference Method

Compared to the finite element method, the finite difference (FD) method has the advantage of conceptual simplicity. It is furthermore easier to program and faster to run than a full FE scheme [Strang, 1986, Ch. 5.4, p.428].

The idea of the finite difference method is to choose a finite number of points in the spatial domain, and to approximate the required derivatives by differences between the values at these points. The points are usually arranged in a regular grid, with distance h_i along the dimension i . A definition of derivative operators is required for the construction of the linear system matrix A , and it is mostly based on central differences. Higher order, and multi-dimensional derivative approximations can be constructed from the basic one-dimensional central difference approximation.

$$\frac{\partial u(x)}{\partial x_i} \approx \frac{u(x + h_i e_i) - u(x - h_i e_i)}{2h_i}, \quad (2.38)$$

with e_i being the unit vector in dimension i . Please compare [Strang, 1986, Ch. 5, p.370] for a more detailed example.

The properties of the actual resulting linear operator A depend on the choice of the regularization and the time discretization scheme.

2.4.2 Finite Element Method

The finite element (FE) method is extremely popular in engineering applications. While its implementation is usually more complex than that of FD, it has the great advantage that it can be applied in irregular geometries. Also, since it can employ adaptive sampling meshes, it can reduce the number of parameters in non-critical areas, thus achieving the same approximation accuracy as FD, however with less degrees of freedom (i.e. with a lower-dimensional transformation model).

From the theoretical point of view the concept of FE is interesting, since it is more general than FD. As a matter of fact, FD can be seen as specific instance of FE [Strang, 1986, Ch. 5.4, p.428].

In the context of registration, the concept of FE is interesting, since it provides a link between the variational derivation and the group of the so called parametric methods. Many parametric methods can be also obtained by variational derivation, with the spatial discretization performed by an appropriate FE method.

In contrast to the FD method, which performs the discretization by choosing a finite number of sampling points, the finite element (FE) method operates by choosing a finite number of basis functions, and approximating the exact infinite dimensional solution by a linear combination of those basis functions [Strang, 1986, Ch. 5.4, p.428]. There are two alternatives for deriving the FE formulation, which are equivalent for the class of problems arising in image registration³ - the *Galerkin* and the *Rayleigh-Ritz* approach [Strang, 1986, Ch. 5.4, p.432]. We will focus on the Rayleigh-Ritz formulation in the following.

The linear PDE which results from the time discretization has the general form

$$Au = -f . \quad (2.39)$$

Please note that here A denotes the complete linear operator resulting from the time discretization, and not only the derivative of the regularization term A_R , e.g. $A = \text{Id} + \lambda\tau A_R$. This PDE has the corresponding energy formulation (variational formulation)

$$\min \frac{1}{2}u^\top Au + f^\top u . \quad (2.40)$$

The FE scheme approximates u by a finite linear combination of basis functions (trial functions) b_i , weighted by parameters p_i

$$u(x) \approx \sum_i b_i(x)p_i . \quad (2.41)$$

By collecting the basis functions into the operator $B = [b_1, \dots, b_n]$, and by denoting the parameter vector by $p = [p_1, \dots, p_n]^\top$, we can define the approximation by

$$\tilde{u} := Bp . \quad (2.42)$$

For implementation, the actual operator B is ultimately rendered finite dimensional by sampling of the bases b_i at discrete locations.

By inserting \tilde{u} into the energy function, we get

$$\min \frac{1}{2}(Bp)^\top ABp + f^\top Bp . \quad (2.43)$$

The necessary condition for a critical point of the above energy is given by taking the derivative with respect to the parameters p . This results in the finite dimensional linear system

$$B^\top ABp = -f^\top B . \quad (2.44)$$

The remaining important point is the actual choice of the bases b_i . There are many different possibilities presented in the FE literature, and a review is beyond

³This is the class of symmetric problems which can be posed as energy minimization.

the scope of this work. Instead, we will briefly list the important properties of the bases, and give some examples used for image registration.

The choice of the basis functions depends on many problem specific properties, such as the geometry of the problem domain Ω , the used regularization term E_R , the chosen time discretization/optimization scheme (which influences the form of A), or the properties of the transformation (e.g. its expected form and complexity).

The requirements for the basis functions can be summed up in the following way, compare [Strang, 1986, Ch. 5.4, p.433]:

- The bases have to be sufficiently expressive to approximate the true solution with required accuracy
- The resulting linear operator $B^T AB$ should be simple to assemble, and have advantageous properties for further computation, i.e. it should be sparse and well-conditioned

One large and popular class of basis functions for FE schemes features bases which are overlapping only in a small part of the domain. The advantage of these bases is that they result in a sparse matrix B , and correspondingly, the linear operator $B^T AB$ is also sparse. A simple example of these functions is given in 2D by bases which are composed of linear functions, defined on triangles of a mesh, which covers the spatial domain. The linear functions share the same weight of 1 at the central, joint node, which is represented by the basis, and equal to 0 at all other nodes (cf. Fig. 5.2c for a 1D example). A second example are the B-Spline FFD bases, which are a popular choice in image registration, and employ a regular mesh cf. Section 5.2.1.

The second popular group of bases, are not necessarily non-overlapping functions, but functions which “are special to the problem” [Strang, 1986, Ch. 5.4, p.433]. The bases can be chosen, such that they are eigenfunctions to the regularization operator A , which leads to a diagonal operator $B^T BA$.

As mentioned earlier, one advantage of FE over FD is to employ problem specific meshes. On the other hand, the generation of problem specific meshes is one of the implementational burdens of FE, which does not exist in FD. A review of the possible techniques is again beyond the scope.

Some examples for the use of FE in image registration are found in following works [Gee et al., 1994], [Ferrant et al., 1999], [Ferrant et al., 2000], [Ferrant, 2001], [Droske and Rumpf, 2004], [Tustison et al., 2006].

2.5 Some Examples of Elastic Approaches

To provide an example of a complete registration system based on the variational derivation, we briefly summarize two early approaches from this group.

The first use of the variational approach for registration of medical images is presented in [Broit, 1981]. In this work, a framework for registration of 2D and 3D images developed, which performs a global registration with an affine transformation, followed by a deformable registration. The deformable work is motivated by the intuition to perform a simulation of an elastic body using a linear elasticity model. To this end, an energy function corresponding to (1.4) is defined, which uses the cross-correlation criterion as the difference measure, and

the linear elasticity regularization term. The optimality criterion is derived as the Euler-Lagrange equation which corresponds to the Navier-Lamé equation

$$\underbrace{\mu\Delta u + (\lambda + \mu)\nabla(\operatorname{div}(u))}_{A_{\text{R}}u} = -f(u) . \quad (2.45)$$

The estimate of the transformation is computed by an iterative process, which is derived by the linearization of the force term by a quadratic approximation of the difference measure.⁴ The spatial discretization, which transforms the linear PDEs from the single iterations into linear equation systems, is performed by the finite difference scheme.

Simultaneously with the work in [Broit, 1981], the seminal method of Horn and Schunck for the estimation of optical flow (i.e. registration of 2D optical images) was presented [Horn and Schunck, 1980, Horn and Schunck, 1981], which also applies a variational approach. One specific of the original Horn and Schunck method was that only one iteration in time was performed. This was soon identified as a drawback and altered by subsequent works, cf. e.g. [Nagel, 1983]. Apart from this singularity, the method of Horn and Schunck corresponds to the standard variational approach as described above. It employs an energy consisting of the sum of squared differences, and the diffusion regularization term. The linearization step is performed by a quadratic approximation of the difference measure, which corresponds to the Gauss-Newton optimization method, cf. [Zikic et al., 2010e] for a more detailed description.

Since the early work on variational methods for image registration, this scheme has been commonly employed in the medical domain, for example in [Amit, 1994, Christensen, 1994, Bro-Nielsen, 1996, Modersitzki, 2004] just to name a few.

2.6 Discussion

The goal of this chapter was to provide an overview of the major concepts of the variational derivation scheme for registration algorithms. We hope that we also managed to highlight some connections between the variational derivation and the other groups of registration approaches, which will be discussed in the following chapters.

The group of variational elastic registration methods has been very popular in the field of medical image registration, but also in other fields where motion estimation is required, such as the optical flow problem. Its original advantages were that it enabled a simple global modeling of the problem on the energy level, while allowing efficient local estimation of the solution based on the corresponding Euler-Lagrange PDEs. The relation to PDEs which are commonly used to model physical problems allows to transfer the physical intuition directly to image registration methods.

While having many advantages, for certain problems in the field of medical imaging, such as computational anatomy applications, the elastic registration framework proved too restrictive since it is limited to comparably small deformations.

This limitation has a number of causes:

⁴Interestingly, the quadratic approximation of the force term makes this approach very related to the optical method Horn and Schunck [Horn and Schunck, 1981], and related to the force computation in the original demons method.

- The major conceptual reason why elastic approaches are not suitable for treatment of large deformations is that the internal regularization energy opposes large and complex deformations. Intuitively, it is not possible to deform an elastic material beyond a certain limit. While it is possible to recover large linear transformations given an appropriate boundary treatment and corresponding regularization, recovering variable and complex deformations which are small in some parts and large in others is not possible [Christensen, 1994]. Reducing the influence of regularization can help to a certain degree, but reportedly it is not possible to reduce the regularization sufficiently to recover the required deformations, without running into numerical problems.
- A further point which might prevent some implementations of elastic methods from recovering large and complex deformations, is that an appropriate update scheme must be employed, and the appropriate reference frame from continuum mechanics theory must be taken into account [Christensen, 1994]. For example, if the additive update scheme (which corresponds to the assumption that displacements are elements of a linear space) is employed together with the Eulerian reference frame, then the so called material derivative should be used to compute the updates. Otherwise, no re-orientation of the point-wise forces is performed during iterations, which might limit the capture range of the resulting methods. This issue can be solved by consistent treatment of the relevant concepts in the design of the registration algorithm, cf. e.g. [Christensen, 1994, Christensen et al., 1996].
- At last, variational elastic registration is susceptible to get trapped in locally optimal solutions. This is the case since the method is gradient-based, and it operates on a highly non-convex and high-dimensional space. In practice, this drawback can be alleviated to a certain extent by employing a scale-space focusing strategy.

The limitation of elastic approaches to small deformations led to the development of fluid registration methods or methods for large deformations, which is discussed in the next chapter.

LARGE DEFORMATION REGISTRATION

The line of work on registration methods capable of recovering large deformations was motivated by the limitation of elastic methods to small displacements. The reasons for this limitation were discussed in the previous chapter, with the major obstacle being posed by the internal regularization energy, which opposes large and complex deformations.

With respect to applications, methods from the large deformation framework are commonly used in the context of *computational anatomy* [Grenander and Miller, 2006]. The field of computational anatomy (CA) studies the variation of shapes in a certain population. This is done 1) by defining a representative average image of the population, and 2) by representing the shapes corresponding to the single images by the deformations which map these images onto the average image. The CA framework uses the notion of a *deformable template*, which describes the action of a group of diffeomorphisms on the fixed representative image. The comparison between the instances on the orbit of possible images can then be performed by analyzing the corresponding diffeomorphisms, which relate the particular instance to the representative image. For CA applications, inter-subject registration is required, which yields large and complex deformations, due to large anatomical differences between the subjects (e.g. for brain images). Since in most cases, no physically correct priors can be employed in these settings, the assumption of transformations being diffeomorphic is made in order to ensure the required regularity.

The initial work on registration methods for large deformations was presented in [Christensen, 1994, Christensen et al., 1996]. The idea in this work was to model the problem not as a simulation of an elastic body reacting to forces, but as a simulation of a *viscous fluid*. In contrast to elastic bodies, fluids do not accumulate an internal energy when deformed. In consequence, large deformations can be achieved. The viscosity property of the fluid assures that the single updates are sufficiently smooth, ensuring the required regularity.

The initial physical motivation of fluid simulation was soon generalized, as it was demonstrated that the fluid simulation can be interpreted as the optimization of the difference measure in the group of diffeomorphisms [Trounev, 1998, Dupuis et al., 1998]. To this end, a novel representation of deformations as so called *flows* was developed, which models deformations as a composition of a series of small update deformations, which correspond to the so called velocity fields. The

flow representation can be modeled as an ordinary differential equation (ODE).

As a further development, based on the parametrization of deformations as flows, a novel regularization term was proposed, which minimizes the geodesic length of a given deformation in the appropriate manifold [Trounev, 1995, Dupuis et al., 1998, Beg et al., 2005].

Furthermore, the work on the large deformation framework raised a number of interesting issues which were previously not focused on. These include the focus on obtaining diffeomorphic transformations, the consideration of the appropriate treatment of deformations in either the Eulerian or Lagrangian reference frame, and the related question of the appropriate update schemes.

In the following we discuss the different developments in the field of registration for large deformations. We start with the original fluid registration in Section 3.1. The parametrization of deformations as *flows*, i.e. composition of small deformations corresponding to the driving velocity fields, is discussed in Section 3.2, and we treat the regularization by the geodesic length of the deformation in Section 3.3. Finally, we will also briefly discuss the different approaches for the parametrization of flows via time-varying and stationary velocity fields in Section 3.4.

3.1 Fluid Registration

The initial work on registration for recovering large deformations in medical settings was presented in [Christensen, 1994, Christensen et al., 1996]. The approach was motivated by the goal to retrieve large deformations, and the observation that classic elastic variational methods were not able to achieve this. The idea followed in [Christensen et al., 1996] was to model the registration as the simulation of a viscous fluid deforming under a set of forces, which are derived from the difference measure. The fluid model was modified to be not mass conserving, thus allowing shrinking and expansion of imaged objects. The simulation of the viscous fluid was governed by the Navier-Stokes equation

$$\mu\Delta v + (\lambda + \mu)\nabla(\operatorname{div}(v)) = -f(u) . \quad (3.1)$$

The force term f was computed as the derivative of the SSD difference measure.

Furthermore, the point was addressed that the treatment of a spatial deformation requires an appropriate use of a spatial reference frame as discussed in the field of *continuum mechanics*. In [Christensen, 1994], the Eulerian reference frame is employed. In this case, the velocity cannot be simply computed as the partial derivative of the displacement, but the so called material derivative has to be employed, which yields the following formulation

$$v(x, t) = \frac{\partial u(x, t)}{\partial t} + \nabla u(x, t)v(x, t) . \quad (3.2)$$

So in summary, the following PDE system has to be solved

$$\mu\Delta v + (\lambda + \mu)\nabla(\operatorname{div}(v)) = -f(u) \quad (3.3)$$

$$\frac{\partial u(x, t)}{\partial t} = v(x, t) - \nabla u(x, t)v(x, t) . \quad (3.4)$$

This results in solving (3.1) in every iteration, followed by the Euler integration step

$$u(x, t_{i+1}) = u(x, t_i) + \tau (v(x, t) - \nabla u(x, t)v(x, t)) , \quad (3.5)$$

for the computation of the new displacement estimate.

In [Christensen, 1994, Christensen et al., 1996], it was observed that for large curved transformations, the estimates would become singular, and a regriding scheme was applied to avoid this problem. The regriding scheme was applied if the Jacobian of the transformation became smaller than a certain value (e.g. 0.5), i.e. $|J| < c$. In these cases, the current transformation was applied to the source image. Then, the problem was treated as a new one, with the warped source as the source image, and the current transformation set to identity. The final result is the concatenation of all the single transformations, which were computed in the course of the algorithm. This scheme results in diffeomorphic deformations, however, the computation of the Jacobian in every step is costly.

It is interesting to note that the Navier-Stokes equation in (3.1) is analogue to the Navier-Lamé equation for linear elasticity, cf. Eq. (2.45), with the difference, that it is operating on the velocity v instead of the displacement u . This observation has lead to a generalization of the concept of fluid methods. Any linear regularization term can in principle be applied to the updates instead of the complete displacement field. Since in this way, no internal energy is accumulated, the resulting approaches are able of recovering large deformations and are classified as fluid.

The fluid registration algorithm proposed in [Christensen, 1994] had the disadvantage to be comparably slow at the time of publication, requiring massive parallel hardware for acceptable runtimes. This property caused the development of faster alternatives. One of these alternatives was the demons method by [Thirion, 1996], which we will discuss in Chapter 4. A second alternative was the approach presented in [Bro-Nielsen and Gramkow, 1996], which follows the same derivation as [Christensen, 1994], but provides a faster numerical method for solving the arising PDEs. The proposal made in [Bro-Nielsen and Gramkow, 1996, Bro-Nielsen, 1996] was to solve the linear PDE in each iteration by convolving the force term by the Green's function to the corresponding Navier-Lamé operator. Also, in [Bro-Nielsen and Gramkow, 1996], a relation of the fluid method to the fluid version of the demons approach is discussed.

3.2 Parametrization of Deformations as Flows

The description of diffeomorphic deformations as flows was proposed in [Trouvé, 1995, Trouvé, 1998, Dupuis et al., 1998]. It presents a generalization of the viscous fluid algorithm. For the parametrization as a *flow*, the transformation is treated as an element of a group of diffeomorphisms \mathcal{G} . The diffeomorphism ϕ is encoded as the solution of the ordinary differential equation (ODE)

$$\frac{\partial \Phi(t)}{\partial t} = v(t) \circ \Phi(t) , \quad (3.6)$$

for $t = 1$, that is $\phi = \Phi(1)$, for a given time-dependent velocity field $v(t)$, and with the initial deformation $\Phi(0) = \text{Id}$. This representation employs the fact

that the group of diffeomorphisms is a Lie group, that is, it has the structure of a differentiable Riemannian manifold. The velocity field $v(t) \circ \Phi(t)$ is defined as an element of the Lie algebra, that is, the corresponding tangent space to the Riemannian manifold at $\Phi(t)$. It has been shown in [Trouvé, 1995, Dupuis et al., 1998], that if the tangent spaces are chosen as sufficiently smooth Sobolev spaces, the resulting transformation ϕ is diffeomorphic.

A possible intuition to (3.6) is to encode the deformation not directly, but to express it as a sequence of accumulated small updates [Trouvé, 1998]. These updates correspond to the velocity fields $v(t)$. For sufficiently small update displacements $v(t)$, their norm can be seen as the first-order approximation to the distance between Id and $\Phi(t) = \text{Id} + v(t)$. This approximation is meaningful if $v(t)$ is sufficiently small, and in this case, $\Phi(t)$ is also invertible. This concept is extended to larger deformations, by interpreting ϕ as a concatenation of small deformations $\phi(t_i) = \text{Id} + v(t_i)$, where the time t_i is assumed to be discretized. The transformation can then be defined as $\phi = \Phi(t_n)$, with $\Phi(t_0) = \text{Id}$ and

$$\Phi(t_{k+1}) = \phi(t_{k+i}) \circ \Phi(t_k) . \quad (3.7)$$

The sequence $\Phi(t_k)$ represents a path from $\text{Id} = \Phi(t_0)$ to $\phi = \Phi(t_n)$, which is generated by the update displacements $v(t_0), \dots, v(t_n)$. Please note that there are infinitely many such sequences, which result in the transformation $\phi = \Phi(t_n)$.

The sequence $\Phi(t)$ denotes a continuous series of updates, which develops over an undiscretized time, starting at $t = 0$ with $\Phi(0) = \text{Id}$, and evolving until $t = 1$, when it becomes identical to the transformation ϕ , so that $\phi = \Phi(1)$. With this definition, the continuous version of the definition for the generating sequence Φ , which corresponds to the discrete version in (3.7), is given by the ODE

$$\frac{\partial \Phi(t)}{\partial t} = v(t) \circ \Phi(t) , \quad (3.8)$$

with $\Phi(0) = \text{Id}$. The generating sequences $\Phi(t)$ are now time-dependent functions solving (3.8).

The question remains, how the velocity fields are defined in the case of registration. This is done by setting the flow equation from (3.6) equal to the negative gradient of a difference measure E_D , so that we have

$$\begin{aligned} \frac{\partial \Phi(t)}{\partial t} &= v(t) \circ \Phi(t) \\ v(t) &= -\nabla_{\mathbb{H}} E_D(\Phi(t)) \end{aligned} . \quad (3.9)$$

We have here $\nabla_{\mathbb{H}} E_D = A^{-1} \nabla_{L^2} E_D$, where A is the differential operator defining the Sobolev space by $\langle \cdot, \cdot \rangle_{\mathbb{H}} = \langle A \cdot, \cdot \rangle_{L^2}$. Often, SSD is employed as the difference measure [Trouvé, 1998, Beg et al., 2005], however, other measures are applicable as well, compare for example [Chefd'hotel et al., 2002, Lorenzen et al., 2006, Avants et al., 2008].

As already mentioned, the solution to the ODE from (3.6) for v is not unique, since the same final diffeomorphism $\phi = \Phi(1)$ can be reached by several flows described by v . This is a negative property for computational anatomy applications, since a single image is supposed to be represented by one unique diffeomorphism mapping it to the average image. To this end, a regularization term was proposed, which penalizes the geodesic length of the flow generating

the diffeomorphism. This term provides a remedy to the problem of uniqueness, by distinguishing the “shortest” flow, while at the same time not opposing large deformations.

3.3 Regularization by Minimization of Geodesic Length

This regularization scheme [Trouvé, 1995, Trouvé, 1998, Dupuis et al., 1998, Beg et al., 2005] assumes that the transformations are elements of the group of diffeomorphisms \mathcal{G} , and penalizes the *geodesic distance* $d_{\mathcal{G}}(\text{Id}, \phi)$ between the identity transformation Id , and the resulting deformation ϕ , in \mathcal{G} by

$$E_{\text{R}} = \frac{1}{2} d_{\mathcal{G}}(\text{Id}, \phi)^2 . \quad (3.10)$$

The question is now, how is the distance $d_{\mathcal{G}}$ defined?

Again, we can follow the finite dimensional intuition from [Trouvé, 1998], and then make the transition to the infinite-dimensional setting. The distance between Id and ϕ can then be approximated by the length of the shortest generating sequence $\Phi(t_n)$, which results in the transformation ϕ , that is $\Phi(t_n) = \phi$. The length of the sequence $\Phi(t_n)$ is given by the sum of the generating displacements $v(t_i)$, so that we get

$$d_{\mathcal{G}}(\text{Id}, \phi) \approx \min_{\Phi(t_n)} \sum_{i=0}^n \|v(t_i)\|_{\mathbb{H}} . \quad (3.11)$$

By making the step to infinitely small update displacements, we arrive at the continuous and non-approximative formulation

$$d_{\mathcal{G}}(\text{Id}, \phi) = \inf_{\Phi: \Phi(1)=\phi} l(\Phi) \quad \text{with} \quad l(\Phi) = \int_0^1 \|v(t)\|_{\mathbb{H}} dt , \quad (3.12)$$

where the updates $v(t)$ can be interpreted as the velocity at time t .

The combination of a difference measure (originally SSD) and the geodesic length regularization term yields the energy term which has been originally studied in [Trouvé, 1995, Trouvé, 1998, Dupuis et al., 1998], and became popular under the name *Large Deformations Diffeomorphic Metric Mapping (LDDMM)* [Beg et al., 2005]. It can be stated as the minimization of the energy

$$\min_{v: \Phi(1)=\phi} \int_0^1 \|v(t)\|_{\mathbb{H}}^2 dt + \|I_{\text{T}} - I_{\text{S}}(\phi)\|_{\text{L}^2}^2 . \quad (3.13)$$

3.4 Time-Varying and Stationary Velocity Fields

Over the last years, two major flavors of the parametrization of diffeomorphisms as flows have been established [Hernandez et al., 2009]. The first follows the original derivation and assumes that the velocity fields are time-dependent. The second group of approaches assumes a single stationary velocity field. While stationary velocity fields restrict the set of possible diffeomorphisms which can be expressed, these approaches have the advantage to be more efficient, and the resulting parametrization can be used for linear statistics on diffeomorphisms in computational anatomy applications.

3.4.1 Non-stationary Velocity Fields

The original approach of using the flow formulation from (3.6) employs a time-dependent velocity function, so that we have $v : \Omega \times t \rightarrow \Omega$, where a single velocity field $v(t)$ is an element of an appropriate Sobolev space \mathbb{H} [Trouvé, 1995, Dupuis et al., 1998].

The challenge posed by this formulation is the increase in required computational resources which become necessary since the problem is extended by a temporal dimension. As an example, if the temporal discretization of the flow formulation is performed with n discrete time steps, then n velocity fields are required to represent one diffeomorphism. In comparison, the representation of ϕ as $\phi = \text{Id} + u$ requires only one displacement field and has correspondingly less parameters. The complexity of the associated optimization problem rises accordingly. For large 3D data sets, which often arise in medical applications, this additional computational burden can be significant.

Because of the complexity of the resulting optimization problem, in [Trouvé, 1995, Trouvé, 1998], the assumption is made that numerical computation of the LDDMM model in (3.13) is not practical, and a sub-optimal algorithm is proposed, which performs gradient descent in the group of diffeomorphisms, defined by the tangent space, which is chosen to be a Sobolev space of certain smoothness. It is discussed in [Trouvé, 1998] that for the appropriate choice of the Sobolev space, this corresponds to the fluid method from [Christensen et al., 1996]. The same strategy is also pursued in [Chefd'hotel et al., 2002]. Starting with $\phi(t_0) = \text{Id}$, these approaches basically evolve the flow system from (3.9) until convergence. This approach avoids the necessity of storing all intermediate velocity fields, since for the evolution it is sufficient to represent only the most recent transformation estimate $\Phi(t_k)$.

On the other hand, in [Beg et al., 2005], a method is developed which actually performs the optimization of the geodesic distance term from (3.13). In this approach, all temporal velocity fields are used, and they are optimized by a gradient descent, such that the geodesic regularization term is minimized. Starting with the initial estimate of $\phi = \text{Id}$, and the corresponding parametrization $v_i = 0$, the single velocity fields are modified, such that the energy in (3.13) is optimized.

3.4.2 Stationary Velocity Fields

As an alternative to the time-varying representation, parametrization of flows with stationary velocity fields have been proposed. The major difference is that in this case, the diffeomorphisms is encoded by an ODE which employs a static velocity field, that is

$$\frac{\partial \Phi(t)}{\partial t} = v \circ \Phi(t) . \quad (3.14)$$

Please note that in contrast to (3.8), the velocity v does not depend on the time in (3.14). This stationary parametrization cannot represent all possible diffeomorphisms, but only a one-parameter subgroup. However, the expressiveness of the model was demonstrated to be sufficient for registration [Vercauteren et al., 2008, Ashburner, 2007] and seems to provide similar performance as the time-varying velocity representation. This approach substantially reduces the

computational requirements compared to the parametrization by time-varying velocity fields.

It is important to note that in (3.14), the diffeomorphism ϕ , which is an element of a non-linear manifold, is represented by a single velocity field v , which is an element of the linear tangent space. This makes it possible to perform the computation of deformation statistics in a linear space, which greatly facilitates the process.

There are two major approaches for the representation by stationary velocity fields. The first approach is the so called *geodesic shooting* strategy, which is based on the idea of the conservation of momentum [Vaillant et al., 2004, Miller et al., 2006]. The second approach is based on the Log-Euclidean framework proposed in [Arsigny et al., 2006]. The Log-Euclidean framework is employed in different ways in [Ashburner, 2007], [Hernandez et al., 2007], and in the context of demons approaches in [Vercauteren et al., 2007b, Vercauteren et al., 2008, Vercauteren et al., 2009].

Overviews of and further details about the approaches for stationary velocity field parametrization of flows are provided in [Hernandez et al., 2008, Hernandez et al., 2009, Younes et al., 2009].

3.5 Discussion

The methods for large deformation registration are not precisely a set a specific methods but rather a wide category. This category encompasses different methods such as the original fluid registration, different variations of the LDDMM approach including the greedy implementation [Trouvé, 1998], parametrization with time-varying velocity fields [Beg et al., 2005], and stationary velocity fields [Ashburner, 2007].

As a further example, the fluid-type demons method, which is discussed in the next chapter, is able to recover large deformations, so it could be also presented in the context of large deformation registration methods. Actually, it has been demonstrated, that the fluid-type demons method corresponds to the greedy implementation of the LDDMM approach as performed in [Trouvé, 1998, Chéfd'hotel et al., 2002], and thus also to the fluid registration approach, cf. also [Bro-Nielsen and Gramkow, 1996].

The category of methods for large deformation registration is of great importance for the field of computational anatomy, and is currently a very active subject of research.

DEMONS REGISTRATION APPROACHES

The so called demons method is an approach for deformable registration originally proposed in [Thirion, 1995a, Thirion, 1995b, Thirion, 1996, Thirion, 1998]. The original motivation for the approach was to provide an efficient method for deformable registration, which was in contrast to the available methods at the time, most of which were based on simulation of physical processes (such as elastic, or fluid methods), which were computationally expensive. Due to its efficiency and simplicity of implementation, the demons method has become a popular choice in numerous applications. Originally, in [Thirion, 1998], not a single method, but a general registration framework was proposed, which is quite similar to the generic algorithm in (1.9), and proposes to perform the registration by computing a set of forces, and smoothing the resulting updated displacement field in each iteration. The framework considers a variety of possible transformation models and different ways of computing the forces, resulting in a number of variants. The first of these variants, “Demons 1: a complete grid of demons” emerged as the most popular of the alternatives and has been established as *the* original demons method.

In its original form, the demons algorithm iteratively estimates a transformation $\phi = \text{Id} + u$ in the following way: In each iteration, the current displacement estimate u_i is updated by a set of forces f , and subsequently the regularization is performed by Gaussian filtering of the updated field

$$f = \frac{(I_T - I_S(\phi)) \nabla I_S(\phi)}{(I_T - I_S(\phi))^2 + \|\nabla I_S(\phi)\|^2} \quad (4.1)$$

$$v = \tau f \quad (4.2)$$

$$u_{i+1} = G_\sigma * (u_i + v) . \quad (4.3)$$

In the original publication, the scaling with the step size in (4.2) is not performed, i.e. we have $\tau = 1$.

The efficiency of the original demons approach is caused by two factors. The first factor concerns the regularization strategy. Regularization by Gaussian filtering in (4.3) features a simple implementation, it is numerically robust so that it allows large time steps, and finally – at the time of original publication – the computation was significantly faster than solving PDE systems arising from other approaches [Bajcsy and Kovačič, 1989, Christensen et al., 1996].

The second, more subtle factor concerns the structure of the force term in (4.1). The analysis in [Cachier et al., 1999, Pennec et al., 1999, Vercauteren et al., 2009] demonstrates that the original force term (4.1) represents an approximation to an update step of a second-order - and thus efficient - optimization of the sum of squared differences (SSD) as the difference measure. This improves the convergence speed of the approach, so that a comparably small number of iterations is required to achieve good results.

Originally, the demons approach was heuristic in nature. However, since its introduction, several interpretations have been proposed which provide a theoretically sound derivation of the algorithm. These interpretations include the formulation of the demons method as an alternating optimization of an elastic energy, and the demonstration that a certain time discretization scheme is equivalent to the demons regularization strategy by smoothing. In the following sections, we discuss these interpretations, as well as the generalizations and extensions of the demons approach, which have been proposed in the literature. These include the generalization to a fluid-type method, in order to enable the computation of large deformations, the extensions to arbitrary difference measures and regularization terms, as well as modifications assuring that the resulting transformations are diffeomorphic.

4.1 Interpretation as Alternating Optimization

A useful interpretation [Cachier et al., 2003], which ties in the demons approach with the variational optimization framework in (1.4) is to see it as optimization of the energy

$$E = E_D(I_T, I_S \circ \phi_1) + \rho \|\phi_1 - \phi_2\|^2 + \lambda E_R(\phi_2) , \quad (4.4)$$

by alternating optimization of the expressions containing the difference measure E_D and the regularization term E_R

$$\phi_1 = \arg \min_{\phi_1^*} [E_D(I_T, I_S \circ \phi_1^*) + \rho \|\phi_1^* - \phi_2\|^2] \quad (4.5)$$

$$\phi_2 = \arg \min_{\phi_2^*} [\lambda E_R(\phi_2^*) + \rho \|\phi_1 - \phi_2^*\|^2] . \quad (4.6)$$

Within this interpretation, the computation of the forces in (4.1) corresponds to a step of the optimization of (4.5), and the regularization step in (4.3) corresponds to one step of (4.6).

4.2 Interpretation as Time Discretization Scheme

In [Pennec et al., 1999], it is argued that the Gaussian smoothing can be considered as a rough simulation of an elastic regularization model. A theoretically founded interpretation of Gaussian smoothing as a time discretization scheme for a variational approach is provided in [Cahill et al., 2009, Cahill, 2009]. We discussed this interpretation in Section 2.3.3. With this interpretation, the demons method can be seamlessly interpreted in the classical elastic variational framework. In Section 2.3.3, we provide a detailed derivation of this argument, including its extension to arbitrary regularization terms besides diffusion.

4.3 Generalization of Difference Measures

One limitation of the original demons approach was that it was based on the assumption of identical intensities in the input images - thus implicitly using SSD as the difference measure - which limits it to mono-modal scenarios.

The relation of the original force computation from (4.1) to the SSD becomes clearer by rewriting (4.1) to

$$f = \frac{(I_T - I_S(\phi)) \nabla I_S(\phi)}{(I_T - I_S(\phi))^2 + \|\nabla I_S(\phi)\|^2} = \frac{-\nabla E_{\text{SSD}}(\phi)}{(I_T - I_S(\phi))^2 + \|\nabla I_S(\phi)\|^2}, \quad (4.7)$$

This reveals that the original force computation can be seen as a modification of the gradient of the SSD. As already mentioned, it was demonstrated that (4.7) corresponds to an update step of an approximation of a second-order optimization scheme on the SSD [Pennec et al., 1999]. Please note that (4.7) does not lead to meaningful forces for other measures than SSD.

Soon after the initial proposal of the demons method, this approach has been generalized to other difference measures, cf. e.g. [Chefd'hotel et al., 2002, Guimond et al., 2002, Cachier et al., 2003, Modat et al., 2010]. Most of these approaches perform this generalization simply by computing the forces by employing the steepest descent scheme, that is

$$f = -\nabla E_D(\phi). \quad (4.8)$$

Please note however, that the force term in (4.8) corresponds to employing the steepest gradient descent optimization of the difference measure. This is known to have a slow convergence rate on ill-conditioned optimization problems, requiring a large number of iterations to achieve satisfying accuracy. An exception to (4.8) is discussed in [Modat et al., 2010], where the forces are computed based on the update of the NL-CG method.

In order to provide an efficient formulation of the forces for arbitrary difference measures, in [Zikic et al., 2011] we propose a general preconditioning scheme, which can be applied to any image-based difference measures, leading to improved convergence speed. We will discuss this approach in more detail in Chapter 6.

Finally, we note that – according to our observation – the actual application of the demons framework in multi-modal settings is still rather rare, compared to the mono-modal settings based on SSD and the original force formulation.

4.4 Fluid Demons

In the original demons framework, in each iteration the Gaussian smoothing is applied to the complete displacement field after the incremental update by the force term, cf. (4.3). An alternative option is to apply the smoothing only the the force term, before the update is applied. This would result in the following demons approach (assuming steepest descent for force computation)

$$f = -\nabla E_D \quad (4.9)$$

$$v = \tau G_{\sigma_{\text{fl}}} * f \quad (4.10)$$

$$u_{i+1} = u_i + v. \quad (4.11)$$

The above version of the demons algorithm was considered by [Bro-Nielsen and Gramkow, 1996], briefly after the original introduction of the demons method by [Thirion, 1995a]. Furthermore, [Bro-Nielsen and Gramkow, 1996] established the relation of this demons version to the viscous-fluid approach of [Christensen, 1994]. For this reason, this variant of the demons approach, in which only the update is regularized is referred to as *fluid*. In [Bro-Nielsen and Gramkow, 1996] it is proposed to solve the Navier-Lamé equations which arise in the fluid approach by applying the appropriate Green’s function. It is then further noted that the Gaussian smoothing, which is applied in the demons method, can be seen as a rough approximation of the Green’s function to the Navier-Lamé operator. This insight is developed further in [Trouvé, 1998] and [Chefd’hotel et al., 2002, Chefd’hotel, 2005], where it is noted that a process equivalent to fluid demons, with regularization performed by smoothing of the forces, corresponds to a steepest descent in an appropriate Sobolev space. The Sobolev space can be based on different linear operators, or alternatively, corresponding Green’s functions. This point of view makes it possible to generalize the regularization of updates in the fluid demons approach to any linear, and spatially constant regularization term, thus replacing the smoothing in (4.10) from $v = \tau G_{\sigma_{\text{fl}}} * f$ by

$$v = \tau \Psi_{\mathbb{H}} * f , \quad (4.12)$$

where the smoothing kernel depends on the employed space \mathbb{H} .

As a further extension, a combination of the elastic and fluid schemes for the demons approach has been discussed for example in [Pennec et al., 1999], [Cachier et al., 2003], or [Stefanescu et al., 2004].

4.5 Generalization of Regularization

Fluid approaches differ from elastic-type methods in the point that no regularization term is employed. As discussed in Chapter 3, the regularity in these approach comes from a choice of a suitable Sobolev space. The choice of the Sobolev space influences the kernel with which the forces are convolved in Eq. (4.12). The selection of the Sobolev space which leads to the Gaussian kernel is discussed in [Chefd’hotel et al., 2002]. A variety of other choices is available, cf. e.g. [Bro-Nielsen and Gramkow, 1996, Trouvé, 1998, Chefd’hotel et al., 2002].

The interpretations in Sections 4.1 and 4.2 demonstrate that the elastic-type demons method can be seen as a particular numerical scheme for the optimization of an energy containing a regularization term, and that the original demons approach corresponds to employing the diffusion regularization term.

The interpretation of demons as a specific time discretization scheme from Section 4.2 presents a convenient approach to generalizing elastic-type demons to other regularization terms. One way of performing this step would consist of computing the appropriate Green’s functions for the linear operators resulting from the different regularization terms. This approach would be very much in line with the work from [Bro-Nielsen and Gramkow, 1996] for fluid approaches. Although not presented as a demons method, the work in [Beuthien et al., 2010] discusses implementing the elastic-type registration schemes, by convolution of the right-hand side with the Green’s function corresponding to the linear operator, which is based on the employed regularization term. Thus this work can

be seen as a demons method, and thus presents a generalization of elastic-type approaches to arbitrary linear and spatially constant regularization terms.

A slightly different approach is followed by [Cahill, 2009], where curvature, and elastic regularization are implemented by a scheme which employs successive convolutions by a Gaussian.

A further possible option for a modification of a regularization term is to make its influence spatially varying. One such approach is discussed in the context of demons in [Stefanescu et al., 2003, Stefanescu et al., 2004], where a combination of a fluid and an elastic strategy is employed. The regularization strength in the fluid part is determined by the confidence about the underlying force estimates, basically leading to suppression of forces in homogeneous image areas, while the strength of the elastic regularization is determined by a heuristic simulation of the stiffness of the structures depicted in the input images. Interestingly, this implementation of the demons approach does not perform the regularization steps by convolution, something which is very untypical for methods usually referred to as demons. The reason for this numerical approach is that the convolution becomes inefficient, if a Gaussian of non-constant width is employed. Instead, an inhomogeneous diffusion PDE is solved, as discussed in [Weickert, 1998, Weickert and Schnörr, 2001].

4.6 Update Schemes and Diffeomorphic Demons

A further extension of the demons method considers the update scheme. In contrast to the assumption that the displacements are elements of a linear vector space, it has been discussed in a series of works that a more appropriate model is to treat the transformations as elements of a group [Trouvé, 1998], [Miller and Younes, 2001], [Chefd'hotel et al., 2002]. If the displacement is treated as an element of a linear vector space, then the addition of updates as in (4.3) is the appropriate operation. On the other hand, if the transformations are treated as a group, then the composition of two transformations is the appropriate operation, which should ensure that the result is still an element of the group. So, in the demons framework, the update scheme in (4.3) (or in 4.11 for the fluid case), can be replaced by a composition, i.e., instead of $u_{i+1} = u_i + v$, we get

$$u_{i+1} = u_i \circ (\text{Id} + v) . \quad (4.13)$$

This strategy has been employed for example in [Stefanescu et al., 2004], [Vercauteren et al., 2007b, Vercauteren et al., 2009]. In [Vercauteren et al., 2009], an increase in accuracy for the compositional scheme is reported.

Furthermore, the choice of the update scheme is an important question, if it is desired that the demons method generates transformations which are diffeomorphic. While there are several possible ways to achieve this goal, they all employ the compositional scheme, since in the infinite-dimensional setting, diffeomorphisms form a group, with the composition as the appropriate operator, and the composition of two diffeomorphisms yields a diffeomorphism.

One strategy for generating diffeomorphisms is to employ the compositional scheme with sufficiently small updates [Chefd'hotel et al., 2002, Stefanescu et al., 2004]. This approach uses the fact that in the finite-dimensional case, which must be ultimately employed in the implementation, transformations resulting from sufficiently small and smooth updates are diffeomorphic. The drawback of this

approach is that it limits the maximal length of the point-wise updates, which can slow down the convergence speed of the method. Due to our experience, this limitation is not too severe in practical applications. To assure diffeomorphisms with sufficiently smooth updates in a discrete setting, one must limit the maximal point-wise length of the update, i.e. $\max \|v(x)\|$, to half the element spacing, which is still an acceptable and meaningful value.

An alternative approach is presented in [Vercauteren et al., 2007b], where the Log-Euclidean framework from [Arsigny et al., 2006] uses stationary velocity fields to ensure that the updates are diffeomorphic.¹ In this approach, the computed updates are assumed to lie in the tangential space to the Riemannian manifold of diffeomorphisms. Then, based on the velocity field (i.e. the update step) v , the corresponding diffeomorphic version is computed by applying the exponential operator, which maps from the tangential space to the manifold of diffeomorphisms, resulting in the update scheme

$$u_{i+1} = u_i \circ \exp(v) . \quad (4.14)$$

A further work based on the Log-Euclidean framework is presented in [Vercauteren et al., 2008]. Instead of encoding only the updates by the static velocity fields, in this work, the resulting deformation at every iteration is represented by a static velocity. This is advantageous for studies in computational anatomy. More precisely, given the current deformation ϕ_i , represented by the static velocity field u as $\phi_u = \exp(u)$, and the update transformation $\phi_v = \exp(v)$, the task is to compute the static velocity field $w = Z(u, v)$, which represents the diffeomorphism $\phi_{i+1} = \exp(w)$, so that

$$\exp(w) = \exp(u) \circ \exp(v) . \quad (4.15)$$

The static field w is approximated from u and v by $w \approx Z(u, v)$, where Z is an approximation to the Baker-Campbell-Hausdorff (BCH) formula. This approximation holds only for sufficiently small updates v . The use of the BCH formula in this context was proposed in [Bossa et al., 2007].

Please note that in all of the above approaches, only the update step is modified, and the remainder of the standard demons framework stays intact.

4.7 General Demons Framework

Summing up the generalization from previous sections, the following general demons framework can be stated:

$$f = -\gamma_{\text{opt}} (\nabla E_{\text{D}}(\phi)) \quad (4.16)$$

$$v = \tau \Psi_{\mathbb{H}} * f \quad (4.17)$$

$$w = \text{update}(u_i, v) \quad (4.18)$$

$$u_{i+1} = \Psi_{\mathbb{R}}(\tau) * w . \quad (4.19)$$

¹Although the use of static velocities in the Log-Euclidean framework is not capable of representing all possible diffeomorphisms, it has been demonstrated that it is sufficiently expressive for registration purposes.

4.8 Further Extensions

Over the last years, numerous further extensions of the demons framework have also been proposed.

One popular topic is the symmetrization of forces for the case of SSD as the difference measure. In [Wang et al., 2005] and [Rogelj and Kovacic, 2006], it has been proposed to compute the force term based on the average of image gradients from both, the source and the target image. In [Vercauteren et al., 2007a, Vercauteren et al., 2009], a link between these approaches and the efficient second order optimization (ESM) scheme [Malis, 2004] is established.

Further extensions include a scheme aiming at inverse-consistency [Vercauteren et al., 2008], and using the demons method for registration on spherical surfaces in [Yeo et al., 2010].

4.9 Discussion

Since its advent as a heuristic method, the demons approach has been provided a solid theoretical foundation (in fact several interpretations have been provided), and generalized from a specific method to a general and powerful framework. It can be seen as an efficient numerical method for implementing elastic-type and fluid-type registration models, for arbitrary difference measures, regularization terms, and different smoothness assumptions for the underlying groups of transformations.

One open challenge in the demons framework is the practical and theoretically justified use of efficient schemes for the optimization of statistical difference measures such as mutual information in multi-modal scenarios. We address this point by proposing a general preconditioning scheme in Chapter 6.2, which enables an efficient optimization and is theoretically justified in the demons framework.

PARAMETRIC REGISTRATION METHODS

The main idea shared by the approaches usually grouped as parametric registration methods is to employ transformation models with a lower number of parameters, than the dense and high-dimensional parametrizations, which are resulting for example from demons algorithms, or from discretization of variational approaches by finite difference schemes.

The use of lower-dimensional models automatically provides increased robustness due to the inherent regularization. Furthermore, these models can simplify or even in the first place practically enable the use of some optimization methods for statistical difference measures, such as Gauss-Newton optimization for mutual information, which is not practicable in variational settings, with dense, high-dimensional transformation models [Zikic et al., 2010a, Zikic et al., 2011].

