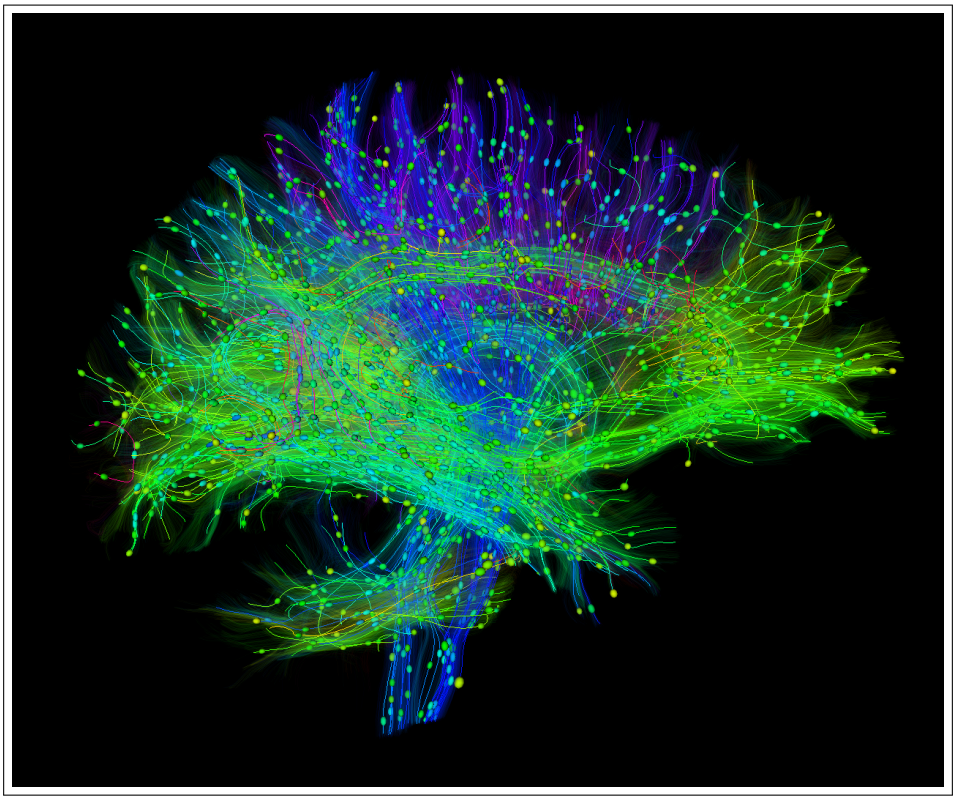
UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

PhD Dissertation

Density-based Similarity in the Registration of Diffusion-Weighted Images



By
Henrik Grønholt Jensen

Academic advisors
Sune Darkner & Mads Nielsen

University of Copenhagen
April 2018

This dissertation is submitted for the degree of Doctor of Philosophy at the University of Copenhagen (UCPH).

Faculty:	Science
Department:	Computer Science
Author:	Henrik Grønholt Jensen
Title:	Density-based Similarity in the Registration of Diffusion-Weighted Images
Academic advisers:	Prof. Mads Nielsen, Assoc. Prof. Sune Darkner
Submitted:	April 6, 2018

Funding for this research has been provided under an open grant from the Department of Computer Science, UCPH.

ISBN XXX-XX-XXXX-XXX-X

To my love,
Katrine Hommelhoff Jensen,
and our unborn daughter.

May she be as beautiful
as her mother,
and waaaay less stubborn
than her father.

I am a brain, Watson. The rest of me is mere appendix. Therefore, it is the brain I must consider.

—Arthur Conan Doyle, The Case-Book of Sherlock Holmes

Abstract

Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI or DWI) is a non-invasive scanning protocol aimed at inferring the structure of biological tissue by tracking the movement of water molecules. As molecules diffuse along and around obstacles, in-vivo images of the diffusion can be used to reconstruct the minuscule anatomy that would otherwise be invisible in standard MRI. The applications of DWI ranges from tumor detection to tracing the neuronal pathways connecting the brain.

DWI is also a complex modality and difficult to both validate and compare. The data is directional and exhibits a non-linear behaviour for high-resolution images. It requires longer scanning times and high magnetic gradients, resulting in an increased amount of noise from motion and external factors. DWI also has no gold standard datasets for comparable quantitative validation. Group studies often present results through private segmentations from trained experts or by qualitative visual evaluation. This is a significant problem as DWI is becoming an issue of Big Data due to increasing amounts of open and freely available datasets. As such, our first contribution is a critical review of image registration and validation of group-wise alignment of DWI. We investigate common approaches to compare DWI data in terms of voxel- and connectivity-based methods.

Image registration is the process of spatially aligning images in a way that allow us to define a shared coordinate system between them. For DWI, the reorientation of the directional information presents a difficult challenge. In many cases, DWI is simply registered using standard 3D algorithms and without considering their non-linear relationship. Our second contribution is a density-based scale-space formulation for DWI that gives access to information-theoretic similarity measures, based on the full diffusion profile. The presented framework is a global registration method that optimizes the mutual information between DWI with explicit reorientation of the gradient vectors. We show that the directional scale is important for aligning DWI.

Nonrigid image registration allows for local warping of images and is an essential part of spatial alignments. While few studies compare the performance of nonrigid methods used for DWI, it is clear that the gold standards of nonrigid registration algorithms are those designed to include the angular information as a direct part of the registration. Our third and final contribution is a nonrigid extension of the framework for density-based DWI registration. We present the full analytical solution and demonstrate it on simulated and synthetically warped DWI, finding empirical evidence of a preservation of the underlying DWI structure during registration. As designing such algorithms is non-trivial and a significant computational challenge in terms of time, memory and numerical precision, we also append a section on how the framework was implemented to make it computationally feasible.

Resumé

Diffusionsvægtet magnetisk resonansbilleddannelse (DW-MR eller DWI) er en ikke-invasiv scanningsprotokol, der sigter mod at udlede strukturen af biologisk væv ved at spore bevægelsen af vandmolekyler. Da molekyler diffunderer langs og omkring forhindringer, kan in-vivo billeder af diffusionen bruges til at rekonstruere den mikroskopiske anatomi, der ellers ville være usynlig i standard MR. Anvendelserne af DWI spænder fra tumordetektion til sporing af de neuronale veje, som forbinder hjernen.

DWI er også en kompleks modalitet og vanskelig at både validere og sammenligne. Data er retningsbestemt og udviser en ikke-lineær adfærd for billeder i høj opløsning. Det kræver længere scanningstider og høje magnetiske gradienter, hvilket resulterer i en øget mængde støj fra bevægelse og eksterne faktorer. DWI har heller intet guldstandard datasæt til sammenlignelig kvantitativ validering. Gruppestudier præsenterer ofte resultater gennem private segmenteringer fra uddannede eksperter eller ved kvalitativ visuel evaluering. Dette er et væsentligt problem, da DWI bliver et Big Data problem på grund af stigende mængder åbne og frit tilgængelige datasæt. Derfor er vores første bidrag en kritisk gennemgang af billedregistrering og validering af gruppevis sammenligning af DWI. Vi undersøger fælles fremgangsmåder til at sammenligne DWI-data med hensyn til voxel- og konnektivitetsbaserede metoder.

Billedregistrering går ud på at transformere billeder på en måde, der giver os mulighed for at definere et fælles koordinatsystem mellem billederne. For DWI udgør reorienteringen af retningsvektorerne en vanskelig udfordring. I mange tilfælde registreres DWI simpelthen ved hjælp af standard 3D-algoritmer og uden at overveje deres ikke-lineære forhold. Vores andet bidrag er en tæthedsbaseret skalarumsformulering til DWI, der giver adgang til informationsteoretiske similaritetsmål baseret på den fulde diffusionsprofil. Det præsenterede program er en global registreringsmetode, der optimerer den gensidige information mellem DWI med eksplicit omorientering af gra-

dientvektorer. Vi viser, at retningskalaen er vigtig for at sammenligne DWI.

Ikke-rigid billedregistrering muliggør lokal transformation af billeder og er en væsentlig del af rumlige tilpasninger. Selv om få undersøgelser sammenligner præstationer af ikke-rigide metoder, der anvendes til DWI, er det klart, at guldstandarderne for ikke-rigid registreringsalgoritmer er dem, der er designet til at inkludere retningsinformationen som en direkte del af registreringen. Vores tredje og sidste bidrag er en ikke-rigid udvidelse af programmet for tæthedsbaseret DWI-registrering. Vi præsenterer den fulde analytiske løsning og demonstrerer den på simuleret og kunstigt deformeret DWI, hvor der findes empiriske beviser for bevarelse af den underliggende DWI struktur under registrering. Da det at designe sådanne algoritmer ikke er trivielt, og da der er betydelige beregningsmæssige udfordringer med hensyn til tid, hukommelse og numerisk præcision, tilføjer vi også et afsnit om, hvordan programmet blev implementeret for at gøre det beregningsmæssigt realistisk.

Contents

List of Figures

Abbreviations

ADC	Apparent Diffusion Coefficient
CR	Correlation Ration
CSF	Cerebrospinal Fluid
dODF	Diffusion Orientation Density Function
DSI	Diffusion Spectrum Imaging
DTI	Diffusion Tensor Imaging
DWI	Diffusion-Weighted (magnetic resonance) Imaging
FA	Fractional Anisotropy
fODF	Fiber Orientation Density Function
FRT	Funk-Radon Transform
GFA	Generalized Fractional Anisotropy
HARDI	High-Angular Resolution Diffusion Imaging
HCP	Human Connectome Project
LIR	Leibniz Integration Rule
LOI	Locally Orderless Images
LOR	Locally Orderless Registration
MD	Mean Diffusivity
MI	Mutual Information
MRI	Magnetic Resonance Imaging

NMI	Normalized Mutual Information
ODF	Orientation Density Function
PV	Partial Volume (effect)
PW	Parzen-Window
QBI	q-Ball Imaging
QSI	q-Space Imaging
ROI	Region Of Interest

Preface

The human brain is a fascinating and incredibly complex structure. The very idea of being able to study how the living brain is connected is tantalizing, as it will one day let us know how thoughts are formed, decisions made, and memories learned. On the visible horizon, it can help us treat more diseases, heal more injuries and improve life quality. My journey towards a PhD project on the comparison of medical images started during my graduate studies, where medical image analysis offered a chance to combine my passion for biology and data analysis.

During the years as a PhD fellow, a lot has happened. I have had the opportunity to work alongside both inspiring people, people who gave guidance and advice, those who lend an ear to my frustration, and those who were simply there.

First and foremost, this PhD would never have happened without Sune Darkner, who I thank for believing in my abilities, and for giving me the time I needed when my life got upended. I would also like to thank Mads Nielsen for supplying academic advice and ideas to the project.

I am truly honored to have had the opportunity to visit Carl-Frederik Westin (CF), Lauren O'Donnell and their research group as a visiting scholar at Harvard Medical School. Their advice and helpful comments were a great help and motivation.

I am grateful to DIKU for awarding me this scholarship and allowing me to cultivate an excellent office environment. Francois Lauze for being a foremost adviser on mathematical issues and for helping in formalizing our papers. Akshay Pai for helping with code and ideas. Kim Steenstrup for feedback and helping me see the PhD to its end. Jens, Anton, Mahdieh, and the others whose company I have enjoyed as we shared office. And the many awesome PhD students and permanent staff at the Image Group.

Finally, I want to particularly send thoughts to my close friends Truls, Kasper, and Jón for their support and friendship. To my family for their endless belief in my abilities and pulling me out of holes. And Katrine, who have cared for me, trusted in me and made me laugh when it seemed hard to stay afloat.

I have come to know basic research as a struggle between the bursts of optimism and the setbacks and dead ends that often follow. Like countless before me, I have learned that research is not just passion and long hours, but a test of endurance and how far you are willing to push yourself past the point where sane minds might stop. Some have it easy, some have it fun, most do not. Yet in the struggle, there happens a unique sculpting of self-discipline as you battle your way through all the frustration, stress, and disillusions. And in the end, it is a funny thing, that in the years seeking knowledge, you learn to appreciate your ignorance of the path ahead.

I hope you enjoy your reading.

Henrik G. Jensen,

April, 2018.

1 Introduction

More than half a century ago, the medical imaging community started on the challenging journey towards the ability to study the in-vivo human brain in a non-invasive and non-harmful way, by using magnetic resonance imaging (MRI) [Carr and Purcell, 1954]. Soon after, with the emergence of diffusion-weighted MRI (DW-MRI, or simply DWI) [Stejskal and Tanner, 1965], researchers began to probe the microstructural tissue through the displacement of water molecules, and in 1986 it was introduced in clinical practice as a diagnostic tool for neurological disorders [Le Bihan et al., 1986]. With the introduction of diffusion tensor imaging (DTI) [Basser et al., 1994], computational methods for DWI became a highly popular area of research and grew to become one of the great challenges of the 21st century. Today, we can with increasing certainty visualize the neurological pathways connecting the brain - even as multiple fibers cross each other in a single voxel of a 3-dimensional image. However, between the inter-variability of brains and the low spatial resolution of DWI, the field has never settled on a gold standard or ground truth dataset for automatic validation. New methods, that aim to model the diffusion and use it to reconstruct the underlying anatomy, often rely on trained experts to manually segment the data for a quantifiable result. And yet, the need for automated evaluation has steadily been growing. Large open datasets are emerging, containing thousands of DWI scans, multiple image modalities, and biometric data. Increasing computational efficiency and high-resolution acquisitions are resulting in incredibly detailed DWI reconstructions, making it harder for trained experts to discern and delineate subtle differences. To properly validate new findings in DWI, the results should be both reproducible across multiple subjects and verifiable by other state-of-the-art reconstruction methods. Nevertheless, group-wise analysis remains one of the great challenges in DWI [Mangin et al., 2016].

1.1 Motivation

The work and motivation of this dissertation revolve around automated group-wise alignment of DWI data and obtaining a better understanding of the state and future of DWI analysis. One of the major problems in comparing DWI scans is that they are relatively noisy and difficult to normalize between subjects and scanners. Combined with the increasing interest in multimodal imaging, aligning scans based on information-theoretic similarity measures is an obvious choice, as these have convenient invariance properties and can model unknown statistical relationships. However, these measures are coupled to scalar images, and we believe that there is be much to be gained by extending them to include the orientational information of DWI.

1.2 Outline

This dissertation is divided into six chapters. Excluding the first, Chapter 2 introduces the core topics by providing a brief layman's introduction to image registration and diffusion-weighted MRI. The basic steps in a typical registration framework are covered, and the DWI terminology relevant for this dissertation is explained along with some of the different types of data acquisitions and computational diffusion models. In Chapter 3, the state of group-wise alignment, registration, and validation of DWI is extensively reviewed. Approaches to voxel-based registration of popular DWI acquisitions are presented with a focus on how registration methods are designed for the directional DWI data. An equal focus is on how tractography is compared and used in group studies that focus on the alignment. Validation is discussed throughout the review. Chapter 4 presents a new scale-space formulation for information-theoretic similarity measures in DWI registration. Global affine registration with Mutual Information is used to examine the properties of the scale-space induced by the intensity and spatio-directional kernels. In Chapter 5, the scale-space formulation is extended to a nonrigid registration framework. This density-based formulation is applied to DWI data in an approach that explicitly accounts for the reorientation of the full diffusion profile, while also defining the statistical relationship between DWI through Normalized Mutual Information. Finally, Chapter 6 summarizes the work performed and reflect upon current and future challenges.

1.3 Contribution

There are three primary contributions in this dissertation: (i) A review of DWI, (ii) an information-theoretic similarity measures for DWI, and (iii) a full nonrigid registration framework for DWI.

I A comprehensive review of registration and validation for group-wise alignment in DWI of the human brain

The first and most recent contribution is an extensive review of image registration methods and group-wise alignment of the human brain in DWI. The manuscript related to this contribution gives a high-level critical overview of recent approaches to the challenging task of group-wise analysis and validation of DWI. We cover both voxel-based and fiber tract-based registration in order to form a wholesome impression of the current state of DWI at a group level, and the challenges and opportunities the community face with a growing amount of publicly available high-quality data. This contribution can be found in Chapter 3.

Related journal manuscript:

Critical Issues in the Registration and Validation of Group-wise Alignment of Diffusion-Weighted Imaging. Manuscript accepted for submission to Human Brain Mapping (HBM), 2018.

II A scale-space formulation for DWI that introduces information-theoretic similarity measure to group-wise DWI registration

The second contribution is a scale-space formulation for DWI that offers explicit control of the orientation, spatial, intensity and integration scale. This new framework extends the Locally Orderless Registration formulation by [Darkner and Sporring, 2013] to DWI, and it is likely the first definition of a DWI-based cost function that optimizes explicitly over both the spatial and directional domain based on a non-linear information-theoretic similarity measure. Such measures have been shown to be robust to the extensive noise and artifacts present in the sensitive DWI scans. In the corresponding paper, we illustrate the application of the density estimate by affine image registration of DWI using Mutual Information. This contribution can be found in Chapter 4.

Related conference paper:

Locally Orderless Registration for Diffusion Weighted Images. Conference paper. Accepted at Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.

III A nonrigid density-based registration framework and scale-space formulation for DWI with explicit reorientation

The third contribution is a nonrigid voxel-based registration framework based on the scale-space formulation for DWI. It is the culmination of our aim to create a density-based framework that allows for nonrigid registration of two DWI over their full diffusion profiles, while at the same time optimizing a similarity measure that reflects the non-linear statistical relationship in DWI data. The contribution lies mainly in the analytical formulation but also in the implementation itself, given the computational complexity of such a method. While demonstrating the framework on simulated examples and synthetically warped real data, we provide empirical evidence of how the underlying structure of the DWI data is preserved during registration. This contribution can be found in Chapter 5.

Related conference abstract:

Density-Based Nonrigid Registration of Diffusion-Weighted Images. Conference abstract. Accepted at International Society for Magnetic Resonance in Medicine (ISMRM), 2017.

Related journal manuscript:

Information-Theoretic Registration with Explicit Reorientation of Diffusion-Weighted Images. Manuscript intended for submission to Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018.

2 A Preamble on Image Registration and Diffusion-Weighted Imaging

The purpose of this chapter is to give a brief layman's introduction to image registration and to diffusion-weighted MRI images. It can be skipped if one is already familiar with these concepts. However, it also serves the purpose of reducing ambiguity by specifying the terminology and wording used in the rest of the dissertation.

2.1 Image Registration

Image registration refers to a set of transformations that aligns two or more images into a shared coordinate system. This process is also referred to as spatial normalization or as data harmonization.

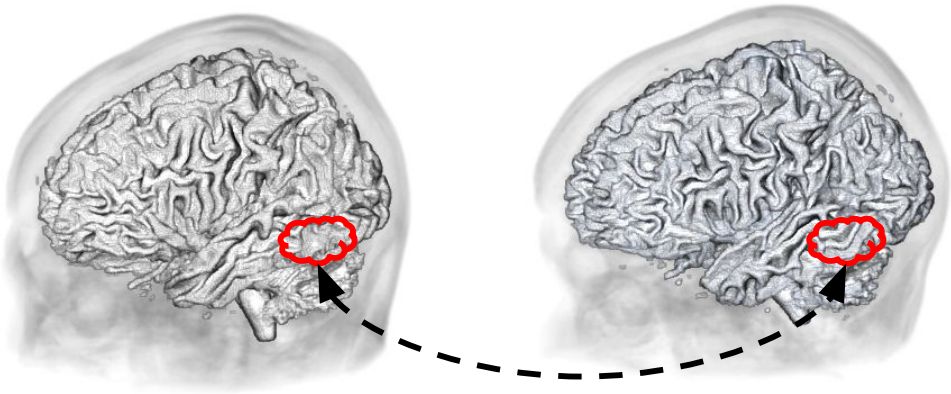


Figure 2.1: It can be difficult to identify similar structures across complex images, which is where image registration comes in. This illustration was created from the white matter segmentations of two HCP subjects.

Image registration is an ill-posed problem [Sotiras et al., 2013]. In medical image analysis, registration is often associated with aligning existing anatom-

ical labels (an atlas) to the image of a patient for navigation and automatic labeling. It can also be used to create a population average (a template), which can help differentiate between groups of patients, e.g. sick and healthy. Alternatively, registration can show how much a tumor has evolved by aligning images of the same patient taken at different points in time [Brown, 1992]. In such a case, the focus would be on how much the images have to be transformed to be aligned. Knowing the difference or similarity between images can also help in diagnosis by looking at the temporal changes within a patient, or by comparing the patient with other healthy or sick populations [Maes et al., 1997]. While image registration is a well-researched field for 2D and 3D images, it is at an early stage when it comes diffusion-weighted MRI as it is both spatial and directional.

We will refer to the image being transformed as the *moving* image $\boldsymbol{I}$ and the stationary image as the *target* image $\boldsymbol{J}$. Our goal is to align these images under a transformation Φ given a regularity condition $\mathcal{S}(\Phi)$ and a similarity $\mathcal{F}(\boldsymbol{I} \circ \Phi, \boldsymbol{J})$, such that our cost function $\mathcal{M}(\boldsymbol{I}, \boldsymbol{J}, \Phi)$ is minimized:

$$\mathcal{M}(\boldsymbol{I}, \boldsymbol{J}, \Phi) = \mathcal{F}(\boldsymbol{I} \circ \Phi, \boldsymbol{J}) + \mathcal{S}(\Phi) \tag{2.1}$$

The circle-operator should be read as the function composition $\Phi(\boldsymbol{I})$ or "Φ applied to $\boldsymbol{I}$". While the focus of this dissertation is on the similarity measure $\mathcal{F}$, this formulation is the foundation of our work, and we briefly cover the individual parts of eq. (2.1).

The Transformation

In image registration, a transformation model Φ can be divided into a global and a local transformation. The global transformation will be a linear model that is either rigid or affine. The local or nonrigid model is often non-linear and defined as a set of local affine transformations. See Figure 2.2 for a quick overview.

For a global transformation model applied to a 3D image, we add 3 parameters or degrees of freedom for each type of transformation, which means that the rigid transformation has 6 and the affine 12 degrees of freedom. A nonrigid registration framework is a combination of global and local registration, and we write the transform $\phi(\boldsymbol{x})$ of a vector or point $\boldsymbol{x}$ as

$$\phi(\boldsymbol{x}) = \phi(\boldsymbol{x})_{\text{global}} \circ \phi(\boldsymbol{x})_{\text{local}}. \tag{2.2}$$

The transformation gives us a set of new coordinates, which we use to interpolate the data in the target image, evaluate the similarity, and update the

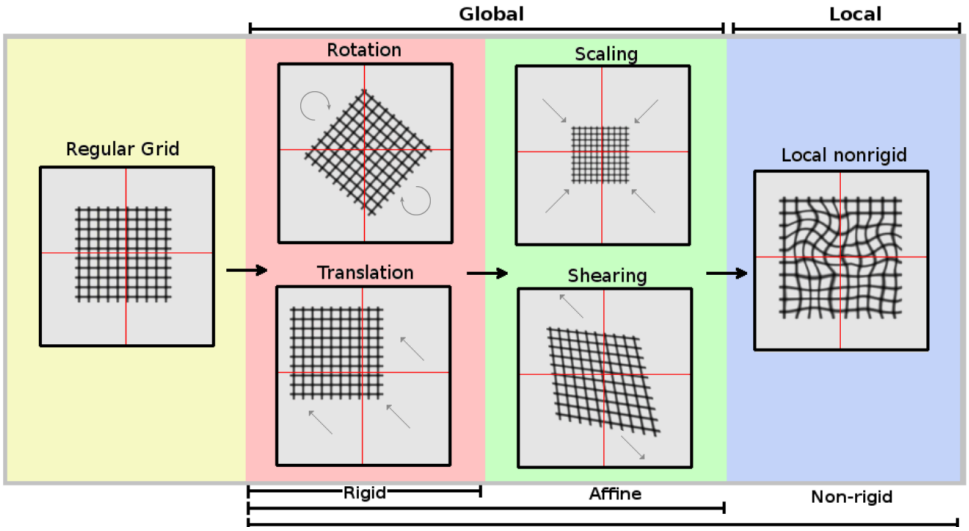


Figure 2.2: Different degrees of transformation. Increasing in complexity from left to right, each figure illustrates a deformation applied to a regular grid, which in turn represents the coordinate system of an image at the initial scale (e.g. an intensity or pixel value at every square). From [Jensen, 2014].

transformation in an iterative optimization process. A successful result is a spatial map, that for any point $\boldsymbol{x}$ in $\boldsymbol{I}$ returns the corresponding point in $\boldsymbol{J}$.

In Chapter 4, we present the scale-space model and information-theoretic similarity measure based on a global affine transformation. In Chapter 5, it is extended to a nonrigid local transformation.