Starting from the energy minimization model

$$\min_{u \in \mathbb{H}} E(u) = \min_{u \in \mathbb{H}} E_D(u) + \lambda E_R(u) , \quad (5.1)$$

parametric methods employ a finite-dimensional approximation $u(p)$ to the displacement u . This approximation is governed by a finite number of parameters, which are combined in the parameter vector $p = (p_1, \dots, p_n)^\top$. The approximation $u(p)$ is then substituted into the energy formulation, thus transforming it into an optimization problem with respect to the parameters

$$p' = \arg \min_p E(u(p)) = \arg \min_p E_D(u(p)) + \lambda E_R(u(p)) . \quad (5.2)$$

As the result of the parametrization, the possible deformations are restricted to a subset of the space of all possible deformations. The first aspect of this restriction is the positive aspect that the subspace of parametrized transformations is always more regular than arbitrary transformations (at least for meaningful transformations models employed for registration). The second aspect is whether the restricted space is sufficiently expressive. This of course depends on the registration task at hand, however it has been demonstrated that models such as FFDs are well capable of estimating the transformations in deformable medical scenarios (cf. e.g. [Murphy et al., 2011]), and can be also used for optical flow problems, which require some strong spatial variation (cf. e.g. [Glocker et al.,

2008b]). Please note that rigid registration methods naturally fall within the class parametrized approaches.

In image registration, linear transformation models are by far the most popular choice for modeling $u(p)$. They are conceptually simple, and their linearity with respect to the parameters is an extremely convenient property for derivation of gradient-based optimization algorithms.

Linear parametric models represent the displacement as a linear combination of arbitrary basis functions $B_k : \Omega \rightarrow \Omega$, resulting in

$$u(p)(x) = \sum_k p_k B_k(x) . \quad (5.3)$$

The parameters $p_k \in \mathbb{R}^d$ can be seen as representative displacement vectors. In many cases, the influence of the parameter vector p_k can be assigned to a specific location in space. We refer to this location as the control point $c_k \in \Omega$.

With a matrix B , obtained the stacking discretized (by sampling on a regular grid corresponding to the image domain), and vectorized versions of the basis functions B_k as $B = [B_1, \dots, B_N]^\top$, and with the parameter vector p containing the single entries p_k , the definition in (5.3) can also be written as

$$u(p) = p^\top B . \quad (5.4)$$

This notation further highlights the linearity of the model.

It is interesting to note that the essence of the parametric approach is very similar to the finite element approach, which we discussed in the context of variational approaches, cf. Section 2.4.2. The difference in the derivation is that in the variational approach, the FE scheme is applied after the linearization of the problem, so that the linear transformation model is inserted into the quadratic energy, which corresponds to the linear PDE. On the other hand, in the parametric approach, the transformation model is inserted directly into the non-linear equation, and then, the optimization is carried out. In general, any FE model can be used for parametric registration, and vice versa. In contrast to the general application of FE, for image registration mostly conceptually simpler transformation models with regular mesh grids are employed.

Furthermore, in order to relate the parametric to variational approaches we find it instructive to analyze the derivative of the energy with respect to the parameters, and its connection with the continuous derivative of the energy with respect to the continuous displacement function, as defined in (2.8). With the parametrization from (5.4), the derivative energy derivative w.r.t. the parameters reads

$$\frac{\partial E}{\partial p} = \frac{\partial E}{\partial u} \frac{\partial u}{\partial p} \quad (5.5)$$

$$= \nabla E^\top B . \quad (5.6)$$

The above reveals that the derivative of the energy w.r.t. a parameter p_i is simply the projection of the dense continuous derivative ∇E onto the corresponding basis function B_i , i.e.

$$\frac{\partial E}{\partial p_i} = \nabla E^\top B_i . \quad (5.7)$$

Analogous to the description of deformable registration as accumulation of smoothed energy gradients, we can intuitively formulate the process of estimating the corresponding parameters as the accumulation of the projections of the energy gradient. In Section 5.3, we will briefly discuss how the form from (5.6) can provide a further way to relate variational approaches to the methods resulting from the parametric derivation.

In the field of intensity-based deformable registration of medical images, parametric registration has been first used by [Rueckert et al., 1998, Rueckert et al., 1999a], where the B-Spline based Free-form deformation (FFD) was employed as the transformation model. Since this work, B-Spline FFDs have become a very popular choice and have been commonly employed, cf. e.g. [Schnabel et al., 2001, Rohlfing and Maurer, 2003, Rohlfing et al., 2003, Glocker et al., 2008a, Klein et al., 2010b]. B-Splines have also been employed for image registration in other settings, cf. e.g. [Szeliski and Shum, 1996, Szeliski and Coughlan, 1997]. Other examples of parametric transformations in the field of registration include radial basis functions [Rohde et al., 2003], wavelets [Yoshida, 1998, Wu et al., 2000], or trigonometric functions [Ashburner and Friston, 1999, Ashburner and Friston, 2004].

The joint property of the most common transformation parametrizations for image registration is that they all employ a linear model. Therefore, the definition of the transformations can be boiled down to a description of the basis functions and the corresponding parameters. In Section 5.1, we perform a characterization of linear transformation models, and discuss some examples in Section 5.2.

5.1 Characterization of Linear Transformation Models

The standard linear transformation models can be characterized based on two sets of basic properties. The first set describes the spatial configuration of control points, and can be summarized by the following attributes.

Spatial Configuration of Control Points:

- *Localization* describes whether the influence of the parameters through basis functions is *localized* at certain spatial positions (e.g. Free-form Deformations (FFD)), or whether it is *global* (e.g. Trigonometric basis functions such as Discrete Cosine Transform (DCT) or Discrete Sine Transform (DST)).
- *Regularity*. This attribute applies only to localized models and describes whether the location of the control points corresponds to a *regular grid* or a mesh (e.g. Free-form Deformations), or whether the distribution of the control points can be arbitrary (e.g. Thin-plate splines (TPS), Adaptive bases), resulting in *irregular* control point distributions. Regular grids have the advantage of conceptual, and implementational simplicity. On the other hand, irregular approaches allow for adaptive resolution in the image domain, such that a finer resolution can be used to capture complex motions only in the areas where required, and lower resolution in

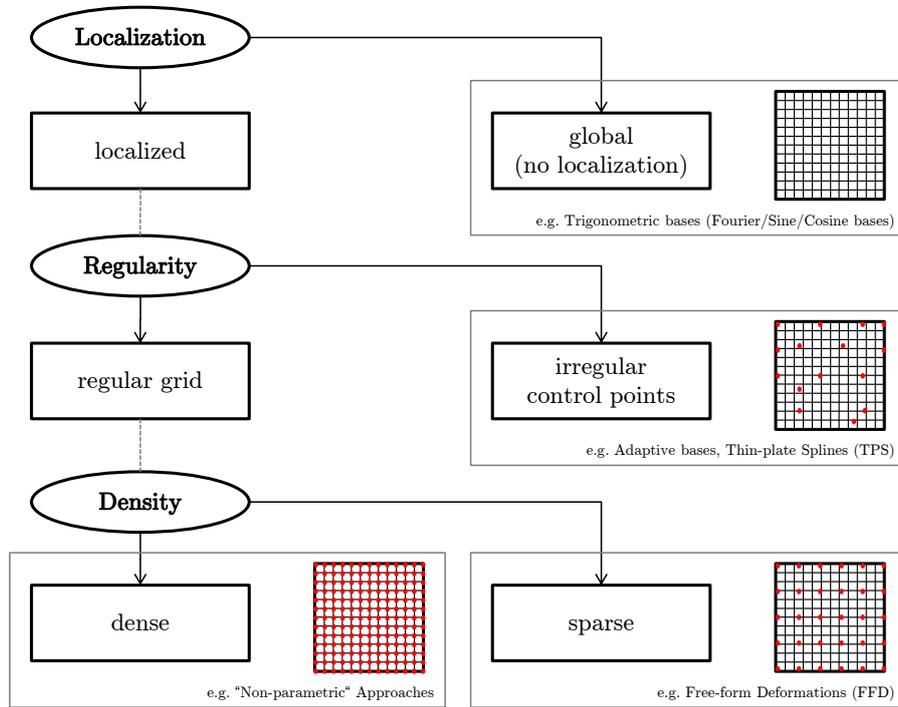


Figure 5.1: Characterization of linear transformation models based on the spatial configuration of the control points.

homogeneous areas. This can lead to a significant reduction of DOFs for same accuracy levels, compared to regular grids. Also, models based on irregular localization of sampling points are often employed in feature-based registration, since in images, features are in general not located on regular grids.

- *Density*. In the case of regularly localized models, this attribute expresses the density of the control point grid. Approaches with *dense* models, in which the grid resolution is equivalent to the image resolution, correspond to variational methods, while *sparse* models such as FFDs require less parameters.

The above properties, describing the spatial configuration of the control points, can be organized in the form of a binary tree – please compare also Figure 5.1 for a graphical representation.

The second set of attributes for characterization of linear transformation models describes the properties of basis functions, and includes the following attributes.

Properties of Basis Functions:

- *Support of Basis Functions* describes the spatial support for a single basis function. Possible options are *compact* support (e.g. B-Spline FFDs), and

global support (e.g. TPS, DCT). The advantage of functions with compact support is that their evaluation requires computations on less points from the image domain, which leads to lower computational costs. This is especially significant for a high number of DOFs, and higher-dimensional problems (that is 3D, compared to 2D).

- *Shape of Basis Functions* expresses whether the single basis functions B_k employed in the model have different forms, or whether they are identical, translated version of each other, i.e. whether for a representative basis function B_0 , and a control point c_k , we have $B_k(x) = B_0(c_k - x)$. An example for identically shaped bases are B-Spline FFDs. On the other hand, trigonometric bases, such as DCT, are an example for a parametrization in which every basis has a different shape.

In contrast to the properties of the spatial configuration of the control points, the different values for the attributes describing the basis functions can be combined in an arbitrary way.

5.2 Examples of Linear Transformation Models

In this section, we present a short overview of some popular transformation models. Please note that this selection is by no means comprehensive. We provide some exemplary 1D illustrations of basis functions in Figure 5.2.

5.2.1 B-Spline Free-form Deformations (FFD)

A very popular transformation model for registration of medical images is the Free-form deformation model based on B-splines. The model consists of B-spline kernels located at control points in a regular grid. The regularity of the grid mesh results in a conceptually simple model, and easy implementation. Due to the limited support, the linear operators resulting from gradient-based optimization are relatively sparse, and due to the definition as tensor product efficient implementation is possible by pre-computing the single one dimensional bases.

The FFD model, which was introduced in [Sederberg and Parry, 1986], was popularized for intensity-based registration of medical images by [Rueckert et al., 1998, Rueckert et al., 1999a]. The parametrization of non-linear transformations by FFDs based on cubic B-Splines is a common technique for registration of medical images. The method has become very popular and widely used in the field since its introduction, cf. e.g. [Schnabel et al., 2001, Kybic and Unser, 2003, Rohlfing et al., 2003, Glocker et al., 2008a, Klein et al., 2010b]. In [Rueckert et al., 2006], the FFD-based registration has been modified, such that it yields diffeomorphic transformations by construction.

In terms of the characterization from Section 5.1, the FFD model can be described as localized with a regular and sparse control point grid, featuring identically shaped basis functions with compact support. The only thing which remains to be specified is the actual form of the B-Spline basis function.

The d -dimensional B-Spline basis B_0 is based on the one-dimensional basis

functions b , defined as

$$b(x) = \begin{cases} b_0((x+2h)/h) & \text{for } -2h \geq x < -h \\ b_1((x+h)/h) & \text{for } -h \geq x < 0 \\ b_2(x/h) & \text{for } 0 \geq x < h \\ b_3((x-h)/h) & \text{for } h \geq x \geq 2h \end{cases}, \quad (5.8)$$

with h being the distance between two control points along the respective dimension, and with

$$b_0(x) = (1-x)^3/6 \quad (5.9)$$

$$b_1(x) = (3x^3 - 6x^2 + 4)/6 \quad (5.10)$$

$$b_2(x) = (-3x^3 + 3x^2 + 3x + 1)/6 \quad (5.11)$$

$$b_3(x) = x^3/6. \quad (5.12)$$

Now, the d -dimensional B-Spline basis B_0 is constructed as the tensor product of b functions corresponding to the single dimensions, that is

$$B_0(x) = \prod_{i=1}^d b(x_i). \quad (5.13)$$

The actual bases B_k , located at points c_k , are simply shifted version of B_0 and can be defined as

$$B_k(x) = B_0(x - c_k). \quad (5.14)$$

Please note that the presented definition of B-Splines is chosen to fit the high-level classification that we employ in this work. Usually, the B-splines are defined in a slightly different but equivalent manner, without the explicit definition of B_0 . This alternative derivation is closer related to the actual implementation schemes, please compare [Rueckert et al., 1999a, Rohlfing et al., 2003]. More details on B-Splines can be found in [Unser et al., 1993a, Unser et al., 1993b] while [Holden, 2008] gives an overview of the historical development.

5.2.2 Trigonometric Functions

Trigonometric functions are a further popular choice for parametrization of transformations. The basic idea is to model the displacement field as a linear combination of a finite number of low-frequency trigonometric functions. The actual choice of the trigonometric function depends on the boundary conditions which are assumed for the image domain. The homogeneous zero boundary condition (Dirichlet condition) requires the Discrete Sine Transform (DST) [Ashburner and Friston, 2004], while an unconstrained boundary can be modeled with Discrete Cosine Transform (DCT) [Ashburner and Friston, 1999]. A combination of DST and DCT can be used in order to impose a sliding boundary condition (Neumann boundary condition). For more details on the boundary condition treatment with trigonometric bases, see [Ashburner and Friston, 2004]. A further possibility is to employ the Discrete Fourier Transformation (DFT) [Amit, 1994, Christensen and Johnson, 2001], or Discrete Cosine Transformation (DCT) [Ashburner and Friston, 1999] basis functions.

In contrast to localized bases such as FFDs or RBFs, the trigonometric bases B_k are global and represent a periodic signal of frequency k . With respect to the other categories of characterization from Section 5.1, the trigonometric bases have global support, and each basis has a different shape.

Theoretically, by using an infinite number of trigonometric bases, any real continuous function can be exactly represented. For discretized settings, employing sufficiently many bases is equivalent to a regular sampling of a certain resolution. In registration applications, however, mostly only a certain number of low-frequency basis functions is used for parametrization. This has the advantage to lower the number of parameters, and it provides a restriction of displacements to a space containing only functions of a corresponding smoothness by construction.

A further motivation for the use of trigonometric functions is that in some cases, they are the eigenfunctions to the linear operators which result from the use of certain regularization operators, and optimization schemes. This property, facilitates the solution, and combined with fast trigonometric routines such as the fast-fourier transform (FFT), provides an efficient solver for these linear systems, cf. e.g. [Modersitzki, 2004].

5.2.3 Radial Basis Functions (RBF)

The class of Radial Basis Functions (RBF) is defined as the set of all real-valued functions, whose value depends only on the distance of the argument to the origin. This can be formalized by stating

$$B(x) = B(\|x\|) \quad , \quad \forall x \in \Omega \quad . \quad (5.15)$$

As a consequence, all level sets of RBFs are spherical, and the value of the level set depends on the radius. Popular examples of RBF bases are for instance the Gaussian function, or the Thin-plate spline (TPS) [Wahba, 1990a, Bookstein, 1989a].

One of the major advantages of RBFs is that they do not require a regular grid, which allows the reduction of the number of DOFs, and a locally adaptive resolution. As mentioned earlier, the irregular localization of control points is beneficial in feature-based registration, which is the reason, why RBFs have been a popular choice in this area, cf. [Rohr, 2000, Rohr et al., 2001a, Rohr, 2001].

A widely used example of an RBF are the so called Thin-plate splines (TPS). The TPS function is the fundamental solution to the biharmonic equation, thus minimizing the bending energy [Bookstein, 1989a]. In 2D, the TPS basis reads

$$B(x) = \|x\|^2 \ln(\|x\|) \quad , \quad (5.16)$$

and in 3D, we have

$$B(x) = -\|x\| \quad , \quad (5.17)$$

compare also [Rohr, 2001] for further details.

An example of the use of RBF bases with compact support for medical image registration is given in [Rohde et al., 2003]. The so called *Adaptive Bases*, which are used in this work are defined by

$$B_k(x) = [\max(1 - r, 0)]^4 (3r^3 + 2r^2 + 6r + 4) \quad , \quad (5.18)$$

with $r = \|x - c_k\|$.

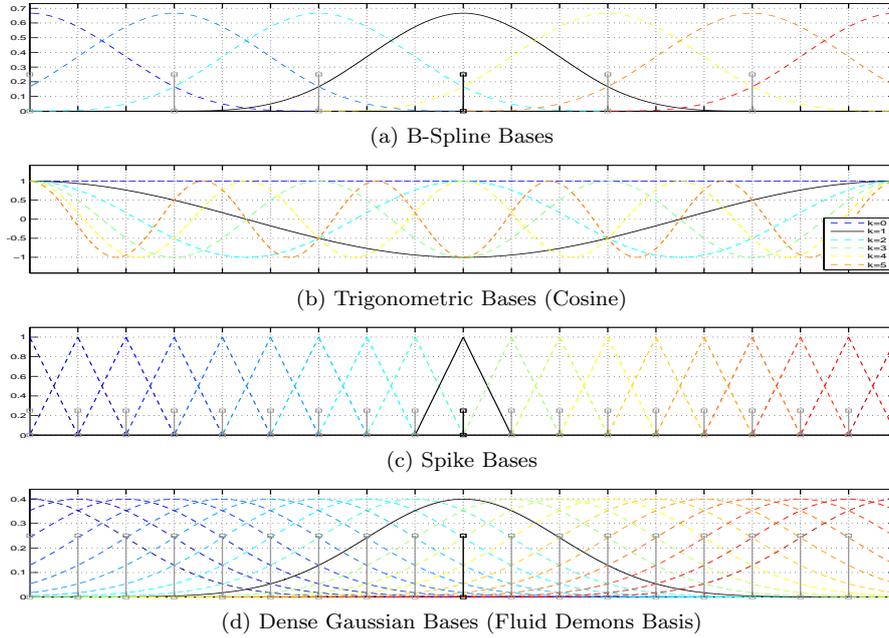


Figure 5.2: 1D illustrations of some parameterizations. Parameters (vertical lines) and corresponding basis functions are given. Please note that for the trigonometric bases (b), the influence of parameters is not localized in space.

5.3 Relation to Other Approaches

When it comes to the relation between variational and parametric approaches, we have already discussed the spatial discretization step in the variational derivation as a link between these two groups of methods in 2.4. In this context, especially the Finite Element method is of interest, since it shares the philosophy of most classical parametric methods to reduce the number of parameters.

In this section we would like to briefly discuss a further possible way of thinking about the connection between the methods which result from variational and the parametric derivations. We will do this by establishing a relation between the variational approaches which employ a Sobolev space to represent displacements, and the parametric approaches with linear transformation models. As we already argued, the setting of Sobolev spaces is related to semi-implicit time discretization for classical variational approaches (Sec. 2.3.2), as well as to large deformation registration (Sec. 3.2), and the demons framework (Sec. 4.4). Therefore, the establishment of a relation between Sobolev spaces and parametric approaches provides a link to the other groups of methods. This section is based on our work in [Zikic et al., 2010f].

Our argument is based on the interpretation of the energy gradient w.r.t. parameters as the projection of the continuous gradient w.r.t. the displacement onto the corresponding bases, as stated in (5.6). For a displacement u , belonging to a Sobolev space based on the metric tensor L^*L , the corresponding energy gradient w.r.t. u can be expressed in terms of the L^2 gradient ∇E as [Neuberger,

1997]

$$(L^*L)^{-1}\nabla E . \quad (5.19)$$

In many cases, this can alternatively be expressed in terms of the Green's function Ψ corresponding to (L^*L) by

$$(L^*L)^{-1}\nabla E = \Psi * \nabla E = \tilde{\Psi} * \tilde{\Psi} * \nabla E , \quad (5.20)$$

where we assume that $\Psi = \tilde{\Psi} * \tilde{\Psi}$. For example, Ψ can be a Gaussian for an appropriate choice of the Sobolev space [Chefd'hotel, 2005] with the appropriate standard deviation σ . Then, $\tilde{\Psi}$ is a Gaussian with standard deviation $\sigma/2$. Please note that we do not request Ψ and $\tilde{\Psi}$ to be the same type of function.

In general, $\tilde{\Psi}$ will have a form of a smoothing kernel, i.e. it can be employed as a spatially localized and regular basis shape, on a dense control point grid. Also, the discrete correspondence to the convolution in (5.20) can be written as

$$\tilde{\Psi} * \nabla E \hat{=} \nabla E^\top [B_1, \dots, B_N]^\top = \nabla E^\top B , \quad (5.21)$$

with $B_i = \tilde{\Psi}(x - c_i)$ referring to the basis function $\tilde{\Psi}$ located at the control point c_i . If we assume that (5.21) gives us the update for the parameters of a dense transformation model dp^\top (cf. (5.6)), then the corresponding update displacement would read

$$u(dp) = dp^\top B = \nabla E^\top B B . \quad (5.22)$$

Analogous to (5.21), we see that the above form (5.22) is a discrete formulation corresponding to the Sobolev gradient of (5.20), which determines the optimization scheme, and thus the registration process. In summary, we can state that a dense parametric transformation model with basis functions $B_i = \tilde{\Psi}(x - c_i)$, corresponds to a discretized version of a variational approach in the Sobolev space based associated to the Green's function $\Psi = \tilde{\Psi} * \tilde{\Psi}$.

As a specific example, we can state that a fluid demons, based on filtering by a Gaussian G_σ , which corresponds to the use of a specific Sobolev space [Chefd'hotel, 2005], with G_σ as the Green's function, corresponds to a parametric approach using $G_{\sigma/2}$ as the basis function on a dense control point grid.

With this intuition, we can now easily compare a B-Spline FFD approach (without regularization) to fluid demons. Since a B-Spline basis can be very accurately approximated by an appropriate truncated Gaussian, the difference between the FFD and the the fluid demons can be characterized by the density of the control point grid. While the demons uses a dense grid, the FFD usually employs a sparse control point grid. Thus, we can see the FFD approach approximately as an appropriately tuned demons, in which a certain number of basis functions is omitted.

A similar argument can be made for the approach resulting from variational derivation with semi-implicit time discretization, through its interpretation to gradient descent in Sobolev spaces (cf. Sec. 2.3.2).

5.4 Discussion

The group of parametric registration approaches summarizes a large number of methods which employ a comparably low-dimensional transformation model

directly in the energy formulation. This approach has a conceptually straightforward derivation and due to the reduced dimensionality in comparison to high-dimensional parametrizations it results in robust methods, which are potentially computationally more efficient, and which pose less restrictions with respect to the choice of the optimization method due to the reduced problem size.

It is to be noted however, that the methods resulting from parametric derivation are not fundamentally different to methods obtained by variational derivation. All methods have in common, that ultimately a finite-dimensional formulation must be obtained for implementation, a step, which is performed by parametrization. The major difference is at which point in the derivation the transition to the finite-dimensional settings is performed, please compare also Figure 1.1. In this context, the Finite Element method for spatial discretization provides a strong link between the variational derivation and the parametric methods. Also, the interpretation of parameter updates as projections of the dense energy gradient to the basis functions can provide an interesting insight into the relation between the different parametric and variational methods. In this way, many of the classic variational approaches, such as the fluid demons method, or the variational elastic method with explicit or semi-implicit time discretization can be also interpreted as parametric approaches and equivalent implementations can be derived in this manner [Zikic et al., 2010f].

Part II

Methodological Contributions

EGALITARIAN PRECONDITIONING: A GENERAL PRECONDITIONING SCHEME FOR IMAGE-BASED DIFFERENCE MEASURES IN DEFORMABLE REGISTRATION[†]

We propose a preconditioning scheme for efficient optimization of arbitrary image-based difference measures in the context of deformable registration.

Our approach is based on the observation and analysis of a fundamental negative property of steepest gradient descent (SGD) for registration problems: For image-based difference measures, SGD locally exhibits slower convergence speed in areas corresponding to low-gradient regions of the source image. We refer to this property as *Local Gradient Bias*, and our approach aims at avoiding the above effect in a principled way, for all difference measures.

The proposed scheme is simple and computationally efficient: It performs an approximate normalization of the point-wise vectors of the difference gradient to unit length. We perform a theoretical analysis of the condition of difference measures in registration problems, and demonstrate that our scheme improves the condition of the original problem. Because of the simplicity of the scheme, its application improves the convergence speed while adding only negligible computational cost, thus resulting in shorter effective runtimes. Since the key idea behind our approach is to remove the inequalities between the updates in the different image regions, we call the resulting approach *Egalitarian Preconditioning*.

The proposed preconditioning is of particular interest for high-dimensional deformable registration with statistical difference measures such as mutual information (MI). In these settings, the range of applicable standard methods for efficient optimization is strongly limited, due to the structure and the size of the resulting optimization problem. Here, our approach presents a conceptually simple, yet a theoretically justified alternative, which can easily be integrated into any gradient-based registration scheme. We demonstrate the application of the proposed preconditioning for registration in the group of diffeomorphisms,

[†]This chapter is based on the following publications: [Zikic et al., 2006a, Zikic et al., 2010d, Zikic et al., 2010a, Zikic et al., 2010e, Zikic et al., 2011].

and for the demons method. For the demons scheme in particular, our approach can be seen as an *efficient* generalization of the demons scheme to arbitrary difference measures.

6.1 Introduction

The motivation for our work and its potential usefulness are given by the lack of efficient optimization methods for statistical difference measures for high-dimensional transformation models. Many standard optimization methods such as Newton-type approaches, which are applicable for measures such as SSD, can not be applied for statistical measures such as MI, due to the size and structure of the required operators. Other approaches such as L-BFGS remain applicable for some settings, but cannot be applied for example in the demons setting. We discuss this motivation in more detail in Section 6.1.1.

The key idea of our approach is based on the observation of the effect, to which we refer as Local Gradient Bias (LGB). The LGB effect describes the fact that the the steepest gradient descent on image-based difference measures results in spatially varying convergence speed. More specifically, the local convergence speed of the registration at the point x depends on the underlying intensity gradient in the source image $\nabla I_S(x)$, and consequently different points in the image domain have a different influence. For more details on the LGB effect and its properties, please see Section 6.1.2

Since the intensity gradient in one of the images cannot provide a meaningful prior on the magnitude of the corresponding displacement, we consider this behaviour not to be justified. Therefore, our goal is to find a theoretically sound approach which removes the inequality of the influence between the different points in the image domain. We show that our goal results in faster convergence, and that it can be formulated as a preconditioning scheme.

Since the term *Egalitarianism* denotes a “philosophy advocating the removal of inequalities among people” (definition from Merriam-Webster Dictionary), and our goal is the removal of inequalities in the influence of pixels, we refer to the resulting preconditioning scheme as *Egalitarian Preconditioning*. For an illustration of the local gradient bias, and the effect of the proposed preconditioning, please compare Figure 6.1.

6.1.1 Optimization of Statistical Difference Measures for High-dimensional Transformation Models

As discussed previously, given the target image I_T and the source image I_S , the task of registration is to compute the transformation ϕ , such that ϕ maps between the corresponding points in the input images. Intensity-based deformable registration is generally formulated as an optimization of an energy E combining a difference measure E_D and a regularization term E_R by

$$E(\phi) = \gamma E_D(\phi) + \lambda E_R(\phi) , \quad (6.1)$$

with positive scalar factors γ and λ . For high-dimensional deformation models, the difficulty of efficient optimization of (6.1) depends largely on the choice of E_D . While many standard methods can be used for point-wise difference measures such as the sum of squared differences (SSD), alternatives for efficient

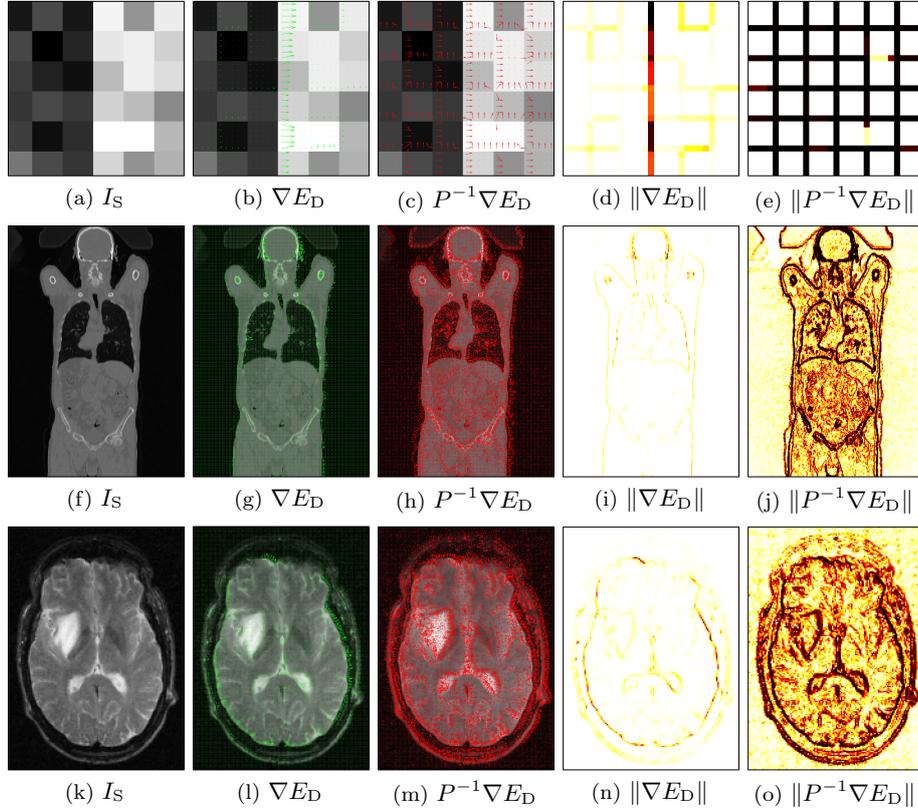


Figure 6.1: Some examples of the local bias effect and the proposed egalitarian preconditioning. We display the source image I_S , and the gradients of the difference measure ∇E_D and their norm $\|\nabla E_D\|$, as well as the preconditioned versions $P^{-1}\nabla E_D$ and $\|P^{-1}\nabla E_D\|$. The first row demonstrates the effect on a synthetic example, the second row for a CT thorax image, and the third for an MR brain image. The energy employed in all cases is the SSD, and the gradient results from shifting I_S by one pixel in relation to I_T .

optimization of statistical measures such as mutual information (MI) [Wells et al., 1996, Maes et al., 1997a], or correlation ratio (CR) [Roche et al., 1998] are more rare and complex. The goal of this work is to provide a simple scheme for efficient optimization of arbitrary difference measures.

For the treatment of multi-modal registration problems in medical image analysis, the use of statistical difference measures such as MI or CR is of particular interest [Pluim et al., 2003a]. These measures operate on the joint probability distribution of the intensities of the two input images, hereby linking all points with the same intensities, which results in a non-local character of the measures. This is in contrast to difference measures which are defined by point-wise comparisons, such as the sum of squared differences (SSD). The non-locality of statistical measures has consequences on the structure of the corresponding Hessian matrix, which is a major building block for many standard optimization methods. While the Hessian of the SSD measure is a sparse

matrix with a small number of non-zero diagonals, the Hessian of MI is dense without any regular sparsity pattern. The size of Hessian matrix H for MI thus becomes prohibitive for practical treatment for high-dimensional settings. With the number of parameters being equal to the number of voxels times the dimension of the images, we get sizes of H in the order of $((3 \cdot 10^6) \times (3 \cdot 10^6))$ for medium size problems. The same problem holds for the Jacobian J of statistical difference measures. This property rules out the Newton-type methods such as Gauss-Newton or Levenberg-Marquardt (employing $J^\top J$) [Nocedal and Wright, 2000, Madsen et al., 2004], which are good choices for point-wise measures such as SSD. Please see also [Modat et al., 2010] for a description of this issue. Please note that the problems related to the size of $J^\top J$ might be less prominent for methods which employ lower-dimensional transformations models, such as e.g. FFD B-Splines [Rueckert et al., 1999a], depending on the resolution of the employed control point grid.

Therefore, efficient optimization methods for statistical difference measures in high-dimensional settings must circumvent the problem of dealing with dense matrices H or $J^\top J$. Standard techniques to achieve this goal include quasi-Newton methods such as L-BFGS (limited memory Broyden-Fletcher-Goldfarb-Shanno), nonlinear conjugate gradient (NL-CG), or preconditioning techniques [Nocedal and Wright, 2000].

While both of these methods can in principle be applied for approaches based on generic optimization of the energy in (6.1), the use of L-BFGS and NL-CG is not theoretically justified for the standard interpretation of the popular *demons* registration method as an alternating optimization process, in which the optimization steps for E_D are interleaved with smoothing operations, which correspond to optimization of E_R [Modat et al., 2010]. Since L-BFGS and NL-CG operate by utilizing the information about the energy gradient from subsequent iterations, this process for E_D is disturbed by the smoothing step, which makes this information inconsistent [Modat et al., 2010]. Presumably in consequence of these properties, optimization of E_D by steepest gradient descent (SGD) is still the standard choice for multi-modal demons registration, and research on efficient and justified optimization schemes for the demons method is a topic of interest [Modat et al., 2010].

A further negative property of L-BFGS and NL-CG in the context of deformable registration from the practical point of view, is the requirement of these methods for a precise step-size estimation, in order to achieve improvement in convergence [Nocedal and Wright, 2000]. For deformable registration, where the evaluation of the energy term is comparably costly, this requirement introduces further computational costs. Finally, L-BFGS requires additional memory to store the n -last energy gradients, where n depends on the required accuracy.

With respect to the above description, the advantages of our method are that it is applicable for all gradient-based approaches, including the demons scheme, it requires no expensive step-size estimation, and no additional memory. Besides, due to its conceptual simplicity, it offers a simple implementation, and efficient computation.

6.1.2 Local Gradient Bias

The approach which we introduce in this work, is based on the observation of an effect, which we refer to as the Local Gradient Bias (LGB). The LGB

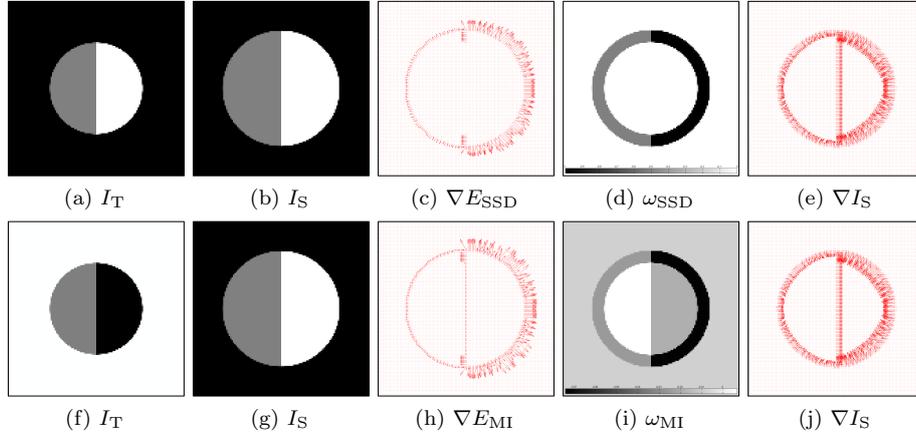


Figure 6.2: Illustration of the structure of the gradient of the difference measure $\nabla E_D = \omega_D \nabla I_S$ for SSD (first row), and MI (second row), on a synthetic example. Please note the local gradient bias effect for both measures, i.e. that the entries of $\nabla E_D(x)$ are considerably lower for points at which the source image has a low intensity gradient $\nabla I_S(x)$.

effect describes the fact that the the steepest gradient descent on image-based difference measures results in spatially varying convergence speed, and that the convergence speed depends on the underlying intensity-gradients of the source image.

The reason for this behavior is the structure of the gradient of image-based difference measures. Since in this case, the energy depends on the transformation through the source image, by applying the chain rule, we get

$$\frac{\partial}{\partial \phi} E_D(I_S(\phi)) = \frac{\partial E_D}{\partial I_S} \frac{\partial I_S}{\partial \phi} . \quad (6.2)$$

Point-wise, we can write the above also as

$$\nabla E_D(x) = \omega_D(x) \nabla I_S(x) , \quad (6.3)$$

with the real-valued function $\omega_D : \Omega \rightarrow \mathbb{R}$. We illustrate the above equations in Figure 6.2. Please note that in this work, by point-wise, we do not refer to one specific entry of a vector, but to the entities assigned to a specific point x in the spatial domain. Therefore, $\nabla E_D(x) \in \mathbb{R}^d$ can be seen as a single displacement vector at location x . The relation in (6.3) has been discussed previously, for example in [Chefd'hotel et al., 2002, Hermosillo et al., 2002]. Please note, that the above derivation holds for all image-based difference measures, and the term which encodes the specific measure is ω_D .

We see from (6.3) that the difference gradient $\nabla E_D(x)$ at a single point depends directly on $\nabla I_S(x)$. Specifically, if not corrected by $\omega_D(x)$, the length of $\nabla E_D(x)$ will be directly influenced by the length of $\nabla I_S(x)$. It is our experimental observation for all common difference measures, that in general ω_D does not have the effect to correct for the length of $\nabla I_S(x)$.

To sum up, the length of the point-wise gradient vector $\nabla E_D(x)$ depends on the length of $\nabla I_S(x)$. And since in steepest-descent-based methods, the update

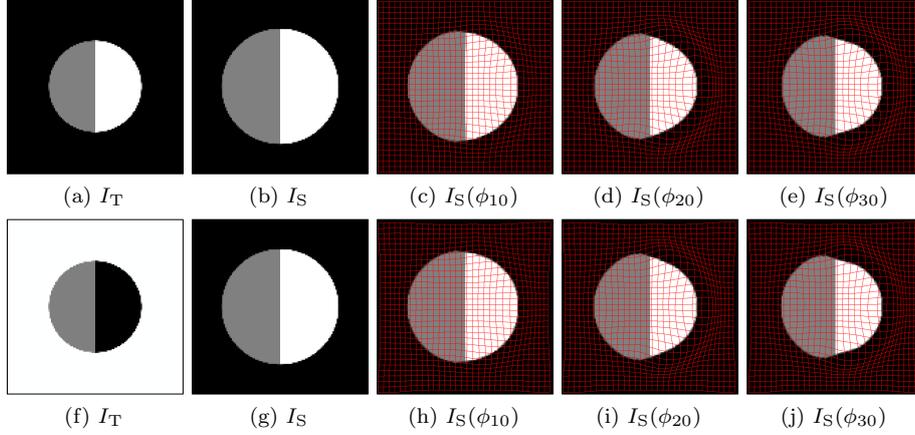


Figure 6.3: Illustration of the Local Gradient Bias effect. The unequal local convergence speed is caused by the steepest gradient descent optimization scheme. We demonstrate this behaviour on the synthetic example from Fig. 6.2 for SSD (first row), and MI (second row). We display the target and source image, as well as the deformation estimates after 10, 20, and 30 iterations, overlaid with the warped source image.

depends significantly on the gradient of the difference measure, the length of the point-wise update $v(x)$ depends also on $\|\nabla I_S(x)\|$. As a consequence, the local convergence speed of the registration depends on the underlying local intensity gradient in the source image, and different points in the image domain have a different influence. Please compare Figure 6.3 for an illustration.

It is our conviction, that the behaviour described above is not justified, since we believe that the knowledge of the intensity gradient at a particular point in one of the images cannot provide a meaningful prior on the magnitude of the corresponding displacement.

To put it differently: Given the information about $\nabla I_S(x)$, and no information about I_T , we cannot know the displacement $u(x)$. Consequently, it does not make sense to make differently large update steps, depending solely on the information in $\nabla I_S(x)$.

Since our early observation of the LGB effect in [Zikic et al., 2006b], through our subsequent work and the analysis, we established that LGB is a fundamental property of steepest descent approaches for image-based difference measures in deformable registration, and were able to observe and describe a number of its properties. Here, we briefly sum up the most important properties of the LGB effect:

- LGB is present for all image-based difference measures. It is specifically also present for statistical measures such as MI. The strength of the effect depends on the difference measure through the term ω_D , and it is for example stronger for SSD (where the effect is actually increased by ω_{SSD}) than for SAD (where $\omega_{SAD} = \pm 1$ has no effect).
- The effect of LGB depends on the structure of the source image, and it can be more pronounced for some images than for others. The effect is

strong if I_S contains intensity gradients of largely different magnitudes cf. Figure 6.1. On the other hand, in a binary image with only two intensities, all non-zero gradients have the same length, and the LGB effect would not be present.

- LGB is specific to steepest gradient descent and it does not occur in Newton-type optimization. We performed an experimental and theoretical analysis of the Gauss-Newton (GN) approach for SSD as difference measure in [Zikic et al., 2010e, Zikic et al., 2010a], and we observe that the major influence of GN for motion estimation is exactly the approximate normalization of the point-wise updates to unit length. We see this as a strong indication that our intuition about the negative properties of LGB is justified, and that the removal of this effect can lead to improved results. Furthermore, in [Zikic et al., 2010d], we introduced a method for deformable registration based on the idea of natural gradients. In this approach, we also observed that the length of the point-wise updates was approximately normalized.
- Our analysis in Section 6.3 reveals that the LGB effect can be seen as a consequence of the bad condition of image-based difference measures. The suboptimal condition of the difference measures can be understood when one considers that not all variations of the displacement field entries about the optimum will result in same energy changes. Intuitively, a problem has a bad condition if the same variation of different parameters results in strongly differing energy changes. For example, for SSD the change in the energy corresponds directly to the edge strength at the dislocated edges.
- Since the LGB effect leads to slower convergence in areas corresponding to lower intensity gradients in the source image, it can have particularly negative effects in medical settings. In many modalities, the inside of the human body typically exhibits weaker gradients than the body/air boundary cf. Figure 6.1. If steepest descent is applied, this can result in slower convergence inside the body, which is often the area of interest.

Led by the above intuition, our goal was to find a theoretically sound approach which removes the inequality of the influence between the different points in the spatial domain.

6.1.3 Related Work

With respect to the demons method, most applications are employing it for mono-modal registration based on SSD, with the original force term, which is based on the Gauss-Newton method. In some works, multi-modal scenarios are approached by preprocessing the intensities of the input images, such that the original mono-modal demons formulation can be applied [Kroon and Slump, 2009]. However, the demons framework has also been applied in some multi-modal settings with statistical difference measures. However, most of these works assume the steepest gradient descent (SGD) update step for the forces [Chefd'hotel et al., 2002, Guimond et al., 2002]. To our best knowledge, the only exception to the use of SGD is the work by Modat *et al.* in [Modat et al., 2010], where the force is computed as an update of the non-linear conjugate gradient (NL-CG) optimization method [Nocedal and Wright, 2000]. Boiled down, the NL-CG

method computes the force term as a specific weighted average of the gradients of the difference measure from the last two iterations [Nocedal and Wright, 2000].

While NL-CG does not require an approximation of the Hessian, which would be problematic for statistical difference measures, it has the following disadvantages (also discussed in [Modat et al., 2010]). First, the usage of NL-CG is theoretically not justified for demons, since the gradient information about the difference measure from consecutive iterations is not consistent due to the interleaved regularization step [Modat et al., 2010]. While, in spite of this point, NL-CG reportedly can improve the convergence speed compared to the SGD, no quantitative comparison is performed in [Modat et al., 2010]. Furthermore, NL-CG requires a precise step size estimation in order to achieve an improvement in convergence [Modat et al., 2010, Nocedal and Wright, 2000]. Since precise step-size search in deformable registration is comparably expensive (order of magnitude of a basic iteration), the increased computational cost can easily outweigh the improvement in convergence speed.

In the context of high-dimensional elastic registration, the L-BFGS method [Nocedal and Wright, 2000] has been employed for example in [Modersitzki, 2009]. The essence of L-BFGS is the approximation of the effects of applying H^{-1} to the energy gradient, without actually setting up or inverting H . Since the approximation in L-BFGS also relies on accumulation of the gradient information from n last iterations, it shares the same disadvantages as NL-CG, which makes it unsuitable for the demons framework. Actually, since L-BFGS keeps a longer history of the previous gradients for the approximation of H^{-1} , this effect is even more pronounced (specific version of NL-CG can be seen as L-BFGS with $n = 2$ [Nocedal and Wright, 2000]). Compared to NL-CG, the additionally required storage for the n last gradients is a further disadvantage.

To our best knowledge, preconditioning schemes have not been studied widely for deformable registration. An exception is [Klein et al., 2011], which proposes a specific scheme for SSD.

In contrast to the above methods, the use of the proposed preconditioning approach is applicable to arbitrary difference measures, and it is justified also in the demons setting, since it does not require information from previous iterations. Also, please note that for the propose scheme, no additional memory is needed. Furthermore, the proposed method does not require a precise step size search, and improves the convergence speed already with a simple fixed step size strategy, thereby directly translating the gain in convergence speed to an effective improvement of runtime.

6.1.4 Outline

The remainder of this Chapter is organized in the following manner. First, we introduce the preconditioning scheme and discuss how it can be employed in different registration frameworks in Section 6.2. After this introduction, we present a theoretical argument, that the proposed scheme actually improves the condition for arbitrary image-based difference measures in deformable registration (Section 6.3). Finally, we present the evaluation for the application of the preconditioning for the demons framework and for registration in groups of diffeomorphisms 6.4.