The Regularization

The regularity condition $\mathcal{S}$ is a bias added to the deformation to prevent overfitting. For image registration, it regularizes the flexibility of the registration by assigning a high cost to undesirable transformations. Too heavy a regularization means that the image cannot be deformed enough for a sufficient alignment, while too little regularization likely means that the transformation will create unrealistic overfitted solutions or even break apart. Regularization is thus a trade-off between generalization and over-fitting. To give an example, the microscopic structures in brain scans requires a detailed alignment as illustrated in Figure 2.3. This entails that we minimize the amount of regularization needed to get a flexible but still robust model. However, with such complex images, we run the risk of matching structures that might exist in one scan but not in the other. There is no easy way to

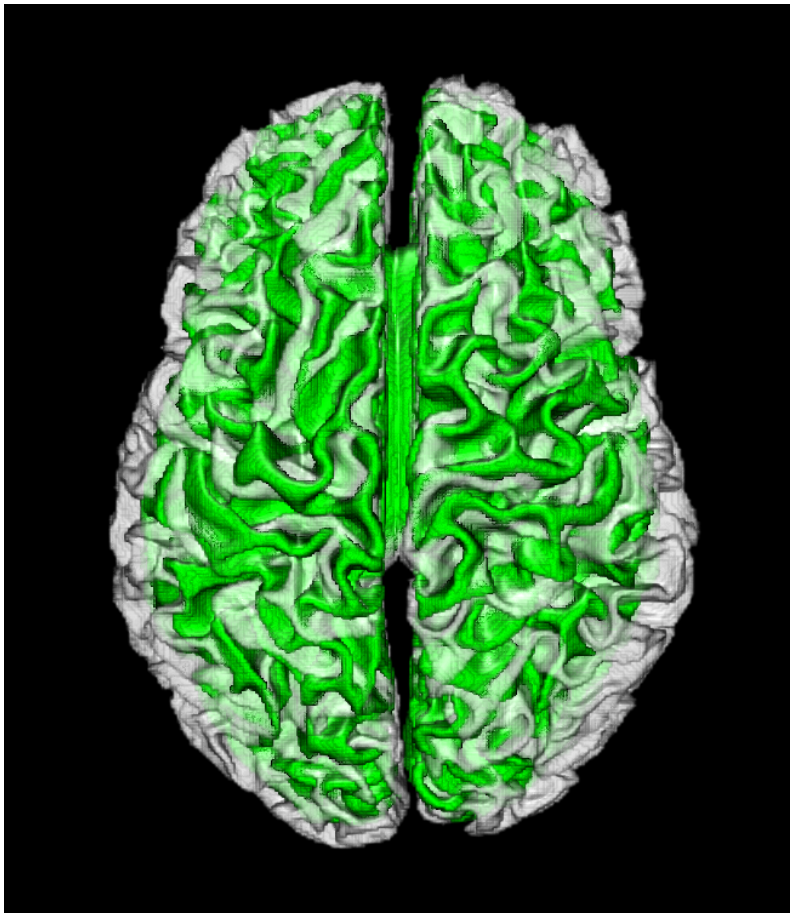


Figure 2.3: The white matter segmentation of two rigidly aligned HCP subjects (translation+rotation), one in green and the other in grey. While brains are overall similar, aligning all the folds of the brain is not only extremely difficult but also likely incorrect. Brains are not diffeomorphic.

define the amount of regularization needed without prior information about the problem, and it is often left to the user to determine the flexibility. Yet, if regularization prevents overfitting then more information in the image should reduce the need for regularization [Hawkins, 2004].

It is our expectation that the voxels in DWI scans, that are rich in angular information, will create a more robust but flexible deformation for the nonrigid model described in Chapter 5.

The Similarity

The similarity measure is defined by a single value, computed from an interpolation or some other pair-wise distance function associated with $\mathcal{F}$. The

choice of similarity measure depends on the type of data. If we compare images of the same modality, it is common to evaluate the pointwise differences in intensities. However, if the goal is to compare different modalities of the same image (e.g. CT vs MRI), it is better to define the similarity in terms of the correlation and co-variance of the voxel intensities, which are more robust to variations in illumination [Maes et al., 1997, Viola and Wells III, 1997]. The definition of similarity is critical to what the final optimal alignment will be.

In this dissertation, we use intensity-invariant similarity measures, previously reserved for multi-modal scalar images, under the hypothesis that inter-subject DWI scans are both noisy and vary in illumination in a complicated or unknown way [Van Hecke et al., 2007]. Such measures have already been used with success in intra-subject (scalar) DWI registration for both susceptibility-induced distortion correction [Bhushan et al., 2012], eddy-current and body motion correction [Rohde et al., 2004], and more.

Schematic Overview

Registration is an optimization, where the cost function is minimized, based on the similarity, given a regularized transformation. But there are other steps involved. Figure 2.4 provides a more detailed overview of a typical pipeline.

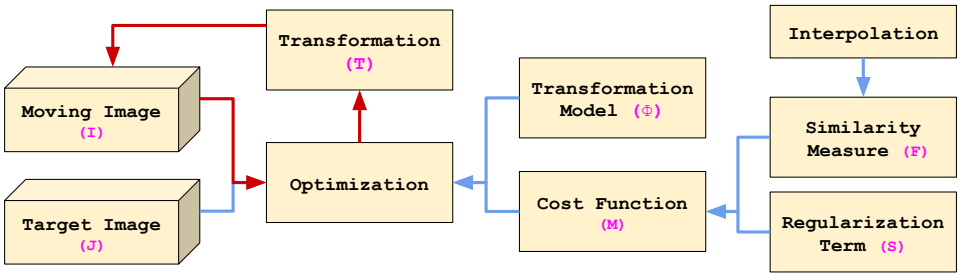


Figure 2.4: A standard pipeline for a registration framework. The iterative process of optimization and transformation is illustrated by the circle of red arrows. The pink highlighted symbols refers to the content of the previous paragraphs.

2.2 Diffusion-Weighted Magnetic Resonance Imaging

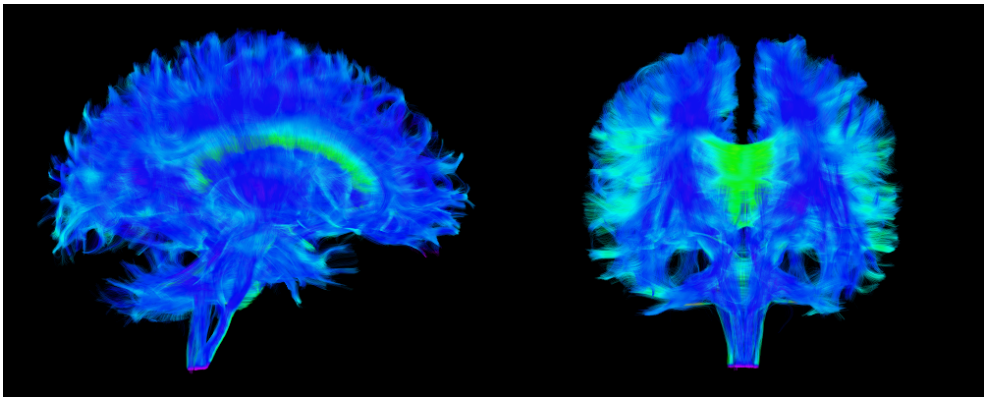
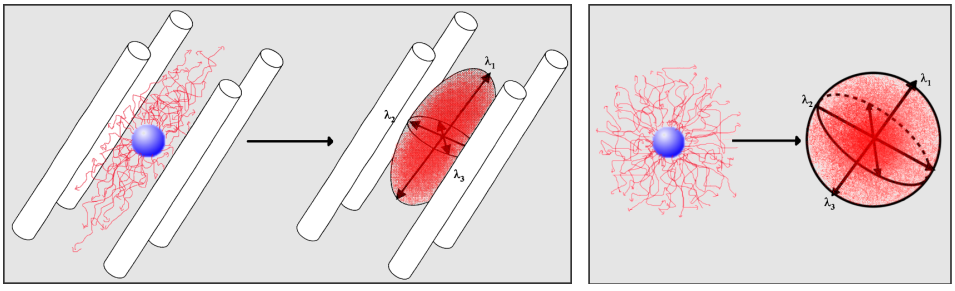


Figure 2.5: Illustration of a DTI-based tractography which we created with the diffusion framework `3D Slicer` [Norton et al., 2017] over HCP subject `103818` at $b = 1000$ [Van Essen et al., 2013].

Diffusion-weighted imaging (DWI) is a technique used to non-invasively trace the movement of water molecules inside and outside cells in the body in order to infer microstructural anatomy. One of the most popular applications of DWI is imaging the white matter in brain scans [O'Donnell and Westin, 2011], where it can be used to trace the neuronal fibers connecting the brain. The hypothesis is that water molecules are more prone to diffuse along rather than across the axons and fibers connecting different areas of the brain, as illustrated Figure 2.6.



(a) **(b)**

Figure 2.6: Illustration of anisotropic and isotropic diffusion. To the left in **(a)**, we show a water molecule (blue) and its potential motion (red lines) surrounded by fibrous tissue (e.g. white matter axons). This results in the anisotropic model shown to the right in **(a)**. **(b)** shows the unhindered molecule and its potential isotropic motion. From [Jensen, 2014].

The anisotropic information in DWI lets researchers trace the white matter fiber pathways by following the hindered directional motion of the diffusion from one voxel to the next. Drawing lines along probable fiber pathways is

called *tractography* and a large portion of DWI studies are concerned with comparing brains on the structurally connected scale. An example of such tractographies can be seen in Figure 2.5. It is important to bear in mind that these *streamlines* are only qualified guesses at actual underlying anatomical fibers and that most likely outline non-existing fibers. However, they serve the useful purpose of outlining the white matter connecting the brain.

To show how we can go from Figure 2.6 to Figure 2.5, we go through the basic and most popular elements of DWI in the following order:

1. The *Apparent Diffusion Coefficient* (ADC) which is the measure of the magnitude of diffusion on average or in a given direction.
2. The *Diffusion Tensor Imaging* (DTI) is the simplest model used to describe the diffusion in a voxel, and remains the most popular tool for DWI in clinical practice.
3. Visualizations of tensor models and tractographies to further illustrate the 4D structure of DWI.
4. Popular quantitative DTI measures such as *Fractional Anisotropy* (FA) and *Mean Diffusivity* (MD).
5. The *Orientation Density Function* (ODF) and complex fiber structures.
6. *HARDI*, *q-Ball Imaging* (QBI), and the *Funk-Radon Transform* (FRT), which can resolve fiber crossings.
7. DTI versus QBI, where we look at a pros and cons of each method.

The following paragraphs briefly covers the above topics.

The Apparent Diffusion Coefficient (ADC)

In structural MRI, the average diffusivity in a voxel is measured and converted to a scalar intensity value at each voxel, which gives an image of different types of tissue. In MRI, it is only the magnitude of the mean diffusion that is relevant (bone has less diffusivity than fat, fat less than spinal fluid, and so on). DWI is MRI with directional information, where we use the directional movement of water molecules to probe the tissue structure at a micrometer scale well beyond the usual millimeter MRI resolution. When we measure diffusion in a specific direction, we measure the diffusion coefficient which describes the ensemble average diffusion or mean-squared displacement on a macroscopic level, caused by erratic random movement of particles on a microscopic level, also referred to as Brownian motion [Einstein, 1905].

This averaged diffusion is not free within a voxel of biological tissue, and we refer to the diffusion measured in any direction as the *Apparent Diffusion Coefficient* or ADC. It is important to note that the ADC is used in two different contexts: One defines it as the observed mean diffusion rate in a given direction, while the other defines it as the overall mean diffusivity (MD). We use the first definition which is less ambiguous, and instead we define the MD as the average ADC. By letting the ADC be the set of diffusivity measurements, we need a model that describes the shape or profile of the diffusion in a voxel.

Diffusion Tensor Imaging (DTI)

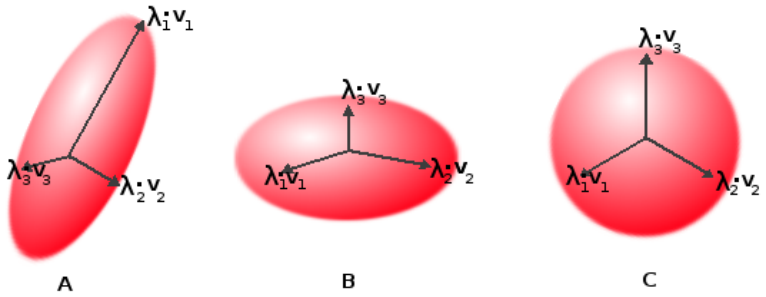


Figure 2.7: Illustration of the three types of diffusion in a single DTI model: (A) $\lambda_1 >> \lambda_2 \approx \lambda_3$, (B) $\lambda_1 \approx \lambda_2 >> \lambda_3$, (C) $\lambda_1 \approx \lambda_2 \approx \lambda_3$. From [Jensen, 2014].

In 1965 Stejskal and Tanner presented a formulation to calculate the diffusion coefficient based on their scanner sequence [Stejskal and Tanner, 1965], which was simplified by [Le Bihan et al., 1986] who introduced the b-value, and finally it was extended to a 2nd order tensor model by [Basser et al., 1994] to describe anisotropic diffusion. The following equation is the foundation of Diffusion Tensor Imaging (DTI), which remains the most clinically used DWI method [Zucchelli et al., 2017]. It is defined as

$$S = S_0 \, e^{-b \, \boldsymbol{g}^T \boldsymbol{D} \boldsymbol{g}} \tag{2.3}$$

where S_0 is the signal without any diffusion gradients, the b-value is all the acquisition-dependent gradient terms gathered in a single scalar parameter by Bihan et al. and assumes a Gaussian distribution, $\boldsymbol{D}$ is a rank-2 tensor (symmetric positive-definite 3x3 matrix), and $\boldsymbol{g}$ are the directions or unit vectors where a minimum of six directions are measured to get a diffusion profile. Relating to the previous paragraph, the tensor $\boldsymbol{D}$ represents the ADCs. Equation (2.3) is based on the fact that Brownian motion of particles can be approximated in any given direction by a zero-mean Gaussian distribution.

This is an idealized model and does not fully reflect the constrained diffusion in anatomy which, unlike a Gaussian model, is not mono-exponential over time. However, this is mostly an issue for high b-values. DTI has clinically attractive quick acquisition times and it is sufficient in many scenarios such as diagnosis, surgical aid, white matter atrophy, etc. Examples of the DTI model can be seen above in Figure 2.7.

The b-value describes the signal sensitivity to diffusion, often given in mm^2/s, and relates to the strength and time of the diffusion gradients. It is a compiled value that makes it easier for non-physicists to work with DWI and hides scanner-specific values. In short, high b-values are more sensitive to diffusion but also have high signal attenuation, meaning the signal disappears with time. Low b-values have better Signal-to-Noise Ratio (SNR) but less angular contrast or directional information. b-values above 1500 mm^2/s exhibit bi-exponential behavior and cannot be modeled with the Gaussian DTI assumption. The bi-exponential behavior is still debatable and is generally attributed to biological constraints on the diffusion [Yablonskiy and Sukstanskii, 2010]. Figure 2.8 illustrates the four different b-values associated with the popular HCP dataset [Van Essen et al., 2013].

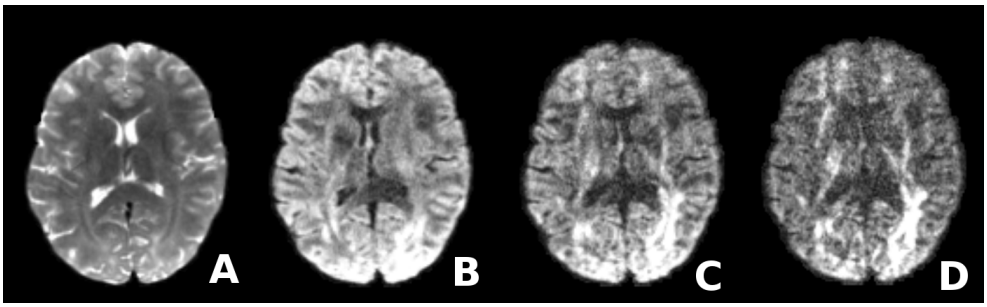


Figure 2.8: Four DWI of the same subject with varying b-values at approximately the same gradient angle. (A) $b = 0\ mm^2/s$, (B) $b = 1000\ mm^2/s$, (C) $b = 2000\ mm^2/s$, and (D) $b = 3000\ mm^2/s$. From [Jensen, 2014].

In Equation (2.3), there are two images with b-values: S_0 with no diffusion gradient ($b < 5\ \text{mm}^2/\text{s}$), and S with a diffusion gradient at around $b = 1000\text{mm}^2/\text{s}$ for DTI. S_0 is necessary to avoid the underlying structural T_2 image to shine through due to signal attenuation and restricted diffusion. When we have (at least) these two images, we can calculate the ADC in every voxel.

We generally refer to the S_0 image as the b_0 image. We will also be working with b-values at 3000 mm^2/s for a better probing of the tissue boundaries.

This is where more advanced diffusion models comes in, such as HARDI and QBI described further below.

Visualization of DTI, tractography and the 4D nature of DWI

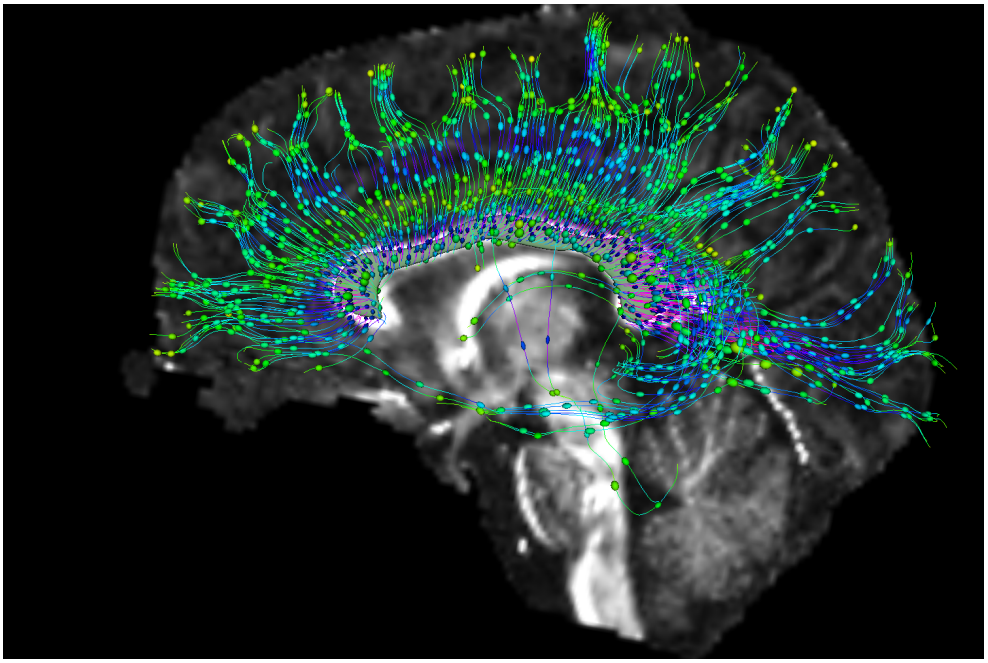


Figure 2.9: DTI-based tractography seeded in the corpus callosum, superimposed on a slice of the b_0 image. The illustration was generated from the same subject as in Figure 2.5, also using `Slicer`.

We have found that structure of DWI data can sometimes be a bit tricky, and so we would like to briefly reiterate on the 4D nature of DWI data. First of all, with the DTI model describing the diffusion profile at each voxel, we can already now start to trace streamlines along the anisotropic diffusion.

In Figure 2.9, we have shown an example of a tractography seeded in the corpus callosum (brain bridge), where potential fiber pathways are being traced by following the dominant direction of the diffusion profiles that are shown as anisotropic spheres. A similar illustration is shown in Figure 2.10 to further illustrate this.

There are two common ways to see DWI data. In the first, a DWI scan consists of multiple scalar 3D images, each corresponding to a measure taken for a specific diffusion gradient direction. It can be thought of as multiple pictures of a semi-transparent object that is illuminated from different angles in each

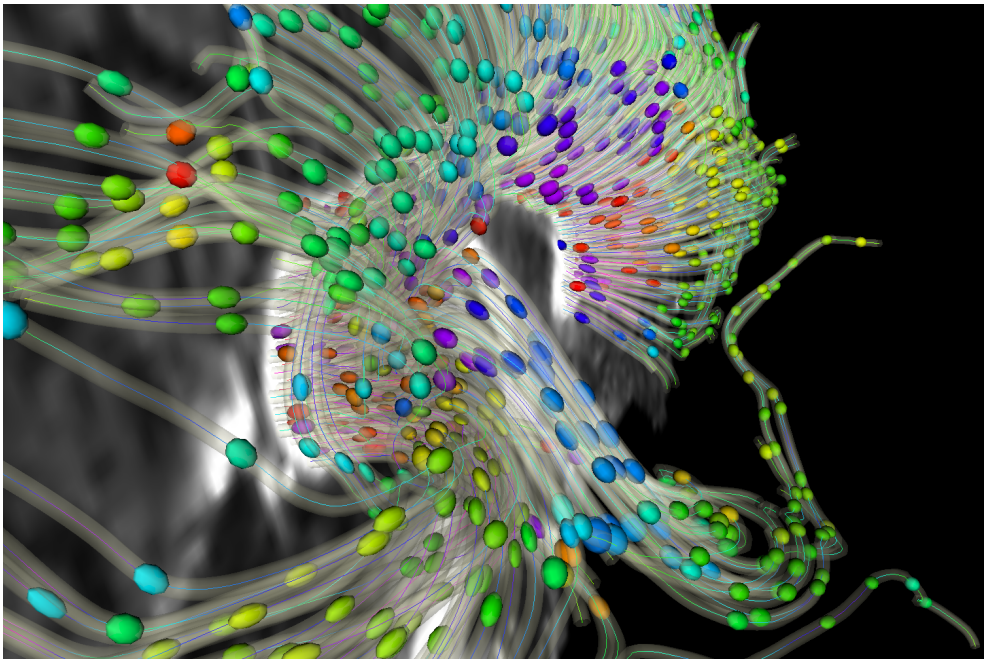
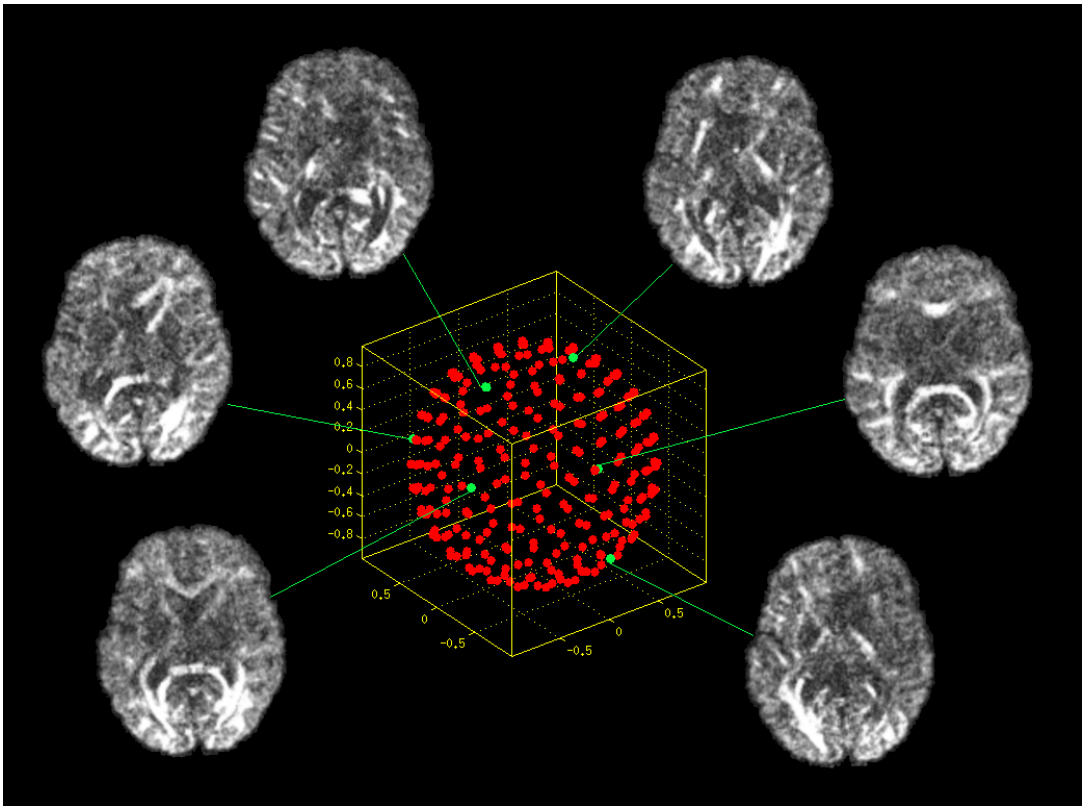


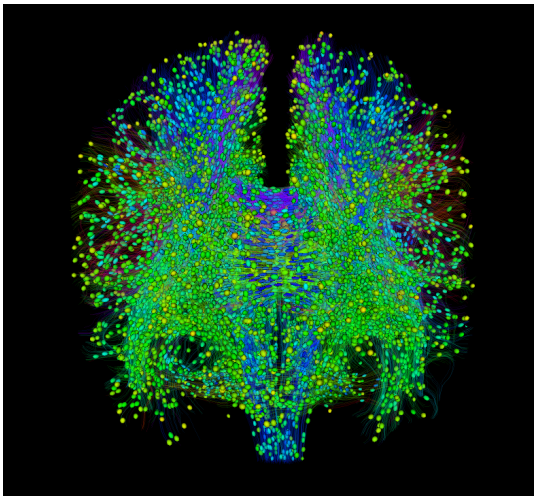
Figure 2.10: DTI-based tractography seeded in the corpus callosum. It is a close-up version of Figure 2.9 and coloured by mean orientation.

image. In the second, a DWI scan is shown as a 3D image where each voxel contains a non-isotropic or uniform spherical diffusion profile, which is how we have illustrated it in Figures 2.9 and 2.10. Both are correct and illustrated in Figure 2.11.