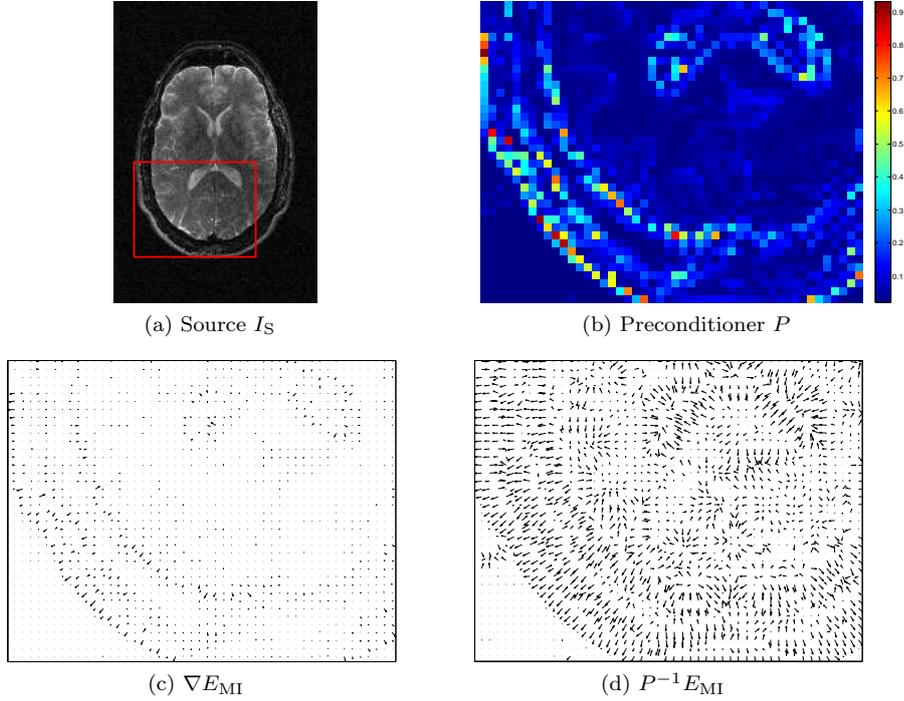


Figure 6.4: Example of the effect of the proposed preconditioning scheme for MI. The point-wise gradient vectors $\nabla E_D(x)$ are normalized to approximately same length by $P^{-1}\nabla E_D(x)$. Our analysis reveals that this simple strategy improves the condition for arbitrary difference measures E_D , which improves the convergence speed of gradient descent methods.

6.2 The Egalitarian Preconditioning Scheme

Given a difference measure E_D , a preconditioned gradient descent approach modifies the descent direction by modifying the gradient $\nabla E_D(\phi)$

$$P^{-1}\nabla E_D(\phi), \quad (6.4)$$

where P is a symmetric positive definite operator [Nocedal and Wright, 2000]. The preconditioning operator P , should present a numerically favorable approximation to the Hessian H_{E_D} . Preconditioning approaches can thus be seen as an approximation to the Newton method, which operates by $H_{E_D}^{-1}\nabla E_D(\phi)$.

The art of preconditioning consists in designing such a P for a specific problem, or as [Saad, 2003] puts it “Finding a good preconditioner ... is often viewed as a combination of art and science.”. The preconditioning term has to make a compromise between the quality of the approximation of the Hessian H_{E_D} , and the computational efficiency of evaluating P^{-1} . On one hand, the possibly accurate approximation to the Hessian is required in order to improve the convergence speed. On the other hand, the efficiency of evaluating P^{-1} is needed since otherwise, despite improved convergence, the effective runtime can increase due to the computational overhead of the inversion in each iteration. As

an example, using $P = H_{E_D}$, which is equivalent to Newton optimization, would result in the perfect preconditioning of the problem, however this is not applicable to high-dimensional registration problems with statistical measures, since H_{E_D} becomes too full, and its processing impracticable. The Gauss-Newton method can also be seen as a preconditioning approach with $P = J^\top J$. However, for statistical measures, the same problem remains. A standard simple alternative for preconditioning is the so called Jacobi preconditioning [Saad, 2003], which approximates the Hessian by its diagonal, i.e. $P = \text{diag}(H_{E_D})$. While this approach is efficient to compute, in our experiments, it did not yield improved results, presumably to the numerical issues involved in computing the second order derivatives for statistical measures.

In this work, we propose a simple preconditioning scheme, which modifies the magnitudes of $\nabla E_D(x)$, such that they become possibly similar, please see Figure 6.4 for an illustration. Point-wise, this can be achieved by defining the action of the preconditioner P as a multiplication of $\nabla E_D(\phi)(x)$ with a positive scalar by

$$P^{-1}\nabla E_D(\phi)(x) = \frac{1}{\|\nabla E_D(\phi)(x)\| + \sigma} \nabla E_D(\phi)(x) . \quad (6.5)$$

This corresponds to a diagonal, positive definite operator P , with the $d \times d$ block $P|_x$, corresponding to $x \in \Omega$ defined by a constant diagonal

$$P|_x = \text{diag} (\|\nabla E_D(\phi)(x)\| + \sigma) . \quad (6.6)$$

Please note that in our context, the term *point-wise* refers to a specific point x in the spatial domain Ω , and not one vector entry. Thus, the point-wise vector $\nabla E_D(x) \in \mathbb{R}^d$ is a sub-vector of ∇E_D , assigned to spatial position $x \in \Omega$, and not simply one entry of ∇E_D . Also, please note that P varies with ϕ .

While P being positive establishes the formal requirements for a proper gradient descent, it has not yet been shown that the application of P will actually improve the condition of E_D . We will demonstrate this property theoretically by the analysis in Section 6.3, and confirm it by experiments in Section 6.4.

6.2.1 Application to the Demons Framework

Due to its efficiency and simplicity of implementation, the demons method [Thirion, 1998] has become a popular choice in numerous applications. We consider a general demons framework which computes the transformation $\phi = \text{Id} + u$ by

$$f = \text{compute_force}(E_D) \quad (6.7)$$

$$g = \tau\gamma G_{\sigma_{\text{fi}}} * f \quad (6.8)$$

$$u_{i+1} = G_{\sigma_{\text{el}}(\tau,\lambda)} * (u_i \circ (\text{Id} + g)) . \quad (6.9)$$

Our focus is on the computation of the force term in (6.7). In the original approach this was performed by

$$f = \frac{1}{(I_T - I_S(\phi))^2 + \|\nabla I_S(\phi)\|^2} \underbrace{(I_T - I_S(\phi)) \nabla I_S(\phi)}_{= -\nabla E_{\text{SSD}}(\phi)} , \quad (6.10)$$

which can be seen as a modification of the gradient of the SSD. It was demonstrated that (6.10) corresponds to a step of a Newton-type optimization scheme on the SSD [Pennec et al., 1999]. Most current approaches which extend the demons framework to other difference measures compute the forces by employing the steepest descent scheme [Chefd'hotel et al., 2002, Guimond et al., 2002], that is

$$f = -\nabla E_D(\phi) . \quad (6.11)$$

The only exception to this approach known to us is [Modat et al., 2010], where the forces are based on the NL-CG method. However, as also discussed in [Modat et al., 2010], there is no theoretical justification for using the NL-CG method for force computation in the demons setting.

In this work, we consider the computation of the demons forces based on a preconditioned gradient descent by

$$f = -P^{-1}\nabla E_D(\phi) , \quad (6.12)$$

with the operator P^{-1} as discussed in Eq. (6.5).

It is interesting to note that in the case of SSD, the proposed scheme from (6.5) approximates the original demons force in (6.10), since for small displacements, we have $|I_T - I_S(\phi)| \approx \|\nabla I_S(\phi)\|$, and

$$\|\nabla E_{SSD}(\phi(x))\| = \|(I_T - I_S(\phi)) \nabla I_S(\phi)\| \quad (6.13)$$

$$\approx \frac{1}{2} (I_T - I_S(\phi))^2 + \frac{1}{2} \|\nabla I_S(\phi)\|^2 \quad (6.14)$$

Thus our approach can be seen as a natural and generalization of the original demons to arbitrary difference measures.

6.2.2 Application to Variational and Parametric Approaches

For classic variational methods (cf. Ch. 2), a preconditioned gradient descent approach can be employed for the energy E from (6.1) by incrementally updating ϕ by

$$v = -K^{-1}\nabla E(\phi) , \quad (6.15)$$

where K is a symmetric positive definite operator, which should present a numerically favorable approximation to H_E . Since $E = \mu E_D + \lambda E_R$, we have to approximate $H_E = \mu H_{E_D} + \lambda H_{E_R}$. While the difference terms E_D are non-linear, most regularization terms are of the quadratic form $E_R = 1/2 \langle A_R u, u \rangle$, with the corresponding Hessian being $H_{E_R} = A_R$. Thus, for these cases, we have to estimate only the preconditioning term $P \approx H_{E_D}$ for the difference measure for the scheme from (6.15), and we get

$$v = -(\mu P + \lambda A_R)^{-1} \nabla E(\phi) . \quad (6.16)$$

Also, the application of the preconditioning to parametric approaches with linear transformations models (as described in Ch. 5) is straight forward. By employing the fact that the energy derivative for the parametrized model corresponds to the projection of the dense derivative onto the respective basis functions (cf. Eq. (5.6)), we can simply incorporate the preconditioning by projecting (6.16) onto the bases B in question, that is by

$$[(\mu P + \lambda A_R)^{-1} \nabla E(\phi)]^\top B . \quad (6.17)$$

6.2.3 Application to Registration in Groups of Diffeomorphisms

In [Zikic et al., 2010a], we have discussed a generalization of approaches for registration in groups of diffeomorphisms, which allows us to incorporate the proposed preconditioning into this setting in terms of a Riemannian metric.

Approaches for registration in groups of diffeomorphisms (cf. Sec. 3.2) are based on computing the derivatives of the energy in the tangent space to the manifold of diffeomorphisms, at the current deformation estimate. Since the corresponding tangent spaces are Sobolev spaces, the derivative has to be computed with respect to the employed Sobolev space \mathbb{H} . This computation of the so called Sobolev gradient [Neuberger, 1997] is based on the definition of the Riemannian metric tensor L^*L in the respective space, and can be performed by projection of the L^2 gradient ∇E as

$$\nabla_{\mathbb{H}} E = (L^*L)^{-1} \nabla E . \quad (6.18)$$

While previous approaches assumed that the metric tensor L^*L is constant during registration and only performs a smoothing [Trouvé, 1998, Dupuis et al., 1998, Chef d'hotel et al., 2002, Beg et al., 2005], in [Zikic et al., 2010a], we make use of the property of the Riemannian metric tensor, that L^*L can vary in the manifold with ϕ , and employ the tensor not only for smoothing but also to additionally modify the descent direction. So, first, instead of using the standard definition for the metric tensor for a Sobolev space \mathbb{H}^k [Chef d'hotel, 2005],

$$L^*L = \sum_{i=0}^k (-1)^i \nabla^i \alpha_i \nabla^i , \quad (6.19)$$

we replace the constant scalars α_i by a potentially variable metric tensors M_i , by

$$L^*L_{\phi} = \sum_{i=0}^k (-1)^i \nabla^i M_i(\phi) \nabla^i . \quad (6.20)$$

With M_i restricted to positive definite operators, this modification does not change the class of functions contained in the Sobolev space. This way, we can incorporate the proposed preconditioning by setting $M_0(\phi) = P$, while for the original approach we had $\alpha_0 \text{Id}$ for $i = 0$. This modification allows us to modify the descent direction additionally to performing the smoothing, which in both approaches is done by the terms corresponding to $i > 0$. Due to the form of P , the change of the descent direction results in faster convergence. In [Zikic et al., 2010a], we actually use a slightly different implementation of (6.6), which however has the same normalizing effect. For the integration on the manifold, we applied the small displacement scheme from [Chef d'hotel et al., 2002].

In summary, our idea in this approach is to modify the metric of the underlying manifold by incorporating the preconditioning term into the definition of the metric, so that the registration process converges more quickly. We present a summary of the results of this approach in Section 6.4.2. For more details, please refer to [Zikic et al., 2010a].

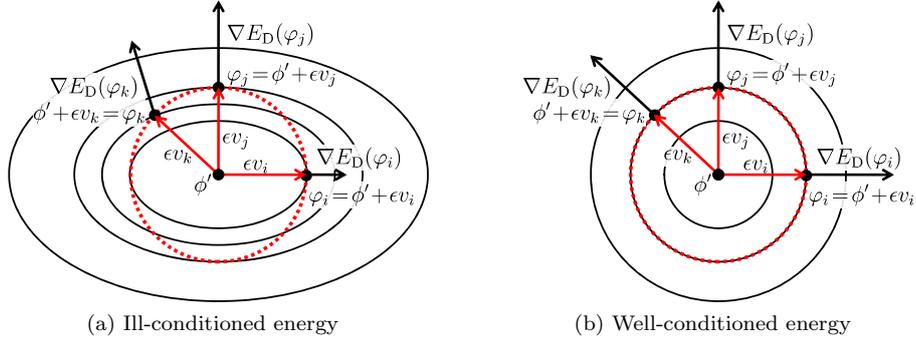


Figure 6.5: Illustration of the model for evaluating the condition of the optimization problem. The condition is described by the maximal variation of $\langle \nabla E_D(\varphi_i), v_i \rangle$, cf. Eq. (6.22) and (6.23). For registration, the directions v correspond to displacement fields, and ϵv represent permissible small warpings of the source image about the optimal deformation, compare also Fig. 6.6.

6.3 Analysis of the Preconditioning Scheme

In this section, we perform a theoretical analysis of the proposed preconditioning scheme, which demonstrates that its application actually improves the condition in comparison to the original problem. The analysis is performed as follows. First, we define the model for the analysis of the condition, which operates by measuring the variation of the energy about the optimum, and is based on the energy gradients (Sec. 6.3.1). Second, for optimization of difference measures, which is an under-constrained problem for high-dimensional deformable registration, the analysis has to be restricted to permissible, meaningful directions (Sec. 6.3.2). We also restrict the class of preconditioning operators to those which perform a point-wise rescaling of the gradient. Finally, we perform the condition analysis for a restricted optimization of difference measures in deformable registration (Sec. 6.3.3). By demonstrating that our scheme approximates the optimal improvement of the condition for the analyzed model (comprising the restrictions, assumptions and approximations made by the model), we hope to show that the proposed scheme improves the condition of the original problem.

6.3.1 Model for Condition Analysis

The condition of an optimization problem can be intuitively seen as the description of the geometry of the energy function about the optimum. The reasoning about the condition is based on the assumption that the quadratic approximation of the energy is valid. For non-quadratic energies, this implies the restriction to the vicinity of the optimum, in which the quadratic Taylor approximation is sufficient.

The condition $\kappa(E)$ for the energy E is usually defined as the ratio of the largest and smallest eigenvalues of the Hessian at the optimum ϕ' . For the perfect condition, we have $\kappa(E) = 1$, which represents a spherical energy, for which the gradients at all points point toward the optimum.

In order to show that the proposed scheme improves convergence for all image-

based difference measures, we stray from the standard analysis and propose to use an alternative model for describing the condition. We introduce this model in the following.

A well-conditioned energy function has a possibly spherical shape. We employ this intuition to develop the following model. For a critical point ϕ' , the perfectly conditioned problem can be formalized by

$$E_D(\phi' + v_i) = E_D(\phi' + v_j) , \quad (6.21)$$

where the energy is varied in all possible directions v with a fixed length $\|v\| = r$. Intuitively, the model states that *for a perfectly conditioned problem, the change of energy about the optimum ϕ' should depend only on the distance to ϕ' and not the probing direction v .*

In order to tie the above condition to the gradient of the function, we will employ the following form

$$\langle \nabla E_D(\phi' + v), v \rangle , \quad (6.22)$$

which describes a spherical function if (6.22) is constant for all directions v with a fixed length $\|v\| = r$. For a visualization, please see Figure 6.5. Also, please note that if v is sufficiently small (which we assume for condition analysis anyhow), then (6.22) can be seen as an approximation to the first-order Taylor approximation of (6.21).

Note that the form from (6.22) is a sufficient condition for a spherical energy, since we request that it holds for all appropriate v . To illustrate this, let us consider the energy $E_D = \phi^\top H_{E_D} \phi$, and as v the rescaled eigenvectors of H_{E_D} . If H_{E_D} is not spherical, then certain eigenvalues of H_{E_D} will differ, and for the corresponding eigenvectors, the values of (6.22) will also be different. This shows that if E_D is quadratic, and (6.22) holds for all v with $\|v\| = r$, then E_D is spherical and thus perfectly conditioned.

If the energy is not perfectly conditioned, then the term in (6.22) is not equal for all possible directions v , e.g. for the eigenvectors of H_{E_D} . Therefore, we measure the quality of the shape of E_D by the variation of the values of (6.22) for all possible directions v , which is bounded by

$$\tilde{\kappa}(E_D) = \frac{\max_v \langle \nabla E_D(\phi' + v), v \rangle}{\min_v \langle \nabla E_D(\phi' + v), v \rangle} . \quad (6.23)$$

We employ $\tilde{\kappa}$ as our model of the condition of the problem. A high $\tilde{\kappa}$ corresponds to high variation of (6.22) and describes an ill-conditioned problem, while a perfectly conditioned function yields $\tilde{\kappa} = 1$.

Our goal will be to determine a preconditioner P , which minimizes or at least reduces the value of

$$\tilde{\kappa}_P(E_D) = \frac{\max_v \langle P^{-1} \nabla E_D(\phi' + v), v \rangle}{\min_v \langle P^{-1} \nabla E_D(\phi' + v), v \rangle} , \quad (6.24)$$

compared to the original problem, that is to no preconditioning, i.e. $P = \text{Id}$.

6.3.2 Restriction of Analysis to Permissible Directions

For optimization of difference measures in deformable registration, we have to restrict the set of allowed directions v , since in general, the optimization of the

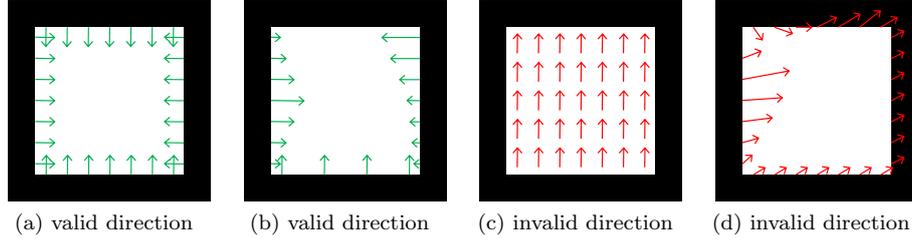


Figure 6.6: Illustration of the restriction to permissible test directions for condition analysis. Directions v are represented by displacements, superimposed onto the source image warped by a scaling transformation. **(a)** represents direction corresponding to the original gradient ∇I_S , and **(b)** illustrates a general valid field, obtained by point-wise multiplication of $\nabla I_S(x)$ by $\alpha(x) \geq 0$. **(c)** is invalid since its application results in no change of energy. **(d)** is invalid as the point-wise vectors $\nabla I_S(x)$ are re-oriented, thus containing a component along the image level set lines, which does not change the energy.

unregularized difference measure subject to a high-dimensional deformation is an under-constrained problem, and the condition is infinitely bad per definition. However, we are interested in improving the condition along the constrained dimensions of the problem.

The optimization of a difference measure is under-constrained since not all deformations result in changes of the energy. For example, this is the case for deformations in homogeneous image areas or along the level set lines of the source image. Therefore, we exclude such under-constrained directions from the analysis, and focus on deformations which do not contain any under-constrained components. Such “pure” displacements v are characterized by being point-wise parallel to the corresponding intensity gradient of the source image, i.e. $v(x) \parallel \nabla I_S(x)$. It is important to note that - with the assumption of group structure for deformations - the original energy gradient has the above permissible structure through $\nabla E_D(\phi)(x) = \omega_D(x) \nabla(I_S \circ \phi)(x)$, cf. e.g. [Chefd’hotel et al., 2002]. For an illustration of the permissible directions, please see Figure 6.6.

Following the above argument, we will restrict the directions v , in a way which is a necessary condition for an actual change in the energy. This can be done by assuring that the point-wise directions $v(x)$ are aligned with the point-wise gradient of the source image $\nabla I_S(\phi')(x)$.

$$v(x) = \tilde{\alpha}(x) \nabla I_S(\phi')(x) = \alpha(x) \frac{\nabla I_S(\phi')(x)}{\|\nabla I_S(\phi')(x)\|}, \quad (6.25)$$

with $\alpha : \Omega \rightarrow \mathbb{R}$. We define the spatial subset $\Omega' = \{x \in \Omega : \|\nabla I_S(\phi)(x)\| \neq 0\}$, and set

$$\alpha(x) \geq 0 \quad \text{for } x \in \Omega' \quad (6.26)$$

$$\alpha(x) = 0 \quad \text{for } x \in \Omega \setminus \Omega'. \quad (6.27)$$

Scaling by α instead of $\tilde{\alpha}$ in (6.25) will facilitate the further analysis. Please note that, given $\nabla I_S(\phi')$, the directions v are completely determined by the choice of α .

In order to restrict the directional vectors v to a sphere, such that (6.25) complies with (6.22), α has to be chosen such that $\|v\| = r$, which gives the following condition

$$r^2 = \|v\|^2 = \sum_{x \in \Omega'} \alpha(x)^2 . \quad (6.28)$$

It is important to note that the choice of α is not unique, and we define the set of all permissible α values as A . Also, due to the construction in (6.25), A is only dependent on $|\Omega'|$, i.e. the number of non-zero point-wise vectors $\nabla I_S(x)$, and not their actual magnitudes or directions.

For the analysis we will employ the directions v as defined in (6.25), and substitute these to the criterion (6.22). Thus, we have to employ the test deformations φ with

$$\varphi = \phi' + v . \quad (6.29)$$

with v and α as defined in (6.25) and (6.28).

Also, we restrict the class of preconditioners P , such that $P^{-1}\nabla E_D$ is permissible in the above sense. This is achieved by restricting P to a point-wise multiplication, such that

$$[P^{-1}\nabla E_D](x) = p(x)^{-1}\nabla E_D(x) , \quad (6.30)$$

with $p \in \mathbb{R}$ and $p > 0$. Our scheme from (6.5) falls within this class with

$$p(x)^{-1} = \frac{1}{\|\nabla E_D(\phi(x))\| + \sigma} . \quad (6.31)$$

It is important to point out that due to this restriction we analyze only preconditioning terms which perform a point-wise multiplication.

6.3.3 Condition Analysis of Difference Measures

Finally, we argue that the proposed scheme actually improves the condition of the problem, by showing that the proposed P reduces the value of (6.24), compared to the original, un-preconditioned problem (i.e. $P = \text{Id}$). Please note that this result implies that the condition of the original problem was not optimal to start with.

For the shape of a preconditioned difference measure, we find by applying a

preconditioned version of (6.22) to directions from (6.25) that

$$\langle P^{-1}\nabla E_D(\varphi), v \rangle = \sum_{x \in \Omega'} \left\langle P^{-1}\nabla E_D(\varphi), \alpha \frac{\nabla I_S(\phi')}{\|\nabla I_S(\phi')\|} \right\rangle \quad (6.32)$$

$$= \sum_{x \in \Omega'} \left\langle p^{-1}\omega_D(\varphi)\nabla I_S(\varphi), \alpha \frac{\nabla I_S(\phi')}{\|\nabla I_S(\phi')\|} \right\rangle \quad (6.33)$$

$$= \sum_{x \in \Omega'} p^{-1}\alpha\omega_D(\varphi) \left\langle \nabla I_S(\varphi), \frac{\nabla I_S(\phi')}{\|\nabla I_S(\phi')\|} \right\rangle \quad (6.34)$$

$$\approx \sum_{x \in \Omega'} p^{-1}\alpha\omega_D(\varphi) \left\langle \nabla I_S(\varphi), \frac{\nabla I_S(\varphi)}{\|\nabla I_S(\varphi)\|} \right\rangle \quad (6.35)$$

$$= \sum_{x \in \Omega'} p^{-1}\alpha\omega_D(\varphi) \frac{\|\nabla I_S(\varphi)\|^2}{\|\nabla I_S(\varphi)\|} \quad (6.36)$$

$$= \sum_{x \in \Omega'} p^{-1}\alpha\omega_D(\varphi)\|\nabla I_S(\varphi)\| \quad (6.37)$$

$$= \sum_{x \in \Omega'} p^{-1}\alpha\|\omega_D(\varphi)\nabla I_S(\varphi)\| \quad (6.38)$$

$$= \sum_{x \in \Omega'} p^{-1}\alpha\|\nabla E_D(\varphi)\| . \quad (6.39)$$

In the above we omit the spatial point x as an argument from $\alpha(x)$, $\omega_D(x)$, $p(x)^{-1}$ and $\nabla E_D(x)$ on the right-hand side for space reasons. The step to (6.35) is based on the assumption that the displacement v is sufficiently small, which was already required for the standard assumption of condition analysis. For Eq. (6.38), we assume that $\omega_D(x) \geq 0$ with the following argument. Since v is leading away from the optimum, we know that it must result in an increase of the energy E_D . For small v , the linear approximation of E_D is dominating, and for $\|v\| \rightarrow \epsilon \approx 0$, the following should hold

$$\langle \nabla E_D(\phi' + v), v \rangle = \langle \nabla E_D(\phi' + v), \tilde{\alpha}(\phi')\nabla I_S(\phi') \rangle > 0 . \quad (6.40)$$

Because of $\|v\| \rightarrow \epsilon$, also the following approximation should hold

$$\langle \nabla E_D(\phi' + v), \tilde{\alpha}(\phi' + v)\nabla I_S(\phi' + v) \rangle > 0 . \quad (6.41)$$

By expanding $\nabla E_D = \omega_D\nabla I_S$, we get

$$\langle \omega_D(\phi' + v)\nabla I_S(\phi' + v), \tilde{\alpha}(\phi' + v)\nabla I_S(\phi' + v) \rangle > 0 . \quad (6.42)$$

In order to guarantee (6.41) for all ∇I_S , and all $\tilde{\alpha}$, we have to request that

$$\langle \omega_D(x)\nabla I_S(x), \tilde{\alpha}(x)\nabla I_S(x) \rangle > 0 , \quad (6.43)$$

with all quantities above evaluated at $\phi' + v$ and with $\|v\| \rightarrow \epsilon$. Since we defined that $\tilde{\alpha}(x) \geq 0$ in (6.26), in order for (6.43) to hold, it is necessary that $\omega_D(x) \geq 0$.

By inserting (6.39) into the condition model from (6.24), we obtain

$$\tilde{\kappa}_P(E_D) = \frac{\max_{\alpha} \sum_{x \in \Omega'} \alpha(x) \|p(x)^{-1}\nabla E_D(\varphi)(x)\|}{\min_{\alpha} \sum_{x \in \Omega'} \alpha(x) \|p(x)^{-1}\nabla E_D(\varphi)(x)\|} . \quad (6.44)$$

The above expression can be simplified. For the maximization problem in the numerator in (6.44), we find that

$$\max_{\alpha} \sum_{x \in \Omega'} \alpha(x) \|p(x)^{-1} \nabla E_D(\varphi)(x)\| = \sqrt{\sum_{x \in \Omega'} \|p(x)^{-1} \nabla E_D(\varphi)(x)\|^2} \quad (6.45)$$

by treating the constrained maximization problem by Lagrange multipliers (please see Appendix 6.A for details.). Furthermore, the minimization sub-problem from the denominator of (6.44) boils down to the minimal value of $\|p(x)^{-1} \nabla E_D(\varphi)(x)\|$, since the minimum of a positive linear combination of positive values is equal to the smallest of these values.

In summary, we get that the optimal value of (6.44) equals

$$\tilde{\kappa}_P(E_D) = \frac{\sqrt{\sum_{x \in \Omega'} \|p(x)^{-1} \nabla E_D(\varphi)(x)\|^2}}{\min_{x \in \Omega'} \|p(x)^{-1} \nabla E_D(\varphi)(x)\|}. \quad (6.46)$$

Note that (6.46) is independent of α or overall scaling of ∇E_D . The minimum of (6.46), and thus the optimal condition, is obtained if the magnitudes of all non-zero point-wise entries are equal: Assuming a constant value for $\|p(x)^{-1} \nabla E_D(\varphi)(x)\|$, for all $x \in \Omega'$, an increase of $\|p(x)^{-1} \nabla E_D(\varphi)(x)\|$ for any $x \in \Omega'$ will increase the numerator of (6.46), while the denominator remains unchanged.

Please note that the constant value of $\|p(x)^{-1} \nabla E_D(\varphi)(x)\|$ for all $x \in \Omega'$ corresponds to a gradient field obtained by the proposed preconditioning (6.5) for $\sigma \rightarrow 0$. On the other hand, for $\sigma \rightarrow \infty$, we arrive at the original energy shape (i.e. $P = \text{Id}$), please compare also Figure 6.7. This shows that for finite values of σ , the proposed scheme improves the condition of the original problem, and for $\sigma \rightarrow 0$ it actually approximates the optimal case for the examined model. More specifically, this shows that for our model two things hold:

- the proposed preconditioning improves the condition for any image-based difference measure
- Among all strategies which perform a rescaling of the point-wise gradient vectors $\nabla E_D(x)$, our proposed strategy approximates the optimal scheme for the analyzed model with $\sigma \rightarrow 0$.

Of the two statements, we are more interested in the first one, since our goal was to demonstrate the improvement of convergence with our approach. Regarding the claim about the approximation to the optimal strategy, it is important to remember that this holds only for our restricted model involving all the approximations assumed, and therefore cannot be expected to translate to the general setting. Still, since we consider the model assumptions to be reasonable, we see the above result as a strong indication that a noticeable improvement of convergence speed can be expected in real applications.

6.4 Evaluation

In the following, we present the evaluation of the proposed “egalitarian” preconditioning scheme in two registration settings: for the demons method in Section 6.4.1, and for the registration in groups of diffeomorphisms in Section 6.4.2.

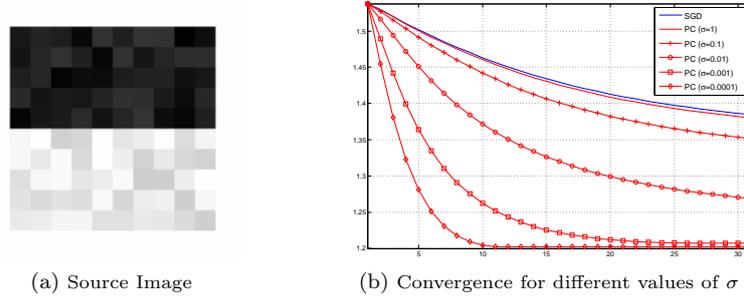


Figure 6.7: Influence of the value of σ parameter. Test run with MI, on a synthetic 3D data set of the pattern as shown in (a). Small values of σ result in a stronger preconditioning, and for large σ , the proposed method behaves as steepest descent.

6.4.1 Evaluation for the Demons Method

This section summarizes the evaluation of the application of the proposed preconditioning in the demons framework, as presented in [Zikic et al., 2011]. We first provide an example of the influence of the σ parameter in Section 6.4.1.1, and then evaluate the fluid and elastic demons versions on 3D brain data in Section 6.4.1.2.

6.4.1.1 Influence of the σ Parameter

The performance of the proposed method depends on the setting of the σ parameter from Eq. (6.5). For large values of σ , the preconditioning effect disappears, and the proposed method behaves as SGD. On real data, too low values of σ will enhance noise, which will lead to inrobust performance. To visualize the effect of the σ parameter in a multi-modal setting, we perform a test with MI, on a synthetic 3D data set, with the source image I_S shown in Figure 6.7(a). The corresponding target image was set to $I_T = 1 - I_S \circ \phi_{GT}$, and the ground truth ϕ_{GT} was generated by a B-Spline FFD. Figure 6.7(b) shows the increasing preconditioning effect for decreasing values of σ . In all our experiments, the σ value is given relative to $\max_x \|\nabla E_{D1}(x)\|$ in first iteration by stating σ' , s.t. $\sigma = \sigma' \max_x \|\nabla E_{D1}(x)\|$. The actual choice of σ depends on the used difference measure, the level of noise in the input images, and the chosen amount of regularization. From our experience, smaller σ values are suited for: (1) SSD rather than MI (probable reason: approximations in implementation of MI introduce “noise”); (2) images from same rather than different modalities; (3) strong regularization, which effectively counteracts noise.

6.4.1.2 Tests on Brain Images

We test on simulated 3D MRI brain images (T1,T2,PD) from the BrainWeb project [Cocosco et al., 1997], with noise-level of 3%, intensity non-uniformity of 20%, and element spacing of 1mm/vx. We perform the 6 possible registrations between the different modalities with different algorithm settings.

The target image is created by applying a ground truth displacement u_{GT}

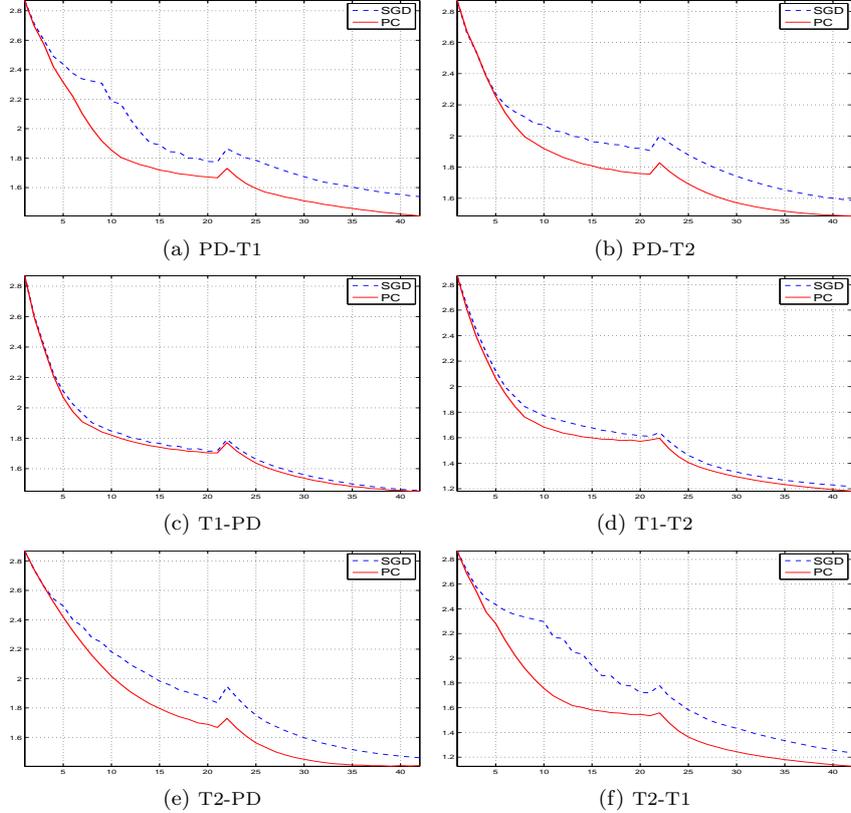


Figure 6.8: **Fluid Demons.** Comparison of proposed method (PC) and steepest descent (SGD) for fluid demons. As in all following figures, we show the convergence plot measuring the development of the mean end-point error in [mm] over iterations.

to one of the input images. The quality of the results is assessed by the mean end-point error in the region of interest Ω_M (the head), measured in millimeters.

The ground truth fields are generated in two steps. This two step approach is performed to generate random deformations, which do not have high-frequency displacements in homogeneous areas, which can not be recovered accurately by any method known to us, and can overshadow the results of the evaluation. While seemingly complex, such an approach is commonly used, cf. e.g. [Modat et al., 2010]. First, a combination of cubic B-spline FFDs with different resolutions is created. This field is used to warp the one of images, and then, a registration with the DROP software [Glocker et al., 2008a] is performed. The resulting deformation is employed as the ground truth field in the experiments. The second step produces deformation fields which are mostly smooth in homogeneous image regions, and thus reduces the amount of this regularization-related error in the experiments. We employ DROP since it is based on gradient-free optimization and thus can be expected to be less affected by the condition of difference measures.

The implementation of the MI follows [Hermosillo et al., 2002] with chosen

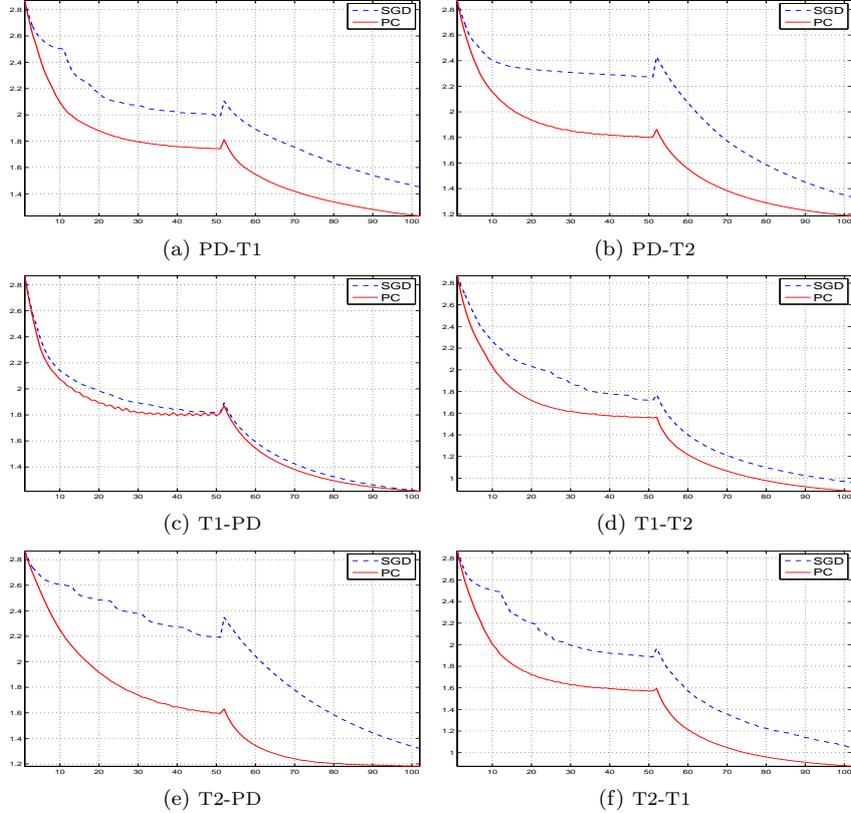


Figure 6.9: **Elastic Demons #1 (Small Displacement Setting)**. Comparison of proposed method (PC) and steepest descent (SGD) for elastic demons with point-wise updates limited to 1mm. This setting is relevant for demons approaches require small updates to generate diffeomorphisms.

histogram size of 40 and the standard deviation of 0.1 for the Parzen windowing. The tests are performed on two levels to simulate a realistic application setting.

As for σ , the difference weight γ is defined as $\gamma = \gamma' / \max \|\nabla E_{D1}(x)\|$ by setting γ' . This facilitates the use of different energy measures. We perform no explicit step size search. The only modification of the step size τ is performed if the maximal update exceeds a certain given value μ . Then, τ is modified, s.t. $\max \|f(x)\| = \mu$.

Fluid Demons. The first test is performed with fluid demons. For the coarse and fine level, the settings are $\sigma_H = 4, 4[\text{mm}]$, $\gamma' = 1, 2$, $\tau = 1, 1$, $\mu = 1, 1[\text{mm}]$. For the proposed method, we set $\sigma' = 0.1, 0.1$. The results of the test are summarized in Figure 6.8.

Elastic Demons #1. (Small Displacement Setting) The first test with elastic demons is performed with limited update steps, with the limits $\mu = 1, 1[\text{mm}]$. This approach is of particular interest, as several demons approaches require small updates in order to generate diffeomorphisms, cf. e.g. [Chefd'hotel

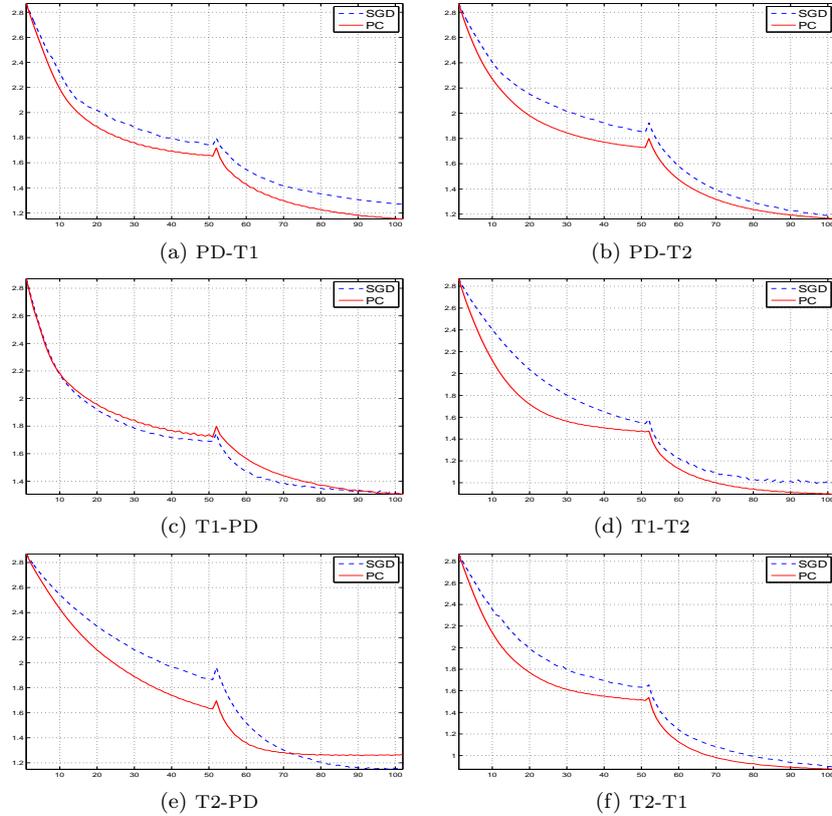


Figure 6.10: **Elastic Demons #2**. Comparison of proposed method (PC) and steepest descent (SGD) for elastic demons. Point-wise updates are uncontrolled for PC (however, always smaller than 3mm), and for SGD had to be limited 10 and 2mm on the coarser and finer level respectively, to avoid divergence.

et al., 2002, Vercauteren et al., 2008], and the proposed method could thus yield a faster convergence for these approaches. The remaining parameters were: $\sigma_{\text{fl}} = 2, 2[\text{mm}]$, $\sigma_{\text{el}} = 1.4, 0.7[\text{mm}]$, $\gamma' = 1, 2$, $\tau = 1, 1$. For the proposed method, we set $\sigma' = 0.4, 0.4$. The results are given in Figure 6.9.

Elastic Demons #2. The steepest descent method can yield significantly better results if larger updates are allowed. In this case, the result is improved since the majority of updates are assigned larger and more meaningful values, however, the maximal updates become too large and lead to local oscillations, which results in divergence.

In this experiment, the proposed method (PC) did not have to be limited, and the maximal occurring updates were always below 3 mm. For SGD, a limit of 10 and 2 mm had to be imposed for the coarse and the fine level, to avoid divergent behavior on several data sets. The remaining joint parameters were: $\sigma_{\text{fl}} = 2, 2[\text{mm}]$, $\sigma_{\text{el}} = 1.4, 0.7[\text{mm}]$, $\tau = 1, 1$. For SGD, we had $\gamma' = 2, 1.3$, and for PC we used $\sigma' = 0.4, 0.4$, and $\gamma' = 1.5, 1.3$. See Figure 6.10 for the results.

Discussion. The results show that in general, the proposed method requires a smaller number of iterations to reach the same level of error as SGD. We observe this improvement in convergence speed in almost all tests, compared to a carefully tuned SGD method. Please note that we use the same preconditioning parameter $\sigma' = 0.1$ in all fluid tests, and $\sigma' = 0.4$ in all elastic tests.

With respect to the actual computation time, the proposed scheme introduces only a minimal overhead. As an example, for the fluid demons test (cf. Fig 6.8), SGD takes 114 sec, compared to 116 sec for the PC version (C++ implementation, CPU: Intel®Core™2 Duo P8700 2.53GHz, RAM: 4GB). For elastic demons, which features an additional costly smoothing step, the difference in runtime is even less prominent.

The effect of the preconditioning depends mostly on the distribution of gradient magnitudes and is more prominent for images with multiple distinct clusters of gradient magnitudes (CGM). For example, in Figure 6.7, I_S has two CGMs: one with weaker gradients between most blocks, and one with stronger gradients resulting from the high-contrast horizontal line. Without the “stronger” CGM, the preconditioning influence would be less prominent. Consequently, the effect for real data also depends on its CGMs. We included all results to show that in some cases the effect can be limited - e.g. for the combination of T1 and PD. Our statement is that our simple technique performs at least as well as SGD in general, and better in most cases, without requiring tedious tuning.

The results for SGD are better in Figure 6.10 than in Figure 6.9 due to an extreme tuning of the step sizes. These are on the limit of the robust behavior for SGD and reaching such performance requires manual tuning for each data set. Our approach reaches its result with more conservative settings.

A further interesting observation is that the performance of our method seems to be less dependent on attribution of image modality to source or target: e.g. the results of PD-T1 and T1-PD should lead to same errors (Figure 6.9: a,b). This is not too surprising, since for SGD, the process depends heavily on I_S through $\nabla E_D(x) = \omega_D(x)\nabla I_S(x)$ (cf. Sec. 6.1.2), and thus on the choice of I_S . In our approach this dependence is strongly reduced by the preconditioning.

6.4.2 Evaluation for Registration in Groups of Diffeomorphisms

In the following we summarize our evaluation of the application of the proposed preconditioning for registration in groups of diffeomorphisms as presented in [Zikic et al., 2010a].