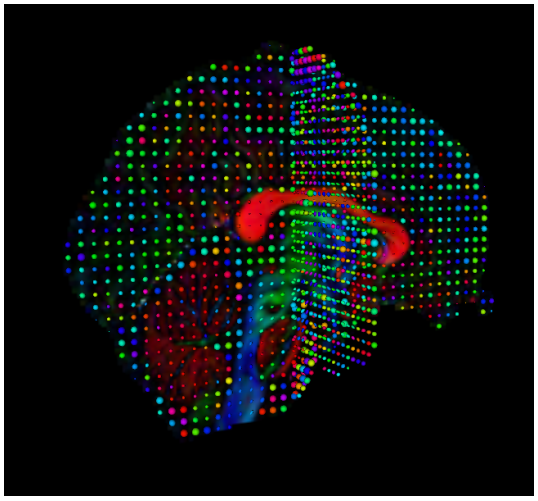
There are two major categories within the field of DWI. The first is tractography - the study of how regions connect and the structure of the white matter. The second category is focused on the individual voxels in DWI scans and the quantitative measures associated with them. These voxel-based methods are often closely tied to the rich portfolio of 3D-based image analysis methods, as they adapt the well-known methods to the directional microstructural information. Such voxel-based methods are often well-suited for exploratory research where biased spatial cohesion can be misleading. Quantitative diffusivity measures are also a lot faster than fiber tract methods, which is often a major factor in clinical practice. Each category has its strengths and limitations, which we investigate in detail in Chapter 3.



(a) Illustration of a set of diffusion measurement directions, each associated with a directional magnetic gradient and a 3D image of the brain. The unit sphere represent the directions the is measured in. From [Jensen, 2014].



(b) Diffusion profile representation 1. Created from a tractography similar to Figure 2.9 but seeded within the entire brain mask.



(c) Diffusion profile representation 2. Here the diffusion is coloured by the orientation of the largest eigenvector.

Figure 2.11: Three different visualizations of DWI data.

Popular quantitative DWI measures

Scalar-based quantitative DWI measures, such as Mean Diffusivity (MD) and Fractional Anisotropy (FA), are 3D representations of the diffusion profiles and highly popular in many areas, from registration [De Santis et al., 2014] to multiple clinical applications like the detection of stroke [Schlaug et al., 1997]. In terms of registration, the attractive factor is that any standard scalar-based 3D registration framework can be used and that the FA provides sharp edges around the white matter, which is often the area of interest. These measures do not provide any directional information, though a color scheme is often used to indicate the principal orientation of a voxel, as shown in Figure 2.12. Under the DTI model, fractional anisotropy (FA) measure provides a quick

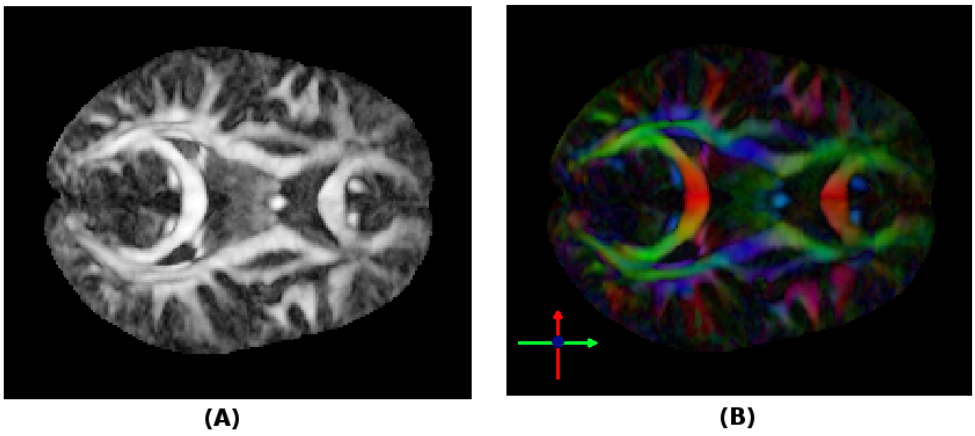


Figure 2.12: (**A**) Fractional anisotropy (FA), and (**B**) FA coloured by the directional of the principal eigenvector. From [Jensen, 2014].

way to investigate the degree of anisotropy in diffusion and especially in white matter, which has a high average anisotropy. It is a normalized scalar value given by [Peter Basser and Pierpaoli, 1996] which is defined as

$$FA = \sqrt{\frac{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{2.4}$$

where λ is the eigenvalues of the ellipsoidal 2nd order diffusion tensor, and MD is the mean diffusivity defined as the mean of the eigenvalues

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \tag{2.5}$$

MD is straightforward as it characterizes the diffusion or overall displacement of molecules.

FA is useful but it can be misleading. Within a voxel, diffusion can move in multiple directions and result in either a disc-shaped or near-isotropic profile, which will result in low FA values. It can also have lower than expected values caused by various different factors such as cell death, change in myelination, increase in extra- or intracellular water, etc., and different combinations of eigenvalues can generate the same FA [Alexander et al., 2000].

By traditional mention, we also have the less common Axial Diffusivity (AD) defined by the largest eigenvalue of the tensor ($AD = \lambda_1$) and Radial Diffusivity (RD) defined by the remaining eigenvalues ($RD = (\lambda_2 + \lambda_3)/2$). AD gives high values at the white matter and in the cerebrospinal fluid (CSF) due to highly anisotropic diffusion or simply a high amount of diffusion, and has been used to detect axonal injuries [Budde et al., 2007]. A combination of the lowest eigenvalues, RD represents less organized matter and CSF, and has been used to detect demyelination [Budde et al., 2007].

The Orientation Density Function (ODF) and complex diffusion profiles

The Orientation Density Function (ODF) describes the diffusion profile and becomes relevant at more complex models than the ellipsoidal 2nd order tensor. A well-known limitation of the simple DTI model is that it cannot resolve the orientations of the diffusion of crossing fibers in a voxel, as this will result in a disc-shaped planar ODF for the diffusion moving along two crossing fibers (see Figure 2.7B), and a misleading isotropic shape in the case of three orthogonal crossing fibers. The ability to model crossing fibers and other complex shapes have prompted higher resolution acquisitions and more advanced methods, such as HARDI and q-ball imaging or higher order tensor models. We return to these in the next paragraph, but first, we look at why some complex ODFs can be hard to reconstruct in a way that models the underlying fibers. From here, we will refer to the diffusion profile as the ODF.

Estimating complex ODFs are at the core of both voxel-based DWI and tractography. It also remains one of the most difficult and largely unresolved parts of DWI. The diffusion process in the majority of voxels is unlikely to be governed by only unidirectional fibers. In fact, according to the survey by [Jones, 2010], between one third and 90% of the voxels in a brain scan are estimated to contain more than one fiber population - depending on the data quality and the models used. The two main obstacles are resolution and antipodal symmetry. Only the bulk of the diffusion and the likely orientation of the fibers are captured within a voxel, and these are measured antipodally symmetric, which means that the measurements do not reveal the direction of the diffusion. This can be resolved for fiber crossings as the direction does

not matter, but it makes it extremely difficult to resolve similar structures such as kissing or curving fibers, as shown in Figure 2.13.

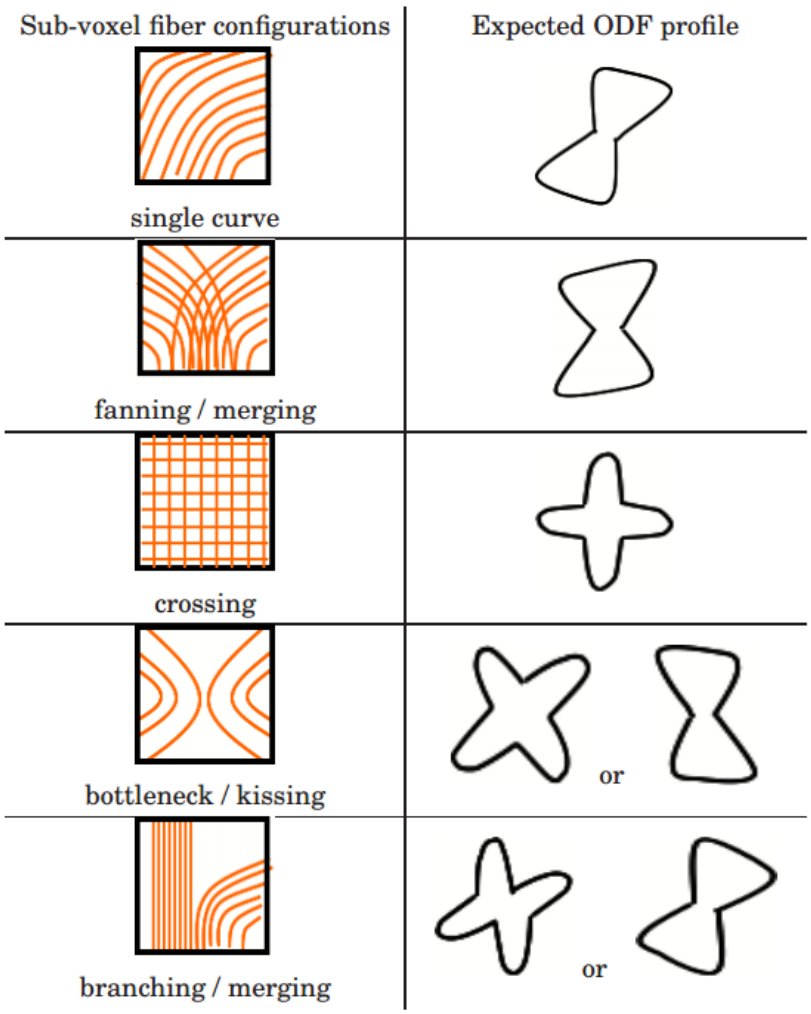


Figure 2.13: Difficult fiber configurations in a voxel (shown in 2D for simplicity). The sketches are borrowed from [Descoteaux, 2008], who got it from J. Campbell of McGill University in Montreal, Canada. We also recommend the thesis of [Descoteaux, 2008] as a great in-depth introduction to DWI.

Today, the 2nd order tensor in DTI is mostly used for quantitative DWI measures, improved alignments or similar. Most established frameworks, that model ODFs and track fibers, have the minimum capability of being able to resolve two-way fiber crossings.

HARDI, q-Ball Imaging, and the Funk-Radon Transform

The more complex ODFs are often described using q-space analysis methods, which covers a wider range of imaging methods[1]. Of these, we limit the focus to q-ball imaging (QBI) [Tuch, 2004], which is based on the top popular sampling method of High-Angular Resolution Diffusion Imaging (HARDI) [Tuch et al., 1999]. The basic idea with HARDI and QBI is to sample the Orientation Density Function (ODF) directly using a fixed b-value and to let the high angular contrast improve the SNR critical to high b-values. This is often referred to as the *raw signal*.

The concept of directly sampling the diffusion signal comes from q-space methods, in which the Gaussian model of free diffusion is replaced by rewriting the Stejskal-Tanner equation in a Fourier transform formulation to get the displacement profile from a wave vector q. We still use at least one S_0 or b_0 image and multiple gradient directions. However, we are now working with the attenuated (low) signal, which is inversely proportional to the diffusion.

QBI saves acquisition time and makes HARDI a clinically popular method for the higher non-Gaussian b-values as it involves reconstructing the ODF directly based on the raw HARDI shell. This is done using the Funk-Radon Transform (FRT), which is a tomographic inversion of the signal - introduced with QBI [Tuch, 2004] and illustrated in Figure 2.14. The FRT inverts the measured signal by integrating over the great circle, or equator, to get the likelihood of diffusion in the perpendicular direction. In this way, the tomographic inversion of a "disc" would be a "cigar", and so on. In Figure 2.14, it is illustrated as the following: if f is a spherical function at a unit vector $\boldsymbol{u}$, the FRT is the integral of $f(\boldsymbol{w})$ over the great circle $\mathcal{C}(\boldsymbol{u})$ perpendicular to $\boldsymbol{u}$. In the notation of Tuch, the mathematical definition is explicitly given by

$$\begin{aligned} \mathcal{G}[f(\boldsymbol{w})](\boldsymbol{u}) &= \int_{w \in \mathcal{C}(\boldsymbol{u})} f(\boldsymbol{w}) d\boldsymbol{w} \\ &= \int \delta(\mathbf{u}^T \boldsymbol{w}) f(\boldsymbol{w}) d\boldsymbol{w} \end{aligned} \tag{2.6}$$

where $\boldsymbol{w}$ is constrained to be of unit length. Intuitively, we re-scale each q-space sample on the sphere by the integral of the corresponding great circle. For QBI this results in some amount of smoothing of the signal. The smoothing depends on the integral of the perpendicular great circles that must be interpolated, which in turn depends on the number of uniformly distributed HARDI samples. The proof is given in Appendix A of [Tuch, 2004] and in

[1] For detailed concepts and equations of q-space analysis methods, we recommend the [Callaghan, 1993] (one of its pioneers).

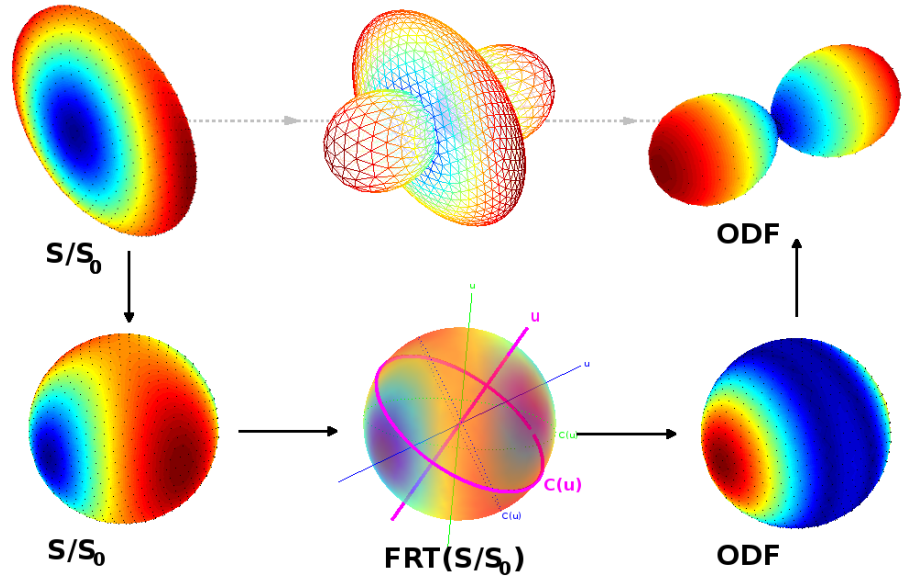


Figure 2.14: The QBI algorithm with ODF reconstruction. Top row: (left) The normalized HARDI signal; (right) the approximated ODF modeled; (middle) and the relation between the two. Bottom row: (left) The signal projected onto sphere; (middle) the FRT illustrated where the lines $\boldsymbol{u}$ represent the gradient directions being re-scaled by the integral of the great circle $\mathcal{C}(\boldsymbol{u})$ for each direction; (right) resulting on the projected ODF. From [Jensen, 2014].

further detail in Appendix 7.7 of [Descoteaux, 2010]. To avoid confusion, we define the diffusion ODF (dODF) as the FRT just described, and the fiber ODF (fODF) as the case where we fit a specific fiber-oriented model to an ODF - i.e. modeling peaks in the dODF. When we talk about voxels in a DWI, we will either refer to them as HARDI for the raw signal and ODFs for the inverted signal or DTI tensor. Additionally, when the term 'reconstructed signal' is used, it usually refers to the dODF.

Finally, we note that DTI is in the range of 6-30 diffusion gradient images, whereas q-space methods, such q-ball imaging, are typically 60-90 or higher. Additionally, at low b-values ($< 1500 mm^2/s$) with fast acquisition time, the wave vector q will resemble free isotropic diffusion, in which the Gaussian model of the DTI remains an acceptable model.

Advantages and Limitations of Popular Diffusion Models

We dedicate the remainder of this introduction to summarize some of the important advantages and limitations of the popular ways of mapping the microstructural anatomy through diffusion imaging. The following summarisations originate from [Assaf and Cohen, 2009]. For references to the various claims, we direct the reader to their excellent work and citations.

Diffusion Tensor Imaging (DTI)

Advantages:

1. DTI is very fast and can be performed in as little as 3-4 minutes, though research scans take up to 20 minutes.

2. It has quantitative and rotationally invariant parameters such as fractional anisotropy (FA) and mean diffusivity (MD) which is used in relation with many different neurological disorders and diseases, e.g. when a neurodegenerative process prompts a reduction in FA and increase in MD.

3. It has qualitative features that can be used to visualize the underlying neuronal fiber tracts, either by 2D orientation-colored FA maps or by tracing lines through anisotropic ellipsoids that represent fiber pathways (tractography).

4. It provides an efficient microstructural probe for clinical patient care and surgical aid.

Limitations:

1. The Partial Volume (PV) effect. Voxels contain several types of tissue such as grey, white matter, and cerebrospinal fluid (CSF). CSF contamination is a well-known artifact where the fractional anisotropy is reduced as voxels on the white matter boundary overlap with the isotropic diffusion of CSF.

2. The averaging effect. Voxels contain more than one predominant fiber pathway. It is well known that certain areas of the white matter have reduced anisotropy due to multiple fibers crossing and that this leads to poor tensor estimations in these regions.

3. Non-Gaussian diffusion. A Gaussian model is unlikely to truly reflect the diffusion in complicated tissue, across membranes, with various viscosities and obstacles of different sizes.

q-Space Imaging

Advantages:

1. It is a quantitative and model-free approach with no Gaussian model of the diffusion, measuring the quantitative displacement.
2. Combined with QBI and HARDI, it provides a simple mathematical relation allowing for fast and robust computation of the ODF.
3. It allows for solving the averaging effect by modeling multiple crossing fiber pathways inside a voxel, given a good model and enough directional measurements relative to the complexity.
4. It can be used to measure the bi-exponential behaviour of high b-values.

Limitations:

1. Acquisition times are longer than conventional DTI. This further leads to noise as more time in the scanner results in added patient motion.
2. To analyze the output of the scan additional modeling of the ODF is required, such as the Funk-Radon transform in QBI.
3. q-space theory requires an additional acquisition parameter, a short gradient pulse, that is difficult to achieve on conventional clinical scanners.

All in all, DTI is still the most applied DWI modality in clinical practice but the more specific q-space methods are more interesting from a basic research perspective.

3 Methods in Diffusion-Weighted Image Registration and Validation

This chapter gives an extensive introduction to and review of different approaches to performing registration, group-wise studies of DWI, and explains how such studies are validated. The lack of gold standards has promoted a large amount of research which is based on closed segmentations from trained experts and visual validation of new approaches. This is a problem in a world that is seeing a positive increase in openly available DWI datasets and still finds itself without public test sets or evaluation protocols. This chapter is the most recent version of a manuscript intended for the Human Brain Mapping journal. The manuscript was initially prompted by a study of current DWI-based registration methods. It was then decided to expand it to a large in-depth study of group-wise DWI studies and validation protocols, as we found an astounding lack of inter-study comparisons in DWI. To aptly cover these topics, we had to look at both voxel-based registration (VBR) and tract-based registration (TBR), which ended as two major chapters or categories in the review. The review is currently without intended illustrative figures and the final submission will have a more selective focus. However, we believe that this broad coverage of existing methods, and their approaches to presenting their results, will benefit the community - in particular, researchers young to DWI.

As it is standard procedure for any review to either be invited or to request an invitation, the following letter was submitted to the editors of the Human Brain Mapping (HMB) journal - along with the title, abstract, and highlights that can be found in the beginning of the review. We would like to thank Prof. Ron Kupers, from the Department of Neuroscience & Pharmacology at the University of Copenhagen, for his feedback and suggestions on the letter for HBM.

Letter to Human Brain Mapping:

Dear editor

I am inquiring as to whether Human Brain Mapping would consider to review our manuscript, entitled "Critical issues in the registration and validation of group-wise alignment of Diffusion-Weighted Imaging". The reason for our pre-submission inquiry is the length of the manuscript, which is about 23 double-column pages long. In this manuscript, we review the current state of voxel-based and tract-based registration for group-wise diffusion-weighted imaging (DWI). Due to the increasing amount of open and freely available datasets, e.g. the HCP, UK Biobank, etc., DWI is becoming an issue of Big Data. We argue that there is an imminent need for guidelines for standardizing the presentation of DWI results to make them comparable across studies. This is exemplified by the sheer lack of comparative studies on DWI registration and group-wise validation methods, and by the frequent use of local tools for segmentation and labelling that are not shared publicly. We make a plea for online evaluation protocols to help the neuroimaging community out of the current validation quagmire. The manuscript is tailored to both neuroimaging scientists and medical doctors who seek a better understanding of popular approaches to registration of DWI data.

Reception:

The letter was subsequently well-received, and we have been invited to submit the review. The reply came from Rebecca Strauss from the Editorial Receiving Office at John Wiley & Sons.

The overall organization of the review:

1. **Introduction** to the paper, along with a list of related reviews, terminology and overall categories of registration.

2. **Voxel-based registration (VBR)** covering different levels of extensions from standard scalar images to DWI

    a) **Validation of VBR** in terms of known issues, approaches and solutions.

3. **Tract-based registration (TBR)** covering both parcellation-based and fiber tract based registration. Within each, we look at refining methods and anatomically unbiased methods.

    a) **Validation of TBR** for both parcellation and tract clustering, along with common approaches to presenting and validating new methods.

4. **Discussion of evaluation protocols** for group-wise DWI studies, together with a hopefully useful FAQ.

5. **Appendix** on popular available VBR registration frameworks.

This ends the brief introduction to the purpose and structure of the review, which continues on the following page.

Critical Issues in the Registration and Validation of Group-wise Alignment of Diffusion-Weighted Imaging

Manuscript by Henrik G. Jensen

Abstract

Diffusion-weighted imaging (DWI) has become the primary non-invasive method used to study the brain's microstructure and connectivity. In both clinical and basic research, DWI studies represent a rapidly expanding multidisciplinary field with increasing amounts of publicly available high quality and multimodal data. With increasing computational efficiency and power, it has become feasible to study large cohorts in detail using multiple atlases and modalities, simultaneous group registration, joint spatial and orientation optimization, etc. This review presents an overview of recent approaches to DWI group-wise analysis of the human brain, and addresses discrepancies in their validation. More specifically, we focus on the comparison of DWI scans and on the image registration procedures underlying the shared coordinate system, or map, between heterogeneous anatomical structures. We provide an overview of current voxel- and tract-based registration methods. For voxel-based registration, we show how diffusion data are transformed under existing scalar-based frameworks, and how these have been designed explicitly for DWI. With respect to fiber tract registration, we discuss group-wise alignments procedures for examining end-to-end connectivity and structural fiber tract features. With a rapidly growing amount of publicly available DWI data, we hope to foster discussions on the potential of a more unified approach to group-wise evaluation, reproducibility, and inter-study comparisons.

Highlights

- A comprehensive high-level overview of registration and group-wise evaluation of DWI in the past decade.
- A definition of voxel-based registration categories based on DWI.
- A definition of fiber-based registration categories in terms of connectivity and structural analysis.
- An overview of validation methods for both voxel- and fiber-based registration.
- A discussion of current and future challenges in quantitative evaluation of DWI data.

1. Introduction

Diffusion-weighted magnetic resonance imaging (DWI) adds directional diffusion information to the scalar MRI. By observing how molecules diffuse, it is possible to infer microstructure in human brain anatomy, study the neuronal connectivity, and add valuable information to aid the diagnosis of diseases [10]. DWI offers the potential for creating informative maps of the brains structure but it also presents a significant computational challenge. There is a constant development within the field towards designing new models, and corresponding image acquisition sequences, that are able to reflect the anatomical microstructure of micrometer axons based on images with a comparative macroscopic resolution [105].

Knowledge of the brain is based on observations made on the individual level, correlated with prior observations, and aggregated to prove a hypothesis. Data is gathered from multiple sources in order to model pathologies, anatomical characteristics, locate new biomarkers, identify abnormalities, etc. Group studies are, at their core, about the ability to identify correspondence between features in sets of data and, in order to perform automated groupwise analysis in medical imaging, a shared coordinate system between subjects is needed. This coordinate system can either created (i) by transforming images into a shared space, i.e. create templates, or (ii) by transforming existing cartographic information (an atlas) to fit the images and extract specific localized data for comparison. At the heart of this is image registration.

Registration offers the potential to identify subtle changes, aid physicians, and categorize the neurological state by including the architectural topology of a patient, as well as combining image modalities to extract and correlate more information. For instance, a common application is to combine structural connectivity information from DWI with functional images in order to correlate the anatomy of the brain with how it works. Image registration is particularly challenging when studying complex images such as DWI from different subjects, and even

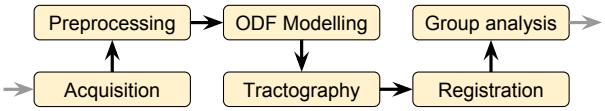


Figure 1: Example of a DWI analysis pipeline focused on tract-based registration.

when following the longitudinal development of the same subject. However, while DWI provides more information to drive a registration, it also involves an intrinsic geometry, orientation, and density, which is challenging to account for during transformation and alignment.