We compare the standard and the proposed approach for the \mathbb{H}^1 Sobolev space. Specifically, for the original approach, the update is based on the gradient as

$$-\tau(\text{Id} - \alpha\Delta)^{-1}\nabla E(\phi) , \quad (6.47)$$

while the generalized Sobolev flow employs the preconditioning term P , and is based on

$$-\tau(P(\phi) - \alpha\Delta)^{-1}\nabla E(\phi) . \quad (6.48)$$

We employ regularization energy as a regularization term. To remain of the manifold of diffeomorphisms, we employ the composition of small updates as the integration method, as discussed in [Chefd’hotel et al., 2002]. Please note

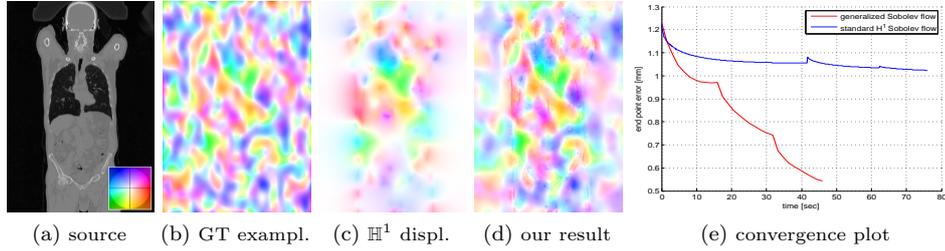


Figure 6.11: Random convergence study in a mono-modal setting with SSD. Results in (e) are the mean of 100 trials, w.r.t. computation time. Displacements in (b)-(d) are color-coded, c.f. (a). The proposed method clearly outperforms the standard \mathbb{H}^1 flow in terms in speed.

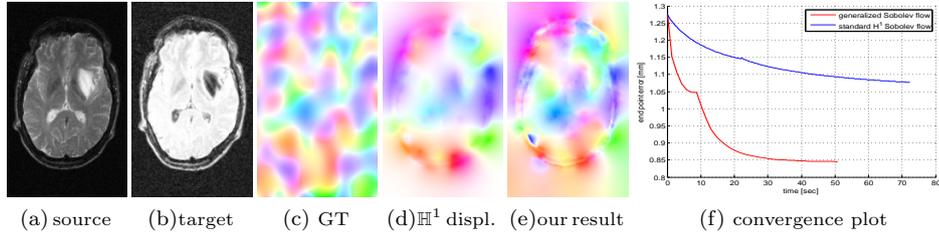


Figure 6.12: Random study in a multi-modal setting, using MI, demonstrates applicability to statistical similarity measures. (b)-(e) depict entities from one trial. Again, the proposed scheme exhibits faster convergence speed, and improved runtime.

that in [Zikic et al., 2010a], we actually use a slightly different implementation of (6.6), which however has the same normalizing effect.

We perform 2D random studies in a controlled environment with known ground truth to demonstrate the improvement in convergence and precision, which result from the proposed approach. Per study, we perform 100 trials in each of which the source image is warped by a random ground truth deformation ϕ_{GT} , generated by B-Spline FFDs, with maximal displacements of 5mm. Method parameters (α , τ) for the standard \mathbb{H}^1 approach are carefully tuned for best possible performance. We monitor the mean euclidean distance between ϕ_{GT} and the estimated deformation (end-point error) in every iteration. A standard multi-level scheme is employed.

The first study is performed on a CT image with SSD as similarity measure, and the results are summarized in Figure 6.11.

We demonstrate the applicability of the proposed approach to statistical similarity measures by a study with MI in a multi-modal scenario (Fig. 6.12). To this end, we employ an MR-T2 image (from <http://www.insight-journal.org/RIRE/>), with intensities rescaled to $[0, 1]$ as I_S , and register it to $I_S = \tilde{I}_S \circ \phi_{GT}$ which includes a non-linear modification of intensities by $\tilde{I}_S(x) = I_S(x) \cdot (1 - I_S(x))$, in order to simulate a multi-modal scenario.

We observe a clear improvement in terms of convergence speed and the

actual runtime for the proposed method. The effectively resulting accuracy is also drastically improved, especially in low gradient regions. This is consistent with our choice of the metric with $M_0 = P$. We observed the same behavior in experiments for SAD and CC as similarity measures.

While the single iterations of the proposed method take longer than for the standard \mathbb{H}^1 approach, due to the extreme improvement of convergence rate, far less iterations are needed, which results in a significant reduction of the overall runtime. For example, the results in Fig. 6.11 feature 30 iterations for the proposed, and 550 for the standard \mathbb{H}^1 method.

It is important to note that the decrease of the energy is very similar for both approaches. Based on the inspection of energy logs alone, the standard method might be considered converged even if the actual error is still significant, cf. Figures 6.11, 6.12. This premature convergence is a serious pitfall for real applications in which the actual error cannot be measured.

6.5 Summary and Discussion

In this work, we present a simple and theoretically justified preconditioning scheme for arbitrary difference measures in deformable registration.

Our approach is of particular interest for cases where other standard optimization methods become too complex, or are not applicable. Important examples of such scenarios are multi-modal registration problems with statistical difference measures (e.g. MI) and high-dimensional deformation models, and especially the multi-modal demons registration.

Our theoretical analysis demonstrates that the proposed scheme improves the condition of the problem, and that it actually even approximates the optimal case for the examined model. More specifically, we show that 1) point-wise multiplication of $\nabla E_D(x)$ can be seen as an important class of preconditioners for deformable registration, and 2) that the strategy to normalize the point-wise lengths $\|p^{-1}(x)\nabla E_D(x)\|$ to the same value is optimal for this class. Our actual strategy is only an approximation to this optimal case, since it performs a damped normalization, which makes the process robust to noise.

Due to its simplicity, our scheme is not only easy to implement for any difference measure, but its application also has only a negligible computational overhead, such that the improvement in convergence speed is directly transferred to an improvement of runtime. The fact that the proposed method does not require any information from previous iteration steps makes it particularly suited for demons registration - this is in contrast to NL-CG or L-BFGS.

The performed experiments show an improvement of the convergence speed, and actual runtime, compared to steepest-descent-based schemes, which are currently the standard approach for the addressed scenarios.

Regarding future work, the proposed scheme constitutes the framework which can be adapted to specific settings. We see an interesting opportunity to perform this task by employing a more generalized term for the σ parameter. In our work, we used a constant σ in the complete image domain and mostly focused on its role to counteract noise. This was done since our primary goal was the development of the overall framework. However, it is probably more appropriate to use σ to express the overall confidence about the update at a certain point (noise is simply one special characteristic influencing confidence). It is possible

to generalize σ by making it spatially varying, and furthermore it might be possible to learn an a priori σ for specific settings. This approach would go in the direction of our original work in [Zikic et al., 2006b], where the confidence about an update was estimated from its local neighbourhood.

6.A Appendix: Derivation of Optimal Model Energy with Lagrange Multipliers

With $e_i = \|p(x_i)^{-1}\nabla E_D(x_i)\|$, and $\alpha_i = \alpha(x_i)$, for $x_i \in \Omega'$, the term from (6.45) can be written as

$$\max_{\alpha} \sum_{i=1}^N e_i \alpha_i \quad \text{with} \quad 1 = \sum_{i=1}^N \alpha_i^2 \quad (6.49)$$

We apply the method of Lagrange Multipliers to (6.49), which corresponds to solving $0 = \nabla L$ with

$$L(\alpha, \lambda) = \sum_{i=1}^N e_i \alpha_i + \lambda \left(1 - \sum_{i=1}^N \alpha_i^2 \right) \quad (6.50)$$

By taking the derivative w.r.t. the parameters, we get that $0 = \nabla L$ corresponds to

$$0 = e_i - 2\lambda\alpha_i \quad (6.51)$$

and

$$0 = 1 - \sum_{i=1}^N \alpha_i^2 \quad (6.52)$$

Substituting $\alpha_i = 1/(2\lambda)e_i$ (from (6.51)) to (6.52), we get

$$\lambda = \pm \frac{1}{2} \sqrt{\sum_i e_i^2} \quad (6.53)$$

Substituting (6.53) back to $\alpha_i = 1/(2\lambda)e_i$, and with the requirement that $\alpha \geq 0$ from (6.26), we get

$$\alpha_i = \frac{e_i}{\sqrt{\sum_i e_i^2}} \quad (6.54)$$

Now, we can substitute (6.54) to the original maximization of energy from (6.49), and see that the value does not depend on α . So, we get

$$\max_{\alpha} \sum_i e_i \alpha_i = \sum_i e_i \frac{e_i}{\sqrt{\sum_k e_k^2}} \quad (6.55)$$

$$= \frac{1}{\sqrt{\sum_k e_k^2}} \sum_i e_i^2 \quad (6.56)$$

$$= \frac{\|e\|^2}{\|e\|} \quad (6.57)$$

$$= \|e\| \quad (6.58)$$

By inserting the original definition $e_i = \|p(x_i)^{-1}\nabla E_D(x_i)\|$, we get for the expression from (6.45)

$$\max_{\alpha} \sum_i e_i \alpha_i = \|e\| = \sqrt{\sum_{i:x_i \in \Omega'} e_i^2} = \sqrt{\sum_{x \in \Omega'} \|p(x)^{-1}\nabla E_D(\varphi)(x)\|^2} \quad (6.59)$$

DEFORMABLE 2D-3D REGISTRATION[†]

Alignment of angiographic 3D scans to 2D projections is an important issue for 3D depth perception and navigation during interventions. Currently, in a setting where only one 2D projection is available, methods employing a rigid transformation model present the state of the art for this problem. In this work, we introduce a method capable of deformably registering 3D vessel structures to a respective single projection of the scene. Our approach addresses the inherent ill-posedness of the problem by incorporating a priori knowledge about the vessel structures into the formulation. We minimize the distance between the 2D points and corresponding projected 3D points together with regularization terms encoding the properties of length preservation of vessel structures and smoothness of deformation. We demonstrate the performance and accuracy of the proposed method by quantitative tests on synthetic examples as well as real angiographic scenes.

7.1 Introduction

Angiographic imaging is a widely used technique for visualization of vessel anatomy in diagnosis and treatment. During most abdominal catheterizations, contrasted 2D projections from one view are acquired by a C-arm for catheter guidance and treatment monitoring. A 3D angiographic scan such as Computed Tomography Angiography (CTA) or Magnetic Resonance Angiography (MRA) is usually acquired preoperatively to assess the region of interest and identify possible complications for the treatment. This 2D/3D setting is sketched in Figure 7.1. In clinical practice, the available 3D information is currently not brought to the interventional room. In some interventions, 3D intraoperative data is available from rotational angiography. This data set, however, is currently not used for guidance or navigation, where 2D projections are favored since they capture the temporal changes compared to a static 3D scan. Only guided by images from one view, it is often very difficult for the physician to find a path through the patient’s vessel system. This is mainly due to overlap of vessel structures and breathing deformation.

[†]This chapter is based on the joint work with Martin Groher, with the related publications: [Zikic et al., 2008b] and [Groher et al., 2009].

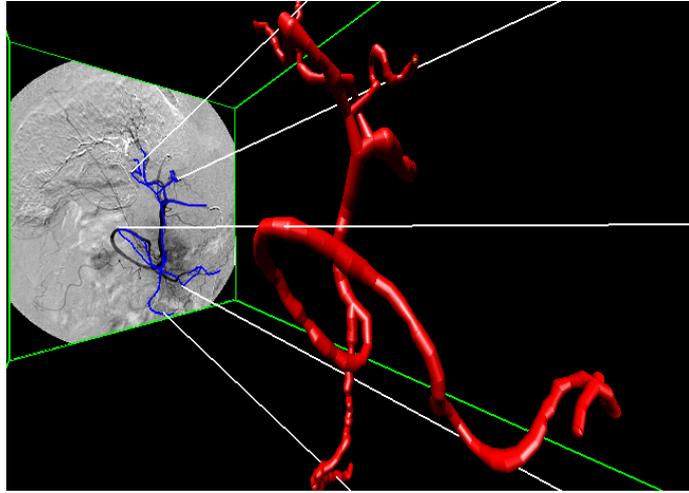


Figure 7.1: Visualization of the 2D/3D C-arm/patient scenario. The 3D vasculature is to be deformably registered to the 2D projection image. The blue line in the 2D Digitally Subtracted Angiogram (DSA) visualizes the projected centerline of the rigidly registered 3D vasculature. This projection is visibly deformed to the projection in the DSA due to breathing motion.

An accurate registration of 3D to 2D vasculature would allow for intraoperative 3D roadmapping or catheter tip backprojection. With this 3D enhancement of angiographic interventions, an increase in depth perception can be achieved while the amount of injected contrast agent and radiation dose can be reduced.

In abdominal or thoracic regions that are subject to deformation, this registration cannot be established by a mere rigid or affine transformation model (for an example, see Figure 7.9). Instead, it is necessary to create a 3D deformation field that locally deforms the 3D vasculature such that its projection matches the 2D vasculature.

For 3D-3D registration of vascular images, methods have been developed to compute the deformation field from sparse correspondences that are determined manually or through rigid pre-alignment [Aylward and Jomier, 2004, Charnoz et al., 2005]. However, the computation of a dense 3D deformation field from sparse 2D-3D feature correspondences is in general an ill-posed problem: The displacement of a point along the projection ray cannot be computed without additional constraints, compare Figures 7.2 and 7.3.

Currently, methods for 2D-3D alignment of vascular images use a rigid transformation model discarding local motion. Such algorithms tend to be robust against deformation changes of vessel structures but do not solve for these deformations, leaving a considerable amount of misalignment, which can be, as reported for e.g. liver, up to 3 cm [Rohlfing et al., 2004].

In order to overcome the shortcomings of the rigid approach, we propose a method for computing a meaningful deformation of a 3D structure from a single 2D projection. Our method combines the correspondence-based approach and the ideas from intensity-based registration, where the registration problem is defined as a minimization of an energy consisting of a difference measure

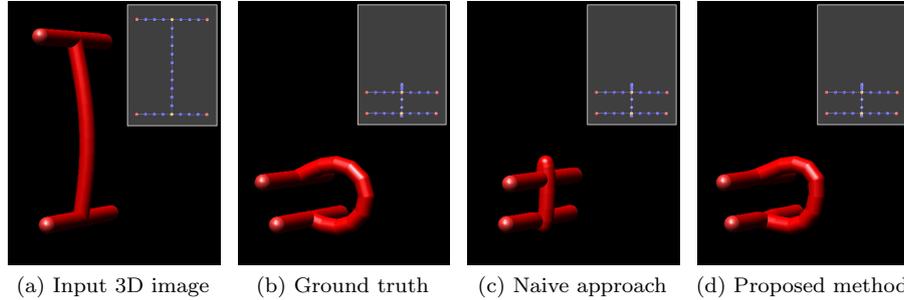


Figure 7.2: Visualization of the method results on a simple synthetic example, with the window-in-window presenting the 2D projection of the respective 3D structure. **(a)** is the input 3D vessel graph to be deformed while **(b)** shows a vessel graph, which was used to generate the input 2D projection image. Hence, **(b)** presents the ground truth for the deformation of **(a)**. The camera is positioned on the right side of the images, such that the shape change in ray direction is not observable from the 2D projection image. **(c)** With the naïve approach using only the distance measure from a single projection, it is not possible to recover the full 3D deformation since there are no constraints along the projection rays. **(d)** Employing the length preservation and diffusion regularization terms present additional constraints and thus allows for correct deformation also in the direction along the projection rays.

and regularization terms, which incorporate the a priori knowledge about the problem, see Figure 7.2.

The difference term used in our approach penalizes the distance between the projection of 3D points from the input vasculature, represented as nodes of a centerline graph, and the corresponding points from the 2D projection image (Figure 7.4).

Minimizing only the difference term results in what we refer to as the *Naïve approach*, which is not able of recovering the deformation in the projection direction and thus can lead to unnatural results. In order to be able to compute the 3D displacement, additionally to the difference, we employ a combination of two regularization terms, which model assumptions about vessel structures and thus yield more realistic deformations.

The first term describes the assumption that the length of vessels does not change heavily inside the human body and penalizes large changes of the vessel length. This term is important since it presents constraints in 3D space and thus reduces the number of solutions for one node from infinitely many to two solutions along the projection ray, if one of the neighbors is assumed fixed (Figure 7.3). Also, in our experiments the minimization of this term by steepest gradient descent results in the nearest solution to the initial position of the respective point. Figure 7.4 illustrates the idea of using the difference term together with length preservation.

However, graphs that are extracted from real vessel structures can have many nodes and large deformations. Here, the length preservation term has the drawback that the behavior is too local. Although the length preservation itself

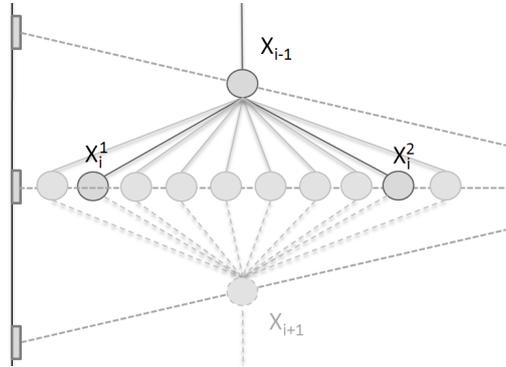


Figure 7.3: Illustrates the reduction of admissible solutions for one node along the projection ray, by using length preservation. For the fixed node eX_{i-1} , the position of the node eX_i is constrained to two possible solutions, eX_i^1 and eX_i^2 , if the distance between eX_{i-1} and eX_i is assumed constant. Without the length constraint, every position along the ray presents a solution, some of which are visualized above.

is performed successfully, in these cases the property that the nearest solution to the initial position is computed introduces unnatural bends in the vessels, thus leading to unwanted results, compare Figure 7.5. In order to counteract this effect, we impose a smoothness condition on the resulting displacement field. To this end, we employ the Diffusion regularization term [Weickert and Schnörr, 2001], which is often used in intensity-based image registration.

So in summary, our method enables meaningful 3D deformations of 3D vessel structures based on a single 2D projection of the same structure. To the best of our knowledge this is the first time that this problem is addressed in the field of medical image processing.

7.1.1 Relation to Prior Work

There is a considerable body of research on *rigid* 2D-3D registration of vascular images, which however mostly addresses rigid structures, for example in neuro surgery, see [Alperin et al., 1994, Feldmar et al., 1995, Kita et al., 1998, Liu et al., 1998, Kerrien et al., 1999, Chung et al., 2002, Hipwell et al., 2003, Chan et al., 2004, Florin et al., 2005]. For the case of abdominal or thoracic 2D-3D image alignment, there exist some methods, which are supplemented by gating information or robustness against deformations [Turgeon et al., 2005, Groher et al., 2007b, Jomier et al., 2006, Groher et al., 2007a]. However, although robust to local transformations, these methods still use a rigid transformation model and do not account for the occurring deformation.

Within the context of registration of two or more 2D projections to an atlas or statistical model of bone anatomy, 2D-3D deformable registration has been addressed by [Fleute and Lavalée, 1999, Benameur et al., 2003, Yao and Taylor, 2003, Zheng, 2006, Tang and Ellis, 2005]. These methods do not focus on vessel anatomy and do not cope with a single view scenario.

Regarding the usage of the constraint of length preservation of vessel struc-

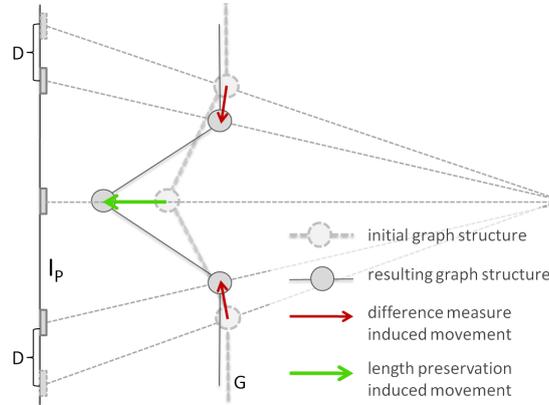


Figure 7.4: 2D illustration of the effect of the difference measure and the length preservation term on the motion of the vessel structure. The difference term D measures the distance in the 2D projection image I_p . The length preservation penalizes the change of length of the 3D graph G .

tures for image registration, to our knowledge, there is no previous work in the literature.

In the robotics and graphics community, computing the 3D pose of a model from a 2D image is regarded as an inverse kinematics problem (see e.g. [Grochow et al., 2004] and references therein), which is somewhat related to our topic. However, the model which is used in these approaches often just has a very limited number of degrees of freedom (DOF) unlike our model, where each feature point introduces 3 DOF.

7.2 Method

The basic idea of the proposed method is to use a difference term and supplement it by regularization terms which incorporate a priori knowledge about the problem and thus impose constraints along the projection rays, which are needed in order to render the problem well-posed.

Having modeled the problem this way, the solution is computed by using an optimization method of choice.

In Section 7.2.1, we first briefly describe the setting for the algorithm and the performed pre-processing steps. We go on by presenting notation and introducing structures we use in Section 7.2.2. Section 7.2.3 will introduce our core model, which is split into difference term (Section 7.2.4), length preservation (Section 7.2.5), and regularization (Section 7.2.6). An algorithmic summary of our model is given in 7.2.7. For a better understanding of the model, we assume correspondence information to be given in sections 7.2.3 - 7.2.7. In Section 7.2.8 we extend our algorithm by an iterative assignment of correspondences using a closest-point criteria. We summarize the overall algorithm in Section 7.2.9 before experimental setups and results are presented in Section 7.3.

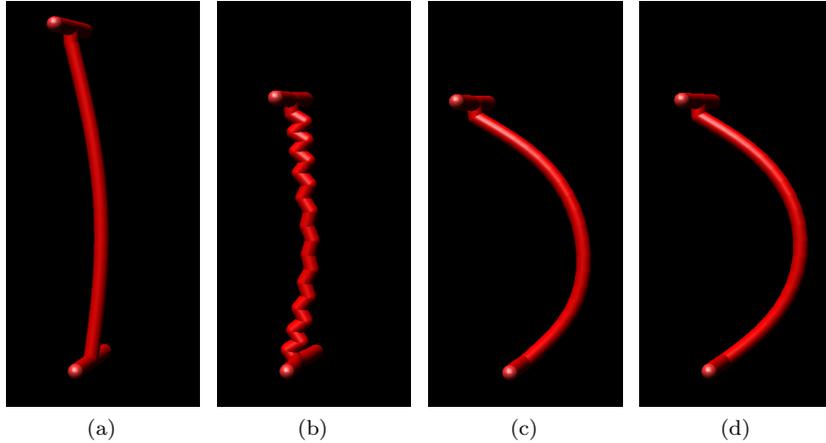


Figure 7.5: Demonstrates the effect of joint usage of the length preservation and the diffusion regularization term. **(a)** Input 3D image. **(b)** Result with Length Preservation. **(c)** Result with Length Preservation and Diffusion Regularization clearly produces a more natural result. **(d)** Ground truth.

7.2.1 Setting and Preprocessing

As input for our method we use an extracted model of 2D and 3D vasculature, as well as a feature-based rigid pre-alignment in a calibrated setting¹ yielding a projection matrix. All of these steps have been previously presented in the literature and are not within the scope of this work. A graph model is created in both 2D and 3D from a region growing step yielding vessel segmentations, followed by topological thinning and bifurcation detection as described in [Selle et al., 2002, Palágyi et al., 2001]. A rigid 2D-3D registration is computed by distance minimization of 2D and projected 3D centerline curves as has been successfully applied to vessels (see e.g. [Feldmar et al., 1995, Groher et al., 2007b]) solving for a projection matrix. 2D and 3D vessel systems are rather different due to local and global contrast injection protocols and segmentation errors. Thus, assigning point correspondences is not straight forward even if a projection matrix is known. In the first sections (7.2.3 - 7.2.7), however, we assume corresponding information to be given. In Section 7.2.8 we incorporate the computation of 2D-3D correspondences into our algorithm by iteratively updating an *assignment matrix*, which stores correspondence probabilities of 3D and 2D points as well as outlier information.

7.2.2 Preliminaries and Notation

We model vessel structures as directed graphs $G^d = (V^d, E^d)$, with a set of n nodes $V^d \subset \mathbb{R}^d$ and the connecting edges $E^d \subset V^d \times V^d$. Here $d \in \{2, 3\}$ denotes the dimension of the graph. For the following, refer also to Figure 7.6.

¹Meaning that intrinsic parameters of the intraoperative imaging device are given. Also, image distortion can be assumed to be absent due to flat-panel detector technology.

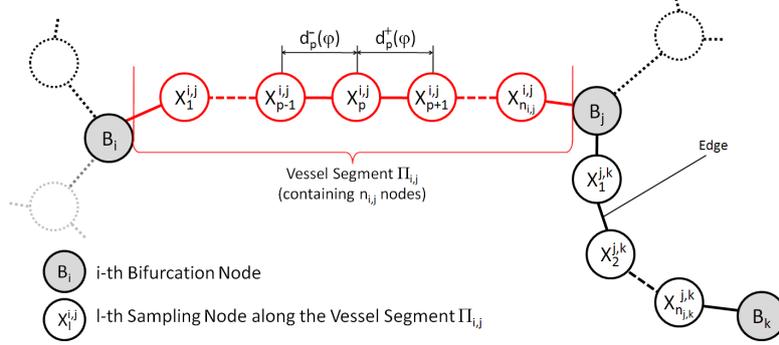


Figure 7.6: Illustration of the used graph structure.

The nodes are classified either as bifurcation nodes V_b^d or sampling nodes V_s^d , such that $V^d = V_b^d \cup V_s^d$ and $\emptyset = V_b^d \cap V_s^d$. While the bifurcation nodes express the topology and the rough geometry of the vessel tree, the sampling nodes are used to describe the geometry of the vessel segments in more detail. The bifurcation nodes are abbreviated by $\mathbf{e}B$ and are identified with their spatial coordinates, such that $\mathbf{e}B \in V^d$. We denote the sampling nodes by $\mathbf{e}X$ in 3D and $\mathbf{e}x$ in 2D and again identify them with their spatial coordinates. The correspondences between the 3D and 2D points are represented by $C \subset V^3 \times V^2$. We define a vessel segment $\Pi_{i,j}$ as a path between two neighboring bifurcation nodes $\mathbf{e}B_i$ and $\mathbf{e}B_j$, containing all sampling nodes and edges between $\mathbf{e}B_i$ and $\mathbf{e}B_j$. The number of nodes in $\Pi_{i,j}$ is $n_{i,j}$ and the number of edges is respectively $n_{i,j} + 1$. The sampling nodes are indexed relative to the vessel segment $\Pi_{i,j}$ starting from 1 to $n_{i,j}$, compare Figure 7.6.

The deformation function is encoded by a set of 3D displacement vectors $\varphi \in \mathbb{R}^{3 \times n}$ centered in the n corresponding graph nodes. The displacement at the i -th node $\mathbf{e}X_i$ is denoted by φ_i , such that the final position of the node is $\mathbf{e}Y_i = \mathbf{e}X_i + \varphi_i$.

We also employ a dense version of the displacement function, which we denote by φ_{TPS} . We obtain φ_{TPS} from φ by interpolation using Thin-Plate Splines (TPS) [Wahba, 1990b]. Interpolating as well as approximating TPS have been successfully applied in deformable registration of medical images, see for instance [Bookstein, 1989b, Rohr et al., 2001b], the latter one being a good introductory reference. In our work we merely use the TPS as an interpolation scheme, i.e. to create a dense displacement field for assigning displacement values to nodes for which no displacement vectors are defined.²

For projections we use a standard pinhole camera model with the principal ray in the direction of the positive Z-axis.

²In order to simplify the implementation, correspondences are computed only for sampling nodes, and thus also the energies are only evaluated there. This technical detail is due to the need to consider predecessor and successor nodes in some parts of the algorithm. Omitting the bifurcation nodes, which often have more than only two neighboring nodes, facilitates the implementation.

7.2.3 The Model

The deformable registration process is now described as a minimization of the energy function \mathcal{E} with respect to the displacements φ of the vessel nodes in order to get the estimate φ' of φ , that is

$$\varphi' = \arg \min_{\varphi} \mathcal{E} , \quad (7.1)$$

with the energy function

$$\mathcal{E} = D + \alpha S_L + \beta S_S , \quad (7.2)$$

where the energy $\mathcal{E} : (G^3, G^2, C, \varphi) \mapsto y \in \mathbb{R}$ consists of a difference term $D : (G^3, G^2, C, \varphi) \mapsto y \in \mathbb{R}$ and regularization terms $S_L : (G^3, \varphi) \mapsto y \in \mathbb{R}$ for length preservation of the vessel segments, and $S_S : \varphi \mapsto y \in \mathbb{R}$ for smoothness of the displacement field. For brevity, we will drop the function arguments in the remainder. The positive scalars α and β control the influence of the respective terms.

In the following, we present the energy terms from Equation (7.2). We also give the respective derivatives which are used in the gradient descent optimization scheme.

7.2.4 Difference Measure

The difference measure D which drives the registration process penalizes the distance between the projection of 3D points from the input graph and the corresponding 2D points from the input projection image.

Given point correspondences C with a single correspondence $C_i = (\mathbf{e}X_i, \mathbf{e}x_i)$ and a projection function $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, we can define the distance measure

$$D = \frac{1}{n} \sum_{i=1}^n \|\mathbf{e}x_i - f(\mathbf{e}X_i + \mathbf{e}\varphi_i)\|^2 . \quad (7.3)$$

Here, $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is a projection function

$$f(\mathbf{e}X) = (\mathbf{e}p_1^\top \mathbf{e}\hat{X} / \mathbf{e}p_3^\top \mathbf{e}\hat{X}, \mathbf{e}p_2^\top \mathbf{e}\hat{X} / \mathbf{e}p_3^\top \mathbf{e}\hat{X})^\top , \quad (7.4)$$

where $\mathbf{e}p_1^\top$, $\mathbf{e}p_2^\top$ and $\mathbf{e}p_3^\top$ constitute the row vectors of the projection matrix $\mathbf{P} \in \mathbb{R}^{3 \times 4}$, and $\mathbf{e}\hat{X} = [\mathbf{e}X^\top, 1]^\top$ is the homogeneous 4-vector representation of the 3D point $\mathbf{e}X$.

For the minimization according to the model (7.2), the derivative of D with respect to φ_k is needed. By using $\mathbf{e}Y_k = \mathbf{e}X_k + \varphi_k$ the gradient is given by

$$\frac{\partial D}{\partial \varphi_k} = -\frac{2}{n} (\mathbf{e}x_k - f(\mathbf{e}Y_k))^\top \mathbf{e}J_k , \quad (7.5)$$

where $\mathbf{e}J_k \in \mathbb{R}^{2 \times 3}$ is the Jacobian of f with respect to $\mathbf{e}\varphi_k$, given by

$$\mathbf{e}J_k = \frac{1}{(\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k)^2} \begin{bmatrix} p_{11}\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k - p_{31}\mathbf{e}p_1^\top \mathbf{e}\hat{Y}_k & p_{21}\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k - p_{31}\mathbf{e}p_2^\top \mathbf{e}\hat{Y}_k \\ p_{12}\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k - p_{32}\mathbf{e}p_1^\top \mathbf{e}\hat{Y}_k & p_{22}\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k - p_{32}\mathbf{e}p_2^\top \mathbf{e}\hat{Y}_k \\ p_{13}\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k - p_{33}\mathbf{e}p_1^\top \mathbf{e}\hat{Y}_k & p_{23}\mathbf{e}p_3^\top \mathbf{e}\hat{Y}_k - p_{33}\mathbf{e}p_2^\top \mathbf{e}\hat{Y}_k \end{bmatrix}^\top \quad (7.6)$$

where p_{ij} denotes the entries of the projection matrix. For a detailed derivation of $\frac{\partial D}{\partial \varphi_k}$ please refer to the Supplementary Material.

7.2.5 Length Preservation Constraint

Since vessel structures are in general enclosed by soft tissue, for example inside liver, and breathing motion is limited to a certain magnitude, the change of length of the vessels is limited. We model this observation by imposing a soft length preservation constraint on the single vessel segments. Thus, we do not impose constant lengths, which would be a too restrictive and unnatural assumption in the given setting. Since the vessel length is defined in 3D space, this constraint is able to induce a deformation orthogonal to projection rays, compare Figure 7.4.

We define the terms $d_i^-(\varphi)$ and $d_i^+(\varphi)$, which measure the length of the edges connected to the sampling node $\mathbf{e}X_i$ for a given set of displacements φ by

$$d_i^-(\varphi) = \|\mathbf{e}Y_i - \mathbf{e}Y_{i-1}\|^2, \text{ and} \quad (7.7)$$

$$d_i^+(\varphi) = \|\mathbf{e}Y_i - \mathbf{e}Y_{i+1}\|^2, \quad (7.8)$$

where we once again set $\mathbf{e}Y_i = \mathbf{e}X_i + \varphi_i$, compare also Figure 7.6. The initial length of the edges connected to $\mathbf{e}X_i$ is now given by $d_i^-(\mathbf{e}0)$ and $d_i^+(\mathbf{e}0)$ where $\mathbf{e}0$ is the zero displacement field.

Now we can define a length preserving cost function as

$$S_L = \frac{1}{n} \sum_{i=1}^n \left| \frac{d_i^-(\mathbf{e}0) - d_i^-(\varphi)}{d_i^-(\mathbf{e}0)} \right|^2 + \left| \frac{d_i^+(\mathbf{e}0) - d_i^+(\varphi)}{d_i^+(\mathbf{e}0)} \right|^2, \quad (7.9)$$

which penalizes the relative deviation from the initial length of the two edges which are directly influenced by the i -th node.

The derivative of S_L with respect to φ_k reads

$$\frac{\partial S_L}{\partial \varphi_k} = \frac{-8}{n} \left[l_k^-(\mathbf{e}Y_k - \mathbf{e}Y_{k-1}) + l_k^+(\mathbf{e}Y_k - \mathbf{e}Y_{k+1}) \right]^\top, \quad (7.10)$$

with

$$l_k^- = \frac{d_k^-(\mathbf{e}0) - d_k^-(\varphi)}{d_k^-(\mathbf{e}0)} \quad \text{and} \quad l_k^+ = \frac{d_k^+(\mathbf{e}0) - d_k^+(\varphi)}{d_k^+(\mathbf{e}0)}. \quad (7.11)$$

The evaluation of the derivative of the length preservation term is performed independently on single vessel segments Π , since for the computation, ordered correspondences and nodes with a left and right neighbor each are needed.

7.2.6 Diffusion Regularization

In order to impose a smoothness constraint onto the displacement field, we employ the regularization term S_S .

We implemented one of the common choices frequently used in intensity-based registration, the so-called Diffusion regularization term (compare e.g. [Weickert and Schnörr, 2001]). Any other standard regularization term, like Bending Energy, compare e.g. [Rueckert et al., 1999b], can be used instead.

For defining the smoothing regularization energy, we employ the interpolating Thin-Plate Spline model φ_{TPS} [Wahba, 1990b] to represent a continuous version of the displacement function, which is explicitly represented at the graph nodes by the vectors φ_i

$$\varphi_{\text{TPS}}(\mathbf{e}X) = \begin{bmatrix} \left(a_0^{(x_1)} + A^{(x_1)\top} \mathbf{e}X + \sum_{k=1}^n \omega_k^{(x_1)} \|\mathbf{e}X_k - \mathbf{e}X\| \right) - \mathbf{e}X^{(x_1)} \\ \left(a_0^{(x_2)} + A^{(x_2)\top} \mathbf{e}X + \sum_{k=1}^n \omega_k^{(x_2)} \|\mathbf{e}X_k - \mathbf{e}X\| \right) - \mathbf{e}X^{(x_2)} \\ \left(a_0^{(x_3)} + A^{(x_3)\top} \mathbf{e}X + \sum_{k=1}^n \omega_k^{(x_3)} \|\mathbf{e}X_k - \mathbf{e}X\| \right) - \mathbf{e}X^{(x_3)} \end{bmatrix}, \quad (7.12)$$

with $A^{(x_k)} = [a_1^{(x_k)}, a_2^{(x_k)}, a_3^{(x_k)}]^\top$, where the scalar values a_i and the vectors ω_k constitute the parameters of the TPS, which are computed to match the n given displacement values at the nodes of the graph, located at points $\mathbf{e}X_k$.

The Diffusion Regularization cost function is then defined as

$$S_S = \frac{1}{n} \sum_{i=1}^n \left\| \nabla \mathbf{e}\varphi_{\text{TPS}}^{(x_1)}(\mathbf{e}X_i) \right\|^2 + \left\| \nabla \mathbf{e}\varphi_{\text{TPS}}^{(x_2)}(\mathbf{e}X_i) \right\|^2 + \left\| \nabla \mathbf{e}\varphi_{\text{TPS}}^{(x_3)}(\mathbf{e}X_i) \right\|^2. \quad (7.13)$$

The derivative of S_S is

$$\frac{\partial S_S}{\partial \varphi_k} = \frac{\partial S_S}{\partial \varphi_{\text{TPS}}(\mathbf{e}X_k)} \quad (7.14)$$

$$= -\frac{2}{n} \Delta \varphi_{\text{TPS}}(\mathbf{e}X_k) \quad (7.15)$$

$$= -\frac{2}{n} \left[\Delta \varphi_{\text{TPS}}^{(x_1)}(\mathbf{e}X_k), \Delta \varphi_{\text{TPS}}^{(x_2)}(\mathbf{e}X_k), \Delta \varphi_{\text{TPS}}^{(x_3)}(\mathbf{e}X_k) \right], \quad (7.16)$$

where the Laplace operator Δ with $\Delta \varphi_k^{(d)} = \partial_{x_1 x_1} \varphi_k^{(d)} + \partial_{x_2 x_2} \varphi_k^{(d)} + \partial_{x_3 x_3} \varphi_k^{(d)}$ is evaluated analytically by computing $\Delta \varphi_{\text{TPS}}^{(x_j)}(\mathbf{e}X)$ to

$$-\sum_{i=1}^3 \sum_{k=1}^n \omega_k^{(x_j)} \left(\frac{-1}{\|\mathbf{e}X_k - \mathbf{e}X\|_\epsilon} + \frac{(\mathbf{e}X_k^{(x_i)} - \mathbf{e}X^{(x_i)})^2}{\sqrt{(\mathbf{e}X_k - \mathbf{e}X)^\top (\mathbf{e}X_k - \mathbf{e}X) + \epsilon^3}} \right). \quad (7.17)$$

The ϵ in the above equation is a small positive scalar, resulting from an approximation to the second norm $\|\mathbf{e}X\|_\epsilon = \sqrt{\mathbf{e}X^\top \mathbf{e}X} + \epsilon$, in the TPS model from Eq. (7.12) in order to ensure its differentiability.

7.2.7 Optimization Scheme

By using all components of the cost function \mathcal{E} together with their gradients, we can give an algorithm based on gradient descent optimization. Since our parameter space has a rather high dimension, the gradient descent optimization can get stuck in local minima. In order to avoid this, we incorporate a relaxation technique into our optimization. In an outer loop, the smoothness parameter β is gradually decreased while the inner part minimizes the energy with a given β in each iteration. The intuition behind this process is to impose a certain rigidity to the transformation if we are far away from the optimum, which is gradually relaxed the nearer we come to the global minimum in order to allow more local transformations. The relaxation process is controlled by a parameter T_{update} , which is usually set between 0.9 and 0.99. See Algorithm 7.1 for the pseudo code of our method.

As optimization method for Algorithm 7.1 we tested a steepest gradient descent as well as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer [Gill et al., 1986]. The latter one brought a considerable speed-up, which will be

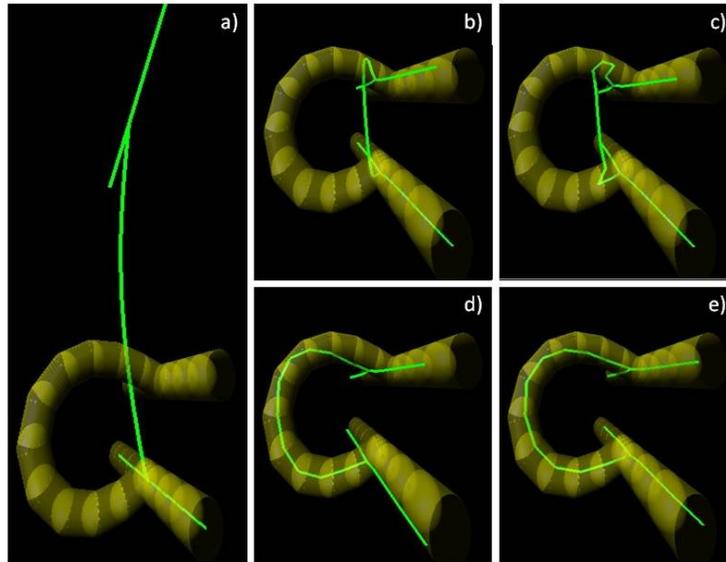


Figure 7.7: Exemplary iteration process of the proposed 2D-3D deformable registration method with steepest gradient descent optimization. In each figure the yellow transparent surface represents the ground truth, the green line represents the centerline of the deformed 3D graph. The 2D graph to which the 3D graph is to be registered is not shown. The input setting is visualized in (a). The intermediate result is shown after 85 (b), 1012 (c), 3674 (d), and 7670 (e) iterations.

discussed in Section 7.4. For BFGS optimization we used the publicly available C++-routines of Jorge Nodcal³.

The values for the coefficients α and β are chosen empirically in the current implementation. For more details on parameter selection and definition of convergence criteria, please refer to Section 7.3.1.

Figure 7.7 shows an exemplary iteration process of the 2D-3D registration algorithm with steepest gradient descent optimization on a synthetic example. Note that the upper and lower part of the “C” shape are bent in the wrong direction after 85 iterations (Figure 7.7(c)), but the algorithm recovers from this situation as can be observed in Figures 7.7(d) and 7.7(e).

7.2.8 Solving for 2D-3D Correspondences

We now extend our model to cope with a more general setup where corresponding information between 2D and 3D graph points is not known a priori.

There are many approaches to solve the correspondence problem while computing a transformation, possibly the most popular being the Iterative Closest Point (ICP) [Besl and McKay, 1992, Zhang, 1994], which has been adapted and applied to medical data rather extensively [Penney et al., 2001, Granger and Penec, 2002, Estepar et al., 2004]. We choose the model of [Gold and

³<http://www.alglib.net/optimization/lbfgs.php>

Algorithm 7.1 Deformable 2D-3D Registration with Additional Constraints

Given the input graphs G^3 and G^2 , ordered point correspondences $C_i = (\mathbf{e}X_i, \mathbf{e}x_i)$, and a projection matrix \mathbf{P} ,

- 1: initialize parameter $T_{update}, \beta_{init}, \beta_{final}$
 - 2: $\beta \leftarrow \beta_{init}$
 - 3: **repeat**
 - 4: {perform optimization}
 - 5: **repeat**
 - 6: calculate $\nabla\mathcal{E} = \nabla D + \alpha\nabla S_L + \beta\nabla S_S$
 - 7: update displacements via a gradient descent step
 - 8: update the 3D TPS and deform whole graph
 - 9: **until** convergence
 - 10: {update β }
 - 11: $\beta = T_{update} * \beta$
 - 12: **until** $\beta < \beta_{final}$
-

Rangarajan, 1996], which is a *soft* version of the assignment as proposed by the ICP, and can be nicely integrated into our model.

Given a variable m_{ij} , where

$$m_{ij} = \begin{cases} 1 & \text{if } \mathbf{e}x_i \text{ corresponds to } \mathbf{e}X_j \\ 0 & \text{otherwise} \end{cases}, \quad (7.18)$$

our energy can be extended to

$$\tilde{\mathcal{E}} = \tilde{D} + \alpha S_L + \beta S_S - \eta P, \quad (7.19)$$

where the new difference term is defined as

$$\tilde{D} = \sum_{i=1}^n \sum_{j=1}^N m_{ij} \|\mathbf{e}x_i - f(\mathbf{e}X_j + \varphi_j)\|^2 \quad (7.20)$$

and a penalization term

$$P = \sum_{i=1}^n \sum_{j=1}^N m_{ij}, \quad (7.21)$$

which avoids trivial solutions. Our optimization problem thus becomes

$$(\varphi', \mathbf{M}') = \arg \min_{\varphi, \mathbf{M}} \tilde{\mathcal{E}}, \quad (7.22)$$

with additional parameters in the assignment matrix $\mathbf{M} = (m)_{ij}$.

In the standard ICP, m_{ij} is a binary variable, which can only take the values of 0 and 1. If this variable is directly put as a parameter into our optimization, many local minima arise and the optimization is likely to fail. Gold and Rangarajan proved that the transition of m_{ij} from a binary to a random variable $m_{ij} \in [0, 1]$, which takes the values

$$m_{ij} = \nu \exp \frac{-\|\mathbf{e}x_i - \mathbf{e}x_j\|^2}{2\nu}, \quad (7.23)$$

will avoid the introduction of too many local minima [Gold et al., 1998].

To guarantee a 1-to-1 mapping, the assignment matrix \mathbf{M} has to be doubly stochastic, i.e. the constraints $\sum_{i=1}^{n+1} m_{ij} = 1$ and $\sum_{j=1}^{N+1} m_{ij} = 1$ must be met, which can be approximated by row and column normalization [Gold et al., 1998].

Outliers are handled by introducing a *slack* row and column to ensure the constraints while being able to assign small probabilities to all entries of a row or column respectively that do not sum up to 1.

To further "convexify" the cost function, a deterministic annealing schedule is proposed, which renders the final energy

$$\tilde{\mathcal{E}} = \tilde{D} + \alpha S_L + \beta S_S - \eta P + \tau L \quad (7.24)$$

subject to the above constraints where, $L = \sum_{i=1}^n \sum_{j=1}^N m_{ij} \log m_{ij}$ is a so-called *entropy* term and τ the annealing parameter. The energy is minimized by an alternating scheme, where each iteration first applies the softassign given a transformation followed by an update of the transformation given correspondences [Gold and Rangarajan, 1996]. An outer loop controls the annealing by gradually decreasing the annealing parameter τ .

This method of "softly" assigning correspondences combined with deterministic annealing has been successfully applied in monodimensional rigid point-based registration [Gold et al., 1998], and has been extended to non-rigid registration as proposed by [Chui and Rangarajan, 2003], who, to this end, introduce a TPS transformation model. Similar to our approach, they gradually decrease the smoothing regularization parameter (β) according to the annealing scheme in order to allow for more local transformations if the algorithm approaches the global minimum.

7.2.9 Overall Algorithm

We adapt the update of correspondences to the projective case, yielding the final algorithm, which solves for both 2D-3D transformation and correspondences, see Algorithm 7.2. The parts, which have been taken from Algorithm 7.1 without change are held in light gray.

For criteria on convergence please refer to Section 7.3.1. It should be emphasized that the row and column normalization includes the slack row and column to allow for outliers. Please note that the update of the assignment matrix takes the projection into account and thus evaluates the Euclidean distance in 2D. Moreover, the 2D correspondences used to compute the transformation are a mixture of all 2D points depending on their respective probabilities (compare [Chui and Rangarajan, 2003]).

7.3 Results and Evaluation

In order to validate our results, besides visual inspection, we compute two different quantitative error measures.

The first measure is the 3D euclidean distance between the nodes of a given graph and the corresponding ground truth (GT) structure. We call this measure position error, which is given by

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{e}X_i - \mathbf{e}X_i^{GT}\|, \quad (7.25)$$

Algorithm 7.2 Deformable 2D-3D Registration with Unknown Correspondences

 Given the input graphs G^3 and G^2 , and a projection matrix \mathbf{P}

- 1: initialize annealing parameters $T_{init}, T_{update}, T_{final}, \beta_{final}$
 - 2: $\tau \leftarrow T_{init}$
 - 3: $\varphi^{(0)} \leftarrow \mathbf{e0}$
 - 4: $\beta \leftarrow T_{init} \cdot \beta_{final}$
 - 5: **repeat**
 - 6: {update assignment matrix according to Equation (7.23)}
 - 7: $m_{ij} = \frac{1}{\tau} \exp \frac{-\|\mathbf{e}x_i - \mathbf{P}(\mathbf{e}X_j + \varphi_j^{(t)})\|^2}{2\tau}$
 - 8: **repeat**
 - 9: {row and column normalization}
 - 10: $m_{ij} = \frac{m_{ij}}{\sum_{k=1}^{n+1} m_{kj}}$, $j = 1, \dots, N$
 - 11: $m_{ij} = \frac{m_{ij}}{\sum_{k=1}^{N+1} m_{ik}}$, $i = 1, \dots, n$
 - 12: **until** convergence
 - 13: {compute new correspondences}
 - 14: **if** $\sum_{i=1}^n m_{ij} > thresh$ **then**
 - 15: $\mathbf{e}x'_j = \sum_{i=1}^n m_{ij} \mathbf{e}x_i$, $j = 1, \dots, N$
 - 16: **else**
 - 17: discard point $\mathbf{e}X_j$ as outlier
 - 18: **end if**
 - 19: {update the transformation on set $C_i = (\mathbf{e}X_i, \mathbf{e}x'_i)$, $i = 1, \dots, N$ }
 - 20: **repeat**
 - 21: calculate $\nabla \mathcal{E} = \nabla D + \alpha \nabla S_L + \beta \nabla S_S$
 - 22: update displacements via a gradient descent step
 - 23: update the 3D TPS and deform whole graph
 - 24: **until** convergence
 - 25: {update annealing parameter}
 - 26: $\tau \leftarrow T_{update} * \tau$
 - 27: $\beta = T_{update} * \beta$
 - 28: **until** $\tau < T_{final}$
-

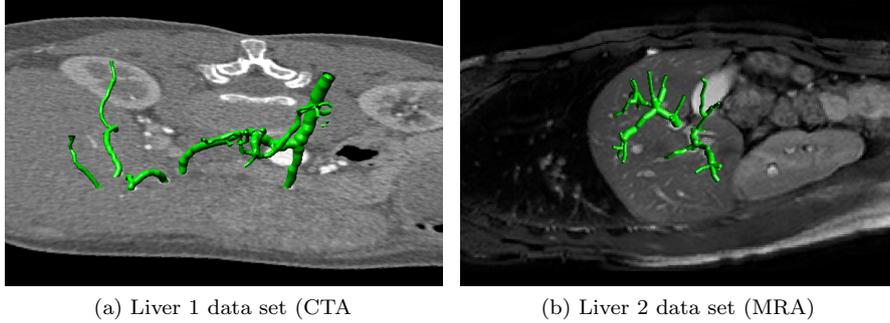


Figure 7.8: Visualization of segmented real vessel structures. **(a)** Liver 1 data set (CTA) taken from a patient suffering from hepatocellular carcinoma. **(b)** Liver 2 data set (MRA) taken from Siebenthal *et al.* [von Siebenthal *et al.*, 2007].

where $\mathbf{e}X_i$ are the positions of all nodes of a graph after registration, and $\mathbf{e}X_i^{GT}$ are the positions of the corresponding nodes from the ground truth graph.

Since this first measure does not take topology into account we also introduce a second measure, which does not penalize the position, but only evaluates the shape. At every node, the angle between the two adjacent edges is computed. This measure is called the shape error, which is given by

$$\frac{1}{n-2} \sum_{i=2}^{n-1} \left| \arccos \frac{\mathbf{e}L_i^\top \mathbf{e}R_i}{\|\mathbf{e}L_i\| \|\mathbf{e}R_i\|} - \arccos \frac{\mathbf{e}L_i^{GT \top} \mathbf{e}R_i^{GT}}{\|\mathbf{e}L_i^{GT}\| \|\mathbf{e}R_i^{GT}\|} \right|, \quad (7.26)$$

where $\mathbf{e}L_i = \mathbf{e}X_{i-1} - \mathbf{e}X_i$, and $\mathbf{e}R_i = \mathbf{e}X_{i+1} - \mathbf{e}X_i$. $\mathbf{e}L_i^{GT}$ and $\mathbf{e}R_i^{GT}$ are defined in an analogous manner. However, this error measure is suitable mostly for cases with a low number of nodes, since for large vessel systems, small random angle errors sum up to large values not describing the quality of the results in an appropriate manner.

We perform the tests on synthetic graphs with artificial deformations in order to test various aspects of the method. To demonstrate the applicability for real applications, we apply the tests to real vessels segmented from angiographic images, deformed by both, artificial and natural deformation fields. Moreover, we conduct a test on patient data coming from a clinical setting.

7.3.1 Parameter Values

The parameter values used for the experiments were determined empirically. The value of $\alpha = 0.01$, controlling the length preservation term, yields good results for all input data sets. For the smoothness term, $\beta_{final} = 0.1$ was used as lower boundary of the relaxation scheme, i.e. the smoothness parameter was not allowed to drop below this value. The initial value for β at the beginning of relaxation was set to $\beta = T_{init} \cdot \beta_{final}$, as described in Algorithm 7.2. The parameter η is not set by the user, since the row and column constraints on the assignment matrix avoid trivial solutions automatically.

Annealing parameters are chosen such that the final value T_{final} is equal to the minimal distance between two nodes in 2D. T_{init} is set to $500 \cdot T_{final}$ and

the update parameter T_{update} , usually between 0.9 and 0.99, is set to 0.93.

The threshold for outlier removal (see Algorithm 7.2) is set to 0.01, i.e. if the sum of all probabilities in a column is below 1% this 3D point is assumed to have no corresponding 2D point.

There are two inner loops listed in Algorithm 7.2, one for optimization, one for row and column normalization of the assignment matrix. Convergence for the optimization is tested by classical termination criteria, i.e. small parameter update, and small difference in energy values between successive iterations. The normalization loop is repeated 60 times, as suggested by [Chui and Rangarajan, 2003].

7.3.2 Tests on Synthetic Data

For these tests, we use two 3D graphs, where one of the graphs is a deformed version of the other. The deformation is performed such that the length is not changed. To this end, we employ a dedicated deformation function, which is not used in our method itself, in order to assure the validness of comparison. The transformation is computed in the following way. Starting from an initial node, all remaining nodes are processed in a sequential manner. At every node, a rotation is performed about an axis passing through the previous node and being orthogonal to the plane formed by the Y dimension and the projection ray. This ensures that most displacement happens along the projection rays, which gives an appropriate test setting for evaluating the role of the regularization terms. The rotation angle is varied at every node due to a trigonometric function, with a variable frequency and amplitude. The computed rotation is applied to the yet unprocessed part of the graph. Same applies also for the deformation used in Section 7.3.3. The graph, which is not deformed serves as input for the method, while the deformed one presents the ground truth solution. The 3D ground truth is not directly used, but we generate a 2D projection of this structure, which is used as input for the method, together with the projection matrix. For three exemplary data sets (Synth1, Synth2 and Synth3) quantitative and visual results are presented in Tables 7.1, 7.2 and Figure 7.11.

7.3.3 Real Data with Artificial Deformation

In order to assess the behavior of the method on natural vessel structures in a quantitative way, we deform the graphs extracted by segmentation from patient data sets with a length-preserving deformation function, as described in Section 7.3.2. This way, we are able to perform our method and measure the distance of the result to a known ground truth in the same way as for synthetic data sets. A projection matrix computed from a rigid CTA-to-DSA registration of the respective patient is used to create the input 2D vessel graph. For the presented tests, we use a liver data set (Liver 1) from a patient who suffers from hepatocellular carcinoma and was treated with Transarterial Chemoembolization, compare Tables 7.1, 7.2 and Figure 7.8a.

7.3.4 Real Data with Natural Deformation

Natural deformation fields for human organs are hard to obtain. In order to verify our method on possibly natural deformations we employ the results presented

by [von Siebenthal et al., 2007]. The displacement fields provided by this work are computed from a series of contrasted 4D MR images of the liver. A deformable registration is performed in [von Siebenthal et al., 2007] between the single 3D images, where the high time resolution together with the strong texture of the contrasted images assures the quality and reliability of the resulting deformation field. We segment the vessel structures from the contrasted MR images used in [von Siebenthal et al., 2007], and generate the input 3D graph for our method. Then, we apply the displacement field from [von Siebenthal et al., 2007] to the 3D graph and thus compute the ground truth for the result. A projection matrix yielding an anterior-posterior image was used for 2D input creation. In the same way as for the synthetic data sets, the 3D ground truth together with initial and deformed 3D input graph are used to quantitatively assess the performance of our method. Despite the small deformation observable in the data set (Liver 2), a clear improvement is achieved. Compare Tables 7.1, 7.2 and Figure 7.8b.

It can be observed that our method performs slightly better for the Liver 1 and 2 data sets if regularization is turned off and correspondences are known (compare Table 7.1). This is due to the fact that for these data sets the deformation in ray direction is minor and regularization prevents the method to fully match in-plane deformation. Interestingly, however, the method performs better for all data sets if regularization is turned on and correspondences are not known (compare Table 7.2). Our proposed regularization prevents the algorithm to assign false correspondences, which can happen easily when using the naïve method.

7.3.5 Test in a Clinical Setup

This test is performed with known projection matrices and a reference deformation field for comparison. An important issue for the creation of this reference deformation field is the correspondence problem on vascular 3D graphs, which we address in an intuitive manner by resampling and length accumulation in the first test and through careful manual inspection in the second and third tests.

For all tests we use two 3D graphs extracted from a preoperative CTA and an intraoperative cone-beam reconstruction of the same patient undergoing a liver catheterization. Both data sets have been acquired in deep inspiration. The intraoperative reconstruction is created from 395 views with projection matrices known from calibration. The 395 projections cover an angular range of 197° in 0.5° steps on a craniocaudal rotation axis.

First, we rigidly register the two 3D graphs. We manually determine point correspondences of all bifurcation points visible in both data sets. Unlike sampling nodes on vessel segments, the location of bifurcations is well-defined at the junction of a vessel graph and hence correspondency can be established. Then, we rigidly register the two resulting 3D point sets using the least-squares method of Umeyama [Umeyama, 1991a].

7.3.5.1 Single Vessel Segment

Here, we test our algorithm in a simplified yet realistic environment by choosing only one vessel segment of the 3D vasculatures, which emanates from a bifurcation for which correspondency is known. We run our registration on this vessel segment only. Thus, we can automatically compute corresponding points using

Test Type		Position Error [mm]		Shape Error [rad]	
Test	Data	μ	σ	μ	σ
Synth 1	Input	4.46	3.58	0.5847	0.777
	Naïve	0.94 (78.9%)	1.06	0.4115 (29.6%)	0.514
	Result	0.87 (80.5%)	0.30	0.1459 (75.0%)	0.143
Synth 2	Input	1.36	1.09	0.3224	0.363
	Naïve	0.37 (72.8%)	0.47	0.2883 (10.58%)	0.307
	Result	0.17 (87.5%)	0.17	0.0780 (75.8%)	0.072
Synth 3	Input	1.42	0.80	0.3463	0.199
	Naïve	0.72 (49.3%)	0.57	0.2448 (29.3%)	0.174
	Result	0.84 (40.8%)	0.63	0.2218 (36.0%)	0.154
Liver 1	Input	7.38	2.23	0.1675	0.168
	Naïve	3.17 (57.0%)	2.85	0.1266 (24.4%)	0.127
	Result	3.70 (49.9%)	3.08	0.1428 (14.75%)	0.157
Liver 2	Input	1.20	0.65	0.0082	0.009
	Naïve	0.99 (17.5%)	0.68	0.0057 (30.5%)	0.008
	Result	0.99 (17.5%)	0.68	0.0062 (24.4%)	0.008
Liver 3	Input	14.88	15.0	0.2188	0.167
	Naïve	8.33 (44.0%)	4.53	0.2274 (-3.9%)	0.221
	Result	8.24 (44.6%)	5.17	0.2847 (-30.2%)	0.223

Table 7.1: Results of error evaluation on several synthetic and real data sets with given correspondences. The position error by euclidean distance, as well as the shape error by angle measurement is assessed. We give the mean error μ and in order to show the significance of the improvement also the standard deviation σ . For the mean, the relative improvement to the input data is given in percent. For visualization of the settings, compare Figures 7.8, 7.9, and 7.11.

a length preserving constraint and hence reduce a possible bias in ground truth computation.

For computing a reference deformation field, we extract two vessel segments, $\Pi_{i,j}$ from the CTA graph and $\Pi'_{k,l}$ from the reconstruction graph that are manually determined to correspond. The chosen segment exhibits a large deformation, which is assessed after rigid 3D-3D registration of the complete vessel systems (see Figure 7.9a). We now want to establish correspondences between all sampling nodes of $\Pi_{i,j}$ and $\Pi'_{k,l}$ given the initial correspondences $\mathbf{e}B_i \leftrightarrow \mathbf{e}B'_k$.

The nodes on $\Pi_{i,j}, \Pi'_{k,l}$ cannot be assumed to have the same sampling since they have been extracted from two different data sets. Thus, we first apply a resampling to $\Pi'_{k,l}$ to have an inter-node distance which is significantly smaller compared to the inter-node distance of $\Pi_{i,j}$. Then, we assign correspondences to the sampling nodes in the following way: For a node $\mathbf{e}X \in \Pi_{i,j}$ determine its curve length to $\mathbf{e}B_i$, $d(\mathbf{e}B_i, \mathbf{e}X)$. Walk through $\Pi'_{k,l}$ starting from $\mathbf{e}B'_k$ until the first node $\mathbf{e}X'$ has been found with $d(\mathbf{e}B'_k, \mathbf{e}X') \geq d(\mathbf{e}B_i, \mathbf{e}X)$, which is assigned as corresponding node to $\mathbf{e}X$. This procedure is repeated for all nodes in $\Pi_{i,j}$.⁴

⁴This sampling technique introduces an error to the assumption of length preservation,

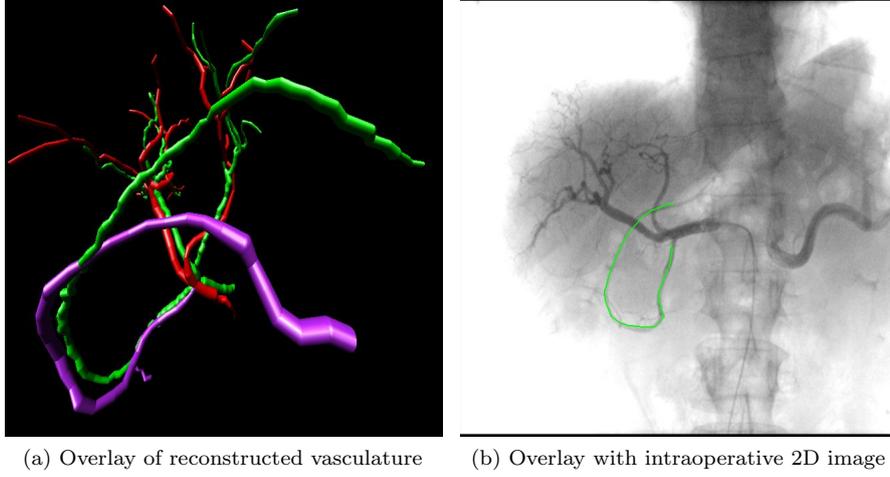


Figure 7.9: Clinical Setup: **(a)** Rigidly registered CTA (green) and intraoperative reconstruction (red and purple) of the vasculature. The purple vessel segment was used for the first clinical test. A large deformation is observable in this vessel compared to the corresponding, intersecting green vessel. **(b)** Projection of vessel segment onto an intraoperative 2D image. The 2D image was used for the intraoperative 3D reconstruction and a projection matrix is known due to calibration.

With the set of correspondences $\{\mathbf{e}X_h \leftrightarrow \mathbf{e}X'_h\}$, $h = 1, \dots, n_{i,j}$ we can compute a 3D Thin Plate Spline to align the two vessel segments $\Pi_{i,j}$ and $\Pi'_{k,l}$. The resulting spline is used to deform $\Pi_{i,j}$ to $\tilde{\Pi}_{i,j}$. With this method - in spite of the discretization error that is introduced - we observed the difference in length of $\Pi_{i,j}$ and $\tilde{\Pi}_{i,j}$ to be smaller than 6%.

As input for our registration method, we use the segment $\Pi_{i,j}$, the segment $\pi'_{k,l}$ which is the projection of $\Pi'_{k,l}$ by the matrix \mathbf{P}_{rec} , and the correspondence information of the 3D segments. \mathbf{P}_{rec} is taken from one of the views used for reconstruction and thus resembles the 2D intraoperative situation, compare Figure 7.9b.

A considerable improvement in position and shape error can be observed when applying our algorithm, despite the large deformation of the vessel segment, compare the entries for Liver 3 in Table 7.1.

A further test has been performed on this vessel segment with unknown correspondences using our extended Algorithm 7.2. The result is summarized in the entries for Liver 3 in Table 7.2. It can be observed that the position error still improves more if regularization terms are incorporated. The shape error slightly increases, but again the impact of the regularization terms is shown when comparing to the naïve method.

Please note that the initial errors for the Liver 3 data set have been acquired after rigid registration.

which is bounded by the inter-node distance of $\Pi'_{k,l}$.

7.3.5.2 Entire Vasculature without Outliers

In this test, we deformed the entire vessel tree of the 3D reconstruction (red and purple vessels in Fig. 7.9a) to match the CTA vasculature (green vessels in Fig. 7.9a). For that we manually identified 3D point correspondences and computed a 3D TPS. Then we registered this deformed vascular system "back" to 395 projections of the non-deformed vasculature using Algorithm 7.2. These projections were created using the vessel tree extracted from the original intraoperative reconstruction and the projection matrices with which it was computed. We use the graph extracted from original 3D reconstruction as ground truth to assess the position error.

Please note that, in order to be able to test the algorithm on a complex data set without outliers, we do not use the CTA vessel system as 3D input graph, but the intraoperative reconstruction, which is deformed to match the CTA vessel system. We conducted a test run, which computed 395 registrations. Moreover, we ran the test once more with both regularization terms set to zero. Both runs were performed with unknown correspondences.

Figure 7.10a shows the position error of the 2×395 registrations. The horizontal line symbolizes the initial error. It can be observed that a naïve registration (without regularization) increases the error whereas our method benefits from the regularization terms to decrease the position error.

For images 180 to 210 and 260 to 280 the results are worse than neighboring results. These cases suffer from a heavy overlap of major vessels in the projections, which hampers the assignment of proper correspondences. This overlap is not present for the images where position errors are smaller after registration. However, in all cases, an improvement can be observed when comparing to the initial position error that was determined after rigid registration.

7.3.5.3 Entire Vasculature in the Presence of Outliers

This test is resembling the clinical scenario most realistically and adds the presence of outliers to the previous test. We used the vessel tree extracted from the preoperative CTA as 3D input. The 2D input graphs are created by projection of the 3D reconstruction that has been acquired intraoperatively. Again, the 395 projection matrices, which have been used in the previous test already, are taken to project the 3D reconstruction. The 3D reference deformation field for error assessment is computed by evaluating a 3D TPS on inverted correspondences used in the previous test.

It should be noted that the two data sets used for registration have been created from global and local injection of contrast agent and thus have a different level of propagation. Usually, dye propagates further down the vessel tree if injected locally through a catheter, which makes small vessels visible in the 3d reconstruction, which cannot be visualized in the CTA data set. Since the extraction of vessel graphs does not change vessel topology, there is a considerable amount of outliers in this test, compare Figure 7.9.

As in the previous test, we computed 2×395 registrations using Algorithm 7.2, one run was conducted using our proposed model, one run with the naïve method, where smoothness and length preservation are excluded from the energy.

Figure 7.10b shows the position error of the 2×395 registrations, again, the horizontal line shows the initial error (after rigid registration). As in the previous

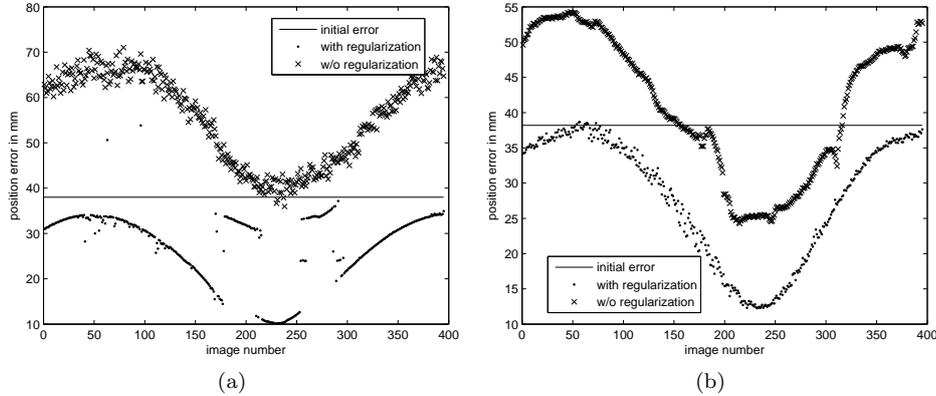


Figure 7.10: Result of the two tests of (a) deformed 3d reconstruction (no outliers) and (b) with original CTA and 395 projections of the intraoperative 3D reconstruction (Sections 7.3.5.2, 7.3.5.3). The dots show the error after registration with our algorithm, the crosses show the error after registration with the naïve method. The horizontal line symbolizes the initial error. It can be observed that a naïve registration increases the error whereas our method benefits from the regularization terms to decrease the position error.

test, a considerable improvement can be observed when using our regularization techniques, whereas the naïve approach increases the error and cannot recover a correct solution.

In the 395 registrations that have been performed in the previous and the current section, the position error can be observed to increase in the first few registrations, decrease until half of the angular run is reached, and increase again until the end of the C-arm rotation. This phenomenon can be explained with the geometric distribution of the liver vessel tree, which is mainly on the frontal plane. Thus, the first and last registrations, which use images of the anatomy from the left or right side of the patient, "compress" the geometric distribution of the vessel tree, which decreases the registration accuracy. However, even in these situations our method can still find a solution, which decreases the position error compared to the initial rigid registration result.

7.4 Discussion

In the previous section we demonstrate that our method performs well in different scenarios. Even in the most general case where complex vascular systems are registered with unknown correspondences and in the presence of outliers, our experiments clearly reveal the impact and importance of our proposed regularization terms together with the difference term minimizing the euclidean error in 2D. We will now first briefly comment the clinical necessity of our proposed deformable registration method, followed by a discussion on the impact of the necessary previous steps on the accuracy, robustness and clinical feasibility. Moreover, runtime issues and possible ambiguities in the computed solution will

Test Type		Position Error [mm]		Shape Error [rad]	
Test	Data	μ	σ	μ	σ
Synth 1	Input	4.46	3.58	0.585	0.777
	Naïve	1.43 (67.74%)	0.94	1.261 (-116%)	1.129
	Result	0.81 (81.83%)	0.40	0.126 (78.4%)	0.135
Synth 2	Input	1.36	1.09	0.322	0.363
	Naïve	0.36 (75.3%)	0.47	0.288 (10.6%)	0.308
	Result	0.31 (77.2%)	0.07	0.023 (77.1%)	0.016
Synth 3	Input	1.42	0.80	0.346	0.199
	Naïve	1.21 (14.8%)	0.95	0.513 (-48.2%)	0.616
	Result	0.89 (37.5%)	0.67	0.271 (21.9%)	0.155
Liver 1	Input	7.38	2.23	0.168	0.168
	Naïve	7.84 (-6.2%)	6.43	0.754 (-350%)	0.636
	Result	3.44 (35.4%)	1.92	0.163 (2.69%)	0.163
Liver 2	Input	1.20	0.65	0.008	0.009
	Naïve	3.8 (-216%)	2.70	0.977 (<-500%)	0.701
	Result	1.15 (4.2%)	0.57	0.007 (9.76%)	0.009
Liver 3	Input	14.88	15.0	0.219	0.167
	Naïve	11.5 (22.7%)	6.40	0.283 (-29.3%)	0.208
	Result	10.71 (28.0%)	6.69	0.268 (-22.3%)	0.204

Table 7.2: Results of error evaluation on several synthetic and real data sets without correspondences. The position error by euclidean distance, as well as the shape error by angle measurement is assessed. We give the mean error μ and in order to show the significance of the improvement also the standard deviation σ . For the mean, the relative improvement to the input data is given in percent. For visualization of the settings, compare Figures 7.8, 7.9, and 7.11.

be addressed.

7.4.1 Vessel Deformation in Clinical Routine

In clinical practice, pre- as well as intraoperative data sets are acquired in deep inspiration while the patient holds his breath. This will usually cause the vessel anatomy to be rather similar and only have a minimal amount of deformation. However, as confirmed by our clinical partners, patients tend to hold their breath in different breathing states before and during the interventions. Moreover, inserted instruments like a guide wire or a catheter can also lead to vessel deformation. These issues can sometimes have a considerable effect on vessel deformation, which has also been shown in our clinical test setup, see Section 7.3.5. Thus, we believe that 2D-3D deformable registration is crucial in this single-view angiographic scenario and will lead to more accurate results in terms of intraoperative navigation and guidance.

7.4.2 Segmentation

Our proposed method operates on 2D and 3D vessel features, which are extracted prior to the registration. The issue of using features rather than intensities for the registration influences our method in terms of accuracy and robustness.

Registration accuracy is directly dependent on the accuracy of the extracted vessel features. Since both, preoperative 3D data sets and intraoperative DSAs have a high intensity gradient (and thus intensity variation) at the vessel boundaries, even simple segmentation techniques like region growing yield results, which can be accurate up to a voxel/pixel precision. As state-of-the-art scanners can reconstruct volumes up to a submillimetric resolution, the metric segmentation error is rather small. In order to run our algorithm on the required features, we need to extract a centerline from the segmentation, which, in our experiments, employs topological thinning as described in [Palágyi et al., 2001]. This algorithm creates an approximation of a skeleton, which resembles the centerline in tubular structures. The approximation error is bounded by the radius of the vessel segments, which is below 2.5 mm for liver arteries. Assuming an error made by the skeletonization approximation of $\frac{1}{4}$ of the radius, the propagated feature extraction error will be less than 0.625 mm corresponding to more than 1 voxel⁵ in 3D and more than 2 pixels in 2D⁶. In summary, neither segmentation nor centerline extraction will introduce a large error. Region growing methods for vascular segmentation are usually very fast and involve a single seed point, which does not create much additional user interaction during an intervention. The process of topological thinning is fully automatic, rendering the feature extraction step feasible for clinical scenarios.

Due to different application of contrast agent in the 2 data sets (global injection in preoperative 3D data, local injection through a catheter in 2D data), our method has to deal with a certain amount of outliers (small vessel segments, which are only visible in 2D, segmented 3D vessel parts, which are not contrasted in 2D). This issue is addressed with the slack row/column in the assignment matrix (Section 7.2.8), which makes the method more robust to outliers. Moreover, as can be depicted from the results in Figure 7.10, length preservation and smoothness regularization penalize solutions where outliers contribute to the computed transformation.

Please note that the error evaluation is not affected by the segmentation error since the ground truth has always been computed from the 3D input vasculature using a deformation field.

7.4.3 Rigid 2D-3D Registration

Our method can be carried out after a rigid 2D-3D registration has been performed since the model utilizes a previously computed projection matrix. Many catheterization interventions provide a preoperative data set from a CTA or MRA scan. In such a case, methods for rigid 2D-3D registration can be used to determine the projection matrix to an intraoperative 2D DSA [Turgeon et al., 2005, Groher et al., 2007b, Jomier et al., 2006, Groher et al., 2007a].

Some hospitals use rotational angiography for acquisition of an intraoperative 3D data set. In this case, a projection matrix that projects the 3D data to match

⁵a typical voxel resolution in CTA data sets is $0.6 \times 0.6 \times 0.6 \text{ mm}^3$

⁶for a typical in-plane pixel resolution of $0.31 \times 0.31 \text{ mm}^2$

a currently acquired 2D DSA can be assembled from calibration data and table position of the imaging device [Gorges et al., 2005].

Please note that our method is not solving for the entries of a projection matrix, but assumes them to be given in advance. Thus, inaccuracies of the estimated matrix directly affect the performance of our method. For intraoperative 3D data sets the estimated projection matrix is usually very accurate (in the range of submillimeters in the image plane) due to a precise offline calibration step [Navab et al., 1996, Livyatan et al., 2002].

2D-3D registration methods for preoperative 3D data sets report in-plane errors below 5 *mm* and out-of-plane errors below 15 *mm* in a single-view scenario [Turgeon et al., 2005, Jomier et al., 2006, Groher et al., 2007a]. We have shown in the experiments that our method can cope with errors up to 3.8 *cm*, which makes a concatenation of rigid and deformable 2D-3D registration feasible.

7.4.4 Runtime

The runtime of our method depends on the number of graph nodes that are used in the optimization step. The 3D graphs used in the clinical evaluation in Section 7.3.5 have 160 and 320 nodes, respectively, which corresponds to a runtime of 3.2 and 4.3 minutes (averaged over the 395 registrations, executed single-threaded on a Intel®Core™2 Duo 3 GHz). Together with 2D segmentation and rigid pre-registration, this sums up to a runtime of ca. 5-7 minutes⁷, which is acceptable in an intraoperative situation as confirmed by our clinical partners.

7.4.5 Ambiguities

Our proposed method converges to the right solution in most cases. However, there are two special initial configurations where the energy function is not guaranteed to drive the gradient descent scheme to the correct solution. Both cases only occur if the deformation changes the sign in vessel curvature along the projection direction.

Typically, a gradient-based optimization algorithm will - given a suitable step size - converge to the “nearest” local optimum. Even if a vessel segment is deformed along the projection direction, the algorithm will compute the right deformation field given a suitable initial position. However, if the vessel segment either deforms from a straight line, or if it changes the sign of curvature during deformation, the proposed method is not guaranteed to converge to the right solution.

A segment, which is a straight line down to the precision of a single voxel is not very probable to be part of real vessel systems. Moreover, vessel systems are rather unlikely to invert their sign of curvature if they undergo a natural deformation due to e.g. patient breathing. Thus, we assume that these special cases do not occur frequently in real scenarios. Furthermore, if the curvature is computed for all vessel segments, straight lines can be detected, and the algorithm can indicate a possible convergence problem.

⁷runtimes for segmentation and 2D-3D rigid registration have been taken from [Groher et al., 2007a]

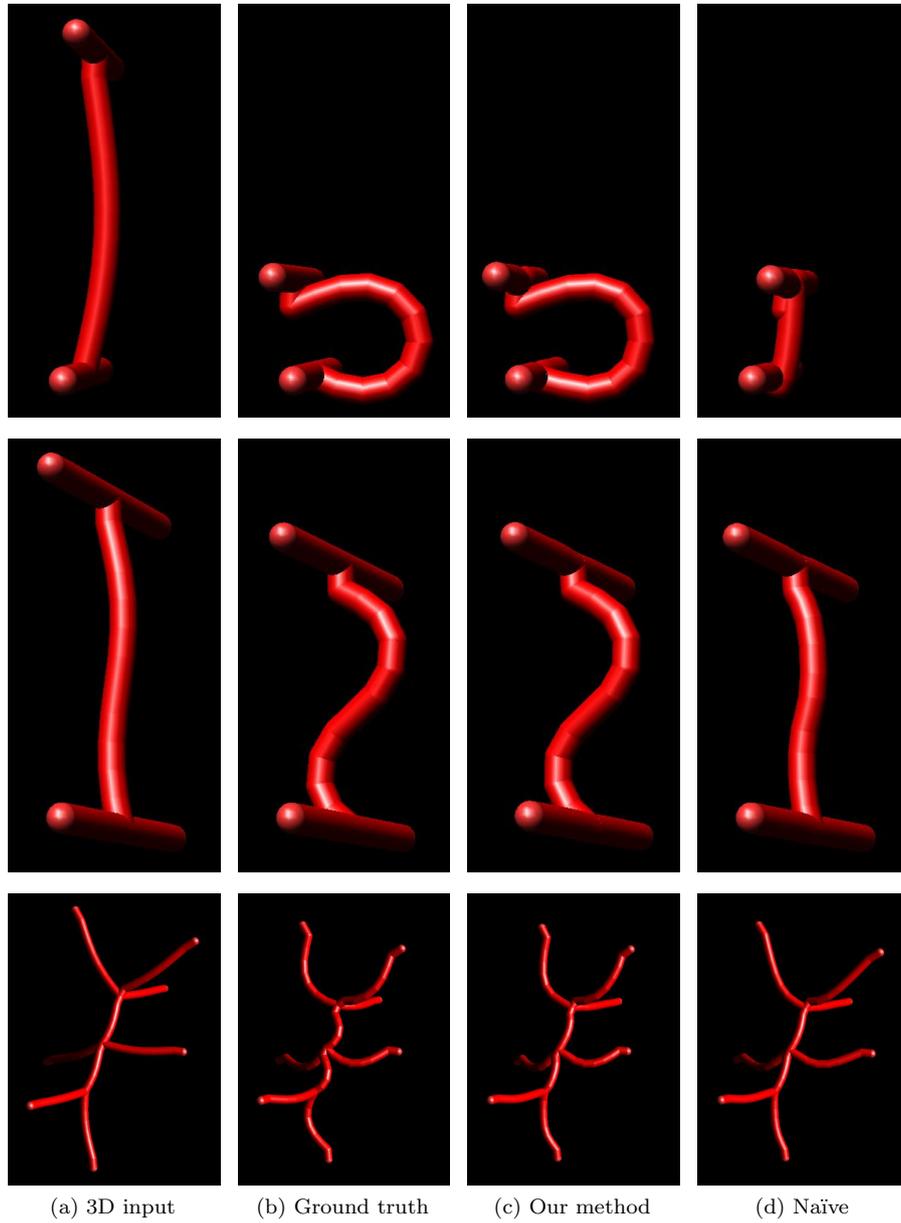


Figure 7.11: Visualization of a selection of tests on synthetic data. Every row presents a single example setting, with the quantitative assessment of the results in Table 7.1 (from top to bottom: Synth 1, Synth 2, Synth 3).

LINEAR REGISTRATION WITH MRF'S AND DISCRETE OPTIMIZATION[†]

We propose a framework for intensity-based registration of images by linear transformations, based on a discrete Markov Random Field (MRF) formulation. Here, the challenge arises from the fact that optimizing the energy associated with this problem requires a high-order MRF model. Currently, methods for optimizing such high-order models are less general, easy to use, and efficient, than methods for the popular second-order models.

Therefore, we propose an approximation to the original energy by an MRF with tractable second-order terms. The approximation at a certain point p in the parameter space is the normalized sum of evaluations of the original energy at projections of p to two-dimensional subspaces. We demonstrate the quality of the proposed approximation by computing the correlation with the original energy, and show that registration can be performed by discrete optimization of the approximated energy in an iteration loop.

A search space refinement strategy is employed over iterations to achieve sub-pixel accuracy, while keeping the number of labels small for efficiency. The proposed framework can encode any difference measure, is robust to the settings of the internal parameters, and allows an intuitive control of the parameter ranges. We demonstrate the applicability of the framework by intensity-based registration, and 2D-3D registration of medical images. The evaluation is performed by random studies and real registration tasks. The tests indicate increased robustness and precision compared to corresponding standard optimization of the original energy, and demonstrate robustness to noise. Finally, the proposed framework allows the transfer of advances in MRF optimization to linear registration problems.

8.1 Introduction

In the last years, the solution of computer vision problems by Markov Random Fields (MRFs) [Geman and Geman, 1984, Li, 2001] and discrete optimization has

[†]This chapter is based on the joint work with Ben Glocker, with the related publications: [Glocker et al., 2009, Zikic et al., 2010c, Zikic et al., 2010b].

become increasingly popular. Different image analysis problems have been solved by this approach, such as segmentation [Boykov and Jolly, 2001] or non-linear registration [Glocker et al., 2008a], just to name some. Recent advances [Boykov et al., 2001, Kolmogorov, 2006, Komodakis et al., 2007, Lempitsky et al., 2007] in discrete optimization methods make this approach very attractive.

However, the MRF-based optimization has not yet been applied to the classical problem of intensity-based registration of images by linear transformations [Maintz and Viergever, 1998, Hajnal et al., 2001a, Zitova and Flusser, 2003]. When analyzed closely, this problem turns out to be rather different in nature from many other problems which have so far been successfully addressed by MRF-based techniques. The differences are twofold.

For many MRF-based methods, the parameters can be assigned to certain points in the image domain at which the parameters are located, and the parameters have a limited local region of influence. In most cases, the parameters are directly identified with pixels and they only influence a direct neighborhood. That is, the parameters are *localized* in space and have *local effects*. This is in strong contrast to the parameters of a linear transformation which act globally and cannot be assigned to a certain point in the image domain. Also, most MRF-based methods share a similar modeling approach. They employ a second-order MRF model which is composed of two terms: A first-order term which encodes a certain cost function (often referred to as *data term*), and a second-order term responsible for the regularization (*prior term*). Due to the mentioned fundamental difference in the nature of the parameters, the estimation of linear transformation parameters requires a different modeling, as will be discussed in Section 8.2.

The second difference originates from the fact that the solution for the linear registration problem cannot be determined by optimizing the single parameters independently of each other. In consequence, the original problem of estimating the linear transformation with n degrees of freedom (DOF) requires an n^{th} -order MRF model. Compared to the second-order models, the current higher-order optimization algorithms are often complex to use or pose additional constraints on the class of energies they can be applied to. Also, the optimization of higher-order terms was until now mostly only demonstrated for regularization terms - a setting that does not translate to the linear registration problem. And finally, as of now, the efficiency of recent methods for second-order models cannot be matched by methods for higher-order models. This is a major point, since the single evaluations of the difference measure for the linear registration problem are computationally demanding.

In order to be able to use the current efficient MRF optimization techniques for the intensity-based estimation of linear transformations, we present an approximation of the original energy by a second-order MRF model. This approximated MRF energy term is optimized by a suitable discrete optimization method inside an iteration loop. In every iteration, the range of the discrete search space for the single parameters is refined, eventually resulting in sub-pixel precision. In order to guarantee convergence, the descent of the original energy is ensured. The resulting algorithm is outlined in Figure 8.1.

The remainder of this chapter is organized in the following way. In Section 8.2, the actual method is presented. After formulating the general problem of intensity-based estimation of linear transformations, we introduce the basic terms of discrete MRF models in Section 8.2.1, and argue that the estimation

```

Algorithm: Basic MRF-based Linear Registration

Input: images  $\mathbf{I}, \mathbf{J}$ 
Output: transformation parameters  $\mathbf{p}$ 
0: set initial parameters  $\mathbf{p}^1$ ;
1: set initial discretization  $\mathbf{L}$ ;
2: set initialize parameter range refinement factor  $\alpha$ ;
3: until convergence do repeat
4:    $\Delta\mathbf{p} := \text{optimize\_E\_approx}(\mathbf{I}, \mathbf{J}, \mathbf{L}, \mathbf{p}^1)$ ;
5:   if ( $E(\mathbf{p}^1 + \Delta\mathbf{p}) < E(\mathbf{p}^1)$ ) //test for decrease of original energy
6:      $\mathbf{p}^1 := \mathbf{p}^1 + \Delta\mathbf{p}$  ;
7:   end
8:    $\mathbf{L} = \text{refine\_label\_ranges}(\mathbf{L}, \alpha)$  ;
9: end

```

Figure 8.1: Outline of the basic algorithm for linear intensity-based registration by optimization of a discrete MRF energy.

of the transformation parameters of the original energy requires a higher-order MRF model. Following this, in Section 8.2.2 we introduce the approximation to the original energy by a tractable second-order model. In order to complete the description of the actual resulting registration method, the Sections 8.2.3 - 8.2.6 provide details on parameterization, discretization of the search space, the employed discrete optimization method, and implementation. Section 8.2.7 summarizes and discusses the properties of the proposed method. In Section 8.3, we evaluate the quality of the proposed energy approximation, as well as the dependency of the method on the settings of the internal parameters. We demonstrate and evaluate an exemplary application to registration of 3D images in Section 8.4.1, and an application to 2D-3D registration of medical images in Section 8.4.2. Finally, the work is summed up in Section 8.5.

8.2 Intensity-based Estimation of Linear Transformations by Markov Random Fields

We consider the task to estimate a linear transformation T which aligns the source image I to the target image J , such that this alignment optimizes a suitable difference measure ξ . \mathcal{T}_p denotes that the transformation T is parameterized by parameters p . Thus, the estimation can be written as

$$\hat{p} = \arg \min_p \xi(I \circ \mathcal{T}_p, J) , \quad (8.1)$$

where \hat{p} are the optimal transformation parameters. Furthermore, we express the parameter p in terms of the initial estimate p' and an update Δp as $p = p' + \Delta p$, so that the optimal parameters are represented by the optimal update as $\hat{p} = p' + \widehat{\Delta p}$.

For future reference we explicitly define the original energy E associated with the problem as

$$E(p) \equiv \xi(I \circ \mathcal{T}_p, J) . \quad (8.2)$$

8.2.1 Discrete Markov Random Fields

Let $\mathcal{V} = \{p_1, \dots, p_n\}$ be a set of n parameter variables p_i , which are also represented by the parameter vector \mathbf{p} . A *clique* is a subset of the parameters from \mathcal{V} , and is denoted either by explicitly stating the set, or by the usage of the multi-index $\mathbf{c} \subseteq \{1, \dots, n\}$, which contains the indices of the involved variables, such that $p_{\mathbf{c}} \subseteq \mathcal{V}$. The set of all defined cliques is denoted as \mathcal{C} . Then a Markov Random Field (MRF) can be defined as a hypergraph \mathcal{G} in which the nodes correspond to the parameters, and the cliques constitute the hyperedges, that is $\mathcal{G} = (\mathcal{V}, \mathcal{C})$. In the case of discrete MRFs, each parameter variable p_i can take a discrete value (also referred to as *label*) from a discrete label set \mathcal{L}_i with $\mathcal{L}_i = \{x \mid x \in \mathbb{R}\}$. The number of labels per parameter is denoted by $|\mathcal{L}_i| = N$. In the context of MRFs, the term *labeling* describes a set of labels assigned to the parameters. A *potential* is a scalar-valued function $\psi_{\mathbf{c}}$, which assigns a certain energy to a labeling of a clique $p_{\mathbf{c}}$ as $\psi_{\mathbf{c}} : p_{\mathbf{c}} \mapsto \psi_{\mathbf{c}}(p_{\mathbf{c}}) \in \mathbb{R}_0^+$. The general discrete MRF energy is a sum of the defined potentials

$$E_{\text{MRF}}(\mathbf{p}) = \sum_{p_{\mathbf{c}} \in \mathcal{C}} \psi_{\mathbf{c}}(p_{\mathbf{c}}). \quad (8.3)$$

The *order* of the MRF model is the maximum size of the involved cliques.

For the following, the model containing exclusively all possible second-order terms will play an important role, thus we define it explicitly as

$$E_{\text{MRF-2}}(\mathbf{p}) = \sum_{\{p_i, p_j\} \in \mathcal{C}_2} \psi_{ij}(p_i, p_j). \quad (8.4)$$

Here, \mathcal{C}_2 denotes the set of all possible second-order cliques, and correspondingly, cliques containing all possible k^{th} -order cliques will be denoted as \mathcal{C}_k .

8.2.1.1 Second- and Higher-order MRF Models

The arguably most common MRF model used in computer vision tasks is the second-order (pairwise) model containing at most cliques of size two (used e.g. for segmentation, image denoising, restoration, deformable registration). Often, second-order MRFs are used synonymously with MRF models, and the modeling consists of assigning the first-order potentials to the data term, and the second-order term to the regularization. Presumably, the major reason for this is that for the second-order MRFs, many efficient and general optimization methods have been proposed over the last years [Boykov et al., 2001, Kolmogorov, 2006, Komodakis et al., 2007, Lempitsky et al., 2007].