There are multiple stages in a DWI analysis pipeline and creating a shared coordinate system for groupwise analysis is only one step. A group average can be used as a prior to improve data modelling, but is also dependent on good acquisitions. An example of a DWI pipeline could look like Figure 1. However, with the increasing growth of substantial open multimodal datasets, that are acquired and preprocessed by the best in the field, large-scale group studies are going to pave the way for our future understanding of the human brain. Some of these projects include the Humman Connectome Project (HCP) [148], the MASSIVE brain dataset [53], the NKI-Rockland sample [103], the UK Biobank [133], and more. Registration is central in all of this in terms of creating atlases, templates, multi-modal combinations etc.

Intended Audience

This is intended for medical imaging researchers and doctors who seek an overview of group studies in DWI, how image registration is relevant to DWI, and common approaches to validating groupwise DWI data. In addition, we aim to provide a better understanding of the the common characteristics between voxel-based and tract-based approaches to group studies, and where the field of groupwise analysis of DWI is headed. The reader is assumed to have a rudimentary understanding of diffusion data such as the difference between DTI and HARDI scans, b-space or q-space, and what a structural T1-weighted or b_0 scan is. To be clear, we will use the term *DWI* for all diffusion-weighted acquisitions/images, *gradient vectors* for scan measurements, and *ODF* for spherical distributions modelling the direction of the diffusion both DTI, HARDI, etc.

Related Work

There exists a substantial amount of guides, reviews, surveys, and comparative papers related to different pipelines for DWI analysis. The following is a compact list of some of the more recent DWI reviews and guides, offering excellent introductions to DWI and/or an overview of common misconceptions, biases, and challenges. While registration is not the core focus of each study, all were selected with registration and groupwise analysis in mind, and each contribute valuable insights. What set this work apart from these papers is that the focus is on the role of registration in groupwise DWI, and how data is used from voxel-based methods to registration based on derived spatial correlations, such as surfaces and fiber tracts etc.

- [84] identifies 25 pitfalls (biases that can lead to lack of accuracy and even substantial errors) along the DWI pre-processing and analysis pipeline with 9 relevant pitfalls dedicated to intra- and inter-subject comparison in group studies.

- [106] gave a comprehensive review of parcellation- and fiber-based white matter segmentation (with a focus on the latter). Focus is given to groupwise analysis and computational feasibility,

- [132] covers the workflow of DTI analysis from pre-processing to analysis and result interpretation. A large amount of relevant literature and prominent methodology are reviewed in a "straightforward hitchhiker's guide to DTI".

- [86] discuss white matter integrity, fiber count, connection strength, and other fallacies in DWI. This review is focused on the different ways to interpret DWI statistics and clarify misconceptions prevalent in the literature. A useful "do's and don'ts" list is given.

- [107] presents a review of DWI methods and measures used in clinical neuroimaging research. Common ways to analyze in vivo white matter are discussed and potential pitfalls are presented for each category.

- [101] provide a boarder review of the DWI pipeline with a focus on acquisition, pre-processing, group analysis methods, and non-DTI models, while also giving and overview of disorders/pathological DWI studies.

- [97] discuss the future of spatial normalization of brain images, how diffeomorphism makes most sense if a template is sufficiently similar, and argues that heterogeneity should be represented by a cohort multi-atlases of combined fMRI and DWI.

- [105] very recently gave a perspective on the current state and future of DWI in relation to tractography, statistical analysis (e.g. distance metrics in group studies for tensors and ODFs), and registration of voxels or fibers.

Registration Terminology

Image registration refers to a process that transforms data into a shared coordinate system. Groups are compared either by matching a *template* such as a group average, a random single subject to the native space of each individual subject, or by transforming each subject to a single shared template space. This is often referred to as *spatial normalization* or *image fusion*. The template will usually have an associated *atlas*, which may be the anatomical segmentation or a probability map. Registration often consists of an initial *global* rigid or affine alignment, and optionally followed by a *nonrigid* alignment. A nonrigid registration can be approximated locally as a set of local rigid or affine transformations. The global deformations are often *diffeomorphic* which means that the transformation between two images is smooth bijection (i.e. a differentiable and invertible map).

Approaches to DWI Registration

Throughout the past three decades, numerous registration methods for DWI have been proposed and new frameworks are continuously released every year. This is largely due constantly improving acquisition schemes, both in terms of data quality and reduction in scan duration. However, it is also owing to a general lack of shared community-accepted platforms for validating registered DWI, different ways of interpreting how Orientation Distribution Functions (ODF) represents the underlying anatomy (e.g. when tracing fibers), and a discrepancy between acquisitions for basic research and most clinical applications. Depending on the representation of the ODF, registration methods focus either on quantitative diffusion measures (scalars derived from the ODF like fractional anisotropy), on models of the ODF to drive the registration (scalars combined with directional statistics), or they trace fiber tracts and group them through clustering or parcellation [156]. These approaches can be split into the following two categories

1. Voxel-based registration methods that focus on the modelling of individual ODFs or quantitative ODF-derived measures. This is also referred to as Voxel-Based Analysis (VBA) or Voxel-Based Morphometry (VBM) and often involves registration of the whole brain using only a brain mask as prior segmentation. To avoid confusion, we will refer to any of these as *Voxel-Based Registration* (VBR) methods.

2. Fiber tract registration or clustering methods that solve the white matter segmentation problem. These will either focus on selected regions-of-interest (ROIs) often from cortical parcellations, or on clustering fiber bundles based on structural similarity. They encompass registration based on derived spatial correlations, and we include the popular tract-based spatial statistics (TBSS) but note that TBSS is not related to fiber tracking methods. All in all, we shall refer to these as *Tract-Based Registration* (TBR) methods.

From a registration point of view, there is a clear overlap between the VBR and TBR as groupwise analysis of tracts often requires some initial voxel-based alignment, although most surface-based registration approaches use 2D vertices, either exclusively or along with voxels [89]. While surface-based registration is important for parcellation methods, the approach does not use any specific DWI-related information and we will omit surface-based methods.

Scope and Section Guide

We review how the angular information in diffusion data is both transformed and used to aid alignment of brain scans, and focus on inter-subject and whole-brain image studies, as these present the greater challenge. We investigate defining paradigms and commonly used tools for comparing DWI scans, both in terms of individual ODFs (VBR) and more spatially correlated models (TBR). In the context of creating a shared coordinate system in groups of DWI, we attempt to answer the following questions

- What are the challenges of extending scalar (3D) image registration to DWI?
- How is tractography being used in registration?
- What are the common ways of validating VBR and TBR?
- What can be done to rank new methods and standardize validation?

The review is organized as follows: VBR in Section 2 covers DWI being deformed according to standard scalar-based registration, and algorithms designed for explicit DWI optimization. TBR is covered in Section 3 with a focus on connectivity and shape-based fiber clustering. Both sections go through several categories of popular methods and recent developments in groupwise DWI analysis - each ending with a focus on typical validation and challenges. We pick up on answering the questions above in Section 4, and give an overall discussion based on what we can summarize - including future perspectives into DWI groupwise validation. Finally, we have attached descriptions of some of the most prominent and popular VBR frameworks in Appendix Appendix A.

2. Voxel-Based Registration of DWI Data

Registration of structural scans e.g. T1- or T2-weighted MRI or quantitative DWI features such as fractional

anisotropy (FA), are by far the most common ways of registering DWI. This is due to (i) manipulating orientation vectors for optimization in a way that is still true to the anatomy and acquisition is very difficult, (ii) the catalogue of traditional 3D registration methods is immense and well-tested in fields with a gold standard, and finally (iii) scalar methods are fast. Two approaches for performing scalar-based DWI registration dominates:

1. Registration of the T1w or T2w MRI scans or the gradient-free b_0 DWI scan and apply the resulting deformation field to the DWI, either by warping the model-free gradient vectors or the model used for ODFs.

2. Registration based on scalars or quantitative measures derived from the ODFs such as Fractional Anisotropy (FA), Eigenvalues, Spherical Harmonics Coefficients, Mean Diffusivity (MD), the Apparent Diffusion Coefficient (ADC), etc.

We introduce another subdivision of methods to the scalar-based registration:

3. Algorithms that are based on either (1) or (2), and that iteratively reorient the gradient vectors or ODF without performing explicit optimization over the reorientation. We define this as *implicit reorientation* along the methodology of [27].

The third category contains methods which do not explicitly calculate the derivatives of the reorientation. These methods use the deformation from scalar-based registration to iteratively reorient the gradient vectors during optimization. An example is two images being aligned by minimizing the difference of orientation invariant feature such as FA. The FA is iteratively recalculated based on the transformed ODFs, which are reoriented using the gradients of the spatial deformation. Yet this is still a by-product of the scalar registration - an important distinction as noted by [4]. The following will be focused on these three categories, but we will also cover *explicit reorientation* methods at the end of this VBR section (one of which is the well-known `DTI-TK` framework). Explicit reorientation methods are distinguished by incorporating the directional gradients into the cost function of the optimization.

Scalar-based registration disregard the information stored in the gradient vectors. However, scalar methods have merit in larger pipelines/frameworks for analyzing diffusion data where the warp or deformation field might have been obtained from elsewhere or for easy swapping of registration algorithms in a modular way. In fact, most popular available frameworks today work in this way Appendix Appendix A.

2.1. VBR: Scalar-based and Implicit Reorientation

In scalar-based registration the orientation of the diffusion gradients or ODFs are not a direct part of the optimization[1]. Independent of the type of scalars we use, the challenge for this type of registration lies in obtaining a correct final reorientation of ODFs, that is coherent with the orientation of neighbouring ODFs. The importance of the reorientation depends on the goal. To underline this, the most recurring discussion, and critical introduction in most DWI VBR papers, falls on whether the warped ODFs can still be used to trace the fiber tracts (thus approximating a "true"/underlining anatomy in the warped space [40]) - and indeed if tracing fibers in the warped space is even the goal[2]. When it comes to inter-subject registration, a rigid registration will not provide an adequate fit between differently shaped objects and, as a minimum, an affine transformation is required.

However, an affine registration (global or local) will result in a shearing, stretching, or non-uniform scaling, and — if this is not accounted for during reorientation of the ODF — it will likely have an impact on the ability to trace fibers through ODFs. Basically, shearing introduces a complex rotational effect that affects the reconstruction of fiber trajectories as each direction of the ODF warped differently [60]. This is generally accounted for by approximating the warp using only the local rotational component from an affine deformation, or in some way accounting for the affine transform, e.g. by re-interpolating the warped ODFs or normalizing the shearing effect.

Reorientation of tensors: FS, PPD, and multi-channel Demons

More than a decade ago, [4] proposed two highly popular approaches to tensor reorientation which attempt to deal with the reorientation of the ODFs and are still used today: the Finite Strain (FS) and Preservation of Principle Direction (PPD) algorithms. FS decomposes the registration into a deformation and a separate tensor rotation component but ignores the affine effects of shearing and non-uniform scaling. PPD assumes that the eigenvectors describe the tissue orientations, and preserves them by approximating the affine transformation (we return to this further below). For DTI, these two approaches to reorientation of ODFs have been fundamental and FS is used in many QBI registration methods. FS has widely been considered an acceptable and fast *approximating* of the deformation, and one of the first uses was seen in [66] with a multi-channel[3] approach, along with the *Demons*

[1] Though there might be the *implicit* reorientation as an iterative deformation of the ODFs and subsequent recalculation of quantitative measures. See for instance [121] as an example of this.

[2] See the discussion Section 2.3 at the end of this section.

[3] Multi-channel refers to optimisation where each voxel is represented by multiple measures/features/modalities like FA, MD, T1w, edge-detection, etc.

registration algorithm, which is worth briefly introducing given its frequent appearance in the literature.

The Demons algorithm [137] is a well-known and popular nonrigid registration method. It is a scalar-based non-parametric approach that is part of the `ITK` toolkit, and variations of it is used in many DWI registration methods. In its original form, it uses gradient descent and sum-of-squared difference (SSD) optimization over a displacement field that is adjusted by a normalized optical flow force[4]. [150] extended the Demons algorithm to a diffeomorphic framework, which was added to the popular `ITK` toolkit and used for DTI registration (see e.g. the comparative study by [155] who compare eight different DTI registration algorithms). The multi-channel demons algorithm of [66] was also extended to include iterative (implicit) tensor reorientation to improve the quality and evaluated it on both quantitative measure and fiber bundles [115]. Demons also appear in more recent works like [20] where HARDI-based ODFs are aligned in a diffeomorphic multi-channel approach (used on a rotationally invariant distance function for Real SH) and reoriented based on FS. [61] also perform diffeomorphic registration on SH coefficients, rotating them (i.e. altering the SH coefficients) in a manner similar to FS. Non-parametric, efficient and diffeomorphic, Demons registration is very popular for complex data that is hard to regularize such as DWI or multi-channel data. While extended to DWI, most of these methods are primarily based on structural registration from b_0 or FA volumes, but as we see in Section 2.2, Demons was also extended to explicit tensor reorientation in the publicly available framework `MedINRIA`.

We return to the PPD tensor reorientation approach. This is a more DTI-specific approach than FS, as PPD approximates the affine deformation by applying multiple rotations. The principal eigenvector is multiplied with the affine matrix, projected back on the unit sphere, and the corresponding rotation is found. The same is done for the next eigenvector, though the second rotation is done perpendicular to the first. Thus, it becomes a sequential set of rotations keeping the ODF structure of each individual voxel. It was used by [85] who first registered T2w scans to create a template, deformed one of the FA images to create a target FA image, and then used this for scalar registration with the PPD method for tensor reorientation. Another early approach, similar to PPD, was taken by [160] who applied tensor reorientation using the deformation fields of co-registered T1w images while iteratively rotating each of the principal directions. Later [27] used PPD of registered DTI volumes to reorient the HARDI-based ODFs in an inverse-consistent fluid registration algorithm that minimizes the symmetrized Kullback-Leibler divergence.

Reorientation of QBI: Registration of HARDI and higher order models

Registration algorithms of q-space data, more specifically HARDI/QBI algorithms, focus on higher angular resolution to model ODFs for basic research over everyday clinical application, such as DTI. Nevertheless, despite having longer acquisition times, QBI is also gaining in clinical popularity [36]. Registration of HARDI-based data is often a scalar registration followed by a reorientation of the ODFs due to the high resolution and complex shape. As with DTI, it is commonplace to use the Jacobian of the deformation to reorient the ODF, performed either by pure local rotation or by local affine transformation.

In the first case (local rotation only), the rotational component is separated from the transformation, which offers an attractive simplification of the transformation as it does not change the ODFs. It is very similar to the FS model of DTI and, despite being a simplification of the actual complex transformation, it still commonly used as seen [61], or more recently in average estimations like [149].

In the second case (local affine transformation), transformations are performed by multiplying the Jacobian with the vectors of the HARDI or ODF model, without any matrix decomposition, followed by a projection or normalization back on the sphere (see for instance [162] or [78]). As a consequence of the affine shearing/stretching/non-uniform scaling, it is often discussed how one preserves the volume fractions[5] of each ODF after the transformation. [71] use the determinant of the Jacobian (from b_0 image registrations) to adjust the length of the transformed vectors, while [122] use weighted SH point spread functions that are reoriented individually from b_0, FA, and T1w, which corresponds to multiplying with the Jacobian and normalizing the transformed vector. [40] extended this work to account for correct transformation of the ODFs in accordance with the expected underlying white matter anatomy, and discussed the risk of creating anisotropy out of isotropic components through affine shearing, which they corrected for by separating the isotropic SH coefficient.

In contrast to the purely orientational tensor in DTI models, shape and density changes in ODFs are more visible issues in volumetric models of HARDI data, such as those based on QBI or similar, where the models have a non-specific number of dominant fiber orientations. Such

[4]The intuition is that a free parameter (the "demon") pushes a point in the direction of the image gradient if the intensity is lower than the target value, and in the opposite direction of the gradient otherwise.

[5]Volume fractions refer to the different components (different fiber orientations, other tissues, water, etc) in a voxel, not to be confused with the partial volume effect which refers the problem of mixing or averaging the volume fractions.

volumetric models are often referred to as the diffusion-ODF (dODF) [82]. However, HARDI data is also being modelled as more purely orientational models, referred to as fiber-ODF (fODF), which focus on the peak orientations of the ODFs. These are often modelled by using Mixture of Gaussians, (constrained) spherical deconvolution[6], or other higher order models to indicate the estimated underlying fibers, similar to DTI. For a better overview on deforming and reorienting dODFs, see [164] who extends on the ideas of [122, 40] for registration and reorientation directly on the HARDI data. For a couple good articles covering approaches from point spread functions to implicit reorientation of fODFs using the Jacobian of the spatial transformation (along with thorough experiments and discussions), see [123] and [121].

Compartment models: Volume fractions and multi-tensor modelling

Spherical harmonics are perhaps the most common way to model QBI and HARDI-based ODFs, before or during registration of q-space data [21], particularly since [37] who provided an analytical and elegant regularized solution to modelling the ODF with a Funk-Radon transform of QBI data in a more numerically efficient manner. However, DTI remains the most popular clinical DWI acquisition protocol ([182]) and various ways have emerged to resolve crossing-fibers for DTI data, and deal with the complex mix of hindered and "free" diffusion within a voxel. These methods are often referred to as *compartment models* ([36]) as they are a mixture of multiple models. Their purpose is to model the volume fractions in a voxel and avoid partial volume effects where e.g. the isotropic part of a voxel mixes with anisotropic compartments. Among these are Gaussian Mixture Models (GMM) which can be viewed as natural extension of the 2^{nd} order tensor with multiple tensors per voxel [144], the Ball-and-Stick Model which uses the GMM for an anisotropic model of fibers (*stick*) and an isotropic Gaussian model for the free diffusion in the voxel not hindered by fibers (*ball*) [18], and the similar but more complex 'Composite Hindered And Restricted ModEl of Diffusion' `CHARMED` by [8], which became the field standard for modelling extra- and intra-axonal compartments from DTI data, extrapolating parameters such as axonal density.

While important for understanding the diffusion processes, these DTI-based compartment methods are usually not extended to registration outside motion correction, although we note that examples do exist, such as scalar-based registration of multi-tensor DTI [135], ball-and-stick DTI [6]), and multi-fascicle DTI [136]. However, for clinical practice a quick acquisition time matters significantly and DTI offers a good cytoarchitectonic probe with the simplified 2^{nd} order tensor model. It is also simpler to use a quick scalar-based registration of a structural atlas instead of trying to deform the multi-component DTI compartment models. The 2^{nd} order tensor can not resolve complex structures, like crossing or kissing fibers, and yet tractography for surgical aid (planning and navigation) is still often based on the single tensor DTI model [47]. On the other hand, more advanced scans that can resolve fiber crossings have been shown to greatly improve the quality for clinical situations [47], and it is not hard to imagine that the field should prepare for a future where these scans replace the single tensor DTI. The tools for modelling complex ODFs of higher resolution scans exists and are constantly improving along with inter-subject HARDI registration, as seen in a recent nonrigid PPD-based multi-compartment example by [32]. In recent years, `CHARMED` has also been extended to HARDI with the `AxCaliber` model [9], and together with `NODDI` (Neurite Orientation Dispersion and Density Imaging) [172] and `DIAMOND` (DIstribution of Anisotropic MicrO-structural eNvironments with Diffusion-weighted imaging) [124], they have formed the foundation for a large range of more complex and computationally intensive diffusion microstructural models (see [69] for an extensive overview).

Scalar-based registration though multichannel methods

It is clear that DTI is primarily used as an cytoarchitectonic probe, meaning that it might not necessarily reflect the true anatomical structure but it is a good indicator, and useful nonetheless as a way of increasing the resolution and improving scalar-based registration through more informative quantitative measures such as FA maps. When we discuss more complex models based on HARDI scans as a clinical tool, it is often in relation to complicated not-everyday cases where longer post-processing times are acceptable. In these cases, registration is an important tool and nonrigid registration can significantly aid manual segmentation and locate biomarkers, as was shown in a recent study on automated classification of Parkinson's Disease [15]. Scalar-based registration is one way of getting around the computationally heavy and analytically difficult advanced ODF models, but it remains a projection of the orientational data and suffers from non-specificity [107], and information is lost when the reorientation of highly angular ODFs is not taken into account during registration. However, scalar-based methods are not without merit. They are fast and a lot of popular VBR frameworks for DWI are scalar-based (see Appendix Appendix A on popular frameworks). Additionally, areas of grey matter and the Cortico-Spinal Fluid (CSF) can

[6]Spherical deconvolution exists in various forms and has its basis in spherical harmonics. It is a technique used to estimate distribution of fibre orientations in an ODF using a set of response functions [140].

likely be more easily distinguished in high spatial resolution structural MRI which can provide added information along with quantitative DWI measures. In recent years a lot of combinatorial methods using rotationally invariant features have been suggested for DWI registration, some multichannel as with the Demons approaches in [20] mentioned above. One example is a registration algorithm by [2] that build feature vectors along the TBSS skeleton (see Section 3.4) at each voxel that contain GFA, eigenvalues, etc. used in combination with the HAMMER (Hierarchical Attribute Matching Mechanism for Elastic Registration [126]) similarity measure. Another framework that also use this multi-component similarity measure is the 'SPherical Harmonic Elastic REgistration' (SPHERE) by [163] which combines edge maps and several orientation invariant ODF features for an initial robust registration followed by an incremental SH order refinement using the earlier mentioned weighted SH PSF reorientations of [122]. SPHERE updates the ODFs based on the scalar maps of each iteration and falls under the category of implicit reorientation. Also in this category are more multichannel Demons approaches as is seen in [19] where a white matter atlas is created in combination with the popular Fiber Orientation Distribution (FOD) model by [140, 5] (as the name implies a method similar to the multi-tensor concepts used to model the fODF, here by constrained spherical deconvolution on HARDI data [139]). F-TIMER (Fast Tensor Image Morphing for Elastic Registration) by [165] combines anatomical landmarks, edges, and tensor information in an approach similar to SPHERE (the two were compared in [163]) - the main difference being that F-TIMER was built for DTI data instead of HARDI as with SPHERE.

Whether we term these more complex models to be multimodal, multichannel or combinatorial methods working on compartment models, FODs or other methods resolving multi-fiber geometry on ODFs, algorithms for utilizing multiple features are popular and not without reason. As long as the models are computationally feasible, adding more data should provide a more robust registration and, while it is not yet used for registration, these scalar ensembles are likely to be among the first methods to be introduced to contemporary popular deep learning frameworks that benefit from more data channels as seen in a tumor segmentation example by [178], can help create entirely new features through convolutional neural networks [143], or for improving sparse DWI acquisition [64].

2.2. VBR: Explicit Reorientation

This section is dedicated to registration algorithms that include the reorientation in the analytical gradients, and by extension makes use of the directional information in optimizing the similarity or cost function of two volumes. We refer to this as *explicit reorientation*. Compared to the previous section, these algorithms are often more computationally heavy, have extensive mathematical derivations, and heavy memory requirements on the derivatives of the transformation and reorientation, which is why these papers most often come with numerical implementation details. On the other hand, they are likely to provide a more robust and possibly faster optimization as shown by [174] who use FS with Polar Decomposition and incorporates the similarity measure between the full DTI profiles in the cost function. Together with [173] their work became the foundation of the DTI-TK registration framework, which remains one of the best and most popular DTI registration frameworks (see Appendix Appendix A). Another widely used framework, that also applies explicit tensor reorientation and the FS model, is MedINRIA by [168], and more specifically the DT-REFinD algorithm. It is a fast diffeomorphic algorithm with a Demons-based objective function that is computationally heavy but argued to show improved results against implicit reorientation and scalar-like computation time as it allows for a Gauss-Newton optimization approach (again we refer to Appendix Appendix A for more details). The explicit reorientation in DT-REFinD was also extended to the log-euclidean domain by [134].

From DTI to HARDI data, explicit reorientation was shown to improve the registration in the works of Du et al. below. This was first done in a Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework with a similarity metric based on Riemannian manifolds where the implementation was compared with both a scalar and tensor version as well [41]. It was extended to a probabilistic LDDMM approach for the HARDI atlas generation in [42], and further to HYDI data (multi-shell HARDI) in [43] where the signal was represented with a Bessel Fourier orientation reconstruction (BFOR) and an atlas was generated. Possibly inspired by their, Zhang et. al also moved to an LDDMM framework and HARDI data in [176] with explicit reorientation of diffusion basis functions (a more generalized symmetric tensor model than of the Watson distribution basis functions used in prior work [175]). Their approach results in warped raw DWI images on which they argue other frameworks, that model ODFs or other DWI features, can be used for quick group analysis.

Another interesting framework for linear and nonlinear registration of both DTI and HARDI was presented in [44] and compared to both DTI-TK and MedINRIA. The method is built on the FS model, although the authors argue for a possible extension to PPD to include the affine transformation, and it uses angular interpolation as a substitute to a specific ODF model — similar to the approach of Zhang et al. above. Their work is an extension of the popular FSL's FLIRT and FNIRT (see

Appendix Appendix A) that includes explicit reorientation and has been made publicly available as an plugin to `FSL`.