Only recently, optimization methods for higher-order models have been proposed. Compared to the second-order approaches, these methods are still not as general and efficient, and not as easy to use. Also, these methods have in common that the higher-order models are applied for implementing regularization terms. While this is applicable to many computer vision problems, it is not obvious how this modeling and corresponding optimization can be transferred to the linear registration problem. Here, we give a brief overview of recent higher-order optimization methods and discuss their applicability to the problem at hand. [Kohli et al., 2007, Kohli et al., 2009] consider an efficient optimization for generalized higher-order Potts' model [Potts, 1952]. The Potts' model is

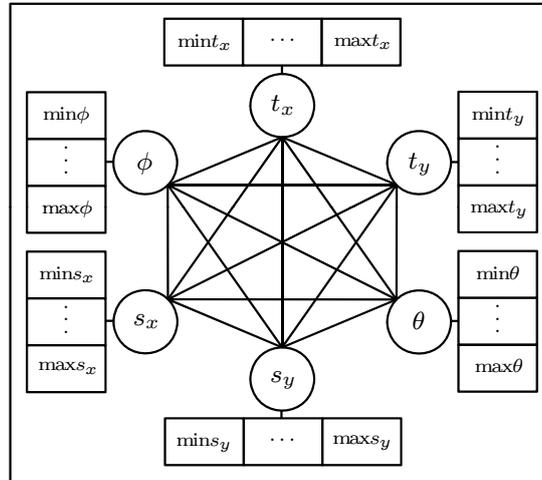


Figure 8.2: Topology of the proposed MRF model for an affine 2D transformation. The transformation parameters are represented by the nodes. The edges encode the costs of the difference measure, associated with the variation of the labeling of the adjacent nodes. The MRF determines the parameters by optimizing the sum of the edge costs.

an energy function applicable for regularization by enforcing label consistency in neighborhoods. However it is not applicable to modeling the linear registration problem with arbitrary difference measures. The same issue regarding the regularization term holds also for the applications considered in [Lan et al., 2006, Potetz and Lee, 2008, Ishikawa, 2009]. [Lan et al., 2006] as well as [Potetz and Lee, 2008] consider efficient versions of Belief Propagation (BP) algorithm for encoding label consistency in 2×2 neighborhoods for image denoising. [Ishikawa, 2009] considers the same application with a different optimization strategy, by combining the reduction of higher-order terms with the fusion-move [Lempitsky et al., 2007] and quadratic pseudo-boolean optimization (QPBO) [Boros and Hammer, 2002]. [Rother et al., 2009] address the optimization of sparse higher-order energies, using soft pattern-based representation of the energy functions, and by transforming the problem into an equivalent quadratic function minimization problem. The linear registration however - in contrast to many others in image analysis - is not a sparse problem. [Komodakis and Paragios, 2009] employ a general framework for solving the higher-order model based on a master-slave decomposition - with many application examples for which the method can be applied. The proposed decomposition profits from a high degree of independence of the single slave problems. However, this independence is not given for the linear registration problem, and the efficient optimization of higher-order slave potentials is not straight-forward in the case of linear registration.

8.2.1.2 MRF Order for Linear Registration

The order of an MRF model expresses the degree of conditional dependence of parameters on each other [Bishop, 2006]. Take for example a simple first-order

model: in this case labeling one parameter with a certain value does not influence the choice for the labeling of the other parameters inside the MRF energy term. Thus, the labeling of the single parameters can be performed independently. Going one step further and looking at a second-order model, we see that the choice of the label for one parameter p_i will influence the choice for the labeling of all the parameters p_j , which form cliques together with p_i , that is $\{p_i, p_j\} \in \mathcal{C}$. Finally, we observe that for an MRF model with n parameters, the most complex model is an n^{th} -order model, in which all the nodes are directly dependent on all the other nodes. Due to the conditional dependence of the parameters for the linear registration problem in (8.2), it follows that in general the estimation of a linear transformation with n DOFs requires an n^{th} -order clique model.

Since direct optimization of this high-order energy is difficult, in the next section we show how to circumvent this by introducing an approximation to the original energy E which involves only tractable second-order terms.

8.2.2 Energy Approximation for Estimation of Linear Transformations by Second-Order MRFs

The key idea of the proposed approximation is to use the second-order cliques to encode a cost of a simultaneous variation of two parameters, while the other parameters are fixed to their current values. This encodes the dependency of the difference measure on the two respective parameters, while the dependency on all the other parameters is ignored. However, by simultaneously taking into account all possible combinations of parameter pairs, we construct the overall energy term such that the selection of one parameter value depends on all the others.

Formalized, this results in an MRF model represented by a fully-connected graph $\mathcal{G}^*=(\mathcal{V}, \mathcal{C}_2)$, where the nodes \mathcal{V} represent the transformation parameters, and the edges \mathcal{C}_2 are the set of all possible parameter pairs (all possible second-order cliques). The potential $\tilde{\psi}_{ij}$ determines the cost of a simultaneous variation of the transformation parameters p_i and p_j while the other parameters are held fixed at their initial values (the so created parameter set is denoted as $p^{\tilde{ij}}$) and can be written as

$$\tilde{\psi}_{ij}(p_i, p_j) = \xi (I \circ T_{p^{\tilde{ij}}}, J) \quad (8.5)$$

Here, the k -th entry of $p^{\tilde{ij}}$ is explicitly defined as

$$p_k^{\tilde{ij}} = \begin{cases} p_i & , k = i \\ p_j & , k = j \\ p'_k & , k \neq i, j \end{cases} , \quad (8.6)$$

which simply means that the parameter vector $p^{\tilde{ij}}$ takes the value of p_i as the i -th entry, and the value of p_j as the j -th entry, and has the corresponding values of the initial parameters p' at all other entries. Substituting (8.5) in (8.4), and normalizing by the number of cliques $|\mathcal{C}_2|$, we define the approximative second-order MRF energy for estimation of linear transformations as

$$\tilde{E}(\mathbf{p}) = \frac{1}{|\mathcal{C}_2|} \sum_{\{p_i, p_j\} \in \mathcal{C}_2} \tilde{\psi}_{ij}(p_i, p_j) . \quad (8.7)$$

Type	2D			3D		
	Rig.	Sim.	Aff.	Rig.	Sim.	Aff.
#params (n)	3	5	6	6	9	12
#cliques ($ \mathcal{C}_2 $)	3	10	15	15	36	66

Table 8.1: The number of cliques (or potentials/edges) in the approximated MRF model in Eq. (8.7), for rigid, similarity with anisotropic scaling, and affine transformations.

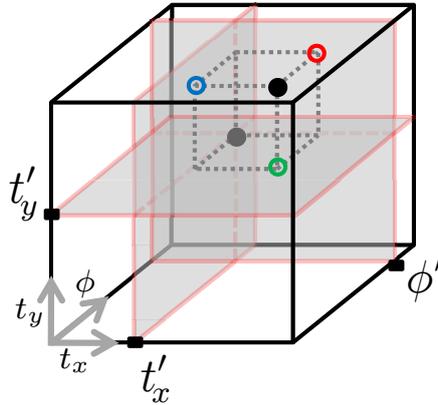


Figure 8.3: Visualization of the energy approximation by a second-order MRF model for rigid 2D registration. Here, t_x , t_y , ϕ denote the translation and rotation parameters, with initial values t'_x , t'_y , ϕ' . The evaluation of the original energy at the parameter point p (black) is approximated by the normalized sum of the energy evaluations at the projections of p to the 2D subspaces (red, green, blue). The subspaces are orthogonal and all pass through the initial point p' (gray).

In general, the number of cliques is $|\mathcal{C}_k| = \binom{n}{k}$, where k is the order of the used terms so that for the second-order model we have $k=2$. We give $|\mathcal{C}_2|$ for relevant cases in Table 8.1.

In summary, the presented approximation allows us to model the conditional dependence of the parameters while efficient optimization methods for second-order models can be used. A simple example of the resulting MRF topology is illustrated in Figure 8.2. In Section 8.3.1, we evaluate the quality of the proposed approximation. By sampling of the original and the approximated energy, we demonstrate that they strongly correlate for reasonable initializations.

8.2.2.1 Interpretation of the Approximation

An interpretation of the approximative energy term (8.7) provides further insight into the proposed method. In Equation (8.7), the single potentials $\tilde{\psi}_{ij}$ as defined in (8.5) are carving out two-dimensional subspaces in the original n -dimensional parameter space. These subspaces are parallel to the respective ij -planes of the

parameter space and are all passing through the point p' representing the initial parameters. By construction, the subspaces are orthogonal to each other.

The approximated energy can now be compared to the original energy in the following way. The original energy $E(p)$ is simply the evaluation of the chosen difference term at the point p . The approximated energy $\tilde{E}(p)$ is computed by projecting the point p to the single subspaces and averaging the energy evaluations at the projected points. If we denote the projections of the point p to the i -subspace by $\Pi_i(p)$, we can write the approximated energy from (8.7) also as

$$\tilde{E}(p) = \frac{1}{|\mathcal{C}_2|} \sum_{i,j} E(\Pi_{ij}(p)) . \quad (8.8)$$

This interpretation is illustrated for the simple case of a rigid 2D transformation in Figure 8.3.

It is important to notice that the quality of the approximation depends on the choice of the initialization p' . Obviously, if we evaluate the approximated energy at the initial point, that is $p=p'$, the approximated energy equals the original energy, since all the projections of p equal to the initial point p' . With increasing distance of p from p' , the error made by the projection increases and deteriorates the approximation. This is demonstrated in Section 8.3.1 where the quality of the approximation is evaluated.

The stated interpretation can be used to reason about extending the proposed approximation, by utilizing cliques of higher orders. The approximation would then read

$$\tilde{E}_k(p) = \frac{1}{|\mathcal{C}_k|} \sum_{\{p_i, \dots, p_k\} \in \mathcal{C}_k} \tilde{\psi}_{i\dots k}(p_i, \dots, p_k) \quad (8.9)$$

$$= \frac{1}{|\mathcal{C}_k|} \sum_{i, \dots, k} E(\Pi_{i\dots k}(p)) , \quad (8.10)$$

and the order k would equal the dimension of the subspaces. For example, using first-order terms would lead to projecting to lines parallel to the space axes and passing through the initial point, and third-order terms would imply projecting to three-dimensional volumes. This generalization might become interesting since using higher-order building blocks can be expected to increase the quality of the approximation. Also, efficient optimization for single third-order terms used as building blocks can be expected to arise earlier than for example for the 12th-order model required for the full 3D affine transformation.

8.2.3 Parameterization

In this section we present the parameterization of linear transformations used in this work. Linear transformations in homogeneous coordinates can be written in the form

$$A = \begin{bmatrix} \hat{A} \\ \mathbf{v}^\top & 1 \end{bmatrix} , \quad (8.11)$$

where $\hat{A} \in \mathbb{R}^{d \times (d+1)}$ and $v \in \mathbb{R}^d$ for d -dimensional problems. Since our work is motivated by medical applications, in the following we do not consider projective

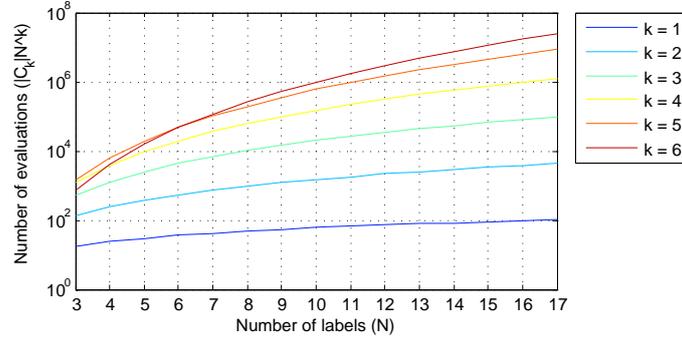


Figure 8.4: Number of evaluations of the difference measure needed to compute the approximated energy at all points in the parameter space. We give an example for a transformation model with 6 parameters, depending on the order $k \in \{1, \dots, 6\}$ of the potentials used for approximation, and with $N = \{3, \dots, 17\}$ samples per dimension. The method proposed in this work has $k = 2$, while the original energy is represented by $k = 6$. Please note that this is only a theoretical upper bound, and that the actual optimization methods require less computations.

transformations and we restrict ourselves to affine transformations [Hajnal et al., 2001a] by assuming that $\mathbf{v} = \mathbf{0}$. This results in 6 degrees of freedom in 2D and 12 DOF in 3D. We employ a parameterization in which the affine transformation is decomposed as

$$A = M_t R_\phi R_\theta^{-1} D_s R_\theta . \quad (8.12)$$

Here, M_t represents a translation, R_ϕ a rotation, and $R_\theta^{-1} D_s R_\theta$ represents the shearing transformation. For the shearing, R_θ is a rotation and D_s is a diagonal matrix, representing anisotropic scaling. We parameterize the single matrices of Equation (8.12) by respective parameters, compare also [Hartley and Zisserman, 2003]. The 3D rotation matrices are parameterized by Euler angles. The resulting parameter vectors for the 2D and 3D case are

$$\mathbf{p} = [t_x, t_y, \phi, s_x, s_y, \theta] \quad (8.13)$$

$$\mathbf{p} = [t_x, t_y, t_z, \phi_x, \phi_y, \phi_z, s_x, s_y, s_z, \theta_x, \theta_y, \theta_z] . \quad (8.14)$$

Please note that we always consider the parameter p in terms of an initial guess p' and an update Δp , such that $p = p' + \Delta p$.

With this representation, restricting the general affine model from (8.12) to a more constrained one is simply performed removing the corresponding nodes and the adjacent edges from the MRF model. Since our implementation performs these modifications automatically if the search range for one parameter is set to zero, there is no need for explicit implementation of the registration for the different transformation types.

8.2.4 Discretization of the Parameter Space

An important part of the proposed method is the discretization of the parameter search space, that is, the definition of the label space \mathcal{L} . On one hand, one would

N	3	5	7	9	11	13	15	17
FastPD	105	190	300	390	500	600	700	800
exhaust.	135	375	735	1215	1815	2535	3375	4335

Table 8.2: Average number of evaluations of difference measure per iteration for FastPD, in comparison to exhaustive search. We give an example for the 6 DOF 3D rigid case with $k=2$.

like to keep the number of labels small for efficient optimization of Equation (8.7). On the other hand, we want to avoid a too coarse sampling of the parameter search space since it might result in inaccurate registration results. To achieve a trade-off between these two objectives, we employ a successive strategy for the refinement of the search space. This results in refined parameter updates in every iteration.

We discretize \mathcal{L} in the following manner. For each parameter p_i we define a search range which is discretized by uniformly sampling the range between preset values $\min p_i = p'_i + \min \Delta p_i$ and $\max p_i = p'_i + \max \Delta p_i$ about the initial value p'_i .

Since the algorithm is guaranteed to reduce the energy in every step (by explicitly assuring that the new solution decreases the original energy, cf. Figure 8.1), we can assume that the new estimate is closer to the locally optimal solution, and reduce the search ranges for the parameters. In iteration k , we rescale the range delimiters $\min \Delta p$ and $\max \Delta p$ by a parameter range refinement factor α^k ($\alpha < 1$) which defines the new possible values for the parameter updates. The iterative label space refinement allows us to keep the number of labels small and we can start with a large parameter range, while being able to achieve sub-pixel registration accuracy. In practice, the number of labels N ranges between 5 and 11 and the maximal possible values for the parameters are used to limit the parameter ranges (e.g. $\pm 180^\circ$ for rotation), which allows for the fixed setting of the ranges. We evaluate the performance of the method for different choices of α as well as the number of labels N in Section 8.3.2.

An interesting observation is that the proposed approximation also strongly reduces the number of similarity evaluations which would be necessary to evaluate the energy at all points in the complete parameter space. The discretized n -dimensional parameter space with N sampling points per dimension contains altogether N^n points. Thus, the computation of the original energy at all points in the parameter space requires N^n evaluations of the difference measure, since it is simply evaluated once at every point in the parameter space. On the other hand, the approximated energy at one point is computed as a sum of $|\mathcal{C}_k|$ evaluations of the difference term at points in the two-dimensional subspaces, while the subspaces require N^k evaluations each. This means that only $|\mathcal{C}_k|N^k$ evaluations have to be performed (with $k=2$ for our second-order terms).¹ From these evaluations, the approximated energy can be computed for every point by simply averaging the $|\mathcal{C}_k|$ scalar values. The number of needed evaluations, depending on the order k of the potentials used as building blocks in the approximation, and the number of labels N , is illustrated in Figure 8.4. This implies that a brute force exhaustive search would be much more efficient for the approximated

¹Actually, slightly less evaluations than $|\mathcal{C}_k|N^k$ are needed, since the single subspaces share points where they intersect.

energy. For small label sets ($N \approx 5$) and low-order ($k = 2, 3$) this might even be a viable option. Please note that $|\mathcal{C}_k|N^k$ is only a theoretical upper bound, and that the actual optimization methods (e.g. FastPD as used in our work) require less computations. For an example, please refer to Table 8.2. Advances in discrete optimization, leading to less evaluations, would have a direct impact on the proposed method.

8.2.5 Optimization

For the optimization of our MRF energy defined in Equation (8.7), we use the recently proposed FastPD algorithm² [Komodakis et al., 2007]. FastPD is able to handle non-submodular functions [Kolmogorov and Rother, 2007], such as general intensity-based difference measures. Furthermore, FastPD performs well both in terms of computational efficiency and optimization quality [Komodakis et al., 2008]. Other optimization methods such as move algorithms based either on QPBO (Fusion Move [Lempitsky et al., 2007]) or iterative graph-cuts (e.g. Expansion Move, Swap Move [Boykov et al., 2001]), or message-passing strategies [Pearl, 1988, Weiss and Freeman, 2001] (e.g. Belief Propagation, TRW-S [Kolmogorov, 2006]) might be considered as well. However, we should note that due to the non-submodularity of the pairwise energy terms, classical graph-cut based methods [Boykov et al., 2001] have to be modified to be able to handle such energies (cf. [Kolmogorov and Rother, 2007]). On the other hand, message-passing algorithms can handle general energies, however with the cost of lower computational efficiency [Szeliski et al., 2008].

8.2.6 Implementation Details

The described registration method is implemented with a standard multi-resolution strategy realized by Gaussian image pyramids with a downsampling factor of two between levels. No downsampling is performed below 32 voxels per dimension. The method also supports random subsampling strategies, which are often used to improve the runtime, cf. e.g. [Pluim et al., 2003b] and references therein. In terms of performance, the evaluation of the given difference measure presents the most computationally demanding part of the method. This is especially the case for large 3D data sets. To this end, we also implemented the evaluation of the 3D difference measures on the GPU which results in a significant speedup of the method.

8.2.7 Method Discussion

Formulating the linear registration as an MRF problem shares the advantages of some standard optimization methods such as Nelder-Mead Simplex or Powell's method [Press et al., 1993], that the difference measure is easily interchangeable, since no derivatives of the measure are required. The integration of novel and more complex measures – where the derivatives might be complicated to compute – is simple to achieve in our framework. With freely available discrete optimization libraries, the implementation of the proposed approach is straight forward.

A possible alternative for decreasing the order of the MRF model is the usage of high-dimensional labels, where one label encodes the values for multiple

²Available at: <http://www.csd.uoc.gr/~komod/FastPD/>.

parameters. This is done for example in [Glocker et al., 2008a] for encoding the local translations for deformable registration. For rigid registration for example, one could encode the translations in x/y/z direction as one parameter with a high-dimensional label consisting of the combinations of values for t_x , t_y , and t_z , and rotation as a second one for ϕ_x , ϕ_y , and ϕ_z . Such an MRF model would be of second-order and have only one clique, which would consist of the two parameters. The problem with this approach is that the size of the label spaces increases too much, such that no efficient solution is possible. For the above example, in which a high-dimensional label represents 3 ordinary labels, the size of the high-dimensional label space would be N^3 . We performed several tests following this approach, however, the computation times were prohibitive.

A different possible future line of work would consist of testing different parameterizations. For example, the label values could be sampled from the ranges in a non-uniform manner, or different parameterizations of the rotation angles could be used.

In this work, we only used second-order potentials as building blocks for the energy approximation. One interesting option would be to also employ first-order terms, for example to encode the prior probability of the single parameters. Please note that this way, for the linear registration problem, the standard MRF philosophy in which the first-order terms encode the data term, and the second-order terms encode the prior knowledge is reversed. Also, a possible future modification as efficient, general, and easy to use optimization methods for third-order MRF models become available would be to build the proposed approximation on third-order instead of second-order potentials. This can be expected to improve the approximation quality. The generalization using k^{th} -order potentials for the approximation is discussed in Section 8.2.2.1.

8.3 Method Evaluation

In this section, we evaluate two important aspects of the proposed framework. In Section 8.3.1, we evaluate the quality of the proposed energy approximation from Equation (8.7). In Section 8.3.2, we test the influence of the internal parameters of the proposed method on the registration results.

All experiments are performed on a standard desktop system with an Intel[®] Core[™] 2 Quad 2.83GHz, 3GB of RAM, and an NVIDIA[®] GeForce[®] GTX 285 GPU with 1GB RAM.

The following stopping criterion is used in all experiments, and for all levels of the image pyramid. The iterations are stopped when the relative decrease of the energy (compared to the energy at the initial state) falls below a threshold. We use a conservative threshold of 0.0001.

8.3.1 Quality of Approximation

In this section we evaluate the quality of the energy approximation, that is how similar the approximated energy \tilde{E} from Eq. (8.7) is to the original energy E in Eq. (8.2). The idea is to compare the original and the approximated energy by sampling these functions and computing the correlation coefficient (CC) between them. This test assesses the error introduced by the approximation, which is independent of the particular method chosen for the optimization.

Test Type	n	N	#eval= N^n	CC Test 1	CC Test 2
2D rigid	3	80	$\sim 0.5 \times 10^6$	0.9590	0.9687
2D similarity	5	14	$\sim 0.5 \times 10^6$	0.8925	0.9169
2D affine	6	9	$\sim 0.5 \times 10^6$	0.8595	0.8976
3D rigid	6	9	$\sim 0.5 \times 10^6$	0.8446	0.8510
3D similarity	9	4	$\sim 2.0 \times 10^6$	0.7864	0.8169
3D affine	12	4	$\sim 16.8 \times 10^6$	0.7961	0.8120

Table 8.3: Evaluation of the quality of the energy approximation \tilde{E} . We compute the correlation coefficient (CC) between sampled versions of the original and the approximated energy. This is done for different transformations with n parameters, and with N sampling points per dimension, resulting in N^n energy evaluations per test. Transformation parameters increase from Test 1 to Test 2. We observe increased approximation quality for simpler models and “smaller” transformations.

We perform this test for 2D and 3D images, and 3 types of transformations: rigid, similarity with anisotropic scaling, and general affine transformations.

In 3D, the tests are performed on an MR-T1 volume of the brain from The Retrospective Image Registration Evaluation Project (RIRE)³ database [West et al., 1997]. The 2D tests are conducted on a slice from the above 3D volume. Since the random tests are mono-modal, we employ the sum of squared differences (SSD) as difference measure. We sample the parameter space equidistantly about the initial parameters p' representing the identity transformation. Two different ranges are tested, **Test 1** representing the larger parameter ranges at the beginning of the registration process, while **Test 2** covers smaller parameter ranges and represents the setting towards the end of the registration. The ranges for the sampling of the transformation parameters for Test 1 are as follows: Translations result from $\pm 10\text{mm}$, rotations lie between $\pm 45^\circ$, scalings range between 1 ± 0.2 , and the shearing angles θ are drawn from $\pm 20^\circ$. For Test 2, the ranges are: $\pm 2\text{mm}$ for translations, $\pm 4^\circ$ for rotations, 1 ± 0.04 for scaling, $\pm 4^\circ$ for shearing angles. The number of sampling points per dimension N is set such that per test at least 500,000 energy evaluations are performed. In order to keep the computation time reasonable, the tests were performed on the third finest levels of the image pyramid, that is on $64 \times 64 \times 26$ images in 3D and 64×64 images in 2D. Altogether, approximately 21 million energy evaluations were performed. Details and results of the tests are summarized in Table 8.3.

The tests show that the proposed approximation correlates well for the tested settings. The average CC over all tests is 0.86, where $\text{CC}=0$ indicates that two signals are linearly unrelated and $\text{CC}=1$ indicate perfect linear correlation. As expected, the correlation is higher for simpler models, that is, increases from affine over similarity to rigid, and it is higher for the 2D than for 3D transformations. Also, the correlation is higher for the smaller parameter ranges.

These findings are supported by inspecting the energy logs of the original and approximated energy, acquired during the random registration experiments performed in Section 8.4.1.1. This results in sampling of the energies along the paths in the parameter space, which are being traced out by the single

³Available at: <http://www.insight-journal.org/rire/>

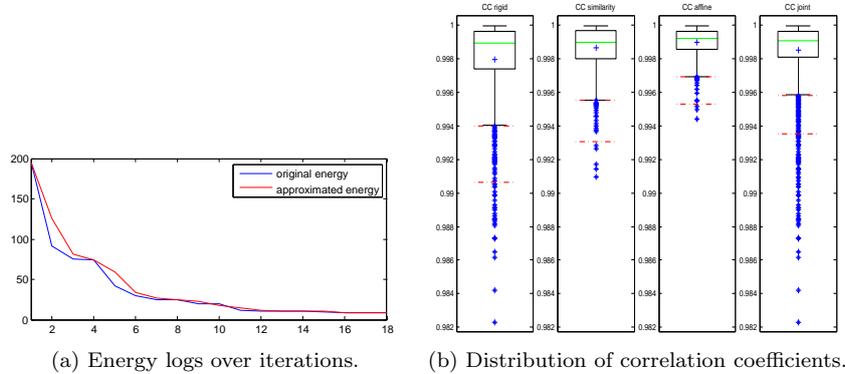


Figure 8.5: Comparison between the original and approximated energy. **(a)** A typical energy log. **(b)** The CC statistics between the logs of all runs performed in Section 8.4.1.1 exhibit a high correlation (mean always >0.98), and indicate that correlation (and thus the approximation quality) increases over iterations (cf. Section 8.3.1).

registrations. We observe higher correlation coefficient in the logs ($CC > 0.98$), compared to the coefficients from the 3D evaluation in Table 8.3 ($CC < 0.85$), caused by a large number of evaluations close to the optimum, which are performed towards the end of the registration. This shows that the proposed approximation improves during the registration process. A sample plot of the logs and the quantitative evaluation is given in Figure 8.5.

8.3.2 Dependency on Parameters

In this section we perform an evaluation on the sensitivity of the proposed method on the internal parameters: N , the number of labels for the discretization of the parameter space, and α the parameter range refinement factor.

We perform the evaluation by varying the setting of the parameters for 1000 random registration tests, and assessing the resulting performance. The random tests consist in registering a 2D MR-T1 source image to a randomly displaced target image. The target image is created by applying a random affine transformation to the source image. The random transformation is generated by uniformly drawing the transformation parameters from the following ranges: translations result from $\pm 10\text{mm}$, rotations lie between $\pm 60^\circ$, scalings range from 1 ± 0.2 , and the shearing angles θ are drawn from $\pm 20^\circ$. We test all possible combinations for N between 3 and 17, and α between 0.4 and 0.8, resulting in 40,000 registration runs. For faster computation, the registration is performed on a downsampled image.

The registration error is computed as the average distance (AD) of the corner points of a centered 200mm rectangle, which are warped once by the ground truth transformation and once by the estimated transformation. Please note that since the AD is measured at the corners, it will in general measure larger errors, than for the region of interest in the center of the volume.

Figure 8.6 summarizes the results of the experiment. Except for $N = 3$, we

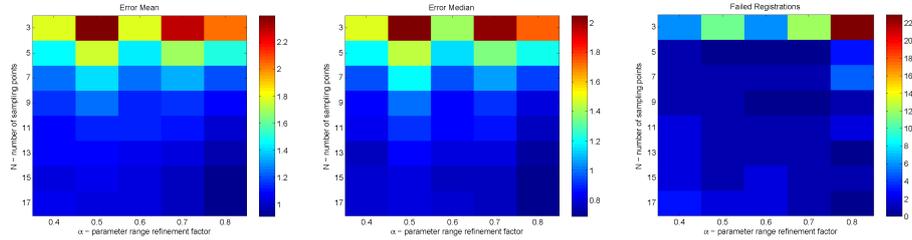


Figure 8.6: Evaluation of the performance of the proposed framework with respect to the internal parameters: the number of sampling points per dimension N , and the parameter range refinement factor α . For each different setting of N and α , 1000 random tests are performed. Except for $N = 3$, we observe no large difference between different parameter settings. The precision is slightly improved for finer discretization of the search space (higher N). This however comes at a higher computational cost.

observe no large difference between the different parameter settings. For $N > 3$, the number of failed runs is similar very low (below 0.3%). For finer discretization of the search space (higher N) the precision is slightly improved, however, this comes at a higher computational cost. With respect to the resolution, the majority of results exhibit a sub-pixel precision. Based on the results of this evaluation, we set the values to $\alpha = 0.6$ and $N = 5$ keep them fixed throughout the registration experiments in Section 8.4.1.

On a side note, we also performed the same 1000 trials by registration with optimization of the original energy with the Nelder-Mead Simplex method [Press et al., 1993] and otherwise same settings. This test resulted with a similar error distribution as the proposed method, however it had 7 failed trials, which is inferior to any of the different settings of the MRF-based method for $N > 3$.

8.4 Applications

8.4.1 Linear Intensity-based Image Registration

In this section, we present results of the direct application of the proposed framework to the problem of linear intensity-based registration between 3D images. We test the proposed method by a random study on 3D brain data in Section 8.4.1.1, and by a series of 3D multi-modal registrations in Section 8.4.1.2. In this section we use 3D brain images (CT, MR-PD, MR-T1, MR-T2) provided by the RIRE data base. The CT image has a resolution of $512 \times 512 \times 29$ and a physical voxel size of $0.65 \times 0.65 \times 4$ mm and the MR images have a resolution of $256 \times 256 \times 26$ and a voxel size of about $1.25 \times 1.25 \times 4$ mm.

8.4.1.1 3D Random Study

The 3D random study is performed by registering a source image to a target image generated by a random transformation. The source image is an MR-T1 volume. The target image is created by applying a random transformation to

Test Type		Mean	Median	Std. Dev	Failed
Rigid	MRF	0.561	0.422	0.401	0%
	Simplex	0.526	0.432	0.403	0%
Similarity	MRF	0.935	0.657	0.805	0.2%
	Simplex	1.222	0.931	1.209	1.2%
Affine	MRF	4.277	3.876	2.279	11.2%
	Simplex	4.594	4.093	2.395	19.6%

Table 8.4: Summary of the results of the 3D random study from Section 8.4.1.1. The proposed MRF-based method is compared to Simplex optimization of original energy. Per transformation type 1000 trials were performed. Results with an AD error (given in mm) of more than 10mm are discarded as failed, and excluded from the computation of the statistics. (see also Figure 8.7). For all runs, the proposed method results in fewer failed runs and exhibits slightly better accuracy.

the source image, with transformation parameters uniformly drawn from the following ranges: Translation $t \in [-30, 30]$ mm, rotation $\phi \in [-45, 45]^\circ$, scaling $s \in [0.8, 1.2]$, and shearing angles $\theta \in [-20, 20]$ mm.

Due to the mono-modality of the test, we use the SSD as the difference measure. The evaluation of the difference term for this test is performed on the GPU and without a sub-sampling strategy. A multi-resolution approach with 3 levels is used, and the registration is initialized by aligning the intensity masses of the two volumes. To speed up the computation and allow for a large number of tests, the registrations are performed only on the coarsest level of the image pyramids where the number of voxels in the first two dimensions is reduced by a factor of four while the number of slices is preserved. Since the tests are performed at lower pyramid levels at which noise is eliminated, we restrain from adding noise to the original images in the first place. The registration error is computed as the average distance (AD), which in 3D is defined on the corner points of a centered 200mm cube. We perform the study for 3 types of transformations: rigid, rigid with anisotropic scaling, and affine. For each type we perform 1000 trials.

We compare the results of the proposed approach to the results obtained by using the Nelder-Mead Simplex optimization method [Press et al., 1993] on the original energy function corresponding to the chosen difference measure. To this end, the same general settings as for our approach are used. In all our experiments, the internal parameters of the Simplex method (step sizes for the transformation parameters) are set to the same values as the corresponding search space ranges of the MRF method. Convergence criteria are thresholds on the difference of function values at subsequent estimates ($1e-20$), and changes in parameters ($1e-15$), as well as the maximum number of iterations (500). We have tested these settings for different registration problems, with consistently good results. The results are summarized in Table 8.4 and the statistics are represented by box plots⁴ in Figure 8.7. Please note that due to the chosen

⁴Box plot notation: The box denotes 50% of data between lower and upper quartiles Q1, and Q3, i.e. the medians of the upper and lower half of the sorted data; Median denoted by green line; Whiskers determined by extreme values within $Q1-1.5\text{IRQ}$ and $Q3+1.5\text{IRQ}$ (interquartile range $\text{IRQ}=Q3-Q1$); The mean and outliers denoted by blue crosses; Red lines

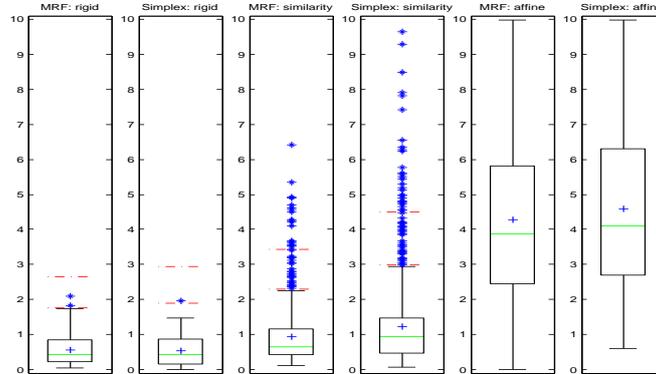


Figure 8.7: Statistics of the AD error in mm for the 3D random study described in Section 8.4.1.1, represented by box plots (please see footnote 4 for description). Please refer to Table 8.4 for a brief description and the numerical representation of the results. With respect to resolution, all errors less than 4mm indicate sub-voxel accuracy.

resolution, errors of less than 4mm are below the physical voxel size. In Figure 8.5, we give an exemplary energy log acquired over the iterations, in which we record the approximated and the original energy. Also, the correlation of the energy logs for all trials is analyzed.

Two observations can be made from the results of the experiment. The first one is that the proposed MRF-based method seems to be more robust, such that more registrations are successful. We classify the registration as successful if the AD is below 10mm. With this threshold, all experiments for both, the Simplex and MRF-based optimization, for the rigid case are successful. However, for the case of similarity with anisotropic scaling, the Simplex method (1.2%) fails 6 times as often as the MRF-based method (0.2%). For the affine case, the Simplex method (19.6%) fails 1.75 times as often as MRF-based method (11.2%). Similar results are also obtained by using different thresholds. The second observation is that for the successful registrations, the MRF-based method yields higher precision. Generally, the MRF-based method exhibits lower mean and median errors, and the corresponding standard deviations are also lower. The only exception is the rigid test, in which the mean error for the Simplex method is lower. However, the more robust median error is lower for the MRF-based method also in this case.

With respect to the runtime, one must note that the MRF-based optimization of the approximated energy is more computationally intensive than the Simplex optimization of the original energy since it requires more evaluations of the difference term. Also, the difference is smaller for simpler models, since the number of evaluations needed for the evaluation of the approximated MRF energy is smaller. For the performed random tests, the average runtimes were 20.4 seconds for the rigid case, 45.6 seconds for similarity, and 85.6 seconds for affine transformations.

at $Q1 - 1.5 \text{ IRQ}$, $Q1 - 3 \text{ IRQ}$, $Q3 + 1.5 \text{ IRQ}$, $Q3 + 3 \text{ IRQ}$.

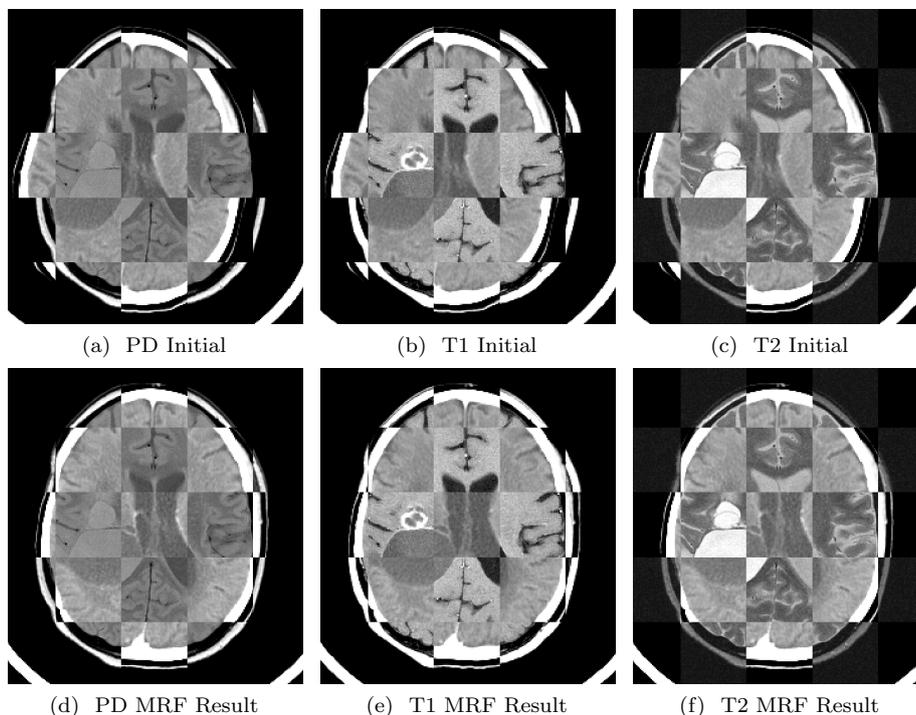


Figure 8.8: Checkerboard visualization of the multi-modal CT-MR registration using MRFs on patient P04.

8.4.1.2 3D Multi-Modal Rigid Registration

In this experiment we demonstrate the performance of the proposed approach for a real multi-modal rigid registration scenario, from the RIRE database. We perform the registration of CT images to MR-PD, MR-T1, and MR-T2 images. As the difference measure, we use the entropy correlation coefficient (ECC) proposed in [Maes et al., 1997b]. A multi-resolution approach with 4 pyramid levels is used. The transformations are initialized by aligning the intensity masses of the two volumes.

We perform two tests, the first one on data of one patient with available ground truth which can be used for training, and an extensive test on 10 patients, for which the evaluation is performed by the RIRE system. For the training data set, ground truth transformations between a CT volume and 6 MR images of one patient (original and rectified scans) are given. We compute the error by measuring the AD at the corner points of the CT volume, using the ground truth.

In order to assess the accuracy of the proposed method, we perform the same tests also by two other methods. The first one is the Simplex optimization on the original ECC energy term, as discussed in the previous sections. The second one is the module of rigid registration from the *Elastix* toolkit⁵ [Klein et al., 2010a].

⁵Available at: <http://elastix.isi.uu.nl/>

Target	Initial	pre-al.	Simplex	Elastix	MRF
MR-PD	67.61	26.52	2.79	2.74	2.90
MR-PD rect.	49.64	22.69	4.10	2.85	3.43
MR-T1	46.18	22.49	3.41	3.02	1.82
MR-T1 rect.	65.00	15.63	2.32	1.93	2.66
MR-T2	56.13	22.00	4.08	2.52	3.60
MR-T2 rect.	62.94	22.87	4.43	4.08	4.28
mean	57.92	22.03	3.52	2.86	3.11
median	59.54	22.59	3.75	2.80	3.17

Table 8.5: Results of the 3D multi-modal CT-MR registrations on training data, described in Section 8.4.1.2. Given is the initial AD (in mm) before registration, after alignment of the intensity masses, and after registration using the Simplex, Elastix, and our MRF-based method.

With respect to Elastix parameters, we did our best to provide a setting as similar as possible to the other two methods, and followed the recommendations from the Elastix manual for other parameters. We use the same number of image pyramid levels, and employ the Normalized Mutual Information (NMI) [Studholme et al., 1999] as the closest fit to the ECC. As the optimization method we selected the adaptive stochastic gradient descent [Klein et al., 2009], for which the internal parameters are estimated automatically. The maximum number of iterations was set to 250. We verified the validity of the settings on the training data. The results show a similar performance by the methods, with best average results obtained by Elastix, followed closely by the proposed methods, and then the Simplex based approach, please compare Table 8.5.

For obtaining comparable runtimes, the evaluation of the difference term is performed on the CPU in this experiment. To decrease the running time, we perform a subsampling strategy for all methods, in which 10% of voxels are uniformly drawn and used for the computation of the difference measure. Our method performed on average 27 iterations per registration (ca. 7 per level), resulting in average runtimes of 190 seconds. Elastix on average took 200 seconds, and the Simplex method converged after 30 seconds. Please note that due to the differences in the implementation of the different methods, this gives only a rough estimate of the order of magnitude for the runtimes. Furthermore, since the focus of this work is not on the speed of registration, different tuning options were not used. For example, while we employ 10% of all voxels in the sub-sampling, it was shown in [Klein et al., 2007] that already 2000-3000 samples can yield accurate results, corresponding to ca. 0.15% of all voxels on the finest level.

For the actual multi-modal test, we used all data sets from the RIRE data base, for which the MR-PD, MR-T1, and MR-T2 data sets are available, resulting in 10 patients with 3 registrations per patient. All registrations were performed with the same parameters as for the training data set. The evaluation is performed by the RIRE system. The tests show a consistent performance of the proposed method. Table 8.6 summarizes the results and shows that the proposed method consistently slightly outperforms the other tested methods in terms of accuracy. See Figure 8.8 for an example of the registration setting. Furthermore, we

performed the registrations with the MRF-based method also with different settings of the internal parameters in order to support the findings from Section 8.3.2. For $N=3$ and $\alpha=0.6$ the error statistics (in mm) are: mean=1.845, median=1.708, max=4.409. On average, 22 iterations were performed per registration (ca. 5 per level), resulting in an average computation time of 86 seconds. For $N=17$ with $\alpha=0.8$, we have: mean=1.746, median=1.695, max=3.849. On average, 33 iterations were performed per registration (ca. 8 per level), resulting in an average computation time of 858 seconds.

In this scenario, the dependency on parameter settings is even less prominent than in Section 8.3.2 so that even the smallest possible label set ($N=3$) leads to accurate and efficient results.

8.4.2 Intensity-based 2D-3D Registration

The second exemplary application is the 2D-3D registration of medical images, which shows the applicability of the proposed framework to variations of the standard registration problem. 2D-3D registration is a challenging task, since a transformation of 3D space is computed from 2D measurements, so that for the 1-view case, the problem is ill conditioned in the projection direction. In this application we also demonstrate the robustness of the proposed method to high noise levels.

The task of 2D-3D registration is to recover a rigid 3D transformation T which relates the coordinate frame of the 3D volume I to the coordinate system of the 2D imaging devices, which generate the 2D projection images J_l . The transformation is estimated by optimizing the difference measure ξ between the perspective projections of the transformed 3D image $P_l(I \circ T_p)$ and the m available 2D images J_l

$$E_{2D3D}(p) = \frac{1}{m} \sum_{l=1}^m \xi(P_l(I \circ T_p), J_l) , \quad (8.15)$$

We apply our MRF-framework to the 2D-3D problem, by using Equation (8.7) and approximate E_{2D3D} as

$$\tilde{E}_{2D3D}(\mathbf{p}) = \frac{1}{|\mathcal{C}_2|} \sum_{\{p_i, p_j\} \in \mathcal{C}_2} \tilde{\psi}_{ij}(p_i, p_j) , \quad (8.16)$$

with

$$\tilde{\psi}_{ij}(p_i, p_j) = \frac{1}{m} \sum_{l=1}^m \xi(P_l(I \circ T_{p^{ij}}), J_l) . \quad (8.17)$$

8.4.2.1 Evaluation

We assess the performance of the proposed approach by comparing it to the optimization of the original energy by the Simplex method. In order to perform the evaluation in a controlled setting, we conduct the 2-view tests on real 3D data, but with synthetically created projections, compare Section 8.4.2.1. This way, the choice of the difference measure plays a smaller role and we can compare the performance of the optimization approaches. This also has the advantage that

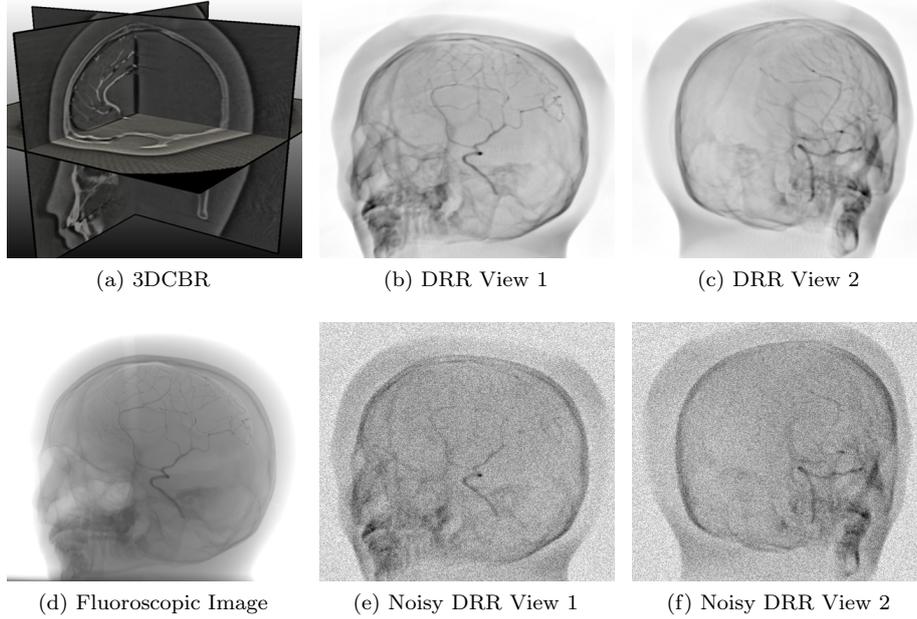


Figure 8.9: Visualization of input data for the 2D-3D tests: **(a)** 3D cone beam reconstruction (3DCBR), which is used in all experiments; **(b,c)** Exemplary DRRs which are computed in the iterations of the algorithm; **(d)** Fluoroscopic image used as target in the 1-view test; **(e,f)** Exemplary DRRs with 20% uniform noise used as targets in the 2-view tests.

we can test on a large number of different views with known ground truth, rather than performing the tests on only a few real views by randomly disturbing the initialization. The robustness and applicability to real settings is demonstrated in experiments by adding noise to the projections. Furthermore, we test our algorithm in a more challenging real 1-view test scenario in Section 8.4.2.1, in which a real fluoroscopic image is used.