2.3. VBR: Validation

In DWI, creating a ground truth or gold standard (a community-accepted way of evaluating the success of a registration) remains an open problem [97, 105]. Due to this lack of ground truth data, qualitative assessment via visual inspection of ROIs is still one of the primary ways to present DWI results and can be found in nearly all DWI studies. It can be useful for inspecting known areas of fiber crossings, sharpness of the registration, for showing clear discrepancies in results, and generally for sanity checking a method. However, such results are rarely reproducible for comparative evaluation in followup studies, and what we really want is approaches to quantify results in a way that is comparable with other methods for external validation. We return to this discussion in Section 4. Here, we briefly list some of the popular quantitative ways studies have evaluated their results, based on the prominent studies covered above and a list of comparative studies.

The following is a summarized list of quantitative groupwise validation methods for VBR which can be applied to most DWI data. Some methods require expert segmentations while others can be performed without medical expertise or without a specific region in mind.

- (Automatic) **Whole-brain voxelwise error**. VBR is most powerful as an unbiased registration approach. Without strong assumptions about spatial correlations, it is possible to perform explorative studies and tests hypotheses regarding atrophy, new biomarkers, etc. A majority of the above citations fall within this category, usually followed by one of the additional validation approaches below. Aside from standard scalar-based measure (RMSE, log-SSD, etc.), some of the popular VBR distance measures for DWI include:
  - *Symmetrized Kullback-Leibler (sKL) divergence*. Information-theoretic cost metric. sKL is popular for both DTI and full ODF profiles [27, 41, 176].
  - *Orientational Discrepancy (OD)*. Measures angular discrepancy in ODFs, more specifically the difference between ODF peaks. [163, 26, 161].
  - *Dyadic Coherence (DC)* and *Tensor Overlap (OVL)*. DC is the variability in the dominant diffusion direction based on principal eigenvectors, while OVL is the overlap of eigenvalue-eigenvector pairs in DTI [85, 173, 27, 67].
  - *SH coefficient distance*. Scalar measures used over SH coefficients provide an ODF distance measure [123, 41].
- (Manual) **Structural markers or labels**. Well-known white matter ROIs (structural delineations of tracts), and at times grey matter structures, are delineated by experts, and a measure of overlap is often used to evaluate the registration. An atlas can replace manual segmentations, if manual segmentations are not available. It is worth noting that using an atlas registration from A for VBR to validate your own registration B is intrinsically limited by A and is mostly used as a sanity check. Examples of label-based validation can be found in [163] who use the Dice ratio, and [41] who use also use sKL.
- (Automatic) **Synthetic warp**. This popular approach, where the data is warped, often randomly, and registered back to itself, is easy to quantify and a great sanity check. However, it should be noted that warping an image in an unbiased way while keeping the warped image anatomically plausible is a significant challenge. It is also often used in papers to visually show that the ODFs are reoriented properly using a controlled deformation. Examples of this approach can be found in nearly all papers, see e.g. [168], [134], and [44].
- (Manual) **Fiber tracts**. In this approach, well-known fiber tracts are segmented manually and used to evaluate the registration. The fiber tracts are estimated in the native space and warped using the resulting VBR deformation field. Spatially correlated measures, such as tracts, are excellent for validation if the labels can be generated, but care should be taken as they rely heavily on the trajectory/tractography model. They are unlikely to compare favourably to TBR in which the fiber tract models themselves are at the center of the registration. Examples can be found in [41] and [163].

There are other groupwise validation methods that depend highly on the data available. These are often seen when studying specific diseases, tumor development, treatment, etc, and often require multiple scans of the same subjects such as intra-subject data. This includes the development in longitudinal scans e.g. identifying stages of Alzheimer's by neuronal atrophy, combining pre- and intra-operative scans, or tracing other changes like the effects of age or a specific treatment (see [86] for this and more).

Once frameworks have been published and committed for public use, the natural next step is to compare such methods with as little bias as possible. Such comparative studies offer an excellent insight into both the state-of-the-art algorithms and well-described validation methods.

- [155] made a comparison of 8 algorithms - 1 linear and 7 nonlinear commonly used DTI registration al-

gorithms. The first 6 were based on FA maps and the last 2 relates to registration with explicit reorientation. Validation included visual inspection of FA, and structural biomarkers. The latter were white matter ROIs measured by the DTI-specific mean FA-weighted angle between principal eigenvectors.

- [157] compared 11 open-source popular DTI registration methods with 7 DTI-specific evaluation criteria. Of these open-source methods, 10 were scalar-based algorithms and 1 was based on explicit DTI reorientation. All registration frameworks are nicely summarized, and the validation spans a useful set of both scalar and orientational similarity measures.

These studies are good examples of popular ways to evaluate and compare registrations, but they also highlight how a method can be shown to outperform another by choosing a biased evaluation criteria. For instance, an algorithm with explicit reorientation such as `DTI-TK` will perform best in test if the error measures are based on the angle between tensors. On the other hand, `SyN` seems better if the evaluation is based on scalars such as FA.

2.4. VBR: Concluding Remarks

The advantages of VBR comes from the more model-free unbiased algorithms which are great for whole-brain exploration, e.g. when searching for population-wise biomarkers. Including directional information in the registration has repeated been shown to improve the registration as opposed to scalar-based algorithms based on structural MRI of quantitative DWI features such as FA or MD.

The two major limitations of VBR, partial volume (PV) and reorientation, are both mainly relate to the warped space that we have from either deforming or interpolating the images. These are related to creating a DWI template or performing tractography on a deformed reconstruction. However, they not primary limitations when it comes to creating a shared coordinate system between scans and analyzing differences and shared features.

1. **Smoothing introduces partial volume (PV) effects**. Smoothing an image is essential to create a stable differentiable deformation or a multi-scale model. If we assume that the initial resolution does not already suffer from PV effects or that a compartment model captures the effects (e.g. on the border of the CSF), smoothing is still bound to make PV effects worse. However, the smoothing is only relevant to the registration. Given a spatial mapping, subjects can be compared across native spaces without adding to PV issues, and a quantitative template can still be created by mapping each subjects scalar values. The major issues from smoothing is the risk that minuscule structures will be disregarded if they do not follow the major dominating features in registration.

2. **Scalar-based registration means challenges in reorientation**. This has been a recurring theme throughout VBR of DWI. If scalars such as FA are used to map voxels then it is unlikely that the fiber trajectories will align as well everywhere — even for theoretical correct reorientation model. Implicit reorientation will give a more correct registration but it is still based on scalar similarity measures. Again, this is likely not a significant issue if we do not care about fiber tracking in the warped space but instead perform model calculations in the native space. Even so, we have seen that aligning tensors and ODFs, based on their shape and directions, with explicit reorientation improves the registration.

3. Tract-Based Registration of DWI Data

In the second part of this review, we look at algorithms that derive spatial correlations from DWI with the intend on modelling the brain based on fiber trajectories or pathways. This is often done either by following the largest eigenvectors of the tensor in DTI, or peaks in the dODF or fODF for q-space methods. The estimated fiber pathways are used to study various topics such as connectivity, shape/length, and quantitative DTI or ODF measures along the tracts. For an introduction to these tract statistics, we recommend [109] who also developed a generalized TBM implementation.

We to use the terms tracts and fibers a bit interchangeably. There is some justified confusion in the imaging field about tract vs fiber in terms of anatomy. Interlocked axons make up a nerve fiber, a nerve tract is a bundle of fibers, and 'bundle' is a scale that can be anything from the smallest to the largest tracts. The term "fibers" is often used as shorthand for fiber tract or fiber bundle. The type of anatomy referred to by these terms often boils down to how the white matter segmentation problem is being solved. That is, do we bundle fibers by connected regions or by their structural properties (we recommend [106] for more discussions on this). Finally, we also use the popular term 'streamlines' for any contiguous set of 3D points generated from tractography that indicate fiber pathways.

White Matter Segmentation

To see where groupwise measures and registration fits into tractography, we have to look at what basis subjects are compared. TBR is all about clustering the (often millions of) fiber trajectories/streamlines into grouped comparable regions - which in turn involves solving the

white matter segmentation problem (globally or locally). Inspired by [106], this is aptly separated into three approaches: parcellation-based connectomes, fiber clustering, and hybrid methods.

- **Parcellation (or ROI) methods** are fiber termini (or cortex) centric, which means that they perform cortical (grey matter) parcellation and group fibers based on what ROIs (or nodes in a connectome graph) they intersect. As fibers are bundled by the brain regions that they connect, these methods are well-suited for graph-based algorithms and comparison with brain function (fMRI scans).
- **Fiber clustering methods** are white matter centric and segment fibers into similar anatomical bundle structures, based on the clustering algorithm and the similarity measure. These approaches are aimed at measuring the central anatomical properties of fiber tracts, and are well-suited for studying neurological diseases and large displacements such as tumors.
- Hybrid methods combine the connectivity-driven information of parcellation-based methods and the clustering of similar white matter anatomy.

Others such as [58] use the terms 'streamline-based' and 'connectivity-based' for essentially the same approaches as the above. Or 'voxel-based' and 'surface-based' as in [28]. Since these methods of segmenting the brain based on fiber tracts often have different goals (connectivity vs. structural anatomy), comparing them is not always straight forward. Without gold standards one paradigm and scale of segmentation can be as valid as another. That said, each method have their strengths and weaknesses, and each relate differently to registration and voxel-based methods. Parcellation-based approaches lean toward an initial nonrigid voxel- or surface-registration as ROIs have to be identified in the cortical regions in a common frame of reference across subjects. Fiber clustering methods on the other hand often segment the white matter prior to a local affine or nonrigid registration of fiber tracts (to get a more sparse representation). Due to the rich portfolio of 3D multimodal registration and the high interest in connectivity, ROI-based methods are the most published methods while fiber clustering has flourished in the more recent decade ([58]). There are two common approaches to perform groupwise analysis in both parcellation and fiber clustering studies

1. **Refinement methods**. Here clustering is often individual followed by an averaging scheme. These are the most used methods and make use of spatial priors (often from histology) or manual delineations. If available, an atlas is registered to the subject space of each individual in order to create a common frame of reference between subjects. Averaging or comparison of specific local anatomy can then be performed in a more computationally efficient localized way, as opposed to handling all subjects at once. This allows for handling more information (e.g. streamlines) but it also limits the parcellation to the initial spacial prior and the accuracy of the registration or the expert delineator.
2. **Unbiased methods**. These methods do not require spatial priors and are less common due to the computational burden of clustering multiple subjects simultaneously. Each subject is either registered to a common space (often a subject in the cohort) after which all subjects are clustered, or all subjects are registered and clustered simultaneously. Such methods are promising as they are more purely data-driven, but they can also be harder to navigate and represent a significant computational challenge. As such, data is often represented as sparse or high-dimensional.

Others, such as [80], referred to these as "top-down" (refinement) and "top-op" (unbiased) methods. There are of course more nuances to these two categories, such as multi-atlases, but it is a subdivision that roughly holds for all group studies, and we use it in the following where we look at state-of-the-art in white matter connectivity and structural analysis.

3.1. White Matter Connectivity

The objective of mapping the brain's anatomy to its function is closely linked to parcellation-based segmentation methods, that seek to chart the brain in-vivo based on what cortical areas are connected. Initially, postmortem maps and neuroanatomical conventions were used to create the first successful surface-based parcellation algorithms[7] as we saw with `FreeSurfer` in [51], and the early DK-atlas presented in [38]. Later it was followed by the popular atlas by [39], and more recently in [90] with their DKT40 atlas. These macrostructural approaches are often rated on their ability to closely identify the same regions across subjects which is highly useful for cortical cartography in a clinical setting. However, as pointed out early in [51], lobes (sulci/gyri) can not always be separated based on macrostructure (folds) or even microstructure (folds within folds). If the data is available, there is evidence that the borders in cortical cartography

[7] We have intentionally stepped around surface-based registration in the voxel-based methods of the previous section as this approach is less relevant for DWI in terms of utilizing directional data. For a comparison of surface-based and voxel-based registration we recommend [89] who performed a large-scale comparative study using `SyN` and `ART` for voxel-based registration, and `FreeSurfer` and `Spherical Demons` for surface-based registration.

should be guided by both "functional fingerprints" and "connectional fingerprints" [116], while structural scans can be used for overall navigation and initial bias (see [46] for a thorough discussion in segmentation based on connectivity and function). In recent years cortical parcellation have gone from being structural maps for functional MRI (fMRI) to incorporating both structural and functional modalities in the parcellation itself, creating entirely new subdivisions of the brain. One such prominent example is [63] who link fMRI to structural MRI, myelin maps and more, creating a new HCP atlas by training a robust classifier to recognize the multimodal fingerprint of each cortical area (by way of clustering and manual experts). This atlas ended up with 180 areas of which 97 were new in comparison with previous reports. They also included but did in fact not use diffusion MRI data.

To center our scope around DWI, we narrow the focus to multimodal parcellations that actively involve connectivity information and white matter segmentation, specifically 'structural connectivity' from DWI data - unlike connectivity as a correlation/interaction term as used in fMRI studies. In terms of [116] this is the hypothesis of functional regions having a specific connectional fingerprint (additionally examined in [30]). Finally, we focus on very recent studies since whole-brain analysis has less manual expert intervention in terms of registration, and it has only recently become computationally feasible. However, we acknowledge early pioneering work in connectivity-based parcellation such as [16] who presented the first in-vivo connectivity-based segmentation of grey matter, based on connections from thalamus to cortex (the registration involved was global affine from `FSL`).

Parcellation: Voxel- or surface-based registration

There are three common approaches to studying the human connectome by way of structural connectivity-based parcellation: (i) voxel-based parcellation when a conventional whole-brain atlas is desired for an initial parcellation often involving a nonrigid voxel-based registration, (ii) surface-based parcellation for a parcellation based on macroscopic cortical features, such as gyri, lobes and commonly shared structural folds, or (iii) a combination of both when surface and subcortical segmentations are needed. The dominating approach seems to be surface-based registration where an initial cortical segmentation is performed using a tool like `FreeSurfer` and related atlases[8]. While we will only cover purely voxel-based parcellation briefly in the first part under refinement methods, it is also present in studies with subcortical segmentations using `FreeSurfer`, as it comes with the option of both voxel and surface registration.

Refinement based on existing parcellations

We first look at voxel-based parcellation, all of which were performed using a scalar-based nonrigid registration frameworks, where an atlas is registered to acquire an initial parcellation. A common choice is `FSL` as we see in [153] where `FNIRT` is used to transform the AAL-90 atlas ([147]), modify the atlas based on connectivity, transform the results back to common space, and create a probabilistic atlas to investigate group difference between healthy and schizophrenic subjects. Another scalar-based registration toolbox is `SPM`[9] which was used in a similar manner on a larger population in [65] to register the AAL-90 atlas to the structural T_1 scans.

For surface-based parcellation, it is fair to say that `FreeSurfer` is the common registration tool of choice. An early example of groupwise registration for anatomical refinement can be found in [28] who modified a `FreeSurfer` atlas parcellation (36 anatomical labels inflated to a sphere) by taking inter-subject connectivity into account. It was also the basis for [22] where the 66 cortical parcellation in the DK-atlas was used to create a nested scale-space by initially subdividing each of the 66 ROIs five times. `FreeSurfer` comes with both voxel- and surface-based nonrigid registration (see [119] or Appendix Appendix A) and the voxel-based registration was also used to extract subcortical grey structures from an atlas in [22]. Their work was based on the prominent correlation studies between structural and functional connectivity in [68] and [70], who used the smallest ROI size compared to the scale-space study (still `FreeSurfer`). In [145] it was used for presurgical planning where `FreeSurfer` and the DK-atlas (here 95 cortical regions) was used in combination with another registration framework[10] more robust to tumors to perform (simultaneous) groupwise clustering. Based on their previous similar parcellation-based tract extraction and adaptive shift clustering in [146], the fibers were clustered using the connectivity-pattern of each voxel along a fiber.

A problem with many of these methods is the lack of availability. A recent bid at publicly available structural connectivity-based parcellation can be found in [93]. Here `FreeSurfer` was used for surface-based cortical registration to parcellate 70 gyri, after which each gyrus was clustered and segmented based on fiber connectivity and group reproducibility. This study presented a combination of structural, connectivity and fiber similarity

[8] `FreeSurfer` was also used to parcellate the popular HCP cohorts [63].

[9] Specifically `SPM5` in this case.

[10] `DRAMMS` or Deformable Registration via Attribute Matching and Mutual-Saliency weighting [110].

while relying on gyri segmentation. It is available in the BrainVISA platform and tested on the ARCHI database. The parcellation has yet to be connected to functional scans. Another interesting example of connectivity-based parcellation with both surface- and voxel-based registration can be found in the recently published 'Human Brainnetome Atlas' by [46]. For each subject, an initial structural surface-based parcellation is created using the DK-atlas in FreeSurfer, where the structural scan is aligned to diffusion space and nonrigidly registered to a common MNI152-space[11]. Here an ROI-based probabilistic white matter segmentation is performed, followed by spectral clustering with cross-validation (and more) for additional subdivision of the DK-atlas. It resulted in 210 cortical and 36 subcortical parcellated regions based on connectivity that were reproducible across the cohort, with added functional connectivity information and with mental processes delineation from the BrainMap database. Of particular note, the authors recognizes the need for what they refer to as 'structural connectivity-based registration', which is their term for using orientational features in the VBR instead of only scalar-based registration. However, they deem it not fully mature and use nonrigid scalar-based registration. The Brainnetome Atlas was followed up very recently with the open source Automatic Tractography-based Parcellation Pipeline (ATPP) in [94][12]. Here a massively-parallel connectivity-based parcellation framework scalable for both clusters and desktop computers is presented and tested by parcellating a variety of brain regions from other papers. As in the Brainnetome project, nonrigid registration to a common space is performed using the scalar-based registration from SPM8 toolkit. It is added that manual inspection of the registration should be performed and is integrated in the pipeline.

Unbiased parcellation

The second category of unbiased parcellation methods is still rare for whole-brain group studies but promising as the results spatially correlated while also giving a more data-driven parcellated atlas without macrostructural bias and predefined number of ROIs. The major challenge lies in the computational burden of clustering streamlines simultaneously in a cohort of subjects in an often hierarchical manner[13]. Finding a common frame of reference can be tricky for close to voxel-sized ROIs due to noise, partial volume effects, and the anatomical variability[14]. While only using a four subjects, an interesting example of a tree or dendrogram solution can be found in [100] who used hierarchical clustering to generate whole brain parcellations, while also comparing several hierarchical clustering methods. Here SyN was used for nonrigid registration of FA maps. A clear advantage of this approach is the dendrograms containing information about how ROIs relate to each other at different levels. This inspired [54] who used a larger cohort of 66 HCP subjects, each of which came with a FreeSurfer surface parcellation that was used to create shared seed points at vertices used to cluster tractograms. The unbiased parcellations were compared against both functional scans and the DK-atlas. Another approach also using parcellated HCP vertices was presented in [114]. 'Supervertices' were used to create a nested multi-scale parcellation over 100 subjects that were split into two groups on three levels going form 500 to 2000 supervertices for a single hemisphere. The resolution was increased and expanded to more modalities including fMRI in [113].

Concluding remarks on parcellation-based methods

Cortical parcellation enables us to study white matter connectivity between brain regions in a subject. It ties structural and functional scans together, and it gives a common frame of reference for groupwise studies. Structural connectivity can be used to create more informative parcellations than the shape and folds of the brain. However, as we saw with refinement methods, the results are often tied to the predetermined lobar borders that are identified in the subject using nonrigid registration (of vertices or voxels) with an atlas. While more unbiased methods exist, atlas parcellation is seems to be the most common. For the popular tools, such as FreeSurfer or FSL, this registration will in most cases involve structural MRI (like T1w or b_0) or derived quantitative measures such as FA, but not the connectivity itself. However, new public tools, like the ATTP pipeline, aim to incorporate DWI more in the parcellation.

The limitations of using structural connectivity for parcellation lies largely in the end-to-end tractography and the quality of the data. For tractography, tracing fibers over long distances, and near the cortical regions, requires the ability to (approximately) resolve crossing/kissing fibers, which is critical as major homogeneous fiber tracts will otherwise dominate as evident with the single tensor DTI [7]. Tracing fibers to the borders of the less anisotropic grey matter also requires higher quality data

[11] The nonrigid registration framework is never mentioned but likely done with the scalar-based SPM8 toolkit as evident in the co-developed ATPP framework below.

[12] To be clear ATPP was developed during the Brainnetome project and share very similar yet optimized approaches.

[13] Without prior knowledge of the number of clusters, some tree-like hierarchy or multi-scale method often works best. The number of clusters remain an open problem [94].

[14] Anatomical variance is also often referred to as *heterogeneity* in the brain. The concept of a template or atlas parcellation of a population is based on capturing the opposite of heterogeneity, namely *homogeneity* which refers to shared features.

[106]. DWI-based parcellations are more difficult to evaluate than parcellations based on macrostructural features, and most such focus on reproducibility as a validation method and proof of reliability (which we return to in Section 3.3).

3.2. White Matter Structure

We move from analyzing the white matter as the wires of the brain, and on to consider the white matter more as the skeleton of the brain. The goal of parcellation-based white matter segmentation was indirectly in defining the graph nodes or ROIs in the cortex based on fiber connections. Here, the goal is to compare fiber tracts based on their structural similarity by the shape of streamlines and tracts, which for whole-brain analysis can mean working with millions upon millions of streamlines across subjects. Such tractography-based spatial alignment algorithms are often defined by the way a fiber is represented and compared to other fibers (*similarity*), how they are grouped (*clustering*), and what registration is used to align them (*registration*). It is worth noting that, in the case of groupwise comparison of fiber tracts, registration is at times only performed on a purely global scale, leaving it to intelligent clustering to segment similar tracts. Streamlines can for instance be identified and clustered with high dimensional spectral clustering as shown in [108]. Others argue that the shape variations in tracts means that localized registration is critical to groupwise comparison - even to a cohort where the fiber bundles have been pre-segmented into known anatomical structures or in a globally aligned template space [59]. And somewhere between clustering and registration, multi-atlas and dictionary algorithms have begun to show significant promise in capturing such shape variations [97]. In all these approaches, complexity and high computational requirements have resulted in algorithms where fiber trajectories are represented in a sparser space (e.g. 1D tracts, 2D sheets major bundles, randomly sampled fibers, fixed length representation of fibers, etc), or in high dimensional spaces such as we see with the popular spectral clustering methods.

In the following, our focus will be on more recent frameworks created to directly cluster streamlines and align fiber tracts in order to segment the white matter for groupwise analysis. We split it into refinement methods, unbiased methods, and multi-atlas approaches.

Refinement of known fiber tracts

Refinement methods are often focused on comparing specific anatomically named bundles e.g. by affine transforms. For such methods, registration plays a more significant role and streamline clustering is primarily used to to find simpler, sparser fiber representations.

In a prominent study [59] "Streamline-based Linear Registration" (SLR) was presented and used to register streamlines of pre-selected bundles in a cohort, and also demonstrated for pairwise whole-brain registration. For bundle-based registration, streamlines were randomly selected (max 400) and affinely registered to a template subject. Their work was a continuation of the 'QuickBundle' method (sparse tract representation) [57], which enables a computationally feasible affine whole-brain fiber registration. QuickBundles and SLR was recently combined to 'RecoBundles' in [58][15] where pre-segmented (or model) bundles can be used to extract similar bundles from a cohort, and for more difficult subjects with large displacements, such as tumors. While all registrations performed remain affine for both bundle and whole-brain transformations, the registration of bundles can be considered locally affine and will thus produce better fits, though limited by the pre-segmentation.

Another registration method came from [98] where streamlines were registered directly. The high computational burden was handled through affine registration of streamlines as feature vectors[16] and ICFs (Iterative Closest Feature/Fiber) for high dimension nearest neighbour matching. ICF is not far removed from the idea behind the sparse fiber model of QuickBundles (in using fiber 'centroids' or averages for the deformation), and it was performed with probabilistic boosting tree classifiers for bundle segmentation in [99] - referred to as 'Hierarchical ICF'. ICF was also applied in a more recent paper by [180] where template bundles were matched to individual subjects, under the assumption that approximate matches can be found in all subjects.

Most of the above represent fiber models with a fixed number of points for all fibers to make the computations feasible. However, other paradigms also exist. An early approach was presented in [159] where the subjects were first rigidly aligned and the bundles — modelled by 'tract density maps' based on a Gaussian Process framework — were nonrigidly registered using a polyaffine groupwise approach. Here, each bundle is an affine registration and all local affine registrations are fused into a single deformation, creating a more unbiased template. The idea of polyaffine registration of bundles was also used earlier in one of the first nonlinear bundle-based registration algorithms by [181] who did pairwise registration and compared the results with a scalar-based nonrigid Demons algorithm on FA maps.

[15] Their framework is publicly available in the open source python toolbox `DIPY` - see [56].

[16] It is a common choice to represent fibers by a fixed number of points for easy vectorization, often ranging from 5 to 20 points independent of length.

Unbiased fiber clustering

Unbiased fiber clustering are characterized as methods that do not rely on prior segmentations of known tracts for groupwise alignment, and can be used to potentially locate new unknown or unexpected fiber constellations. There are at least two different approaches to groupwise alignment of unknown fiber tract segmentations.

The first approach is to perform simultaneous clustering of streamlines from multiple subjects. Without prior knowledge of the location or number of fiber tracts appearing in across individuals, subjects are globally aligned and bundles are identified as shared spatially similar structures - by clustering and fiber registration. This often requires an extra focus on sparse and efficient computations, and we have only recently begun to see such methods potential in results for large groups in a decent resolution, like the HCP. An early noteworthy example of streamline-based joint groupwise clustering, was presented by [108], in which global rigid registration of a cohort was performed to create a high-dimensional white mater atlas based on subsequent spectral embedding and clustering. Streamlines were represented as high dimensional points, and random sampling of streamlines were used to help solve the computation burden. The atlas was manually segmented after the multi-subject clustering, and any new subject was clustered and segmented automatically. The rigid registration was changed to an affine registration based on fiber similarity in [112], who presented one of the first simultaneous groupwise tractography-based registration methods. Finally, the two methods were combined with a nonrigid b-spline registration in a recent extension to automatic detection of major tracts in subjects with large displacements (brain tumors) [111][17]. For another recent example of inter-subject streamline clustering, we also recommend [25] who presented a heuristic[18] multi-subject streamline clustering framework reported to have both good scalability and outlier detection. It was tested on both synthetic data and healthy HCP subjects, and it is available on Github[19].

The second approach is to integrate anatomical ROI data along with structural streamline similarity to group fiber tracts across subjects. As an example, [127] presented an alternative hybrid method for unbiased fiber clustering, where `FreeSurfer` (specifically HCP) cortical and subcortical parcellations were used for unsupervised spectral clustering of streamlines, along with their structural similarity. The spatial registration was global affine but indirectly coupled with the nonrigid `FreeSurfer` pipeline. Their results indicate that using structural ROIs with fiber anatomy (shape) gives a clustering closer to an expert manual labelling. It was recently extended to a hierarchical tree clustering with 'AnatomiCuts' in [128] - made computationally feasible with random sampling of streamlines during clustering.

The multi-atlas and fiber tract registration

On the millimeter scale, the complexity of the brain cannot be captured in a single average but varies too significantly, even in healthy subjects, for an optimal nonrigid spatial mapping. This does not mean that we cannot use registration and spatial mapping to identify areas of homo- and heterogeneity in subjects. However, it presents a significant challenge when it comes to segmentation and large group studies with more than one subgroup and correct atlas. [97] referred to this as "The diffeomorphism delusion", and points to a future of multi-atlas strategies and dictionaries of anatomical patterns. Intuitively, we would get a better segmentation by splitting populations into subgroups with shared or similar unique features and creating a hierarchy of shared coordinate systems. More importantly, two criteria for doing so are within range: (i) large publicly available and growing datasets, and (ii) sufficient computational power and efficiency.

We have put the concept of using a multi-atlas with structural fiber clustering as this is a quickly growing area within multi-atlas approaches for groupwise analysis of DWI. The reason seems to be that streamline clustering for group comparison is already performed using sparse representations of bundles, which makes multi-atlases, dictionaries, and feature learning possible. It is also important that clustering methods tend to only perform global rigid or affine alignments of DWI and, without the flexibility of nonrigid registration, identifying new bundles requires more labels or learned features to capture inter-subject variability [59]. As such, fiber clustering methods have started to replace nonrigid registration with multi-atlases and voting schemes. As an example, [169] used 12 manually labeled subjects with a segmentation of 7 major bundles while assigning arbitrary labels to non-identified bundles, such that each tract in a test subject is individually labelled based on the atlas with the most similar structures. In a related study by [80], multiple atlases of manually labeled tracts were used to transform (scalar-based nonrigid FA) the native space of 198 young normal twins with HARDI scans, and to segmented them with a label-fusion method. The fiber clustering used here was a hybrid approach, similar to [127], with both ROI and fiber shape clustering. However, the tracts selected for the multi-atlas were conservative or well-known estimates that should exist in all subjects, and so this multi-atlas should be considered to be within the same subgroup or diffeomorphism. Similar to these multi-atlas approaches,

[17] All their methods are available online and open source for public use - see the paper for details.

[18] In this case the 'Multiple Species Flocking model'.

[19] https://github.com/amiraCHEK/3D-SSF-work-in-progress.

dictionary learning tries to capture or learn the various configurations of fiber tracts in a sparse representation. In a recent example, [91] demonstrated a sparse fingerprint[20] method, called 'Fiberprint', on 861 subjects from the HCP that could separate subjects (including twins) based on 3.000 fiber trajectories.

Concluding remarks on methods based on white matter structure

Streamlines and tracts give a high level, model-based, and spatially correlated view of the brain. It can be an advantage over VBR, but it is also challenging as it is more computationally expensive, involves a greater risk of biased connectivity assumptions, and sparse representations can easily neglect important density information [59]. The alignment of fiber tracts are limited by the initial tractography, and a high number of streamlines is often required to increase remove outliers and false positives.

In terms of registration and clustering, tracts and streamlines are treated differently than VBR and cortical parcellation methods. First of all, parcellation-based methods group tracts based on their fiber termini, which means that they often rely on the initial voxel or surface registration and not on the tracts themselves. Second, the resulting transformation of a VBR can be used to align streamlines without losing spatial coherence with the underlying anatomy, as these assumptions are already made in the initial native space tractography (unlike the reorientation of ODFs in VBR). Additionally, if an adequate tractography can be performed in the native space, VBR is likely outperformed by using the streamlines or tracts directly for the registration, though the a significant limitation is that fibers are often treated as the same length, with the same number of points representing a fiber, for computational efficiency and vectorization (e.g. [112] with 5 points, [127] with 10, and [59] with 20).

All in all, detailed registration of fiber tracts still struggles with being computationally heavy for groupwise studies of the whole brain. However, this is changing with efficient sparse representations and multi-atlases. Structural fiber clustering has already been shown to be a highly efficient tool in both basic research and clinical practice, and likely the best tool if a study is aimed a specific bundles [105].

3.3. TBR: Validation

Validation of TBR methods, like VBR, is not easy and suffers from the same lack of ground truths and difficulties in comparing results with previous studies, as discussed in Section 2.3. However, TBR involves more assumptions about spatial correlation through tractography, and the validation is often focused on the correctness of the tracts and streamlines, and on the segmentation of well-known fiber bundles or cortical regions. While both categories of TBR methods segment the white matter based on shape or connectivity, their goals and approaches to validation are not always the same. On the other hand, template reproducibility is often a keystone for both types of methods when it comes to results that are based on the end product of a long pipeline - especially when a good choice for the number of clusters in various regions or bundles is unknown.

Validation of parcellation methods

Parcellation based on structural connectivity lies on the boundary between white matter structural fiber clustering and functional correlation in the cortex. Regarding cortical and subcortical parcellation, the first thing to consider is whether parcellation is used to directly to add anatomical bias to white matter segmentation (like with [146]), or if the structural connectivity is used to improve parcellations created from a macrostructural bias. We have focused on the latter as many DWI-based parcellation methods add structural connectivity information in order to refine the cortical regions, and combine it with functional modalities, such as fMRI [46]. For these multimodal methods, the validation is essentially the same as with fMRI parcellations, and here we cannot hope to give a better introduction and discussion on the topic than [7], who systematically compared 10 subject- and 24 group-level state-of-the-art parcellation methods along with common quantitative assessments.

To our knowledge, there does not currently exist any DWI-related comparative surveys of parcellation-based methods. Nor are we aware of any challenges specific to parcellation that include diffusion data. For VBR, this presented significant problem as the performance of various algorithms were difficult to compare. However, for parcellation methods the evaluation is often automatic as most of these methods start from the assumption that subjects share a coordinate system in the form of the DK-atlas in `FreeSurfer` or similar. Even unbiased methods like [113] assume an approximate vertex mapping in HCP subjects (again `FreeSurfer`). Incidentally, visualization of the refined parcellations have also been centered around well-known cortical folds, such as the pre- and post-central gyrus on the left hemisphere defined by `FreeSurfer` [94, 92, 93] - which should make the different methods somewhat more comparable. If we go ahead with the assumption that the registration is sufficiently accurate, and that the cohort is diffeomorphic in the cortical atlas region, we can form an idea of the general validation steps by looking at the two recent public frameworks of [93] and [94], and correlate this with

[20] A fingerprint here refers to a compact feature vector representation that is unique to an image.

[7]. For now, we also assume that the tractography is well-founded.

- **Reproducibility across parcellations**. Spatial consistency is the primary quantitative measure for correlation/robustness, as structural connectivity is used to investigate new potential subdivisions of the cortical areas. It is common to split the cohort in subgroups and use cross-correlation with structural labels (Section 2.3). The *Dice coefficient* is a well-used example an overlap measure between parcels [93, 94], and it is often used with the *Adjusted Rand Index (ARI)*. The ARI measures the agreement of two parcellations without parcel matching which is more effective if two parcellations have a different number of clusters [93]. As an alternative [94] used the similar *Cramer's V*.

- **Overlap with cytoarchitectonic areas**. Here we refer to the well-known areas in the cerebral cortex, that are often used for visual validation of parcellations. Most common are the motor or visual cortex, which have a both strong structural alignment and functional agreement between subjects [7]. They are available in both `FreeSurfer` and HCP, and examples with the motor region can be seen in [93, 94] (pre-/post-central gyrus).

- **Cluster validity**. While both of the above contribute to searching for an optimal number of clusters (still considered an unsolved problem), a common measure used to evaluate cluster homogeneity is the *Silhouette coefficient*, found in both [93, 94]. It measures how much a voxel/vertex belongs to its cluster (i.e. parcel), as compared to other clusters, and it is derived from the correlation in the native connectivity matrix/profile of a parcellation. Additionally, [94] also use topology measures, such as the *Hierarchical index*, which depict how well sub-clusters fall within the same parent cluster.

Most of these and more can be found in [7], who also reflect that the optimal number clusters depend on the area of the brain and the type of study.

The idea of structural connectivity-based parcellation have been met with great interest but also with a lot of skepticism, which mainly focus on the tractography and is beyond the scope of this review. DWI-based parcellation have a certain amount of limitations and pitfalls as discussed in [83]. We recommend a recent review of tractography in the brain by [79], who discuss the challenges of tracking fibers near the cortex and general misconceptions regarding connection strength between long and short range regions. Also pointed out in [7] is the risk of large bundles dominating the evaluation, and the issue of streamlines that tend to end in the gyri and results in a false bias towards following the cortical macrostructural folds.

Validation of fiber clustering methods

Based on the structural similarity (or shape) of the streamlines, the three primary approaches to validating groupwise fiber clustering are (i) synthetic data, (ii) manual segmentation, and (iii) visualization. We briefly outline common approaches in these categories, and return to discussing to manual segmentations, as it is a limiting factor in groupwise studies.

- **Synthetic data/deformation**. Simulated setups are common in groupwise fiber clustering to quantify results, along with manual segmentations. [58] used a synthetic tractogram from the ISMRM 2015 Tractography Challenge, deformed one of bundles to create a second version, and registered the two. [112] did something similar but instead deformed randomly sampled streamlines from a healthy subject. Publicly available tools have also been developed to generate synthetic fiber constellations, such as `FiberFox` which was used to create the ISMRM 2015 tractogram [102][21], and the Numerical Fibre Generator (`NFG`) used in the HARDI reconstruction ISBI 2013 Challenge [29]. There is also the more recent `Phantomas` [23], and `D-BRAIN` [117].

- **Manual delineation**. These are treated as ground truth examples that can be used for quantification of real world data. This means that similarity measures, associated with knowing the ground truth, are often used, such as *specificity*, *sensitivity* and *accuracy* with cross-validation [58]. Or *precision* and *recall* in [99]. It is also common to use overlap measures, such as the *Dice coefficients* and the *Jaccard coefficients* [127, 99]. There is also [180] who counted matching fibers based on a minimum nearest-neighbour distance. Different tractography algorithms are rarely compared on a group level, and focus is on the scores of the individual well-known tracts. For more alternatives similar to parcellation validation, [111] used cluster consistency/reproducibility and fMRI activation, while [127] used homogeneity and completeness.

- **Visualization**. A large part of DWI is visualization. In fact so much so, that it prompted the well-known paper with "Just pretty pictures?" as part of the title [81]. Nonetheless, it serves as an important quality measure for fiber clustering - in particular for inter-subject analysis where tracts can be overlaid for a

[21] Updated on `http://www.tractometer.org`.

quick comparison. As an interesting example, the Optic Radiation (OR) tracts are difficult to trace and hard to manually delineate, and thus popular for visual validation. They can be seen delineated in [99] and [59].

With the growth in large open online datasets, it is increasingly important to address the limits of manual validation, which is rarely shared and will differ from trained expert to expert. As we mentioned, studies of groupwise fiber clustering will in often rely significantly on manual segmentations by neuroanatomical experts with specialized tools such as `Slicer`, `TrackVis`, `DTIStudio`, or similar (see [132] for more). In fact, nearly all methods listed in Section 3.2 have their own referenced experts to create manual delineations[22]. While other DWI-based methods also lack ground truth data, such as VBR and connectivity-based parcellation, the lack of consensus seems more present in fiber clustering, where new methods quantify their results by manually seeded tracts or outlined ROIs as the gold standard[23]. Involving experts for segmentations lends credibility and helps quantify results, and it also adds flexibility in noisy scans or brains with visible differences. On the other hand, automation is the goal of DWI analysis. However, manual annotation does not scale well for large group studies where every ground truth must be annotated, it is expensive and often not shared publicly, it affects reproducibility due to human imprecision [179], and it is increasingly difficult as tractographies become more complex (e.g. multi-tensor DTI or HARDI) and the number of required ROIs increases to restrict the selection [111]. While it is highly applaudable that cutting-edge open-source tools such as `SlicerDMRI` [104] is making it a lot easier for clinical researchers to study difficult individual cases, there remains a significant requirement for automated validation approaches of new (or even existing) fiber tract segmentation methods in DWI. This is especially true for the large publicly available DWI databases, where much is to be gained from groupwise analysis.

We briefly review the typical roles of the manual expert, and go on to give an example of a recent alternative to manual validation.

- **Adding anatomical knowledge**. The primary role of the expert is to delineate ROIs (in 2D) that a tract is expected to pass through (usually two or more), place seed pixels from which streamlines are generated (e.g. to avoid whole-brain tractography), and validate major well-known fiber bundles (e.g. corpus callosum, corticospinal tracts, cingulum, etc. [33]).

- **Pruning tracts**. The wide range of deterministic and probabilistic tracing approaches are known to be plagued by false-positive and false-negative streamlines [62]. Having seeded a fiber bundle, to delineate it an expert will often have to 'prune' the result for anatomically implausible streamlines (outliers).

- **Tuning parameters**. Depending on the task, an expert might have to change the parameters such as the minimum length of a streamline, the maximum angular momentum, and the minimum require anisotropy, until a tract looks as expected.

For automated image analysis, this corresponds to the nontrivial steps of registration with an atlas, clustering the fibers, and learning parameters. To give an example of this, we summarize the important take-home validation points of [25], who presented a method without the aid of manual segmentation (introduced in Section 3.2).

1. **Ground truth**. The accuracy of the results was evaluated by segmenting 37 bundles in each subject using the promising state-of-the-art white matter query language by [158] (`TractQuerier`). The ROIs available from `FreeSurfer` in the HCP data was used as input to the segmentation. The bundles were converted to a probabilistic atlas and quantitatively compared with their own method and another state-of-the-art method.

2. **Synthetic evaluation**. The correctness of the streamline clustering was evaluated using the `NFG` software. This was both used to test complex ODFs and outlier fibers (pruning), again against other state-of-the-art methods.

3. **Visualization**. This goes without saying, as it is one of the attractive features of tractography papers. It is not the best way to validate results, but it is an important indicator for the more experienced readers.

4. **Specific data**. An important thing to notice is that the 10 healthy subjects used from the HCP was actually listed by subject number - giving others an opportunity for comparison.

This is one way to introduce a new method without manual expert aid. On the other hand, one of their main arguments for success is the performance of competing methods, which might be a limiting factor in itself.

The remaining topic on TBR validation is comparative studies. However, it says a few things about the state of tractography and groupwise fiber clustering as there does

[22] Exceptions to manual validation are a few the referenced methods such as [181] who use an in-house developed atlas, and some of the unbiased methods like [25] and [112].

[23] Manual delineation of fiber tracts is also, a bit morbidly perhaps, referred to as 'manual dissection'.

not seem to be any comparative studies of inter-subject fiber clustering. The few we could find are listed below and does not necessarily include streamlines or fiber tracts in the cost function of the registration.

- [156] recently evaluated groupwise tractography using six different VBR algorithms. Comparison was based on the Hausdorff distance, RMSE, MSE, tractwise-FA, and the amount of tract intersection (cosine distance) between deterministic fibers in 10 subjects and a random subject template. All with eight manual ROIs based on FA. Additionally, spatial correlation was used to evaluate probabilistic tractography with 27 automatically masked ROIs. `DTI-TK` was concluded to be the overall best registration algorithm.
- [35] evaluated two nonlinear image registration algorithms, `FNIRT` and `Elastix`, with a redesigned TBSS approach without the projection phase[24]. It offers an example of standard scalar-based registration being used to warp (probabilistic) tractography to a template space for evaluation.

3.4. TBSS - A Voxel-Based Approach on Tract Statistics

We briefly cover the highly popular approach Tract-Based Spatial Statistics (TBSS) model by [129, 130], which has become central to DWI registration (particularly DTI). Despite the name, it is not related to tractography and is (in its original appearance) a spatially correlated VBR method. The name comes from the way that the method traces the white matter 'skeleton' from the mean or centerlines of FA maps, which moves along some of the major tracts of the brain. It is a method that attempts to combine some of the strengths from both VBR and TBR. TBSS is based on two steps. First, a nonrigid alignment of FA images to a common space where the templated FA image is thinned and thresholded to create the skeleton along the lines where to subject cohort have most in common. Second, each subject's spatially aligned FA image is projected onto the skeleton by searching for perpendicular max FA values, and statistics along the skeleton is gathered. We recommend [131] for a good introduction to TBSS.

TBSS is worth mentioning, not only because it is popular and the foundation for many new methods [35, 152], but also because it is considered an alternative to dealing with the the "degree of smoothing"-issue present in VBR (discussed in Section 2.3) [132, 107]. In this, there are a few things to keep in mind: (i) TBSS, as an FA scalar-based algorithm, is a paradigm best-suited for clinical DTI, (ii) it is a white matter analysis framework, so when TBSS is an improvement from VBR, it is so in relation to studies specific to the major fiber bundles, and (iii) TBSS methods are often purely scalar-based registration methods without directional data. However, it was extended to include tractography in [152] (TABSS) which improved the statistics but also limited the results to major commonly shared tracts found in an atlas. TBSS is currently at the core of the rigorous statistical test framework in `FSL` for DWI data, where it was incorporated by the author of the framework. After becoming part of `FSL`, it is fair to say that TBSS has become one of the most popular tools for hypothesis-testing in pathology related to the white matter [14]. In particular since little to no manual expert segmentation is required. For a comparison between manual ROI and TBSS, see [95]. For a list of pitfalls in TBSS and a review of methods suggesting improvements, see [170, 125]. Zalesky et al. points out the risk of misinterpreting a lower FA error for general misalignment. Schwartz et al. additionally points out that some of the limitations in TBSS lies in the skeletonization process, which in turn is limited by the nonrigid registration, and they suggest using more sophisticated registration algorithms. In relation to this, [87] provided a study of the effect of the choice of target image in the registration of TBSS, and concluded that the best approach was to use a groupwise average atlas (created using `FSL`). Finally, [14] gives a thorough review of TBSS strengths, misconceptions, and a list of recommendation about processing, interpretation and future improvements.

4. Discussion

We end this review by discussing the need for standardized evaluation protocols for groupwise DWI studies, and give a few recommendations based on frequently asked questions.

4.1. A Need for Public Evaluation Protocols and Challenges

We have reviewed a large portion of both groupwise VBR and TBR, and we have covered common approaches to validation, identifying some of the issues that comes from the lack of gold standards. There is a lack of standardized evaluation protocols for new frameworks and algorithms - both in terms of unbiased online evaluation and as segmented downloadable dataset. This is not the same as saying that modelling ODFs and tracing fiber tracts are completely without ground truth datasets. Tractography in particular is rich on community challenges. However, there is a gap when it comes to groupwise validation and evaluating shared features across images. For both VBR and TBR, being able to perform state-the-art multi-shell modelling or complex tractographies is a limited application, if there are no validated ways of comparing the results. Recent challenges in related fields

[24] See Section 3.4 for an introduction to TBSS.

have started to address this problem, such as the Continuous Registration Challenge [34], which offers automatic benchmarking on multiple datasets that include many lung and brain modalities - though at the time of writing not DWI. Below, we briefly discuss the current and future state of VBR and TBR validation.

Challenges in VBR validation

A gold standard validation of VBR is unlikely to come from individual studies reporting different error measures or comparing tuned methods against state-of-the-art on default parameters settings. The future ways of validating DWI, in terms of groupwise alignments and across subjects, is more likely to come from the DWI community in forms of challenges/competitions [105, 48]. These are excellent unbiased ways of collecting and testing the most recent cutting-edge methods in the field. However, DWI challenges have been focused on the individual level. Given a subject or phantom, the objective has been to correct noise or predict missing data in acquisitions [75, 24], model ODFs [74], or evaluate the ability to infer structural connectivity (tractography challenges listed in Section 3.3). There is a current lack of VBR validation challenges[25] for DWI that can be used to pave the way for accessible evaluation of new methods. To this day, new registration methods still make their own sequence of evaluations and with little to no reuse of data from other studies - despite the fact that it is common to use public resources such as the HCP data. As representative examples, we found two recently published methods that represent common situations from articles in the previous sections:

1. A template construction method that rely on their own setup with MSE-based peak signal-to-noise ratio, visual inspection, and orientation discrepancy (OD) between fiber estimates (ODF peaks) [161]. Here a q-space patch-based mean-shift algorithm for constructing diffusion templates with sharper signal profiles (ODFs) was presented and evaluated on synthetic (ODFs with noise) and real data (HCP and DTI).

2. A multi-level FFD registration method by [3] that uses dictionary learning to create sparse DWI representations. Their method is evaluated by random synthetic nonrigid deformations, and on two different cohorts that compare their proposed method to two other registration frameworks with a sequence of quantitative measure including RMS, sKL, angular error, GFA, fiber alignment, etc. and visualization of the error post-registration of random subjects.

These methods have two common but different objectives - template evaluation, and comparison with other registration methods. In both cases, it is clear that the reported results would be difficult for others to compare with. It is also clear that in the future of validating VBR for DWI, state-of-the-art methods would benefit from having one or several of the following evaluations available, e.g. in form of challenges or online ranking resources

- **Scans of real world phantoms** simulating anisotropic structures with deformations or similar in shape. As an example, phantoms were designed and used for fiber tracking in [141] and a link to existing phantoms can be found at the bottom of this page[26].
- **Synthetic but realistic warps** of high quality ground truth DWI. If the goal is to create a realistic warp for validation, the synthetic warp can e.g. be tested by performing fiber tracking. Recent work within this category can be found in [45], though at the time of writing it is used for single-subject noise and distortion correction.
- **Computer-generated or artificial DWI structures**, e.g. crossing or kissing fibers in comparable configurations. These are seen in challenges for noise correction or fiber tracking, but not as multiple constellations that can be compared via nonrigid image registration. With some expertise, examples can be created using tools such as `FiberFox`. [102].
- **Manually segmented ROIs in real world scans** from multiple subjects - specifically intended for VBR of DWI. The current gold standards, such as the HCP, are based on automatic segmentations from `FreeSurfer` or similar. Various approaches to ROI validation was discussed recently in [52].

The best way to avoid overfitted and unbiased results is to keep the validation online, similar to many challenges, along with a public ranking. To circumvent problems with high memory requirements and the data format of resulting ODFs, it is likely best to evaluate results based on the deformation field (i.e. displacement of each voxel). Reorientation of the ODFs is irrelevant to the validation if the correct point-to-point mapping is achieved. As brains are inherently heterogeneous, test cases should also contain non-diffeomorphic examples to test the regularization of the registration, and to evaluate if similar structures can be identified despite regional differences.

Challenges in TBR validation

Similar to the validation of VBR, challenges and competitions are the key to standardized evaluation of new

[25] An overview of previous and upcoming challenges in biomedical image analysis can be found at https://grand-challenge.org/all_challenges.

[26] https://www.nitrc.org/projects/diffusion-data

methods. Here, TBR has an advantage over VBR as the tractography community is already rich on public challenges. However, these are not groupwise challenges, and while the methods for tracing streamlines and performing tractography are beyond the scope of this review, it is worth noting that the challenges in tractography and clustering are focused on individual cases. None of the recent challenges address the gap in quantitative validation of groupwise fiber clustering and registration. The FiberCup challenge (ISMRM) was a quantitative evaluation of 10 tractography algorithms on a single realistic diffusion MR phantom [49]. In the ISMRM 2015 Tractography Challenge, this was extended to a realistic brain-like synthetic model with HCP-like quality, which got 96 distinct submissions from 20 different research groups [96]. In the DTI Challenge (MICCAI) investigated four neurosurgical cases with tumor and edema [120], though not with any intent on groupwise analysis. There is also the Tractography-reproducibility Challenge with Empirical Data (TraCED - ISMRM 2017) with a scan-rescan dataset of HARDI multi-shell scans consisting of 20 datasets in total of the same subject over different sessions and scanners [142]. With reproducibility being a factor, it is closer but still not quite aimed at groupwise analysis. And finally, we have the 3-D Validation of Tractography with Experimental MRI (3D VoTEM) tractography challenge (ISBI 2018) [151], consisting of three very interesting ground truth cases: a phantom reconstruction, a squirrel monkey with tracer injections, and a high resolution ex vivo macaque monkey brain. All well designed challenges that will help the tractography field move forward - just not in particular for inter-subject comparison and validation.