General Setting As input data for the experiments we use a 3D cone beam reconstruction (3DCBR) of a phantom head with a cerebral vessel structure, computed from a single sweep of a monoplane stationary C-arm with flat-panel detector (Siemens Axiom Artis dTA). As 2D input we use a fluoroscopic image obtained by the same device, with ground truth transformation obtained by feature-based registration, and verified by careful inspection. An overview of the data is given in Figure 8.9.

The 2D projections required inside the algorithm and synthetically generated targets are computed as digitally reconstructed radiographs (DRRs) by GPU accelerated ray-casting. A conversion operator is used to remap the intensity values to X-ray energies.

All tests are performed with a set of random offset poses. The poses are generated by uniform random sampling of the parameters from certain ranges. For the 2-view test, we have $t \in [-30, 30]$ mm, and $\phi \in [-80, 80]^\circ$. For the 1-view

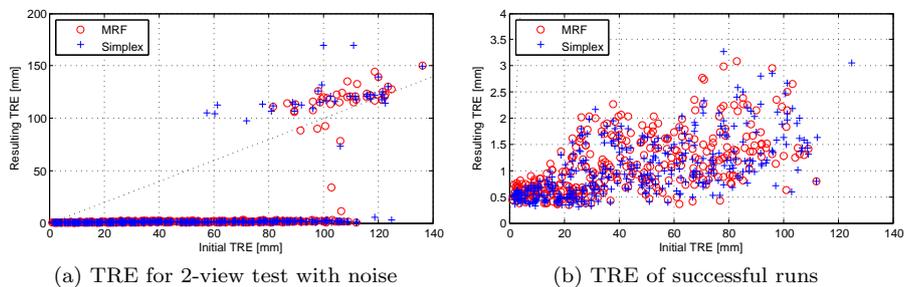


Figure 8.10: Results of 400 random 2-view test runs with 20% uniform noise. The x and y-coordinates of the graph points represent the TRE before and after registration. The diagonal is the line of no improvement. (b) shows a zoom of the area of successful test runs.

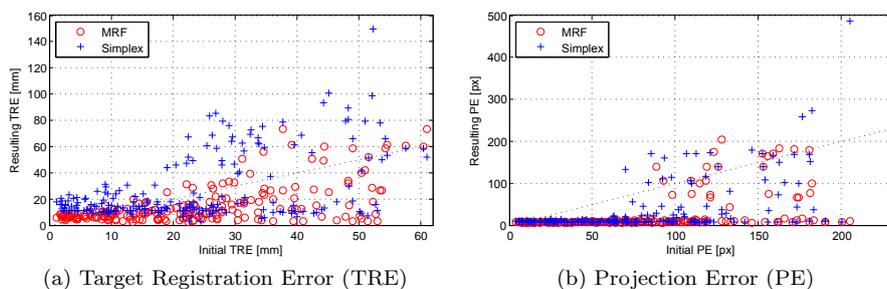


Figure 8.11: Results of the 200 random 1-view test runs with fluoroscopic image. In (a), the TRE is shown, while (b) illustrates the projection error.

test, we use $t \in [-15, 15]$ mm, and $\phi \in [-40, 40]^\circ$. We evaluate the results by the Target Registration Error (TRE). The TRE is computed as the mean of the distances between a set of points transformed by the ground truth transformation and the same point set transformed by the estimated transformation. The point set consists of points of a regular $10 \times 10 \times 10$, 20cm large grid, centered at the phantom head.

As the difference measure we use the local version of the Normalized Cross Correlation [Penney et al., 1998]. We also performed tests with the Gradient Difference measure [Penney et al., 1998], with very similar results.

The registrations are performed with a standard multi-resolution strategy using a Gaussian image pyramid. In all 2D-3D experiments, the initial search space ranges for the MRF-based method are ± 50 mm for the translations and $\pm 90^\circ$ for the rotations. We use $N=7$ and $\alpha=0.66$. For these settings, the average registration run takes about 4 minutes. Again, different parameter values were tested for the parameters, resulting in very similar registration results.

2-View Test with DRRs as Target Image We carry out the 2-view test by performing 400 runs. In each run, a new pair of orthogonal DRRs J_1, J_2

is generated by applying a random pose to the 3DCBR image. In order to test the robustness of the approach in a challenging scenario, uniform noise in the range of 20% of the intensities is added to all generated DRRs, please compare Figures 8.9e and 8.9f. The same test was also performed without noise, yielding comparable results. The tests show that both, the Simplex and the MRF approach perform well, with a small number of failed registrations only for very large initial misalignments, for details please compare Figure 8.10.

1-View Test with Fluoroscopic Image as Target The second series of experiments is performed in a 1-view scenario by registering the 3DCBR volume to a real fluoroscopic image. Here, we use 200 random poses as initial offsets to the 3D transformation. Similar to Section 8.4.1.1, we employ only the two coarsest levels of a five-level image pyramid to speed up the computation (resulting in projection images of size 128×128). Additionally to the TRE, we also evaluate the Projection Error (PE), which is computed by projecting the transformed 3D samples used for TRE computation to the image and computing their mean distance. The PE measures the pixel error visible in the image, such that it is not very sensitive to misalignments in the viewing direction. By comparing the TRE and PE measures in Figure 8.11, we observe that while the PE is approximately the same for both approaches, the TRE (in mm) is slightly better for the MRF-based approach (Simplex: mean=27.48, median=15.43, std.=25.09; MRF: mean=14.7, median=8.86, std.=14.8). With respect to the used resolution, the minimal errors are within sub-voxel accuracy. This suggests that the MRF optimization is able to better recover the translation along the viewing direction.

8.5 Conclusion

We present an MRF-based framework for linear intensity-based registration of medical images. The key idea is to approximate the original energy corresponding to the chosen difference measure by tractable second-order terms. To our best knowledge, this is the first time that an approach for linear intensity-based registration based on MRFs and discrete optimization is presented. The proposed model deviates from the common MRF problem modeling since it involves only pairwise terms, which do not encode a regularization term but are used as building blocks to represent the cost function. While the number of parameters is lower than for classical MRF problems, the parameters are not localized in space, and they influence the problem globally and not only in their direct neighborhood. The experimental evaluation confirms the quality of the proposed approximation and shows that the method is not sensitive to the setting of the internal parameters.

Furthermore, we demonstrate the applicability of the method for standard mono- and multi-modal image registration, as well as 2D-3D image registration. The registration experiments indicate increased robustness and precision of the proposed method in comparison to standard optimization of the original energy function. The 2D-3D experiments show the applicability of the framework also to less well-conditioned problems, such as 1-view 2D-3D registration. These tests also demonstrate the robustness of the method to high noise levels.

Regarding the computation time, we find that our method is less efficient than the Simplex method due to the higher number of evaluations of the difference

term. However, our running times are similar to the stochastic gradient descent when using the same number of sub-samples. Further tuning options, such as reducing the number of sub-samples were shown to decrease the computation time of the stochastic gradient descent with similar accuracy. This might also be an option for our method.

Due to the active development in discrete optimization, the proposed method has a strong further potential as it allows to transfer the advances in MRF optimization to linear registration problems, for example the integration of faster methods, or upcoming methods for third-order models.

Target	Mean Error			Median Error			Maximal Error		
	Simplex	Elastix	MRF	Simplex	Elastix	MRF	Simplex	Elastix	MRF
PD - P01	1.831	2.200	1.870	1.743	2.042	1.781	2.387	2.807	2.501
PD - P02	0.972	1.237	1.001	0.952	1.226	0.996	1.460	1.826	1.479
PD - P03	16.438	16.784	16.407	16.569	16.550	16.472	17.427	18.690	17.495
PD - P04	3.359	4.749	3.226	3.260	4.663	3.133	4.052	6.310	3.884
PD - P05	1.706	1.493	1.746	1.724	1.434	1.733	2.047	1.783	2.029
PD - P06	1.534	1.769	1.587	1.506	1.589	1.525	2.819	3.079	2.902
PD - P07	1.880	2.026	1.891	1.924	2.036	1.938	2.047	2.880	2.026
PD - P08	3.069	3.300	3.063	2.959	3.255	2.962	3.711	4.287	3.716
PD - P09	2.591	2.400	2.574	2.398	2.297	2.385	3.668	3.915	3.634
PD - P10	2.180	1.999	2.190	2.262	2.061	2.263	2.416	2.362	2.416
PD overall	2.067	2.226	2.078	2.005	2.018	1.986	4.052	6.310	3.884
T1 - P01	1.218	1.520	1.215	1.313	1.487	1.090	1.680	2.444	1.688
T1 - P02	0.592	0.723	0.608	0.695	0.705	0.690	0.820	1.069	0.872
T1 - P03	1.306	1.205	1.423	1.424	1.286	1.484	1.880	1.937	2.006
T1 - P04	2.080	1.944	2.126	2.206	2.057	2.274	2.873	2.240	3.003
T1 - P05	0.933	0.917	0.956	0.929	0.961	0.953	1.064	1.215	1.084
T1 - P06	1.253	1.288	1.253	1.186	1.327	1.148	2.159	2.088	2.301
T1 - P07	0.908	1.012	0.907	0.874	1.058	0.863	1.054	1.655	1.136
T1 - P08	1.775	2.020	1.886	1.818	1.937	1.954	1.971	2.364	2.068
T1 - P09	1.569	1.737	1.521	1.405	1.553	1.465	2.495	3.172	2.589
T1 - P10	1.417	1.175	1.296	1.416	1.241	1.302	1.458	1.377	1.395
T1 overall	1.275	1.334	1.286	1.259	1.230	1.154	2.873	3.172	3.003
T2 - P01	2.909	2.099	2.247	2.826	1.957	2.268	4.271	2.707	2.835
T2 - P02	2.132	1.370	1.512	1.938	1.364	1.414	3.647	1.898	2.636
T2 - P03	1.224	2.018	1.160	1.236	1.938	1.113	1.536	3.060	1.518
T2 - P04	2.817	3.541	2.899	2.793	3.569	2.696	3.245	4.192	3.647
T2 - P05	1.614	1.828	1.652	1.743	1.879	1.799	2.376	2.678	2.299
T2 - P06	1.266	1.264	1.276	1.210	1.222	1.240	2.094	2.232	2.324
T2 - P07	1.952	1.986	1.842	1.945	1.875	1.840	2.168	2.750	1.877
T2 - P08	2.673	3.053	2.686	2.575	2.975	2.580	3.099	3.748	3.556
T2 - P09	2.218	2.220	1.899	2.104	2.072	1.735	3.331	3.598	2.737
T2 - P10	2.027	2.110	1.799	2.080	2.106	1.804	2.374	2.900	2.076
T2 overall	2.053	2.085	1.856	1.979	1.950	1.853	4.271	4.192	3.647
OVERALL	1.789	1.870	1.729	1.743	1.815	1.739	4.271	6.310	3.884

Table 8.6: Results of the 3D multi-modal CT-MR registrations on patient data, described in Section 8.4.1.2, with $\alpha=0.6$ and $N=5$. Given is the mean, median and maximal error (in mm) after registration using the Simplex, Elastix, and our MRF-based method. The results for PD - P03 are not taken into account, since all three methods failed in this case.

MINIMAL DEFORMATIONS[†]

Nonlinear registration is mostly performed after initialization by a global, linear transformation (in this work, we focus on similarity transformations), computed by a linear registration method. For the further processing of the results, it is mostly assumed that this preregistration step completely removes the respective linear transformation. However, we show that in deformable settings, this is not the case. As a consequence, a significant linear component is still existent in the deformation computed by the nonlinear registration algorithm. For certain applications, such as construction of Statistical Shape Models (SSM) from deformations, this is an unwanted property: SSMs should be invariant to similarity transformations, since these do not capture information about shape. We propose a method which performs an a posteriori extraction of a similarity transformation from a given nonlinear deformation field. To this end, a closed-form solution minimizing the squared Euclidean norm of the displacement field subject to similarity parameters is used. Experiments on real inter-subject data and on a synthetic example show that for application to SSMs, the theoretically justified removal of the similarity component by the proposed method has a large influence on the model and significantly improves the results.

9.1 Introduction

This work has two major goals. The first consists of pointing out that deformation fields resulting from standard registration schemes often contain a significant amount of linear transformation, and proposing a method to extract this linear component, thus computing *minimal nonlinear deformations*. For a visualization of the setting, please refer to Fig. 9.1. Secondly, we identify the construction of shape models to be an application for which from the theoretical point of view, no similarity transformation should be contained in the deformations which are used to construct the model. If the similarity component is not extracted from the deformation, the first modes of the constructed model may not describe the largest variation in shape of the given samples. We show that by the proposed method, this negative effect can be eliminated, thus resulting in improved shape models.

[†]This chapter is based on the work in [Zikic et al., 2008c, Zikic et al., 2008a].

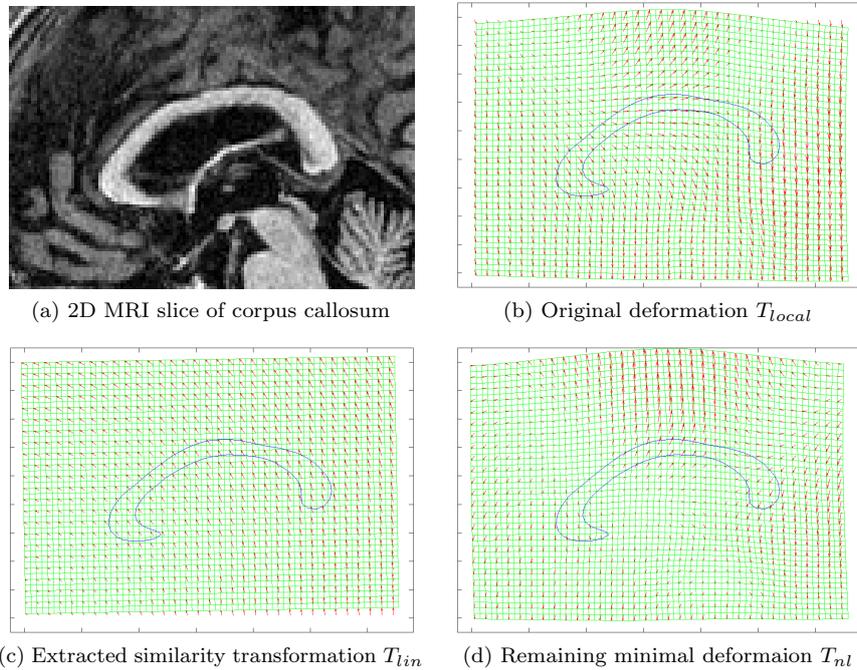


Figure 9.1: Illustration of the discussed setting. On a set of images of corpus callosum (a) (see also Section 9.3.1.2), nonlinear registration is performed, resulting in a deformation field T_{local} , see (b). A similarity component T_{lin} (c) is extracted from the original field T_{local} , resulting in a minimal nonlinear transformation T_{nl} (d).

In the following, we hope to provide an intuitive understanding why standard registration methods in general do not compute minimal deformations, and also why this is an important point when building shape models.

9.1.1 Nonlinear Registration and Resulting Deformations

Nonlinear registration is a technique which has been studied heavily over the last two decades. The goal of nonlinear registration is to estimate the nonlinear transformation which relates two given images. This can also be seen as computing the dense point correspondences between the two images. Virtually all schemes for nonlinear registration proposed in the literature employ a global linear preregistration step, followed by a nonlinear method accounting for local differences between the images. We make the observation that global registration methods cannot fully recover linear transformations in deformable scenarios. As a consequence, the deformation field which is computed by the local nonlinear registration algorithm contains a linear transformation. This occurs with intensity-based, as well as with landmark-based linear registration approaches. The reason for this is that for the exact estimation of the linear part, all point correspondences must be known. However, the correspondences are not given for the linear methods, since - as mentioned above - it is the actual task of the

nonlinear registration method to estimate the dense point correspondences. Theoretically, landmark-based methods can estimate the linear component correctly, given dense landmarks. Establishing such landmarks however, is hardly feasible in practice, and would render the nonlinear registration step superfluous. A very similar discussion is presented in [Yezzi and Soatto, 2003].

9.1.2 Statistical Shape Models

Statistical Shape Models (SSM) are an emerging technique in the field of medical image processing and analysis. SSMs are supposed to capture the information about shape variations of a certain population. SSMs are not only a valuable tool for studying shape variations of organs and pathologies, but also provide a mean to capture prior knowledge and thus aid the process of nonlinear registration in the demanding inter-patient scenario, in which standard methods may easily fail. For creation of SSMs, two different approaches are used: landmark- and deformation-based. For the landmark-based approach, corresponding landmarks have to be computed for all images of the population, compare e.g. [Cootes et al., 1995]. This can be a very challenging and time-consuming task. Thus, the idea was developed to create models of shape using deformations which result from nonlinear registration, compare e.g. [Grenander and Miller, 1998], [Gee and Bajcsy, 1998], [Joshi, 1998], [Rueckert et al., 2003], or [Cootes et al., 2004].¹

With respect to important properties of deformation fields for construction of SSMs, the literature agrees on the fact that the fields should be free of any similarity transformation, that is, they should not contain any amount of scaling, rotation and translation, compare for example [Rueckert et al., 2003, Sec II.A, p. 1016] or [Cootes et al., 2004, Sec. 4, p. 452]. This is a very intuitive goal, since these parameters do not describe a variation in shape between single subjects of a population, but a variation in pose and size. Usually, it is supposed that the global preregistration accounts for this transformation. However, due to the reasons stated above, in general, linear registration methods cannot fully recover linear transformations in deformable scenarios, which leads to deformation fields which contain a linear transformation component. Thus, using these deformation fields presents a potential source of error for construction of SSMs. The goal of the presented work is to eliminate this source of error by computing deformation fields containing no similarity transformation.

Computation of Minimal Deformations. The proposed method decomposes a given deformation into a similarity transformation and a minimal nonlinear deformation part. The minimality of the nonlinear part is computed with respect to the mean squared Euclidean norm of the displacement field representing the deformation. The actual computation is performed by using a closed-form solution. The minimization problem is modeled in such way that the computed minimal deformation is expressed in the reference frame of the target image. Since the method operates on point correspondences, it can be applied to dense deformation fields in the complete image domain as well as only to a region of interest. The method is comparably fast and since it presents

¹In the context of computing shape models from deformations, the term *Statistical Deformation Models (SDM)* is sometimes also used [Rueckert et al., 2003]. The landmark-based models are also referred to as *Active Shape Models (ASM)* [Cootes et al., 1995]. For generality, we employ the term *Statistical Shape Models (SSM)*.

a post-processing step for any given point correspondences, it can be easily integrated into any existing framework for construction of shape models based on deformations.

Contribution We consider the following points to be the major contribution of the presented work.

1. We study the performance of linear registration methods in presence of nonlinear deformations, and show that for real data, after the global preregistration, there is a significant linear component of the transformation which is not retrieved.
2. We propose a method to a posteriori extract the similarity transformation component from any given deformation field.
3. We theoretically show that using non-minimal deformations for constructing shape models can be expected to result in models which do not describe shape variations appropriately.
4. We empirically confirm the results of the theoretical analysis, and show that the removal of similarity components from deformation fields prior to model construction leads to improved shape models in terms of interpretability and compactness.

9.2 Methods

9.2.1 Notation and Basic Definitions

In the context of registration, the transformation which aligns the target and source images I_T and I_S is a function $T : \Omega \rightarrow \Omega$ where $\Omega \subset \mathbb{R}^d$ is the image domain of dimension $d = 2, 3$. In most of the current methods for deformable scenarios, the transformation T is composed of a global, linear transformation T_{global} and the nonlinear local part T_{local} , resulting in

$$T = T_{global} \circ T_{local} , \quad (9.1)$$

where \circ denotes composition. In general, the global part is computed prior to the local component and no joint computation of the two terms is employed.

We model nonlinear transformations as a sum of the identity function Id and a displacement field U , as $T = \text{Id} + U$. For the local nonlinear transformation we also apply the notation $T_{local}(X) = Y$.

9.2.2 Computation of Minimal Deformations

Our goal is to extract the remaining linear transformation component from a given deformation. To this end, we model the deformation as a composition of a linear and a nonlinear part

$$T_{local} = T_{lin} \circ T_{nl} . \quad (9.2)$$

The task now is to estimate T_{lin} and T_{nl} , such that T_{nl} becomes minimal in some meaningful sense.

The model from Eq. (9.2) can be reformulated, such that it allows us to minimize the norm of the displacement field U_{nl} of the nonlinear component T_{nl} with respect to the linear transformation T_{lin} .

$$T_{local}(X) = (T_{lin} \circ T_{nl})(X) \quad (9.3)$$

$$Y = T_{lin}(X + U_{nl}(X)) \quad (9.4)$$

$$T_{lin}^{-1}(Y) - X = U_{nl}(X) . \quad (9.5)$$

Here, in (9.4) we use $T_{local}(X) = Y$, and express the deformation T_{nl} by the displacement U_{nl} .

Thus, we can define a cost function, the optimization of which results in a linear transformation (described by parameters p), such that the norm of the vectors of the displacement field becomes minimal with respect to the mean squared norm. The cost function E for the displacement fields discretized by n points is given by

$$E(X, Y, p) = \frac{1}{n} \sum_i \|X_i - T_{lin}^{-1}(Y_i; p)\|^2 , \quad (9.6)$$

and the respective minimization is

$$p = \arg \min_{p'} E(X, Y, p') . \quad (9.7)$$

In other words, the extraction of any other linear transformation would result in larger displacements (given the reference frame and the type of linear transformation).

We chose to use the squared norm for several reasons. First, this model has an exact and fast closed-form solution. Second, for the application to SSMs, the squared norm is the most common choice in literature, and it is consistent with the PCA-based shape model.

Once the linear transformation T_{lin}^{-1} is computed, the corresponding displacement field is resulting from Eq. (9.5) as $U_{nl}(X) = T_{lin}^{-1}(Y; p) - X$, and the nonlinear remaining part can be constructed by $T_{nl} = \text{Id} + U_{nl}$.

Please note that the minimal deformation T_{nl} is expressed in the reference frame of the target image. This is an important property for the application of the method to SSMs, compare [Cootes et al., 2004]. Please note also that our method does not change the results of the complete registration procedure, but rather computes a different decomposition, that is $T = T_{global} \circ T_{local} = T_{global} \circ T_{lin} \circ T_{nl}$.

Up to this point, the discussion is valid for any invertible linear transformation T_{lin} . In the following we constrain the discussion to a similarity transformation.

9.2.2.1 Least-Squares Optimization

For the computation of the similarity transformation which minimizes the mean squared norm of the displacement field, we employ the closed-form solution of Umeyama [Umeyama, 1991b], which is shown to give the exact result.

It solves the so called *Absolute Orientation Problem*, which consists of finding the similarity transformation which minimizes the mean squared distance between two point sets A and B of arbitrary dimension d , that is

$$e^2(R, t, c) = \frac{1}{n} \sum_{i=1}^n \|B_i - (cRA_i + t)\|^2 , \quad (9.8)$$

where c is the scaling factor, t is the translation vector, and R is the rotation matrix, and n is the number of points.

We can apply this method to our problem directly by identifying B with points X of the image domain, which are the origins of the vectors of the displacement field, and identifying A with $Y = X + U(X)$, that is, the destination points of the displacement vectors. The computed entities R , t , c are used to parametrize the similarity transformation T_{lin}^{-1} .

Please note, that for minimization, our method interprets the deformation as a set of two corresponding point sets. An advantage is that the developed method is not restricted to dense deformations, but can be applied to arbitrary point sets. For SSMs, a meaningful choice is to restrict the computation to regions of interest of the deformation fields [Rueckert et al., 2003].

The computation is comparably fast with the complexity of $\mathcal{O}(dn + d^3)$ and has a memory consumption of $\mathcal{O}(dn + d^2)$. The complexities are linear in n and as we have $d \ll n$, this makes the method very attractive for our application. In practice, the runtimes of our MATLAB implementation are for example 0.13s for a 2D example with $n = 100^2$ points and 6.27s for $n = 100^3$ in 3D.

9.2.3 Statistical Shape Models

Principal component analysis (PCA) is the preferred method for statistical shape models [Rueckert et al., 2003, Cootes et al., 2004]. The attractive properties of the PCA for shape modeling include optimal linear reconstruction of the data set variance, the estimated modes of variation are orthogonal and uncorrelated, and a closed form solution exists for calculating the principal components at a relatively low computational cost.

The shape model is built from m given displacement fields $U = \{U_i\}$ representing the deformations. The d -dimensional deformation fields with n displacement vectors are linearized as column vectors $u_i \in \mathbb{R}^{dn}$.

From u_i , a linear shape model, which approximates a given field u is given by \bar{u} and Φ as

$$u = \bar{u} + \Phi b . \quad (9.9)$$

Here \bar{u} is the mean of all m displacement fields, that is $\bar{u} = \frac{1}{m} \sum_{i=1}^m u_i$. The matrix Φ is constructed from the k first eigenvectors Φ_i of the covariance matrix C , given by $C = \frac{1}{m-1} \sum_{i=1}^m (u_i - \bar{u})(u_i - \bar{u})^\top$. The eigenvalues corresponding to Φ_i are denoted by λ_i . The vectors Φ_i are also referred to as *modes*. Finally, $b \in \mathbb{R}^k$ is the parameter vector, describing the contribution of the principal modes contained in Φ in order to approximate u by the employed linear model. By assuming a Gaussian distribution on the single displacements entries, the variance of the parameter b_i can be given by λ_i [Rueckert et al., 2003].

An important measure for evaluating the constructed model is the so called *reconstruction error* e_{rec} , given by

$$e_{rec}(\bar{u}, \Phi, u, b) = \|u - (\bar{u} + \Phi b)\|^2 . \quad (9.10)$$

The reconstruction error measures the error between a given vectorized displacement field u , and the reconstruction of u by using the parameters b corresponding to u , and the model given by \bar{u} and Φ . The parameter vector b is computed by a projection of u onto the model, that is by $b = \Phi^\top (u - \mu_u)$, where μ_u denotes

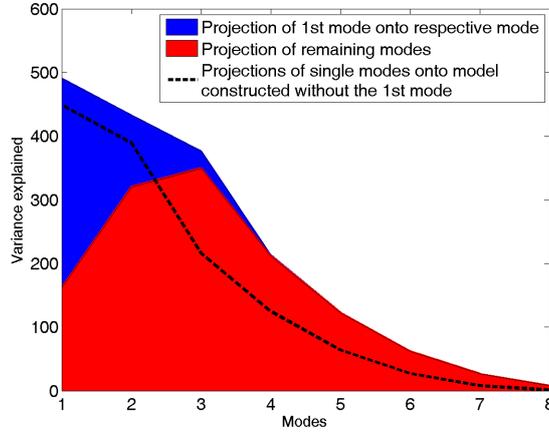


Figure 9.2: Example demonstrating how the largest mode used for generation of data (blue area) is mixed with the remaining modes (red). We measure this by the amount of the explained variance. This behavior is due to the finite number of samples used, in which case the PCA does not reconstruct the actual modes which generate the data, but rather their linear combination. The black curve shows the true variance of the remaining modes (2nd to last).

the mean of u . The variance explained by a single mode Φ_i corresponds to the variance of b_i , which is $\text{Var} [\Phi_i^\top (u - \mu_u)]$.

With respect to the proposed method, the only modification of the standard model construction process is that instead of the original displacement fields U , we use minimal deformation fields U_{nl} , from which the maximum amount of similarity transformation is extracted by our method as described in Sec. 9.2.2.

9.2.3.1 Influence of Similarity Transformations in Deformation Fields on SSMs

In this Section, we argue that if similarity transformation components are not removed from the deformation fields, this will in general lead to shape models in which the first modes do not necessarily describe the largest variations in shape.

It is a general property of the PCA that - when operating on a finite number of samples - it does not compute the actual modes which generate the data, but rather a linear combination of these. This behavior is illustrated in Fig. 9.2 for a general example. Here, a model is constructed from $n = 122$ orthogonal samples. It can be seen how the variance which is actually generated by the first mode during construction of the samples is explained by the first four reconstructed modes.

In particular, this general behavior also occurs for data, which can be seen as generated as a combination of nonlinear deformation and similarity transformations. This is the case for non-minimal deformations. This means that the similarity transformations will not be represented by single modes, but their contribution is distributed over several modes describing nonlinear deformation. Since similarity transformations are global, the corresponding modes have a large variance, such that mostly the first modes of the model will be influenced by



Figure 9.3: Three example instances of the Deformed Spoon.

the effect described above. As a consequence, for non-minimal deformations, in general we can not expect that the first modes of the model describe the largest variations in shape. This underlines the importance of using minimal deformation fields. Our experimental tests on real data (Fig. 9.5) closely resemble the behavior predicted in Fig. 9.2.

9.3 Results and Evaluation

In this part, we discuss and evaluate the results of the application of the proposed method to the construction of shape models.

9.3.1 Test Settings

In this section we briefly present the synthetic and the corpus callosum setting, which are used for evaluation.

9.3.1.1 Synthetic Example

To demonstrate some of the propositions of the presented work on a simple example, a synthetic data set with ground truth deformations is constructed. The set consists of 100 deformed versions of an image of a spoon - a selection is illustrated in Fig. 9.3. The spoons are deformed randomly according to two deformation modes, one altering the cup-size and the other controlling the grip width. The outline of the spoons in the target image is equipped with a dense set of 50 equidistant landmarks. The transformations used for warping the images are also applied to the landmarks. The deformed spoons are registered by a similarity transformation which minimizes the squared norm of the errors on the landmarks.

9.3.1.2 Test on Real Corpus Callosum Data

We test the proposed method on real data which is part of the LADIS (Leukoaraiosis And DISability) study [Pantoni et al., 2005], a pan-European study involving 12 hospitals and more than 600 patients. The data in question consists of 62 2D MR images of the midsagittal cross-section of the corpus callosum brain structure. The data set is equipped with a set of 72 corresponding landmarks in each image chosen by physicians. Attention is paid in order to achieve a possibly accurate and well-distributed set of landmarks. For an example image of the data set, please refer to Fig. 9.1a.

For the creation of SSMS on the corpus callosum data, nonlinear registration is performed by two different methods after a linear preregistration. While the first method (NRM 1) is dedicated and tested for the creation of shape models, the second (NRM 2) is an alternative method used for comparison in Sec. 9.3.2.

Linear Preregistration. The linear preregistration step is performed by a similarity transformation based on the landmarks in the images, minimizing the squared norm.

Nonlinear Registration. The nonlinear registration method (NRM 1) primarily used for computation of the deformation fields is dedicated to the construction of shape models and is shown to generate accurate results for the corpus callosum data in question [Hansen et al., 2007].

The method computes the deformations $T_{local i}$, based on registering the images I_i of the data set to the reference image I_R , by solving the following minimization problem iteratively

$$T_{local} = \min_{T'_{local}} \sum_i \mathcal{D}(I_R, I_i \circ T'_{local i}) + \mathcal{S}(T'_{local i}) , \quad (9.11)$$

where \mathcal{D} is a difference measure between images, and \mathcal{S} is a regularization term on the transformation. The process is iterative in the sense that starting from an initial estimate of the reference image I_R , which is simply an average of all the images, a new, improved estimate is computed in every iteration until convergence. It can be shown that under an assumption of Gaussian distribution of the noise, this choice of reference image is optimal with respect to achieving an unbiased coordinate frame for the shape model [Joshi et al., 2004]. For the corpus callosum data, \mathcal{D} is chosen as sum of squared distances and \mathcal{S} as the L2 norm on the parameter space. The parameterization of the deformation model is performed by free-form deformation (FFD), based on sine-kernels.

Nonlinear Registration II. This alternative method (NRM 2) is used in Sec. 9.3.2, in order to demonstrate that the occurrence of similarity components in deformation fields is independent of the chosen nonlinear method. The method is based on B-spline FFDs and uses discrete optimization using Markov random fields. It differs from NRM 1 in choice of parameters, grid resolution, and the optimization method. For details, please refer to [Glocker et al., 2007].

Method	Param	μ_a	m_a	σ_a	\max_a	μ	σ
NRM 1 (dense)	α	0.816	0.609	0.773	5.224	0.053	1.123
	Δc	0.028	0.022	0.025	0.165	-0.001	0.037
	t_x	2.08	1.51	2.03	12.46	-1.08	2.70
	t_y	2.23	1.58	2.20	12.93	-1.07	2.95
NRM 1 (region)	α	1.361	0.972	1.357	10.112	0.090	1.920
	Δc	0.028	0.016	0.034	0.191	-0.002	0.044
	t_x	2.61	1.83	2.74	18.54	-1.07	3.63
	t_y	2.60	1.66	2.86	19.47	-1.02	3.73
NRM 1 (boundary)	α	1.293	0.967	1.272	9.312	0.080	1.813
	Δc	0.028	0.015	0.034	0.187	-0.001	0.044
	t_x	2.60	1.84	2.69	18.21	-1.07	3.59
	t_y	2.55	1.64	2.83	19.37	-0.99	3.69
NRM 2 (dense)	α	0.698	0.588	0.556	3.955	0.294	0.843
	Δc	0.010	0.008	0.008	0.056	0.001	0.013
	t_x	1.28	1.15	0.91	6.02	1.03	1.19
	t_y	1.39	1.23	0.95	5.24	-1.10	1.28
NRM 2 (region)	α	1.458	1.158	1.365	9.034	0.321	1.972
	Δc	0.019	0.011	0.022	0.157	-0.003	0.028
	t_x	2.10	1.46	2.04	12.80	0.84	2.81
	t_y	2.07	1.58	2.06	13.36	-1.05	2.72
NRM 2 (boundary)	α	1.408	1.159	1.241	8.226	0.254	1.860
	Δc	0.018	0.011	0.021	0.150	-0.002	0.027
	t_x	2.07	1.53	1.92	11.76	0.81	2.70
	t_y	2.01	1.54	1.94	12.25	-0.96	2.62

Table 9.1: Quantification of the similarity transformations extracted for the corpus callosum data set. The *amount* of the transformation is described by the mean (μ_a), median (m_a), standard deviation (σ_a) and maximum (\max_a) of the norm of the parameters. The *variation* is given by the mean (μ) and standard deviation (σ) of the actual parameters. The scaling is expressed as deviation from 1, that is $\Delta c = c - 1$. The rotation is given in degrees, translation in millimeters, and scaling is a unit-less factor.

9.3.2 Quantification of Extracted Similarity Transformation Components

In this section, we quantify the amount of similarity transformations which were extracted from deformations computed for the corpus callosum example. We describe the *amount* of the extracted similarity transformations by computing the mean, median, standard deviation, and maximum of the norm of the computed parameters. The *variation* of the parameters is described by the mean and standard deviation of the actual parameters (not their norm).

In order to show that the existence of the effect is not dependent on the nonlinear registration method, we computed the results by (NRM 1) as well as by an alternative method (NRM 2) described in 9.3.1.2.

Furthermore, the extraction is performed by considering three different regions of the image domain: 1) the complete *dense* field on the whole image domain, 2) only the segmented *region* of the corpus callosum on the reference image, and 3)

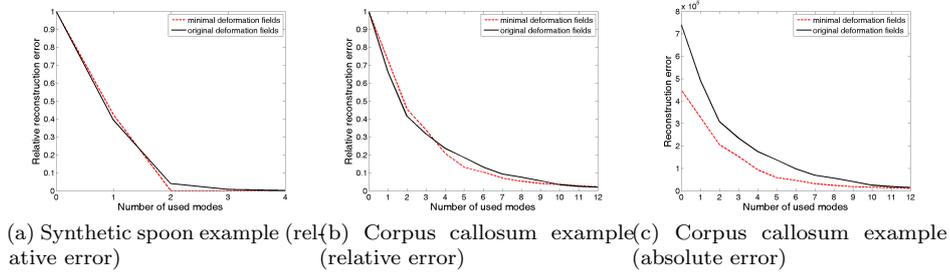


Figure 9.4: Analysis of the *relative* reconstruction error of the original and modified shape model on (a) synthetic and (b) on the corpus callosum data (NRM 1). Please note that the error is scaled by the initial error, such that both plots start with unit error. This is done to have a fair comparison, since the absolute error for the original model is always larger, compare (c). Notice the better reconstruction ability of the modified model when more than the very first modes are used.

only the *boundary* of the corpus callosum on the reference image. As discussed in [Rueckert et al., 2003], the construction of the model based on the region of interest can be a meaningful choice for a specific part of anatomy. For these cases (*region, boundary*), the extracted similarity components are larger than for the complete dense field, compare Table 9.1.

The results in Table 9.1 demonstrate that there is a large and highly varying amount of similarity transformation contained in the computed deformation fields. For a visualization of the extracted similarity transformations (based on the deformation in the corpus callosum region), please refer to the supplementary material at <http://campar.in.tum.de/personal/zikic/cvpr2008/>.

9.3.3 Effects of Proposed Method on SSMs

In the following sections we discuss the impact of using minimal deformations for construction of shape models and compare the resulting models with original SSMs.

In Sec. 9.3.3.1 we show that not only the variance, but also the relative reconstruction error of the modified SSMs is lower. More importantly, in Sec. 9.3.3.2 we inspect the contribution of the single modes to error reduction. We demonstrate that the original modes contain a similarity transformation, which is an unwanted property for SSMs, while the modified SSMs are similarity-free by construction.

9.3.3.1 Reconstruction Ability of the Model

The shape models built from minimal deformations have a reduced variance compared to the original models - this is equivalent to a larger reconstruction error, and shown in Fig. 9.4c. As we show in Sec. 9.3.3.2 and Fig. 9.5, this is due to the original model containing a similarity component. However, we also investigate how much of the *remaining* variance can be explained by using the

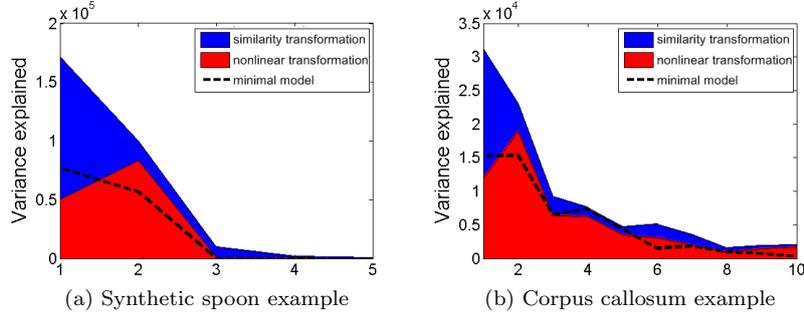


Figure 9.5: Analysis of the reconstruction ability of the single modes of the original and modified shape model on synthetic **(a)** and the corpus callosum data (NRM 1). The variance explained by single modes of the respective model is given. The contribution for the original model is decomposed into the variance explaining the similarity transformation component and the nonlinear component. This compares well to our theoretical analysis in Figure 9.2, supporting the hypothesis that the similarity transformation component is the most significant in the original deformation data.

model and to this end we compute the *relative* reconstruction error. Here, the reconstruction error as defined in Sec. 9.2.3 is divided by the norm of original error, in order to enable the comparison of relative performance of original and modified model.

The reconstruction ability is tested by averaging a series of 10 leave-10-out experiments. It can be seen in Fig. 9.4 that the very first modes of the original model describe more than the corresponding modes of the modified model. Sec. 9.3.3.2 and Fig. 9.5 reveal that this relative improvement is due to the reduction of the error in the similarity component of the original model and that actually more deformation is described by the first mode of the modified model for both test cases. When more than the very first modes are included - which is the interesting case for applications - the reduction of the relative error of the modified model is superior, compare Fig. 9.4.

Please also note that for the simple synthetic example which was generated from two modes, the relative reconstruction error is reduced to zero by the modified model by using the first two modes, in contrast to the original model, compare Fig. 9.4a.

9.3.3.2 Analysis of the Single Modes

To gain further insight into the reconstruction ability of the models, the single modes of the models are examined. For this, the modes of the original model are divided into a similarity and a nonlinear component, as described in Sec. 9.2.2.1. Please note that the modified model does not contain a similarity transformation by construction. In Fig. 9.5 we visualize the variance explained by the single components of the original model, and compare this with the modified model. The results correspond to the theoretical prediction given in Sec. 9.2.3.1 and Fig. 9.2.

Several observations favor the minimal model. First, the modified model is more compact. Second, the first mode of the modified model explains more nonlinear deformation than the first mode of the original model, which mostly explains the similarity component. This means that if non-minimal deformations are used to construct the model, the first mode does not describe the strongest variation in shape.

9.4 Discussion and Conclusion

In this work, we show that similarity transformations are inherently a part of the deformation fields obtained by standard nonlinear registration schemes. This is in contrast to the assumptions usually made. We propose a method which computes minimal nonlinear deformations by extracting the similarity transformation components.

We show that using minimal deformations is crucial for the construction of shape models, since otherwise the models are seen to describe other effects than changes in shape, which can lead to wrong results for the applications.

The proposed method eliminates the negative effects which can occur for SSMs based on non-minimal deformations. Existing SSM frameworks based on deformations can be extended in a straight forward way to make use of our method and thus benefit from using minimal deformations.

LIST OF AUTHOR'S PUBLICATIONS

- [1] Darko Zikic, Maximilian Baust, Ali Kamen, and Nassir Navab. A general preconditioning scheme for difference measures in deformable registration. In *Proc. International Conference on Computer Vision (ICCV)*, 2011. 27, 43, 49, 61, 79
- [2] Wolfgang Wein, Oliver Kutter, André Aichert, Darko Zikic, Ali Kamen, and Nassir Navab. Automatic non-linear mapping of pre-procedure ct volumes to 3d ultrasound. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, Rotterdam, The Netherlands, April 2010.
- [3] Maximilian Baust, Darko Zikic, and Nassir Navab. Diffusion-based regularisation strategies for variational level set segmentation. In *Proc. British Machine Vision Conference (BMVC)*, 2010.
- [4] Martin Groher, Maximilian Baust Darko Zikic, and Nassir Navab. Monocular deformable model-to-image registration of vascular structures. In *Proc. International Workshop on Biomedical Image Registration (WBIR)*, 2010.
- [5] Darko Zikic, Maximilian Baust, Ali Kamen, and Nassir Navab. Generalization of deformable registration in riemannian sobolev spaces. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2010. 25, 27, 49, 61, 67, 72, 83, 84
- [6] Darko Zikic, Ben Glocker, Oliver Kutter, Martin Groher, Nikos Komodakis, Ali Kamen, Nikos Paragios, and Nassir Navab. Linear Intensity-based Image Registration by Markov Random Fields and Discrete Optimization. *Medical Image Analysis (MedIA)*, 14(4):550–562, 2010. 115
- [7] Darko Zikic, Ben Glocker, Oliver Kutter, Martin Groher, Nikos Komodakis, Ali Khamene, Nikos Paragios, and Nassir Navab. Markov Random Field Optimization for Intensity-based 2D-3D Registration. In *Proc. SPIE Medical Imaging*, 2010. 115
- [8] Darko Zikic, Ali Kamen, and Nassir Navab. Natural gradients for deformable registration. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 61, 67

LIST OF AUTHOR'S PUBLICATIONS

- [9] Darko Zikic, Ali Kamen, and Nassir Navab. Revisiting horn and schunck: Interpretation as gauß-newton optimisation. In *Proc. British Machine Vision Conference (BMVC)*, 2010. 27, 31, 61, 67
- [10] Darko Zikic, Ali Kamen, and Nassir Navab. Unifying characterization of deformable registration methods based on the inherent parameterization. In *Proc. International Workshop on Biomedical Image Registration (WBIR)*, 2010. 25, 56, 58
- [11] Ben Glocker, Darko Zikic, Nikos Komodakis, Nikos Paragios, and Nassir Navab. Linear image registration through mrf optimization. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, 2009. 115
- [12] Martin Groher, Darko Zikic, and Nassir Navab. Deformable 2d-3d registration of vascular structures in a one view scenario. *IEEE Transactions on Medical Imaging (TMI)*, 28(6):847–860, 2009. 89
- [13] Ali Khamene, Darko Zikic, Mamadou Diallo, Thomas Boettger, and Eike Rietzel. A novel intensity similarity metric with soft spatial constraint for a deformable image registration problem in radiation therapy. In Guang-Zhong Yang, David Hawkes, Daniel Rueckert, Alison Noble, and Chris Taylor, editors, *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 5761 of *Lecture Notes in Computer Science*, pages 828–836. Springer Berlin / Heidelberg, 2009.
- [14] Selen Atasoy, Martin Groher, Darko Zikic, Ben Glocker, Tobias Waggershauser, Marcus Pfister, and Nassir Navab. Real-time respiratory motion tracking: Roadmap correction for hepatic artery catheterizations. In *Proc. SPIE Medical Imaging*, February 2008.
- [15] D. Zikic, B. Glocker, M. S. Hansen, A. Khamene, and N. Navab. Construction of Statistical Shape Models from Minimal Deformations. In *Workshop Manifolds in Medical Imaging: Metrics, Learning and Beyond in conjunction with Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2008. 141
- [16] D. Zikic, M. S. Hansen, B. Glocker, A. Khamene, R. Larsen, and N. Navab. Computing Minimal Deformations: Application to Construction of Statistical Shape Models. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 141
- [17] Darko Zikic, Martin Groher, Ali Khamene, and Nassir Navab. Deformable registration of 3d vessel structures to a single projection image. In *Proc. SPIE Medical Imaging*, 2008. 89
- [18] Darko Zikic, Steven Sourbron, Xinxing Feng, Henrik J. Michaely, Ali Khamene, and Nassir Navab. Automatic alignment of renal dce-mri image series for improvement of quantitative tracer kinetic studies. In *Proc. SPIE Medical Imaging*, 2008.
- [19] M. Blume, D. Zikic, W. Wein, and N. Navab. A new and general method for blind shift-variant deconvolution of biomedical images. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, October 2007.

LIST OF AUTHOR'S PUBLICATIONS

- [20] Axel Martinez-Möller, Darko Zikic, René Botnar, Ralph Bundschuh, William Howe, Sibylle Ziegler, Nassir Navab, Markus Schwaiger, and Stephan Nekolla. Dual cardiac-respiratory gated pet: Implementation and results from a feasibility study. *European Journal of Nuclear Medicine and Molecular Imaging*, 34(9):1447–1454, September 2007.
- [21] A. Khamene, L. Schwarz, D. Zikic, F. Azar, E. Rietzel, and N. Navab. A unified and efficient approach for free-form deformable registration. In *Proc. International Conference on Computer Vision (ICCV)*, 2007.
- [22] Darko Zikic, Wolfgang Wein, Ali Khamene, Dirk-Andre Clevert, and Nassir Navab. Fast deformable registration of 3d-ultrasound using a variational approach. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4190, pages 915–923, Oct. 2006. [66](#), [86](#)
- [23] Darko Zikic, Ali Khamene, and Nassir Navab. Intensity-unbiased force computation for variational motion estimation. In *Workshop on Image Registration in Deformable Environments, in conjunction with BMVC*, pages 61–70, 2006. [61](#)

REFERENCES

- [Alperin et al., 1994] Alperin, N., Levin, D. N., and Pelizzari, C. A. (1994). Retrospective registration of x-ray angiograms with mr images by using vessels as intrinsic landmarks. *Journal of Magnetic Resonance Imaging (JMRI)*, 4:139–144. [92](#)
- [Amit, 1994] Amit, Y. (1994). A nonlinear variational problem for image matching. *SIAM Journal on Scientific Computing*, 15(1):207–224. [31](#), [54](#)
- [Arsigny et al., 2006] Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A log-euclidean framework for statistics on diffeomorphisms. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4190, pages 924–931. [39](#), [46](#)
- [Ashburner, 2007] Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95 – 113. [38](#), [39](#)
- [Ashburner and Friston, 1999] Ashburner, J. and Friston, K. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266. [51](#), [54](#)
- [Ashburner and Friston, 2004] Ashburner, J. and Friston, K. J. (2004). Spatial normalization using basis functions. In *Human Brain Function*. Academic Press. [51](#), [54](#)
- [Avants et al., 2008] 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. Special Issue on The Third International Workshop on Biomedical Image Registration - WBIR 2006. [36](#)
- [Aylward and Jomier, 2004] Aylward, S. and Jomier, J. (2004). Rigid and deformable vasculature-to-image registration: A hierarchical approach. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 3216 of *Lecture Notes in Computer Science (LNCS)*, pages 829–836. Springer. [90](#)

REFERENCES

- [Bajcsy and Kovačič, 1989] Bajcsy, R. and Kovačič, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46(1):1–21. [41](#)
- [Beg et al., 2005] Beg, M., Miller, M., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision (IJCV)*, 61(2):139–157. [34](#), [36](#), [37](#), [38](#), [39](#), [72](#)
- [Benameur et al., 2003] Benameur, S., Mignotte, M., Parent, S., Labelle, H., Skalli, W., and de Guise, J. (2003). 3d/2d registration and segmentation of scoliotic vertebra using statistical models. *Computerized Medical Imaging and Graphics*, 27:321–337. [92](#)
- [Besl and McKay, 1992] Besl, P. J. and McKay, D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14:239–256. [99](#)
- [Beuthien et al., 2010] Beuthien, B., Kamen, A., and Fischer, B. (2010). Recursive green’s function registration. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. [44](#)
- [Bishop, 2006] Bishop, C. (2006). *Pattern recognition and machine learning*. Springer. [119](#)
- [Bookstein, 1989a] Bookstein, F. (1989a). Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(6):567–585. [55](#)
- [Bookstein, 1989b] Bookstein, F. L. (1989b). Principal warps: Thin-plate splines and decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11:567–585. [95](#)
- [Boros and Hammer, 2002] Boros, E. and Hammer, P. (2002). Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123:155–225. [119](#)
- [Bossa et al., 2007] Bossa, M., Hernandez, M., and Olmos, S. (2007). Contributions to 3d diffeomorphic atlas estimation: Application to brain images. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4791, page 667. Springer. [46](#)
- [Boykov and Jolly, 2001] Boykov, Y. and Jolly, M.-P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *Proc. International Conference on Computer Vision (ICCV)*. [116](#)
- [Boykov et al., 2001] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(11):1222–1239. [116](#), [118](#), [125](#)
- [Bro-Nielsen, 1996] Bro-Nielsen, M. (1996). *Medical Image Registration and Surgery Simulation*. PhD thesis, Department of Mathematical Modelling (IMM), Technical University of Denmark. [12](#), [31](#), [35](#)

- [Bro-Nielsen and Gramkow, 1996] Bro-Nielsen, M. and Gramkow, C. (1996). Fast fluid registration of medical images. In *Proc. International Conference on Visualization in Biomedical Computing (VBC)*, volume 1131, pages 267–276. [35](#), [39](#), [44](#)
- [Broit, 1981] Broit, C. (1981). *Optimal Registration of Deformed Images*. PhD thesis, University of Pennsylvania. [19](#), [22](#), [27](#), [30](#), [31](#)
- [Brown, 1992] Brown, L. G. (1992). A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376. [17](#)
- [Brox et al., 2004] Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In *Proc. European Conference on Computer Vision (ECCV)*, pages 25–36. Springer. [27](#)
- [Cachier et al., 2003] Cachier, P., Bardinet, E., Dormont, D., Pennec, X., and Ayache, N. (2003). Iconic feature based nonrigid registration: the pasha algorithm. *Computer Vision and Image Understanding (CVIU)*, 89(2-3):272–298. [42](#), [43](#), [44](#)
- [Cachier et al., 1999] Cachier, P., Pennec, X., and Ayache, N. (1999). Fast non-rigid matching by gradient descent: Study and improvements of the demons algorithm. Research Report 3706, INRIA. [42](#)
- [Cahill et al., 2009] Cahill, N., Noble, J., and Hawkes, D. (2009). Demons algorithms for fluid and curvature registration. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 730–733. [42](#)
- [Cahill, 2009] Cahill, N. D. (2009). *Constructing and solving variational image registration problems*. PhD thesis, University of Oxford. [25](#), [42](#), [45](#)
- [Chan et al., 2004] Chan, H., Chung, A., Yu, S., and Wells, W. (2004). 2D-3D vascular registration between digital subtraction angiographic (DSA) and magnetic resonance angiographic (MRA) images. In *Proc. of the IEEE International Symposium on Biomedical Imaging*, volume 1205 of *Lecture Notes in Computer Science (LNCS)*, pages 708–711. IEEE. [92](#)
- [Charnoz et al., 2005] Charnoz, A., Agnus, V., Malandain, G., Forest, C., Tajine, M., and Soler, L. (2005). Liver registration for the follow-up of hepatic tumors. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 3750 of *Lecture Notes in Computer Science (LNCS)*, pages 155–162. Springer. [90](#)
- [Chefd’hotel, 2005] Chefd’hotel, C. (2005). *Geometric Methods in Computer Vision and Image Processing: Contributions and Applications*. PhD thesis, L’Ecole Normale Supérieure de Cachan. [44](#), [57](#), [72](#)
- [Chefd’hotel et al., 2002] Chefd’hotel, C., Hermosillo, G., and Faugeras, O. (2002). Flows of diffeomorphisms for multimodal image registration. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 753–756. [16](#), [36](#), [38](#), [39](#), [43](#), [44](#), [45](#), [65](#), [67](#), [71](#), [72](#), [75](#), [81](#), [83](#)
- [Christensen and Johnson, 2001] Christensen, G. and Johnson, H. (2001). Consistent image registration. *IEEE Transactions on Medical Imaging (TMI)*, 20(7):568–582. [54](#)

REFERENCES

- [Christensen et al., 1996] Christensen, G., Rabbitt, R., and Miller, M. (1996). Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing (TIP)*, 5(10):1435–1447. [32](#), [33](#), [34](#), [35](#), [38](#), [41](#)
- [Christensen, 1994] Christensen, G. E. (1994). *Deformable Shape Models for Anatomy*. PhD thesis, Washington University, Sever Institute of Technology. [11](#), [12](#), [31](#), [32](#), [33](#), [34](#), [35](#), [44](#)
- [Chui and Rangarajan, 2003] Chui, H. and Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding (CVIU)*, 89:114–141. [101](#), [104](#)
- [Chung et al., 2002] Chung, A. C. S., William M. Wells, I., Norbash, A., and Grimson, W. E. L. (2002). Multi-modal image registration by minimising kullback-leibler distance. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science (LNCS), pages 525–532. Springer. [92](#)
- [Cocosco et al., 1997] Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B., and Evans, A. C. (1997). Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage*, 5:425. [79](#)
- [Cootes et al., 1995] Cootes, T., Taylor, C., Cooper, D., Graham, J., et al. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding (CVIU)*, 61(1):38–59. [143](#)
- [Cootes et al., 2004] Cootes, T., Twining, C., and Taylor, C. (2004). Diffeomorphic statistical shape models. In *Proc. British Machine Vision Conference (BMVC)*, volume 1, pages 447–456. [143](#), [145](#), [146](#)
- [Droske and Rumpf, 2004] Droske, M. and Rumpf, M. (2004). A variational approach to nonrigid morphological image registration. *SIAM Journal on Applied Mathematics*, 64(2):668–687. [30](#)
- [Dupuis et al., 1998] Dupuis, P., Grenander, U., and Miller, M. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56(3):587–600. [11](#), [19](#), [33](#), [34](#), [35](#), [36](#), [37](#), [38](#), [72](#)
- [Estepar et al., 2004] Estepar, R. S. J., Brun, A., and Westin, C.-F. (2004). Robust generalized total least squares iterative closest point registration. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 3216 of *Lecture Notes in Computer Science (LNCS)*, page 234–241. [99](#)
- [Evans, 1998] Evans, L. C. (1998). *Partial Differential Equations*. American Mathematical Society. [19](#), [20](#), [26](#)
- [Feldmar et al., 1995] Feldmar, J., Ayache, N., and Betting, F. (1995). 3D-2D projective registration of free-form curves and surfaces. In *Proc. International Conference on Computer Vision (ICCV)*, volume 20-23, pages 549–556. IEEE. [92](#), [94](#)
- [Ferrant, 2001] Ferrant, M. (2001). *Physics-based Deformable Modeling of Volumes and Surfaces for Medical Image Registration, Segmentation and Visualization*. PhD thesis, Université catholique de Louvain. [30](#)

- [Ferrant et al., 1999] Ferrant, M., Warfield, S., Guttman, C., Mulkern, R., Jolesz, F., and Kikinis, R. (1999). 3d image matching using a finite element based elastic deformation model. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 99. 30
- [Ferrant et al., 2000] Ferrant, M., Warfield, S., Nabavi, A., Jolesz, F., and Kikinis, R. (2000). Registration of 3d intraoperative mr images of the brain using a finite element biomechanical model. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 1935, pages 19–28. 30
- [Fitzpatrick et al., 2000] Fitzpatrick, J., Hill, D., and Maurer Jr, C. (2000). Image registration. *Handbook of medical imaging - Medical Image Processing and Analysis*, 2:447–513. 17
- [Fleute and Lavalée, 1999] Fleute, M. and Lavalée, S. (1999). Nonrigid 3-d/2-d registration of images using statistical models. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 138–147, London, UK. Springer. 92
- [Florin et al., 2005] Florin, C., Williams, J., Khamene, A., and Paragios, N. (2005). Registration of 3D angiographic and X-ray images using sequential monte carlo sampling. In *Computer Vision for Biomedical Image Applications, First Int'l Workshop, CVBIA '05*, volume 3765 of *Lecture Notes in Computer Science (LNCS)*, pages 427–436. Springer. 92
- [Gee and Bajcsy, 1998] Gee, J. and Bajcsy, R. (1998). Elastic matching: Continuum mechanical and probabilistic analysis. *Brain Warping*. 143
- [Gee et al., 1994] Gee, J., Haynor, D., Reivich, M., and Bajcsy, R. (1994). Finite element approach to warping of brain images. In *Proc. SPIE Medical Imaging*, pages 327–337. 30
- [Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6(6):721–741. 115
- [Gill et al., 1986] Gill, P. E., Murray, W., and Wright, M. H. (1986). *Practical Optimization*. Elsevier. 98
- [Glocker et al., 2007] Glocker, B., Komodakis, N., Paragios, N., Tziritas, G., and Navab, N. (2007). Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *Information Processing in Medical Imaging*, Kerkrade, Netherlands. 149
- [Glocker et al., 2008a] Glocker, B., Komodakis, N., Tziritas, G., Navab, N., and Paragios, N. (2008a). Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis (MedIA)*, 12(6):731–741. 51, 53, 80, 116, 126

REFERENCES

- [Glocker et al., 2008b] Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., and Navab, N. (2008b). Optical flow estimation with uncertainties through dynamic mrfs. *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 49
- [Glocker et al., 2009] Glocker, B., Zikic, D., Komodakis, N., Paragios, N., and Navab, N. (2009). Linear image registration through mrf optimization. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*. 115
- [Gold and Rangarajan, 1996] Gold, S. and Rangarajan, A. (1996). A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18(4):377–388. 99, 101
- [Gold et al., 1998] Gold, S., Rangarajan, A., ping Lu, C., and Mjolsness, E. (1998). New algorithms for 2D and 3D point matching: Pose estimation and correspondence. *Pattern Recognition*, 31:957–964. 101
- [Gorges et al., 2005] Gorges, S., Kerrien, E., Berger, M.-O., Troussset, Y., Pescatore, J., Anxionnat, R., and Picard, L. (2005). Model of a vascular C-Arm for 3D augmented fluoroscopy in interventional radiology. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 214–222. 112
- [Granger and Pennec, 2002] Granger, S. and Pennec, X. (2002). Multi-scale EM-ICP: A fast and robust approach for surface registration. In *Proceedings of the 7th European Conference on Computer Vision-Part IV (ECCV’02)*, pages 418–432, London, UK. Springer-Verlag. 99
- [Grenander and Miller, 1998] Grenander, U. and Miller, M. (1998). Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, 56:617–694. 143
- [Grenander and Miller, 2006] Grenander, U. and Miller, M. (2006). *Pattern Theory: From Representation to Inference*. Oxford University Press. 33
- [Grochow et al., 2004] Grochow, K., Martin, S. L., Hertzmann, A., and Popovic, Z. (2004). Style-based inverse kinematics. In *ACM Transactions on Graphics (Proc. of SIGGRAPH 2004)*. 93
- [Groher et al., 2007a] Groher, M., Bender, F., Hoffmann, R., and Navab, N. (2007a). Segmentation-driven 2D-3D registration for abdominal catheter interventions. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4792 of *LNCS*, pages 527–535. Springer. 92, 111, 112
- [Groher et al., 2007b] Groher, M., Jakobs, T., Padoy, N., and Navab, N. (2007b). Planning and intraoperative visualization of liver catheterizations: New CTA protocol and 2D-3D registration method. *Academic Radiology*, 14(11):1324–1339. 92, 94, 111
- [Groher et al., 2009] Groher, M., Zikic, D., and Navab, N. (2009). Deformable 2d-3d registration of vascular structures in a one view scenario. *IEEE Transactions on Medical Imaging (TMI)*, 28(6):847–860. 89

- [Guimond et al., 2002] Guimond, A., Roche, A., Ayache, N., and Meunier, J. (2002). Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. *IEEE Transactions on Medical Imaging (TMI)*, 20(1):58–69. [43](#), [67](#), [71](#)
- [Hajnal et al., 2001a] Hajnal, J., Hawkes, D., and Hill, D. (2001a). *Medical Image Registration*. CRC Press. [116](#), [123](#)
- [Hajnal et al., 2001b] Hajnal, J. V., Hill, D. L., and Hawkes, D. J., editors (2001b). *Medical Image Registration*. CRC Press. [17](#)
- [Hansen et al., 2007] Hansen, M., Hansen, M., and Larsen, R. (2007). Diffeomorphic Statistical Deformation Models. *Workshop on Non-rigid Registration and Tracking through Learning - NRTL 2007, in conjunction with ICCV 2007*. [149](#)
- [Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press. [123](#)
- [Hermosillo, 2002] Hermosillo, G. (2002). *Variational Methods for Multimodal Image Matching*. PhD thesis, Université de Nice - Sophia Antipolis. [22](#)
- [Hermosillo et al., 2002] Hermosillo, G., Chefd’Hotel, C., and Faugeras, O. (2002). Variational methods for multimodal image matching. *International Journal of Computer Vision (IJCV)*, 50(3):329–343. [16](#), [65](#), [80](#)
- [Hernandez et al., 2009] Hernandez, M., Bossa, M., and Olmos, S. (2009). Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision (IJCV)*, 85:291–306. 10.1007/s11263-009-0219-z. [37](#), [39](#)
- [Hernandez et al., 2007] Hernandez, M., Bossa, M. N., and Olmos, S. (2007). Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. [39](#)
- [Hernandez et al., 2008] Hernandez, M., Olmos, S., and Pennec, X. (2008). Comparing algorithms for diffeomorphic registration: Stationary lddmm and diffeomorphic demons. In Pennec, X. and Joshi, S., editors, *Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2008)*. [39](#)
- [Hill et al., 2001] Hill, D. L. G., Batchelor, P. G., Holden, M., and Hawkes, D. J. (2001). Medical image registration. *Physics in Medicine and Biology*, 46(3):R1–R45. [17](#)
- [Hipwell et al., 2003] Hipwell, J., Penney, G., McLaughlin, R., Rhode, K., Summers, P., Cox, T., Byrne, J., Noble, J., and Hawkes, D. (2003). Intensity based 2D-3D registration of cerebral angiograms. *IEEE Transactions on Medical Imaging (TMI)*, 22:1417–1426. [92](#)
- [Holden, 2008] Holden, M. (2008). A review of geometric transformations for nonrigid body registration. *IEEE Transactions on Medical Imaging*, 27(1):111–128. [54](#)

REFERENCES

- [Horn and Schunck, 1981] Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203. [19](#), [27](#), [31](#)
- [Horn and Schunck, 1980] Horn, B. K. and Schunck, B. G. (1980). Determining optical flow. Technical report, Massachusetts Institute of Technology, Cambridge, MA, US, Cambridge, MA, USA. [19](#), [31](#)
- [Ishikawa, 2009] Ishikawa, H. (2009). Higher-order clique reduction in binary graph cut. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. [119](#)
- [Jomier et al., 2006] Jomier, J., Bullitt, E., van Horn, M., Pathak, C., and Aylward, S. (2006). 3d/2d model-to-image registration applied to tips surgery. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 662–670. [92](#), [111](#), [112](#)
- [Joshi et al., 2004] Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:151–160. [149](#)
- [Joshi, 1998] Joshi, S. C. (1998). *Large deformation diffeomorphisms and Gaussian random fields for statistical characterization of brain sub-manifolds*. PhD thesis, Dept. of Electrical Engineering, Washington University. [143](#)
- [Kerrien et al., 1999] Kerrien, E., Berger, M.-O., Maurincomme, E., Launay, L., Vaillant, R., and Picard, L. (1999). Fully automatic 3d/2d subtracted angiography registration. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 1679 of *Lecture Notes in Computer Science (LNCS)*, pages 664–671. Springer. [92](#)
- [Kita et al., 1998] Kita, Y., Wilson, D., and Noble, J. (1998). Real-time registration of 3D cerebral vessels to X-ray angiograms. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 1496 of *Lecture Notes in Computer Science (LNCS)*, pages 1125–1133. Springer. [92](#)
- [Klein et al., 2009] Klein, S., Pluim, J. P. W., Staring, M., and Viergever, M. (2009). Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision (IJCV)*, 81(3):227–239. [133](#), [132](#)
- [Klein et al., 2011] Klein, S., Staring, M., Andersson, P., and Pluim, J. (2011). Preconditioned stochastic gradient descent optimisation for monomodal image registration. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. [68](#)
- [Klein et al., 2010a] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. (2010a). elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging (TMI)*, 29(1):196–205. [132](#)
- [Klein et al., 2010b] Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. W. (2010b). elastix: A toolbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on*, 29(1):196–205. [51](#), [53](#)

- [Klein et al., 2007] Klein, S., Staring, M., and Pluim, J. P. W. (2007). Evaluation of optimisation methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Transactions on Image Processing (TIP)*, 16(12):2879–2890. 133
- [Kohli et al., 2009] Kohli, P., Ladicky, L., and Torr, P. H. (2009). Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision (IJCV)*, 82(3):302–324. 118
- [Kohli et al., 2007] Kohli, P., Mudigonda, P., and Torr, P. (2007). P3 & beyond: Solving energies with higher order cliques. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 118
- [Kolmogorov, 2006] Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28(10):1568–1583. 116, 118, 125
- [Kolmogorov and Rother, 2007] Kolmogorov, V. and Rother, C. (2007). Minimizing nonsubmodular functions with graph cuts—a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(7):1274–1279. 125
- [Komodakis and Paragios, 2009] Komodakis, N. and Paragios, N. (2009). Beyond pairwise energies: Efficient optimization for higher-order mrfs. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 119
- [Komodakis et al., 2007] Komodakis, N., Tziritas, G., and Paragios, N. (2007). Fast, approximately optimal solutions for single and dynamic mrfs. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 116, 118, 125
- [Komodakis et al., 2008] 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 (CVIU)*, 112(1):14–29. 125
- [Kroon and Slump, 2009] Kroon, D. and Slump, C. (2009). MRI modality transformation in demon registration. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*. 67
- [Kybic and Unser, 2003] Kybic, J. and Unser, M. (2003). Fast parametric elastic image registration. *Image Processing, IEEE Transactions on*, 12(11):1427–1442. 53
- [Lan et al., 2006] Lan, X., Roth, S., Huttenlocher, D., and Black, M. J. (2006). Efficient belief propagation with learned higher-order markov random fields. In *Proc. European Conference on Computer Vision (ECCV)*. 119
- [Lempitsky et al., 2007] Lempitsky, V., Rother, C., and Blake, A. (2007). Log-cut - efficient graph cut optimization for markov random fields. In *Proc. International Conference on Computer Vision (ICCV)*. 116, 118, 119, 125
- [Li, 2001] Li, S. Z. (2001). *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc. 115

REFERENCES

- [Liu et al., 1998] Liu, A., Bullitt, E., and Pizer, S. (1998). 3D/2D registration via skeletal near projective invariance in tubular objects. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 1496 of *Lecture Notes in Computer Science (LNCS)*, pages 952–963. Springer. 92
- [Liviyatan et al., 2002] Liviyatan, H., Yaniv, Z., and Joskowicz, L. (2002). Robust automatic C-Arm calibration for fluoroscopy-based navigation: A practical approach. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 112
- [Lorenzen et al., 2006] Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., and Joshi, S. (2006). Multi-modal image set registration and atlas formation. *Medical Image Analysis*, 10(3):440 – 451. Special Issue on The Second International Workshop on Biomedical Image Registration (WBIR’03). 36
- [Madsen et al., 2004] Madsen, K., Nielsen, H., and Tingleff, O. (2004). Methods for non-linear least squares problems. Technical report. 27, 64
- [Maes et al., 1997a] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997a). Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198. 63
- [Maes et al., 1997b] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997b). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging (TMI)*, 16(2):187–198. 132
- [Maintz and Viergever, 1998] Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36. 17, 116
- [Malis, 2004] Malis, E. (2004). Improving vision-based control using efficient second-order minimization techniques. *Robotics and Automation, 2004. Proceedings. ICRA’04. 2004 IEEE International Conference on*, 2. 47
- [Miller et al., 2006] Miller, M., Trounev, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 24:209–228. 10.1007/s10851-005-3624-0. 39
- [Miller and Younes, 2001] Miller, M. I. and Younes, L. (2001). Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision (IJCV)*, 41:61–84. 45
- [Modat et al., 2010] Modat, M., Vercauteren, T., Ridgway, G. R., Hawkes, D. J., Fox, N. C., and Ourselin, S. (2010). Diffeomorphic demons using normalized mutual information, evaluation on multimodal brain mr images. In *Proc. SPIE Medical Imaging*. 27, 43, 64, 67, 68, 71, 80
- [Modersitzki, 2004] Modersitzki, J. (2004). *Numerical methods for image registration*. Oxford University Press. 31, 55
- [Modersitzki, 2009] Modersitzki, J. (2009). *FAIR: Flexible Algorithms for Image Registration*. SIAM. 14, 68

- [Murphy et al., 2011] 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., Comowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporring, J., de Bruijne, M., Han, X., Heinrich, M., Schnabel, J., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J., Ourselin, S., Münzing, S., Viergever, M., Nigris, D. 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. (2011). Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Transactions on Medical Imaging (TMI)*. 49
- [Nagel, 1983] Nagel, H.-H. (1983). Constraints for the estimation of displacement vector fields from image sequences. In *IJCAI'83: Proceedings of the Eighth international joint conference on Artificial intelligence*, pages 945–951, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. 31
- [Navab et al., 1996] Navab, N., Ben-Hashemi, A., Mitschke, M. M., Holdsworth, D. W., Fahrig, R., Fox, A., and Graumann, R. (1996). Dynamic geometrical calibration for 3d cerebral angiography. In *SPIE Medical Imaging*. 112
- [Neuberger, 1997] Neuberger, J. (1997). *Sobolev gradients and differential equations*. Springer Berlin. 56, 72
- [Nocedal and Wright, 2000] Nocedal, J. and Wright, S. (2000). *Numerical optimization*. Springer. 24, 64, 67, 68, 69
- [Palágyi et al., 2001] Palágyi, K., Sorantin, E., Balogh, E., Kuba, A., Halmai, C., Erdöhelyi, B., and Hausegger, K. (2001). A sequential 3d thinning algorithm and its medical applications. In *Proc. International Conference on Information Processing in Medical Imaging (IPMI)*, volume 2028 of *Lecture Notes in Computer Science (LNCS)*, pages 409–415. Springer. 94, 111
- [Pantoni et al., 2005] Pantoni, L., Basile, A. M., Pracucci, G., Asplund, K., Bogousslavsky, J., Chabriat, H., Erkinjuntti, T., Fazekas, F., Ferro, J. M., Hennerici, M., O'brien, J., Scheltens, P., Visser, M. C., Wahlund, L. O., Waldemar, G., Wallin, A., and Inzitari, D. (2005). Impact of age-related cerebral white matter changes on the transition to disability - the LADIS study: Rationale, design and methodology. *Neuroepidemiology*, 24(1-2):51–62. 149
- [Papenberg et al., 2006] Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J. (2006). Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision (IJCV)*, 67(2):141–158. 27
- [Pearl, 1988] Pearl, J. (1988). *Probabilistic Reasoning*. San Francisco, CA: Morgan Kaufmann. 125
- [Pennec et al., 1999] Pennec, X., Cachier, P., and Ayache, N. (1999). Understanding the demon's algorithm: 3d non-rigid registration by gradient descent. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 27, 42, 43, 44, 71

REFERENCES

- [Penney et al., 2001] Penney, G., Edwards, P., King, A., Blackall, J., Batchelor, P., , and Hawkes, D. (2001). A stochastic iterative closest point algorithm (stochastICP). In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science (LNCS)*, pages 762–769. 99
- [Penney et al., 1998] Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L. G., and Hawkes, D. J. (1998). A comparison of similarity measures for use in 2D-3D medical image registration. *IEEE Transactions on Medical Imaging (TMI)*, 17(4):586–595. 136
- [Pluim et al., 2003b] Pluim, J., Maintz, J., and Viergever, M. (2003b). Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging (TMI)*, 22(8):986–1004. 125
- [Pluim et al., 2003a] Pluim, J., Maintz, J., and Viergever, M. (Aug. 2003a). Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, 22(8):986–1004. 63
- [Potetz and Lee, 2008] Potetz, B. and Lee, T. S. (2008). Efficient belief propagation for higher-order cliques using linear constraint nodes. *Computer Vision and Image Understanding (CVIU)*, 112:39–54. 119
- [Potts, 1952] Potts, R. (1952). Some generalized order-disorder transformation. In *Proceedings of the Cambridge Philosophical Society*, volume 48. 118
- [Press et al., 1993] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1993). *Numerical Recipes in C*. Cambridge University Press. 125, 129, 130, 128
- [Roche et al., 1998] Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The correlation ratio as a new similarity measure for multimodal image registration. *Proc. MICCAI*, 98:1115–1124. 63
- [Rogelj and Kovacic, 2006] Rogelj, P. and Kovacic, S. (2006). Symmetric image registration. *Medical Image Analysis*, 10(3):484 – 493. Special Issue on The Second International Workshop on Biomedical Image Registration (WBIR’03). 47
- [Rohde et al., 2003] Rohde, G., Aldroubi, A., and Dawant, B. (2003). The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging (TMI)*, 22(11):1470–1479. 51, 55
- [Rohlfing and Maurer, 2003] Rohlfing, T. and Maurer, C. (2003). Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *Information Technology in Biomedicine, IEEE Transactions on*, 7(1):16–25. 51
- [Rohlfing et al., 2004] Rohlfing, T., Maurer, C. R., O’Dell, W. G., and Zhong, J. (2004). Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Medical Physics*, 31(3):427–432. 90

-
- [Rohlfing et al., 2003] Rohlfing, T., Maurer, C.R., J., Bluemke, D., and Jacobs, M. (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. 51, 53, 54
- [Rohr, 2000] Rohr, K. (2000). Elastic registration of multimodal medical images: A survey. *KI*, 14(3):11–17. 55
- [Rohr, 2001] Rohr, K. (2001). *Landmark-based image analysis: using geometric and intensity models*. Kluwer Academic Publishers. 55
- [Rohr et al., 2001a] Rohr, K., Stiehl, H., Sprengel, R., Buzug, T., Weese, J., and Kuhn, M. (2001a). Landmark-based elastic registration using approximating thin-plate splines. *Medical Imaging, IEEE Transactions on*, 20(6):526–534. 55
- [Rohr et al., 2001b] Rohr, K., Stiehl, H. S., Sprengel, R., Buzug, T. M., Weese, J., and Kuhn, M. H. (2001b). Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging (TMI)*, 20(6):526–534. 95
- [Rother et al., 2009] Rother, C., Kohli, P., Feng, W., and Jia, J. (2009). Minimizing sparse higher order energy functions of discrete variables. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 119
- [Rueckert et al., 2006] Rueckert, D., Aljabar, P., Heckemann, R., Hajnal, J., and Hammers, A. (2006). Diffeomorphic registration using b-splines. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 702–709. 53
- [Rueckert et al., 2003] Rueckert, D., Frangi, A., and Schnabel, J. (2003). Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging (TMI)*, 22(8):1014–1025. 143, 146, 151
- [Rueckert et al., 1998] Rueckert, D., Hayes, C., Studholme, C., Summers, P., Leach, M., and Hawkes, D. (1998). Non-rigid registration of breast mr images using mutual information. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 51, 53
- [Rueckert et al., 1999a] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. (1999a). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging (TMI)*, 18(8):712–721. 51, 53, 54, 64
- [Rueckert et al., 1999b] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. (1999b). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging (TMI)*, 18(8):712–721. 97
- [Saad, 2003] Saad, Y. (2003). *Iterative methods for sparse linear systems*. SIAM. 69, 70

REFERENCES

- [Schnabel et al., 2001] Schnabel, J., Rueckert, D., Quist, M., Blackall, J., Castellano-Smith, A., Hartkens, T., Penney, G., Hall, W., Liu, H., Truwit, C., et al. (2001). A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. *Medical Image Computing and Computer-Assisted Intervention*, 2208:573–581. [51](#), [53](#)
- [Sederberg and Parry, 1986] Sederberg, T. and Parry, S. (1986). Free-form deformation of solid geometric models. *ACM Siggraph Computer Graphics*, 20(4):151–160. [53](#)
- [Selle et al., 2002] Selle, D., Preim, B., Schenk, A., and Peitgen, H.-O. (2002). Analysis of vasculature for liver surgery planning. *IEEE Transactions on Medical Imaging (TMI)*, 21(11):1344–1357. [94](#)
- [Stefanescu et al., 2003] Stefanescu, R., Pennec, X., and Ayache, N. (2003). Grid enabled non-rigid registration with a dense transformation and a priori information. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer. [45](#)
- [Stefanescu et al., 2004] Stefanescu, R., Pennec, X., and Ayache, N. (2004). Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis (MedIA)*, 8(3):325–42. [16](#), [44](#), [45](#)
- [Strang, 1986] Strang, G. (1986). *Introduction to Applied Mathematics*. Wellesley Cambridge Press. [28](#), [29](#), [30](#)
- [Studholme et al., 1999] Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86. [133](#), [132](#)
- [Szeliski, 2010] Szeliski, R. (2010). *Computer Vision Algorithms and Applications*. Springer. [2](#)
- [Szeliski and Coughlan, 1997] Szeliski, R. and Coughlan, J. (1997). Spline-based image registration. *International Journal of Computer Vision*, 22:199–218. [10.1023/A:1007996332012](#). [51](#)
- [Szeliski and Shum, 1996] Szeliski, R. and Shum, H.-Y. (1996). Motion estimation with quadtree splines. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(12):1199–1210. [51](#)
- [Szeliski et al., 2008] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 30(6):1068–1080. [125](#)
- [Tang and Ellis, 2005] Tang, T. S. Y. and Ellis, R. E. (2005). 2d/3d deformable registration using a hybrid atlas. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 3750 of *Lecture Notes in Computer Science (LNCS)*, pages 223–230. Springer. [92](#)

-
- [Thirion, 1996] Thirion, J. (1996). Non-rigid matching using demons. *International Conference on Computer Vision and Pattern Recognition*, pages 245–251. [35](#), [41](#)
- [Thirion, 1998] Thirion, J. (1998). Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis*, 2(3):243–260. [25](#), [26](#), [41](#), [70](#)
- [Thirion, 1995a] Thirion, J.-P. (1995a). Fast Non-Rigid Matching of 3D Medical Images. Research Report RR-2547, INRIA. [41](#), [44](#)
- [Thirion, 1995b] Thirion, J.-P. (1995b). Fast non-rigid matching of 3d medical images. In *Medical Robotics and Computer Aided Surgery (MRCAS’95)*, page 47–54. [41](#)
- [Trouvé, 1995] Trouvé, A. (1995). An infinite dimensional group approach for physics based models in patterns recognition. Preprint. [11](#), [19](#), [34](#), [35](#), [36](#), [37](#), [38](#)
- [Trouvé, 1998] Trouvé, A. (1998). Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision (IJCV)*, 28(3):213–221. [33](#), [35](#), [36](#), [37](#), [38](#), [39](#), [44](#), [45](#), [72](#)
- [Turgeon et al., 2005] Turgeon, G.-A., Lehmann, G., Guiraudon, G., Drangova, M., Holdsworth, D., and Peters, T. (2005). 2D-3D registration of coronary angiograms for cardiac procedure planning and guidance. *Medical Physics*, 32:3737–3749. [92](#), [111](#), [112](#)
- [Tustison et al., 2006] Tustison, N. J., Avants, B. B., Sundaram, T. A., Duda, J., and Gee, J. C. (2006). A generalization of free-form deformation image registration within the itk finite element framework. In *International Workshop on Biomedical Image Registration (WBIR)*, pages 238–246. [30](#)
- [Umeyama, 1991a] Umeyama, S. (1991a). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(4):376–380. [105](#)
- [Umeyama, 1991b] Umeyama, S. (1991b). Least-squares estimation of transformation parameters between two point patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(4):376–380. [145](#)
- [Unser et al., 1993a] Unser, M., Aldroubi, A., and Eden, M. (1993a). B-spline signal processing. part i. theory. *IEEE Transactions on Signal Processing*, 41(2):821–833. [54](#)
- [Unser et al., 1993b] Unser, M., Aldroubi, A., and Eden, M. (1993b). B-spline signal processing. part ii. efficient design and applications. *IEEE Transactions on Signal Processing*, 41(2):834–848. [54](#)
- [Vaillant et al., 2004] Vaillant, M., Miller, M., Younes, L., and Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *Neuroimage*, 23:161–169. [39](#)

REFERENCES

- [Van den Elsen et al., 1993] Van den Elsen, P., Pol, E., and Viergever, M. (1993). Medical image matching—a review with classification. *IEEE Engineering in Medicine and Biology Magazine*, 12(1):26–39. 17
- [Vercauteren, 2008] Vercauteren, T. (2008). *Image Registration and Mosaicing for Dynamic In Vivo Fibered Confocal Microscopy*. PhD thesis, Ecole des Mines de Paris. 27
- [Vercauteren et al., 2007a] Vercauteren, T., Pennec, X., Malis, E., Perchant, A., and Ayache, N. (2007a). Insight into efficient image registration techniques and the demons algorithm. In *Proc. Information Processing in Medical Imaging (IPMI'07)*, volume 4584 of *Lecture Notes in Computer Science*, pages 495–506, Kerkrade, The Netherlands. Springer-Verlag. PMID: 17633724. 47
- [Vercauteren et al., 2007b] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2007b). Non-parametric diffeomorphic image registration with the demons algorithm. In Ayache, N., Ourselin, S., and Maeder, A. J., editors, *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'07)*, volume 4792 of *Lecture Notes in Computer Science*, pages 319–326, Brisbane, Australia. Springer-Verlag. 39, 45, 46
- [Vercauteren et al., 2008] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008). Symmetric log-domain diffeomorphic registration: A demons-based approach. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 5241 of *Lecture Notes in Computer Science*, pages 754–761. Springer-Verlag. 38, 39, 46, 47, 81
- [Vercauteren et al., 2009] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supp.1):S61–S72. PMID: 19041946. 27, 39, 42, 45, 47
- [von Siebenthal et al., 2007] von Siebenthal, M., Székely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, and Ph. (2007). 4d mr imaging of respiratory organ motion and its variability. *Phys. Med. Biol.*, 52:1547–1564. 103, 105
- [Wahba, 1990a] Wahba, G. (1990a). *Spline Models for Observational Data*. Society for Industrial Mathematics. 55
- [Wahba, 1990b] Wahba, G. (1990b). *Spline Models for Observational Data (C B M S - N S F Regional Conference Series in Applied Mathematics)*. Society for Industrial & Applied Mathematics. 95, 97
- [Wang et al., 2005] Wang, H., Dong, L., O’Daniel, J., Mohan, R., Garden, A. S., Ang, K. K., Kuban, D. A., Bonnen, M., Chang, J. Y., and Cheung, R. (2005). Validation of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy. *Physics in Medicine and Biology*, 50(12):2887. 47
- [Weickert, 1998] Weickert, J. (1998). *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, Stuttgart, Germany. 45

- [Weickert and Schnörr, 2001] Weickert, J. and Schnörr, C. (2001). A theoretical framework for convex regularizers in pde-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264. 45, 92, 97
- [Weiss and Freeman, 2001] Weiss, Y. and Freeman, W. (2001). On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Transactions On Information Theory*, 47(2):736–744. 125
- [Wells et al., 1996] Wells, W., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51. 63
- [West et al., 1997] West, J., Fitzpatrick, J., Wang, M., Dawant, B., Maurer Jr, C., Kessler, R., Maciunas, R., Barillot, C., Lemoine, D., Collignon, A., et al. (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554. 127
- [Wu et al., 2000] Wu, Y.-T., Kanade, T., Li, C.-C., and Cohn, J. (2000). Image registration using wavelet-based motion model. *International Journal of Computer Vision*, 38:129–152. 10.1023/A:1008101718719. 51
- [Yao and Taylor, 2003] Yao, J. and Taylor, R. (2003). Assessing accuracy factors in deformable 2d/3d medical image registration using a statistical pelvis model. In *Proc. International Conference on Computer Vision (ICCV)*, page 1329, Washington, DC, USA. IEEE Computer Society. 92
- [Yeo et al., 2010] Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., and Golland, P. (2010). Spherical demons: Fast diffeomorphic landmark-free surface registration. *IEEE Transactions on Medical Imaging (TMI)*, 29(3):650–668. 47
- [Yezzi and Soatto, 2003] Yezzi, A. and Soatto, S. (2003). Deformation: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2):153–167. 143
- [Yoshida, 1998] Yoshida, H. (1998). Removal of normal anatomic structures in radiographs using wavelet-based nonlinear variational method for image matching. In Laine, A. F., Unser, M. A., and Aldroubi, A., editors, *Wavelet Applications in Signal and Imaging Processing*, volume 3458, pages 174–181. SPIE. 51
- [Younes et al., 2009] Younes, L., Arrate, F., and Miller, M. I. (2009). Evolutions equations in computational anatomy. *NeuroImage*, 45(1, Supplement 1):S40 – S50. Mathematics in Brain Imaging. 39
- [Zhang, 1994] Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision (IJCV)*, 13:119–152. 99
- [Zheng, 2006] Zheng, G. (2006). A novel 3D/2D correspondence building method for anatomy-based registration. In *3rd Int'l. Workshop on Biomedical Image Registration*, pages 75–82. Springer. 92

REFERENCES

- [Zikic et al., 2010a] Zikic, D., Baust, M., Kamen, A., and Navab, N. (2010a). Generalization of deformable registration in riemannian sobolev spaces. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 25, 27, 49, 61, 67, 72, 83, 84
- [Zikic et al., 2011] Zikic, D., Baust, M., Kamen, A., and Navab, N. (2011). A general preconditioning scheme for difference measures in deformable registration. In *Proc. International Conference on Computer Vision (ICCV)*. 27, 43, 49, 61, 79
- [Zikic et al., 2008a] Zikic, D., Glocker, B., Hansen, M. S., Khamene, A., and Navab, N. (2008a). Construction of Statistical Shape Models from Minimal Deformations. In *Workshop Manifolds in Medical Imaging: Metrics, Learning and Beyond in conjunction with Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 141
- [Zikic et al., 2010b] Zikic, D., Glocker, B., Kutter, O., Groher, M., Komodakis, N., Kamen, A., Paragios, N., and Navab, N. (2010b). Linear Intensity-based Image Registration by Markov Random Fields and Discrete Optimization. *Medical Image Analysis (MedIA)*, 14(4):550–562. 115
- [Zikic et al., 2010c] Zikic, D., Glocker, B., Kutter, O., Groher, M., Komodakis, N., Khamene, A., Paragios, N., and Navab, N. (2010c). Markov Random Field Optimization for Intensity-based 2D-3D Registration. In *Proc. SPIE Medical Imaging*. 115
- [Zikic et al., 2008b] Zikic, D., Groher, M., Khamene, A., and Navab, N. (2008b). Deformable registration of 3d vessel structures to a single projection image. In *Proc. SPIE Medical Imaging*. 89
- [Zikic et al., 2008c] Zikic, D., Hansen, M. S., Glocker, B., Khamene, A., Larsen, R., and Navab, N. (2008c). Computing Minimal Deformations: Application to Construction of Statistical Shape Models. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 141
- [Zikic et al., 2010d] Zikic, D., Kamen, A., and Navab, N. (2010d). Natural gradients for deformable registration. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*. 61, 67
- [Zikic et al., 2010e] Zikic, D., Kamen, A., and Navab, N. (2010e). Revisiting horn and schunck: Interpretation as gauß-newton optimisation. In *Proc. British Machine Vision Conference (BMVC)*. 27, 31, 61, 67
- [Zikic et al., 2010f] Zikic, D., Kamen, A., and Navab, N. (2010f). Unifying characterization of deformable registration methods based on the inherent parameterization. In *Proc. International Workshop on Biomedical Image Registration (WBIR)*. 25, 56, 58
- [Zikic et al., 2006a] Zikic, D., Khamene, A., and Navab, N. (2006a). Intensity-unbiased force computation for variational motion estimation. In *Workshop on Image Registration in Deformable Environments, in conjunction with BMVC*, pages 61–70. 61

REFERENCES

- [Zikic et al., 2006b] Zikic, D., Wein, W., Khamene, A., Clevert, D.-A., and Navab, N. (2006b). Fast deformable registration of 3d-ultrasound using a variational approach. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4190, pages 915–923. [66](#), [86](#)
- [Zitova and Flusser, 2003] Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000. [17](#), [116](#)