The obstacles in groupwise validation of TBR are similar to those in VBR mentioned above. TBR results often relies even more on not public manual segmentations, and it is still difficult to compare the methods and results of one inter-subject study with another, without the aid of trained experts. However, phantoms and tools for creating simulated fiber tracts are in rapid development, and with the datasets being developed in the challenges above, we are not far from a challenge with multiple similar DWI volumes and the goal of identifying similar or different features in fiber tracts. For instance, it would be interesting to see if outliers and false-positive connections from individual tractography, such as in the ISMRM 2015 challenge, can be identified by congealing information from multiple similar datasets.

4.2. FAQ on lessons learned and recommendations

Here we list our take on a few of the questions that often come up when DWI registration is discussed.

When do I use VBR? It is the better choice if the goal is to study quantitative DWI measures in certain areas, or perform exploratory unbiased studies of populations. Care should be taken with low spatial resolution DWI, though compartment models have improved ODFs formed from partial volumes.

When do I use TBR? Excluding the obvious cases, where to goal is to study connectivity or specific local fiber bundles, TBR is the better choice if a clear hypothesis exists which will help define the minimum length of streamlines, the clustering scale, and other parameters. However, while TBR requires more computational power, it is becoming exploratory as well.

Can I rely on models applied to warped DWI data? The main attraction of evaluating DWI data in a warped space is to add information that can help resolve complex ODFs during tractography in poor quality data. However, we recommend using VBR for the spatial mapping, which can be subsequently used to transfer tracts from a good quality tractography, or a clustered average of streamlines, as we saw with unbiased TBR methods. Alternatively, we recommend [1] for a study that evaluates deterministic fiber tracking in non-linearly warped DTI against native DTI.

Should I use scalar-based registration for DWI? Scalar-based registration is quick, efficient, and well-documented outside DWI. It is easy to use with tools like `FSL` that can apply the deformation to DWI data, and it likely the better choice for low resolution or very noisy clinical data. However, for most quantitative DWI studies, there is a lot of evidence that points to a better performance by frameworks supporting implicit or explicit reorientation of tensors and ODFs.

Is multi-atlas better than nonrigid registration? Registration is the glue combining multi-model images, correction distortions, and for matching variable structures of a similar origin/shape. However, with increasing computational power and what seems to be a renaissance of machine learning, it is more than likely there will be an increased focus on multi-atlases and feature banks of observed fiber configurations, though defining sparse feature representations still poses a significant challenge in high dimensional and memory intensive medical data. Building hierarchical atlases of subgroups of populations (e.g. different stages of Alzheimer's) might be the next step, where registration is more poly-affine than nonrigid.

What can I do to aid community validation efforts? Share the simulated experiments and synthetically

warped data, and name the HCP (or similar) subjects used. Until new challenges are defined, the challenge for computation imaging researchers often lies in comparing new results with the results of existing studies. The field still suffers from an abundance of visual validation and lack of comparative quantitative results.

Why have DSI not been reviewed? For the same reason that we have not covered multi-shell HARDI acquisitions. To narrow the review a bit, we have focused on the most common clinical or near-clinical methods, and while DSI provide a much more high quality image, it is often a method for basic research with longer acquisitions times.

Appendix A. Popular Registration Frameworks

We briefly summarize some of the most popular registration frameworks in the context of DWI registration. The frameworks are all publicly available, frequently used and support both rigid, affine, and nonrigid registration. They are not all made for DWI registration but selected on the basis that we often see them used when doing DWI registration and spatially normalizing populations in group studies. Again, when we define methods as scalar-based it also includes algorithms that do not directly optimize over reorientation as part of cost function, while full tensor based algorithms do integrate reorientation of the ODF. The following methods described are: `DTI-TK`, `SyN`, `MedINRIA`, `FreeSurfer`, `FSL`, `MRtrix` - with an additional paragraph on recent and promising frameworks.

DTI-TK (full tensor)

The `DTI-TK` remains on of the best and most used registration tool for diffusion imaging. From the homepage: "`DTI-TK` is a spatial normalization and atlas construction toolkit optimized for examining white matter morphometry using DTI data". It is diffeomorphic deformable tensor registration. It was created as part of Gary Zhang's PhD Thesis and the methods were presented throughout [174] (nonrigid, piece-wise affine analytical derivatives for tensor reorientation), [173] (same but more in-depth and with subject registration and fiber-tracking), and [171] (larger study tested against two other scalar-based FA registration approaches). The initial main contribution was incorporating a similarity measure between diffusion profiles into the cost function and analytically optimizing the registration over this in a nonrigid computationally feasible manner (much like contribution of this dissertation). In their framework, the tensor reorientation is based on the FS approach (as PPD is argued to be too expensive and not available on closed-form) combined with Polar Decomposition-based affine parametrization (as opposed to eigenvectors). The method is available for both single-tensor DTI and higher-order information (such as the full tensor profile through spherical harmonics). It is compared with a more recent frameworks such as `DTI-DROID` in [72] which incorporates additional information such as neighbourhood interpolation and various tensor-based quantitative features along with edge detection. Additionally in [154] who combines tract and tensor statistics, and in [55] who claims approximately similar performance but six times faster computation by using discrete cosine transforms and extending part of SPM (Statistical Parameter Mapping) toolkit to tensor reorientation. `DTI-TK` is still being used directly or as a comparative state-of-the-art standard to this day in prominent studies such as [31].

SyN (scalar-based)

`SyN` (Symmetric Image Normalization) was developed by [12] and is a symmetric diffeomorphic image normalization method for maximizing the cross-correlation within the space of diffeomorphic maps. It is part of the popular `ANTs` (Advanced Normalization Tools) package [13] and uses information theoretic similarity measures (an advantage with noisy DWI data). It transforms scalar images by optimizing and integrating a time-varying velocity field, and was shown to be among the most accurate intensity-based registration methods among 14 others in [88]. Its popularity and nice symmetric formulation has led to it being used in many DWI studies with scalar-based registration, and while it does not include ODF reorientation, Avants el al. themselves extended it to `SyNMN` (SyN MultiVariate) [11]. Here it was expanded to include a similar explicit optimization over the full tensor as provided by Zhang et al. in `DTI-TK` (note that both `SyN` and `DT-ITK` is integrated in the Insight and Registration Toolkit (`ITK`) toolkit which supports tensor reorientation). As a more recent prominent example, `SyN` was used in [123] for registering higher order Fiber Orientation Distributions (FODs) modelled by spherical harmonics.

MedINRIA (full tensor)

`MedINRIA` is a highly used diffeomorphic medical image analysis and visualization toolbox with a tensor-based extension by [168] (originally introduced in [167]) that uses the FS gradient into a diffeomorphic DTI registration scheme. This, `DT-REFinD` (Diffusion Tensor Registration with Exact Finite-strain Differential), is a fast diffeomorphic nonrigid Demons algorithm (not far removed from the approach of `DTI-TK`) where the exact gradients of the objective function with the tensor reorientation in-

corporated in an analytical optimization. The Demons approach was described in Section 2.1. With the extension to DTI and calculation of exact gradients, it becomes computationally heavy on large diffusion images but the authors argue for this with improvements in the registration. Additionally, convergence is improved as the exact gradients allows for the use of a Gauss-Newton method for optimization (which is argued to bring it down to scalar image computation time). The algorithm in `MedINRIA` was compared with `DTI-TK` in [155] (along with six other DTI registration algorithms) and performed almost as well at a second-best, and in [177] where the two are close and slightly outperforms each other in different scenarios.

FreeSurfer (scalar- and surface-based)

`FreeSurfer` [50] deserves a mention as part of it is generally involved in most larger DWI studies. It (as anyone in fields related to MRI would be aware of) takes the price for being the most extensive and used framework for image analysis of brain scans. The `FreeSurfer` pipeline is extensive, starting at preprocessing of raw noisy scans and all the way to atlas and surface models (parcellation) and multimodal analysis. When it comes to registration `FreeSurfer` is scalar-based and surface-based registration structural (often T_1) scans to a common template space. It is made for structural and functional MRI but not in particular for DWI registration. On the other hand it is very often used together with popular DWI registration algorithms as the subcortical parcellation provides more information to the cost function, or to evaluate the registration in an ROI manner, or visualize the results. It does play a role in DWI preprocessing and tractography with its sophisticated tool `TRACULA` (TRActs Constrained by UnderLying Anatomy) [166] that also offers longitudinal analysis (i.e. rigid intra-subject registration).

FSL (scalar-based)

`FSL` [77] also deserves to mentioned as it might be the most popular framework for linear and nonlinear registration of medical images, being more light-weight and more registration-specific than e.g. `FreeSurfer`. It is primarily scalar-based registration with the main tools being `FLIRT` (linear registration) and `FNIRT` (nonlinear registration). `FLIRT` works well on different image modalities and intensity variations where a nice array of cost functions are available. It is a generally accepted tool for global intra-subject (rigid) registration of structural scans (T_1-weighted scans, etc.) to DWI space (often b_0 or MD), and also for an initial global alignment prior to inter-subject nonlinear registration [88]. `FNIRT` on the other hand only provides SSD as a similarity measure and might not be optimal for scans with significant intensity variations (as can be common for DWI). Thus, it is commonly used on structural MRI scans and FA maps, and together with another scheme for DWI reorientation. `FSL` does come with tensor reorientation (for both `FLIRT` and `FNIRT`) in its `vecreg` tool based on the scalar transformation though it is not clear what transformation is used (digging into the C++ code it seems to be Alexander's PPD approach) and it is not a broadly used tool.

Like `FreeSurfer`, `FSL` comes with tools for noise correction of DWI scans, and with tools for modelling ODFs and performing tractography. `DTIFIT` creates first-order tensors of the DWI data and quantitative measures, and QBI tools like `qboot` also exist. `BEDPOSTX` (Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques, X is for crossing fibers) is a more advanced (but also time consuming) tool by [17] for modelling higher order ODFs to solve crossing fibers and perform probabilistic tractography (and even with multi-shell options by [76]). It works with `PROBTRACKX` for probabilistic streamlines and more.

MRtrix (scalar-based)

`MRtrix` (specifically `MRtrix3`) is a more recent suite of tools similar to `FreeSurfer` providing preprocessing, image analysis, and visualization while being focused on DWI and white matter analysis (i.e. tractography and fibre analysis). Developed by [138] the ODFs are modelled by constrained spherical deconvolution (CSD) and the registration is performed by scalar maps like FA but, it should be noted, with an iterative reorientation of the Fibre Orientation Distribution (FOD) [123]. Despite its youth, `MRtrix` is already highly popular and used in various recent studies (mainly for tractography as it allows the use of precomputed warps in the registration). Different from `FreeSurfer`, it is less specific to the human brain and used in multiple studies related to tractography in animals.

DR-TAMAS (full tensor)

`DR-TAMAS` (Diffeomorphic Registration for Tensor Accurate alignMent of Anatomical Structures) by [73] is included as an example of a more recent framework that extends upon popular frameworks. It gives the option of using a transformation model of `SyN` from the `ANTs` package, or to use a time-varying velocity-based model (both in `ITK`). It is based on the exact FS in `MedINRIA` for the differentials but optimizes on speed and parallelism. It is optimized for brain registration and incorporates structural data in the registration for better alignment of isotropic areas (CSF, GM) in a combined metric similarity with DTI. Additionally, they ([73]) gives a refreshingly honest and extensive discussion on the use of various similarity measures for evaluation (and more). They argue that their method would be great for generating whole-

brain atlases but does not argue to be better than methods like DTI-TK optimized for anisotropic (i.e. WM) regions. The algorithm is part of the TORTOISE pipeline for diffusion MRI, originally introduced by [118].

References

[1] Adluru, N., Destiche, D. J., Tromp, D. P., Davidson, R. J., Zhang, H., and Alexander, A. L. (2016). Evaluating consistency of deterministic streamline tractography in non-linearly warped dti data. *arXiv preprint arXiv:1602.02117*.

[2] Afzali, M., Fatemizadeh, E., and Soltanian-Zadeh, H. (2013). High angular resolution diffusion image registration. In *Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on*, pages 232–236. IEEE.

[3] Afzali, M., Fatemizadeh, E., and Soltanian-Zadeh, H. (2017). Sparse registration of diffusion weighted images. *Computer methods and programs in biomedicine*, 151:33–43.

[4] Alexander, D. C., Pierpaoli, C., Basser, P. J., and Gee, J. C. (2001). Spatial transformations of diffusion tensor magnetic resonance images. *IEEE transactions on medical imaging*, 20(11):1131–1139.

[5] Anderson, A. W. (2005). Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 54(5):1194–1206.

[6] Arkesteijn, G., Poot, D., Niestijl, M., Vernooij, M., Niessen, W., van Vliet, L., and Vos, F. (2017). Longitudinal analysis of diffusion-weighted mri with a ball-and-sticks model. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 783–786. IEEE.

[7] Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., and Parisot, S. (2017). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage*.

[8] Assaf, Y. and Basser, P. J. (2005). Composite hindered and restricted model of diffusion (charmed) mr imaging of the human brain. *Neuroimage*, 27(1):48–58.

[9] Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., and Basser, P. J. (2008). Axcaliber: a method for measuring axon diameter distribution from diffusion mri. *Magnetic resonance in medicine*, 59(6):1347–1354.

[10] Assaf, Y., Johansen-Berg, H., and de Schotten, M. T. (2017). The role of diffusion mri in neuroscience. *bioRxiv*, page 140459.

[11] Avants, B., Duda, J. T., Kim, J., Zhang, H., Pluta, J., Gee, J. C., and Whyte, J. (2008a). Multivariate analysis of structural and diffusion imaging in traumatic brain injury. *Academic radiology*, 15(11):1360–1375.

[12] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008b). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41.

[13] Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ants). *Insight j*, 2:1–35.

[14] Bach, M., Laun, F. B., Leemans, A., Tax, C. M., Biessels, G. J., Stieltjes, B., and Maier-Hein, K. H. (2014). Methodological considerations on tract-based spatial statistics (tbss). *Neuroimage*, 100:358–369.

[15] Banerjee, M., Okun, M. S., Vaillancourt, D. E., and Vemuri, B. C. (2016). A method for automated classification of parkinson's disease diagnosis using an ensemble average propagator template brain map estimated from diffusion mri. *PloS one*, 11(6):e0155764.

[16] Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Sheehan, K., Ciccarelli, O., et al. (2003a). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*, 6(7):750–757.

[17] Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155.

[18] Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003b). Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magnetic resonance in medicine*, 50(5):1077–1088.

[19] Bloy, L., Ingalhalikar, M., Eavani, H., Schultz, R. T., Roberts, T. P., and Verma, R. (2012). White matter atlas generation using hardi based automated parcellation. *NeuroImage*, 59(4):4055–4063.

[20] Bloy, L. and Verma, R. (2010). Demons registration of high angular resolution diffusion images. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 1013–1016. IEEE.

[21] Booth, B. G. and Hamarneh, G. (2015). Diffusion mri for brain connectivity mapping and analysis. *MRI: Physics, Image Reconstruction, and Analysis*, pages 137–171.

[22] Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J. P., Sporns, O., Do, K. Q., Maeder, P., Meuli, R., and Hagmann, P. (2012). Mapping the human connectome at multiple scales with diffusion spectrum mri. *Journal of neuroscience methods*, 203(2):386–397.

[23] Caruyer, E., Daducci, A., Descoteaux, M., Houde, J.-C., Thiran, J.-P., and Verma, R. (2014). Phantomas: a flexible software library to simulate diffusion mr phantoms. In *ISMRM*.

[24] CDMRI, . (2017). Diffusion mri data harmonisation. https://projects.iq.harvard.edu/cdmri2017. Accessed: 2017-12-10.

[25] Chekir, A., Hassas, S., Descoteaux, M., Côté, M., Garyfallidis, E., and Oulebsir-Boumghar, F. (2017). 3d-ssf: A bio-inspired approach for dynamic multi-subject clustering of white matter tracts. *Computers in Biology and Medicine*.

[26] Chen, G., Zhang, P., Li, K., Wee, C.-Y., Wu, Y., Shen, D., and Yap, P.-T. (2016). Improving estimation of fiber orientations in diffusion mri using inter-subject information sharing. *Scientific reports*, 6:37847.

[27] Chiang, M.-C., Leow, A. D., Klunder, A. D., Dutton, R. A., Barysheva, M., Rose, S. E., McMahon, K. L., De Zubicaray, G. I., Toga, A. W., and Thompson, P. M. (2008). Fluid registration of diffusion tensor images using information theory. *IEEE transactions on medical imaging*, 27(4):442–456.

[28] Clarkson, M. J., Malone, I. B., Modat, M., Leung, K. K., Ryan, N., Alexander, D. C., Fox, N. C., and Ourselin, S. (2010). A framework for using diffusion weighted imaging to improve cortical parcellation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 534–541. Springer.

[29] Close, T. G., Tournier, J.-D., Calamante, F., Johnston, L. A., Mareels, I., and Connelly, A. (2009). A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. *NeuroImage*, 47(4):1288–1300.

[30] Cloutman, L. L. and Ralph, M. A. L. (2012). Connectivity-based structural and functional parcellation of the human cortex using diffusion imaging and tractography. *Frontiers in neuroanatomy*, 6.

[31] Colgan, N., Siow, B., O'Callaghan, J. M., Harrison, I. F., Wells, J. A., Holmes, H. E., Ismail, O., Richardson, S., Alexander, D. C., Collins, E. C., et al. (2016). Application of neurite orientation dispersion and density imaging (noddi) to a tau pathology model of alzheimer's disease. *NeuroImage*, 125:739–744.

[32] Commowick, O., Hedouin, R., Caruyer, E., and Barillot, C. (2017). ℓ_2 similarity metrics for diffusion multi-compartment model images registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 257–265. Springer.

[33] Côté, M.-A., Girard, G., Boré, A., Garyfallidis, E., Houde, J.-C., and Descoteaux, M. (2013). Tractometer: towards validation of tractography pipelines. *Medical image analysis*, 17(7):844–857.

[34] CRC, W. . (2018). Continuous registration challenge. https://continuousregistration.grand-challenge.org. Accessed: 2018-03-09.

[35] de Groot, M., Vernooij, M. W., Klein, S., Ikram, M. A., Vos, F. M., Smith, S. M., Niessen, W. J., and Andersson, J. L. (2013). Improving alignment in tract-based spatial statistics: evaluation and optimization of image registration. *Neuroimage*, 76:400–411.

[36] Descoteaux, M. (2015). High angular resolution diffusion imaging (hardi). *Wiley Encyclopedia of Electrical and Electronics Engineering*.

[37] Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R.

(2007). Regularized, fast, and robust analytical q-ball imaging. *Magnetic resonance in medicine*, 58(3):497–510.

[38] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980.

[39] Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15.

[40] Dhollander, T., Van Hecke, W., Maes, F., Sunaert, S., and Suetens, P. (2010). Spatial transformations of high angular resolution diffusion imaging data in q-space. In *Online Proceedings at http://cmic. cs. ucl. ac. uk/cdmri10/*, pages 73–83.

[41] Du, J., Goh, A., and Qiu, A. (2012). Diffeomorphic metric mapping of high angular resolution diffusion imaging based on riemannian structure of orientation distribution functions. *IEEE Transactions on Medical Imaging*, 31(5):1021–1033.

[42] Du, J., Goh, A., and Qiu, A. (2013). Bayesian estimation of white matter atlas from high angular resolution diffusion imaging. *arXiv preprint arXiv:1310.3233*.

[43] Du, J., Hosseinbor, A. P., Chung, M. K., Bendlin, B. B., Suryawanshi, G., Alexander, A. L., and Qiu, A. (2014). Diffeomorphic metric mapping and probabilistic atlas generation of hybrid diffusion imaging based on bfor signal basis. *Medical image analysis*, 18(7):1002–1014.

[44] Duarte-Carvajalino, J. M., Sapiro, G., Harel, N., and Lenglet, C. (2013). A framework for linear and non-linear registration of diffusion-weighted mris using angular interpolation. *Frontiers in neuroscience*, 7.

[45] Esteban, O., Zosso, D., Daducci, A., Bach-Cuadra, M., Ledesma-Carbayo, M. J., Thiran, J.-P., and Santos, A. (2016). Data on the verification and validation of segmentation and registration methods for diffusion mri. *Data in brief*, 8:871–876.

[46] Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., et al. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral Cortex*, 26(8):3508–3526.

[47] Farquharson, S. and Tournier, J.-D. (2016). High angular resolution diffusion imaging. In *Diffusion Tensor Imaging*, pages 383–406. Springer.

[48] Ferizi, U., Scherrer, B., Schneider, T., Alipoor, M., Eufracio, O., Fick, R. H., Deriche, R., Nilsson, M., Loya-Olivas, A. K., Rivera, M., et al. (2017). Diffusion mri microstructure models with in vivo human brain connectome data: results from a multi-group comparison. *NMR in Biomedicine*, 30(9).

[49] Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., et al. (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion mr phantom. *Neuroimage*, 56(1):220–234.

[50] Fischl, B. (2012). Freesurfer. *Neuroimage*, 62(2):774–781.

[51] Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22.

[52] Froeling, M., Pullens, P., and Leemans, A. (2016). Dti analysis methods: Region of interest analysis. In *Diffusion Tensor Imaging*, pages 175–182. Springer.

[53] Froeling, M., Tax, C. M., Vos, S. B., Luijten, P. R., and Leemans, A. (2017). "massive" brain dataset: Multiple acquisitions for standardization of structural imaging validation and evaluation. *Magnetic resonance in medicine*, 77(5):1797–1809.

[54] Gallardo, G., Fick, R., Wells, W., Deriche, R., and Wassermann, D. (2016). Groupwise structural parcellation of the cortex: A sound approach based on logistic models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 99–112. Springer.

[55] Gan, L. and Agam, G. (2011). Efficient nonlinear dti registration using dct basis functions. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 17–22. IEEE.

[56] Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I., and Contributors, D. (2014). Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8.

[57] Garyfallidis, E., Brett, M., Correia, M. M., Williams, G. B., and Nimmo-Smith, I. (2012). Quickbundles, a method for tractography simplification. *Frontiers in neuroscience*, 6.

[58] Garyfallidis, E., Côté, M.-A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., and Descoteaux, M. (2017). Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage*.

[59] Garyfallidis, E., Ocegueda, O., Wassermann, D., and Descoteaux, M. (2015). Robust and efficient linear registration of white-matter fascicles in the space of streamlines. *NeuroImage*, 117:124–140.

[60] Gee, J. C. and Alexander, D. C. (2006). Diffusion-tensor image registration. *Visualization and processing of tensor fields*, pages 327–342.

[61] Geng, X., Ross, T. J., Gu, H., Shin, W., Zhan, W., Chao, Y.-P., Lin, C.-P., Schuff, N., and Yang, Y. (2011). Diffeomorphic image registration of diffusion mri using spherical harmonics. *IEEE transactions on medical imaging*, 30(3):747–758.

[62] Ghosh, A. and Deriche, R. (2015). A survey of current trends in diffusion mri for structural brain connectivity. *Journal of neural engineering*, 13(1):011001.

[63] Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178.

[64] Golkov, V., Dosovitskiy, A., Sperl, J. I., Menzel, M. I., Czisch, M., Sämann, P., Brox, T., and Cremers, D. (2016). q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE transactions on medical imaging*, 35(5):1344–1351.

[65] Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., and Beaulieu, C. (2008). Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral cortex*, 19(3):524–536.

[66] Guimond, A., Guttmann, C. R., Warfield, S. K., and Westin, C.-F. (2002). Deformable registration of dt-mri data based on transformation invariant tensor characteristics. In *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*, pages 761–764. IEEE.

[67] Guo, Z., Wang, Y., Lei, T., Fan, Y., and Zhang, X. (2016). Dti image registration under probabilistic fiber bundles tractography learning. *BioMed research international*, 2016.

[68] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159.

[69] Harms, R., Fritz, F., Tobisch, A., Goebel, R., and Roebroeck, A. (2017). Robust and fast nonlinear optimization of diffusion mri microstructure models. *NeuroImage*, 155:82–96.

[70] Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040.

[71] Hong, X., Arlinghaus, L. R., and Anderson, A. W. (2009). Spatial normalization of the fiber orientation distribution based on high angular resolution diffusion imaging data. *Magnetic resonance in medicine*, 61(6):1520–1527.

[72] Ingalhalikar, M., Yang, J., Davatzikos, C., and Verma, R. (2010). Dti-droid: diffusion tensor imaging-deformable registration using orientation and intensity descriptors. *International Journal of Imaging Systems and Technology*, 20(2):99–107.

[73] Irfanoglu, M. O., Nayak, A., Jenkins, J., Hutchinson, E. B., Sadeghi, N., Thomas, C. P., and Pierpaoli, C. (2016). Dr-tamas: Diffeomorphic registration for tensor accurate alignment of anatomical structures. *NeuroImage*, 132:439–454.

[74] ISBI, . (2013). Hardi reconstruction challenge. http://hardi.

epfl.ch/static/events/2013_ISBI/index.html. Accessed: 2017-12-10.

[75] ISBI, . (2015). White matter modelling challenge. http://cmic.cs.ucl.ac.uk/wmmchallenge/. Accessed: 2017-12-10.

[76] Jbabdi, S., Sotiropoulos, S. N., Savio, A. M., Graña, M., and Behrens, T. E. (2012). Model-based analysis of multishell diffusion mr data for tractography: How to get over fitting problems. *Magnetic Resonance in Medicine*, 68(6):1846–1855.

[77] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

[78] Jensen, H. G., Lauze, F., Nielsen, M., and Darkner, S. (2015). Locally orderless registration for diffusion weighted images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 305–312. Springer.

[79] Jeurissen, B., Descoteaux, M., Mori, S., and Leemans, A. (2017). Diffusion mri fiber tractography of the brain. *NMR in Biomedicine*.

[80] Jin, Y., Shi, Y., Zhan, L., Gutman, B. A., de Zubicaray, G. I., McMahon, K. L., Wright, M. J., Toga, A. W., and Thompson, P. M. (2014). Automatic clustering of white matter fibers in brain diffusion mri with an application to genetics. *NeuroImage*, 100:75–90.

[81] Johansen-Berg, H. and Behrens, T. E. (2006). Just pretty pictures? what diffusion tractography can add in clinical neuroscience. *Current opinion in neurology*, 19(4):379.

[82] Johansen-Berg, H. and Behrens, T. E. (2013). *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press.

[83] Jones, D. K. (2010). Challenges and limitations of quantifying brain connectivity in vivo with diffusion mri. *Imaging in Medicine*, 2(3):341.

[84] Jones, D. K. and Cercignani, M. (2010). Twenty-five pitfalls in the analysis of diffusion mri data. *NMR in Biomedicine*, 23(7):803–820.

[85] Jones, D. K., Griffin, L. D., Alexander, D. C., Catani, M., Horsfield, M. A., Howard, R., and Williams, S. C. (2002). Spatial normalization and averaging of diffusion tensor mri data sets. *Neuroimage*, 17(2):592–617.

[86] Jones, D. K., Knösche, T. R., and Turner, R. (2013). White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion mri. *Neuroimage*, 73:239–254.

[87] Keihaninejad, S., Ryan, N. S., Malone, I. B., Modat, M., Cash, D., Ridgway, G. R., Zhang, H., Fox, N. C., and Ourselin, S. (2012). The importance of group-wise registration in tract based spatial statistics study of neurodegeneration: a simulation study in alzheimer's disease. *PloS one*, 7(11):e45996.

[88] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802.

[89] Klein, A., Ghosh, S. S., Avants, B., Yeo, B. T., Fischl, B., Ardekani, B., Gee, J. C., Mann, J. J., and Parsey, R. V. (2010). Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage*, 51(1):214–220.

[90] Klein, A. and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience*, 6.

[91] Kumar, K., Desrosiers, C., Siddiqi, K., Colliot, O., and Toews, M. (2017). Fiberprint: A subject fingerprint based on sparse code pooling for white matter fiber analysis. *NeuroImage*, 158:242–259.

[92] Lefranc, S., Roca, P., Perrot, M., Poupon, C., Coulon, O., Le Bihan, D., Hertz-Pannier, L., Mangin, J.-F., and Rivière, D. (2014). Validation of consistent inter-subject connectivity-based parcellation. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 923–926. IEEE.

[93] Lefranc, S., Roca, P., Perrot, M., Poupon, C., Le Bihan, D., Mangin, J.-F., and Rivière, D. (2016). Groupwise connectivity-based parcellation of the whole human cortical surface using watershed-driven dimension reduction. *Medical image analysis*, 30:11–29.

[94] Li, H., Fan, L., Zhuo, J., Wang, J., Zhang, Y., Yang, Z., and Jiang, T. (2017). Atpp: A pipeline for automatic tractography-based brain parcellation. *Frontiers in neuroinformatics*, 11.

[95] Ly, M. T., Nanavati, T. U., Frum, C. A., and Pergami, P. (2015). Comparing tract-based spatial statistics and manual region-of-interest labeling as diffusion analysis methods to detect white matter abnormalities in infants with hypoxic-ischemic encephalopathy. *Journal of Magnetic Resonance Imaging*, 42(6):1689–1697.

[96] Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., et al. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature communications*, 8(1):1349.

[97] Mangin, J.-F., Lebenberg, J., Lefranc, S., Labra, N., Auzias, G., Labit, M., Guevara, M., Mohlberg, H., Roca, P., Guevara, P., et al. (2016). Spatial normalization of brain images and beyond.

[98] Mayer, A. and Greenspan, H. (2008). Bundles of interest based registration of white matter tractographies. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 919–922. IEEE.

[99] Mayer, A., Zimmerman-Moreno, G., Shadmi, R., Batikoff, A., and Greenspan, H. (2011). A supervised framework for the registration and segmentation of white matter fiber tracts. *IEEE Transactions on medical imaging*, 30(1):131–145.

[100] Moreno-Dominguez, D., Anwander, A., and Knösche, T. R. (2014). A hierarchical method for whole-brain connectivity-based parcellation. *Human brain mapping*, 35(10):5000–5025.

[101] Mueller, B. A., Lim, K. O., Hemmy, L., and Camchong, J. (2015). Diffusion mri and its role in neuropsychology. *Neuropsychology review*, 25(3):250–271.

[102] Neher, P. F., Laun, F. B., Stieltjes, B., and Maier-Hein, K. H. (2014). Fiberfox: facilitating the creation of realistic white matter software phantoms. *Magnetic resonance in medicine*, 72(5):1460–1470.

[103] Nooner, K. B., Colcombe, S. J., Tobe, R. H., Mennes, M., Benedict, M. M., Moreno, A. L., Panek, L. J., Brown, S., Zavitz, S. T., Li, Q., et al. (2012). The nki-rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Frontiers in neuroscience*, 6.

[104] Norton, I., Essayed, W. I., Zhang, F., Pujol, S., Yarmarkovich, A., Golby, A. J., Kindlmann, G., Wasserman, D., Estepar, R. S. J., Rathi, Y., et al. (2017). Slicerdmri: open source diffusion mri software for brain cancer research. *Cancer research*, 77(21):e101–e103.

[105] O'Donnell, L. J., Daducci, A., Wassermann, D., and Lenglet, C. (2017). Advances in computational and statistical diffusion mri. *NMR in Biomedicine*.

[106] O'Donnell, L. J., Golby, A. J., and Westin, C.-F. (2013). Fiber clustering versus the parcellation-based connectome. *NeuroImage*, 80:283–289.

[107] O'Donnell, L. J. and Pasternak, O. (2015). Does diffusion mri tell us anything about the white matter? an overview of methods and pitfalls. *Schizophrenia research*, 161(1):133–141.

[108] O'Donnell, L. J. and Westin, C.-F. (2007). Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE transactions on medical imaging*, 26(11):1562–1575.

[109] O'Donnell, L. J., Westin, C.-F., and Golby, A. J. (2009). Tract-based morphometry for white matter group analysis. *Neuroimage*, 45(3):832–844.

[110] Ou, Y., Sotiras, A., Paragios, N., and Davatzikos, C. (2011). Dramms: Deformable registration via attribute matching and mutual-saliency weighting. *Medical image analysis*, 15(4):622–639.

[111] O'Donnell, L. J., Suter, Y., Rigolo, L., Kahali, P., Zhang, F., Norton, I., Albi, A., Olubiyi, O., Meola, A., Essayed, W. I., et al. (2017). Automated white matter fiber tract identification in patients with brain tumors. *NeuroImage: Clinical*, 13:138–153.

[112] O'Donnell, L. J., Wells, W. M., Golby, A. J., and Westin, C.-F. (2012). Unbiased groupwise registration of white matter tractography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 123–130. Springer.

[113] Parisot, S., Arslan, S., Passerat-Palmbach, J., Wells, W. M., and Rueckert, D. (2016). Group-wise parcellation of the cortex through multi-scale spectral clustering. *NeuroImage*, 136:68–83.

[114] Parisot, S., Arslan, S., Passerat-Palmbach, J., Wells III, W. M., and Rueckert, D. (2015). Tractography-driven groupwise multi-scale parcellation of the cortex. In *International Conference on Information*

Processing in Medical Imaging, pages 600–612. Springer.

[115] Park, H.-J., Kubicki, M., Shenton, M. E., Guimond, A., McCarley, R. W., Maier, S. E., Kikinis, R., Jolesz, F. A., and Westin, C.-F. (2003). Spatial normalization of diffusion tensor mri using multiple channels. *Neuroimage*, 20(4):1995–2009.

[116] Passingham, R. E., Stephan, K. E., and Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nature Reviews Neuroscience*, 3(8):606–616.

[117] Perrone, D., Jeurissen, B., Aelterman, J., Roine, T., Sijbers, J., Pizurica, A., Leemans, A., and Philips, W. (2016). D-brain: Anatomically accurate simulated diffusion mri brain data. *PloS one*, 11(3):e0149778.

[118] Pierpaoli, C., Walker, L., Irfanoglu, M., Barnett, A., Basser, P., Chang, L., Koay, C., Pajevic, S., Rohde, G., Sarlls, J., et al. (2010). Tortoise: an integrated software package for processing of diffusion mri data. In *ISMRM 18th annual meeting, Stockholm, Sweden*, page 1597.

[119] Postelnicu, G., Zollei, L., and Fischl, B. (2009). Combined volumetric and surface registration. *IEEE transactions on medical imaging*, 28(4):508–522.

[120] Pujol, S., Wells, W., Pierpaoli, C., Brun, C., Gee, J., Cheng, G., Vemuri, B., Commowick, O., Prima, S., Stamm, A., et al. (2015). The dti challenge: toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery. *Journal of Neuroimaging*, 25(6):875–882.

[121] Raffelt, D., Tournier, J.-d., Crozier, S., Connelly, A., Salvado, O., et al. (2012). Reorientation of fiber orientation distributions using apodized point spread functions. *Magnetic resonance in medicine*, 67(3):844–855.

[122] Raffelt, D., Tournier, J.-D., Fripp, J., Crozier, S., Connelly, A., and Salvado, O. (2009). Non-linear spatial normalisation of high angular resolution diffusion imaging data using fiber orientation distributions. *Diffusion Modelling and the Fibre Cup, MICCAI*.

[123] Raffelt, D., Tournier, J.-D., Fripp, J., Crozier, S., Connelly, A., and Salvado, O. (2011). Symmetric diffeomorphic registration of fibre orientation distributions. *NeuroImage*, 56(3):1171–1180.

[124] Scherrer, B., Schwartzman, A., Taquet, M., Prabhu, S. P., Sahin, M., Akhondi-Asl, A., and Warfield, S. K. (2013). Characterizing the distribution of anisotropic micro-structural environments with diffusion-weighted imaging (diamond). In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 518–526. Springer.

[125] Schwarz, C. G., Reid, R. I., Gunter, J. L., Senjem, M. L., Przybelski, S. A., Zuk, S. M., Whitwell, J. L., Vemuri, P., Josephs, K. A., Kantarci, K., et al. (2014). Improved dti registration allows voxel-based analysis that outperforms tract-based spatial statistics. *Neuroimage*, 94:65–78.

[126] Shen, D. and Davatzikos, C. (2002). Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11):1421–1439.

[127] Siless, V., Chang, K., Fischl, B., and Yendiki, A. (2016). Hierarchical clustering of tractography streamlines based on anatomical similarity. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 184–191. Springer.

[128] Siless, V., Chang, K., Fischl, B., and Yendiki, A. (2018). Anatomicuts: Hierarchical clustering of tractography streamlines based on anatomical similarity. *NeuroImage*, 166:32–45.

[129] Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., et al. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505.

[130] Smith, S. M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T. E., Miller, K. L., Robson, M. D., Jones, D. K., Klein, J. C., Bartsch, A. J., et al. (2007). Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nature protocols*, 2(3):499.

[131] Smith, S. M., Kindlmanny, G., and Jbabdi, S. (2013). Cross-subject comparison of local diffusion mri parameters. In *Diffusion MRI, Second ed.*, pages 209–239. Academic Press, London.

[132] Soares, J. M., Marques, P., Alves, V., and Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in neuroscience*, 7.

[133] Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779.

[134] Sweet, A. and Pennec, X. (2010). Log-domain diffeomorphic registration of diffusion tensor images. In *International Workshop on Biomedical Image Registration*, pages 198–209. Springer.

[135] Taquet, M., Scherrer, B., Commowick, O., Peters, J., Sahin, M., Macq, B., and Warfield, S. K. (2012). Registration and analysis of white matter group differences with a multi-fiber model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 313–320. Springer.

[136] Taquet, M., Scherrer, B., Commowick, O., Peters, J. M., Sahin, M., Macq, B., and Warfield, S. K. (2014). A mathematical framework for the registration and analysis of multi-fascicle models for population studies of the brain microstructure. *IEEE transactions on medical imaging*, 33(2):504–517.

[137] Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with maxwell's demons. *Medical image analysis*, 2(3):243–260.

[138] Tournier, J., Calamante, F., Connelly, A., et al. (2012). Mrtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology*, 22(1):53–66.

[139] Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*, 35(4):1459–1472.

[140] Tournier, J.-D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185.

[141] Tournier, J.-D., Yeh, C.-H., Calamante, F., Cho, K.-H., Connelly, A., and Lin, C.-P. (2008). Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage*, 42(2):617–625.

[142] TraCED, I. . (2017). Tractography-reproducibility challenge with empirical data. https://my.vanderbilt.edu/ismrmtraced2017. Accessed: 2018-01-12.

[143] Trebeschi, S., van Griethuysen, J. J., Lambregts, D. M., Lahaye, M. J., Parmer, C., Bakers, F. C., Peters, N. H., Beets-Tan, R. G., and Aerts, H. J. (2017). Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric mr. *Scientific Reports*, 7.

[144] Tuch, D. S. et al. (2002). *Diffusion MRI of complex tissue structure*. PhD thesis, Massachusetts Institute of Technology.

[145] Tunç, B., Ingalhalikar, M., Parker, D., Lecoeur, J., Singh, N., Wolf, R. L., Macyszyn, L., Brem, S., and Verma, R. (2015). Individualized map of white matter pathways: connectivity-based paradigm for neurosurgical planning. *Neurosurgery*, 79(4):568–577.

[146] Tunç, B., Parker, W. A., Ingalhalikar, M., and Verma, R. (2014). Automated tract extraction via atlas based adaptive clustering. *NeuroImage*, 102:596–607.

[147] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

[148] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79.

[149] Varentsova, A., Zhang, S., and Arfanakis, K. (2014). Development of a high angular resolution diffusion imaging human brain template. *NeuroImage*, 91:177–186.

[150] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009).

Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.

[151] VoTEM, I. . (2018). 3-d validation of tractography with experimental mri. https://my.vanderbilt.edu/votem/. Accessed: 2018-01-12.

[152] Wang, D., Luo, Y., Mok, V. C., Chu, W. C., and Shi, L. (2016a). Tractography atlas-based spatial statistics: Statistical analysis of diffusion tensor image along fiber pathways. *NeuroImage*, 125:301–310.

[153] Wang, Q., Chen, R., JaJa, J., Jin, Y., Hong, L. E., and Herskovits, E. H. (2016b). Connectivity-based brain parcellation. *Neuroinformatics*, 14(1):83–97.

[154] Wang, Q., Yap, P.-T., Wu, G., and Shen, D. (2011a). Diffusion tensor image registration with combined tract and tensor features. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 200–208.

[155] Wang, Y., Gupta, A., Liu, Z., Zhang, H., Escolar, M. L., Gilmore, J. H., Gouttard, S., Fillard, P., Maltbie, E., Gerig, G., et al. (2011b). Dti registration in atlas based fiber analysis of infantile krabbe disease. *Neuroimage*, 55(4):1577–1586.

[156] Wang, Y., Shen, Y., Liu, D., Li, G., Guo, Z., Fan, Y., and Niu, Y. (2017). Evaluations of diffusion tensor image registration based on fiber tractography. *Biomedical engineering online*, 16(1):9.

[157] Wang, Y., Yu, Q., Liu, Z., Lei, T., Guo, Z., Qi, M., and Fan, Y. (2016c). Evaluation on diffusion tensor image registration algorithms. *Multimedia Tools and Applications*, 75(13):8105–8122.

[158] Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., and Westin, C.-F. (2016). The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structure and Function*, 221(9):4705–4721.

[159] Wassermann, D., Rathi, Y., Bouix, S., Kubicki, M., Kikinis, R., Shenton, M., and Westin, C.-F. (2011). White matter bundle registration and population analysis based on gaussian processes. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 320–332. Springer.

[160] Xu, D., Mori, S., Shen, D., van Zijl, P., and Davatzikos, C. (2003). Spatial normalization of diffusion tensor fields. *Magnetic resonance in medicine*, 50(1):175–182.

[161] Yang, Z., Chen, G., Shen, D., and Yap, P.-T. (2017). Robust fusion of diffusion mri data for template construction. *Scientific Reports*, 7:12950.

[162] Yap, P., Chen, Y., An, H., Gilmore, J., Lin, W., and Shen, D. (2010a). Non-parametric deformable registration of high angular resolution diffusion data using diffusion profile statistics. In *ISMRM*, volume 18, page 3968.

[163] Yap, P.-T., Chen, Y., An, H., Yang, Y., Gilmore, J. H., Lin, W., and Shen, D. (2011). Sphere: Spherical harmonic elastic registration of hardi data. *NeuroImage*, 55(2):545–556.

[164] Yap, P.-T. and Shen, D. (2012). Spatial transformation of dwi data using non-negative sparse representation. *IEEE transactions on medical imaging*, 31(11):2035–2049.

[165] Yap, P.-T., Wu, G., Zhu, H., Lin, W., and Shen, D. (2010b). F-timer: fast tensor image morphing for elastic registration. *IEEE transactions on medical imaging*, 29(5):1192–1203.

[166] Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., et al. (2011). Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Frontiers in neuroinformatics*, 5.

[167] Yeo, B. T., Vercauteren, T., Fillard, P., Pennec, X., Golland, P., Ayache, N., and Clatz, O. (2008). Dti registration with exact finite-strain differential. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 700–703. IEEE.

[168] Yeo, B. T., Vercauteren, T., Fillard, P., Peyrat, J.-M., Pennec, X., Golland, P., Ayache, N., and Clatz, O. (2009). Dt-refind: Diffusion tensor registration with exact finite-strain differential. *IEEE transactions on medical imaging*, 28(12):1914–1928.

[169] Yoo, S. W., Guevara, P., Jeong, Y., Yoo, K., Shin, J. S., Mangin, J.-F., and Seong, J.-K. (2015). An example-based multi-atlas approach to automatic labeling of white matter tracts. *PloS one*, 10(7):e0133337.

[170] Zalesky, A. (2011). Moderating registration misalignment in voxelwise comparisons of dti data: a performance evaluation of skeleton projection. *Magnetic resonance imaging*, 29(1):111–125.

[171] Zhang, H., Avants, B. B., Yushkevich, P. A., Woo, J. H., Wang, S., McCluskey, L. F., Elman, L. B., Melhem, E. R., and Gee, J. C. (2007). High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *IEEE transactions on medical imaging*, 26(11):1585–1597.

[172] Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. (2012). Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016.

[173] Zhang, H., Yushkevich, P. A., Alexander, D. C., and Gee, J. C. (2006). Deformable registration of diffusion tensor mr images with explicit orientation optimization. *Medical image analysis*, 10(5):764–785.

[174] Zhang, H., Yushkevich, P. A., and Gee, J. C. (2004). Registration of diffusion tensor images. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE.

[175] Zhang, P., Niethammer, M., Shen, D., and Yap, P.-T. (2013). Large deformation diffeomorphic registration of diffusion-weighted images with explicit orientation optimization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 27–34. Springer.

[176] Zhang, P., Niethammer, M., Shen, D., and Yap, P.-T. (2014). Large deformation diffeomorphic registration of diffusion-weighted imaging data. *Medical image analysis*, 18(8):1290–1298.

[177] Zhang, S. and Arfanakis, K. (2013). Role of standardized and study-specific human brain diffusion tensor templates in inter-subject spatial normalization. *Journal of Magnetic Resonance Imaging*, 37(2):372–381.

[178] Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., and Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224.

[179] Zhang, Y., Zhang, J., Oishi, K., Faria, A. V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G. B., Evans, A., et al. (2010). Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *Neuroimage*, 52(4):1289–1301.

[180] Zimmerman-Moreno, G., Ben Bashat, D., Artzi, M., Nefussy, B., Drory, V., Aizenstein, O., and Greenspan, H. (2016). Whole brain fiber-based comparison (fbc)—a tool for diffusion tensor imaging-based cohort studies. *Human brain mapping*, 37(2):477–490.

[181] Ziyan, U., Sabuncu, M. R., O'donnell, L. J., and Westin, C.-F. (2007). Nonlinear registration of diffusion mr images based on fiber bundles. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 351–358. Springer.

[182] Zucchelli, M., Descoteaux, M., and Menegaz, G. (2017). Noddish: a computational efficient noddi extension for fodf estimation in diffusion mri. *arXiv preprint arXiv:1708.08999*.

4 Locally Orderless Registration for Diffusion-Weighted Images

This chapter serves as an introduction to the first part of our framework for DWI registration: A scale-space formulation that defines nonlinear information-theoretic similarity between DWI scans, while also taking the full ODF into account. Explicit reorientation in the registration of DWI scans has been shown to significantly improve voxel-based DWI registration, as discussed Chapter 3. By defining the analytical gradients of the spatial, directional and intensity scale, we get a smooth and robust explicit reorientation model, which can be used for both DTI and HARDI data. As this is the first part of our registration framework, the registration is purely global and focused on the scale-space model, the gradients, and the similarity measure. It was presented as a conference paper at MICCAI 2015 [Jensen et al., 2015], which can be found at the end of this chapter. We will refer to it as LOR-DWI.

The following sections will briefly reiterate on the background behind the LOR-DWI framework, and expand on some of the background and methodology left out in the paper. It is structured as follows:

Section 4.1 - Locally Orderless Images and Registration. An introduction to the concept of Locally Orderless Images (LOI). We cover the term 'locally orderless' in relation to registration.

Section 4.2 - Choice of Scale-Space Kernels. An introduction to the discrete scale-space kernels used in the implementation of the framework.

Section 4.3 - Choice of Similarity Measure. The motivation behind choosing Mutual Information as the non-linear similarity measure.

Section 4.4 - The Conference Paper. "Locally Orderless Registration for Diffusion-Weighted Images" [Jensen et al., 2015].

4.1 Locally Orderless Images and Registration

The LOR-DWI framework is an extension of DWI registration based on the Locally Orderless Registration (LOR) density estimation framework for image similarity by [Darkner and Sporring, 2013]. This, in turn, is based on the concept of Locally Orderless Images (LOI) originally introduced in [Koenderink and Van Doorn, 1999]. The concept of 'locally orderless' can cause confusion out of context. This is understandable as the actual framework is not completely orderless. Instead one could say that it is pair-wise orderless, due to a joint histogram formulation between two DWI scans. We do not use the locality term but instead a special global Parzen-Window (PW) definition of LOI (as shown in [Darkner and Sporring, 2013]). The reason we use the term 'locally orderless' comes from the original theoretical foundation of the framework. We briefly cover the concept of what is local and what is orderless, while at the same time expanding on some of the concepts of the paper.

The Locally Orderless Image (LOI)

The LOI is defined in [Koenderink and Van Doorn, 1999] as a histogram-valued image in a region of interest (ROI). The values in these images depend on the size of the ROI, the resolution of the image, and the width of the bins in the histogram. In the construction of any such histogram, the order (topology) of the values is lost, and it is here that the *orderless* part comes in. By this definition, the image is orderless in the ROI but not globally. However, there is spatial invariance within the ROI. As an example, this means that an LOI, consisting of vertical black and white stripes, would be indistinguishable from a set of corresponding horizontal stripes. This histogram approach enables us to create density estimates of the LOI, from which we can derive information-theoretic similarity measures, which we further address inSection 4.3. We define the LOI of a DWI as the histogram

$$h_{\beta,\alpha,\sigma,\kappa}(i|\boldsymbol{x}) = \int_{\Omega\times S^2} P_\beta(\boldsymbol{I}_{\sigma,\kappa}(\boldsymbol{x},\boldsymbol{v}) - i) W_\alpha(\boldsymbol{\tau}-\boldsymbol{x}) d\boldsymbol{\tau} \times d\boldsymbol{v} \tag{4.1}$$

where P_β is a Gaussian PW with standard deviation β, W_α is a Gaussian window of integration (i.e. our ROI) of standard deviation α, and $\boldsymbol{I}_{\sigma,\kappa}$ is a smooth, interpolated DWI. $\boldsymbol{I}_{\sigma,\kappa}$ is defined as

$$\boldsymbol{I}_{\sigma,\kappa}(\boldsymbol{x},\boldsymbol{v}) = \int_{S^2} \left(\int_\Omega \boldsymbol{I}(\boldsymbol{\tau},\boldsymbol{\nu}) K_\sigma(\boldsymbol{\tau}-\boldsymbol{x}) d\boldsymbol{\tau} \right) \Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v}) d\boldsymbol{\nu} \tag{4.2}$$

where K_σ is a Gaussian kernel of standard deviation σ and Γ_κ is a Watson kernel for interpolation on the sphere with a concentration parameter κ. So

far, the area of integration is still local within W_α and $h_{\beta,\alpha,\sigma,\kappa}$ is orderless, defined as the scale-space described by Koenderink et. al with an added directional scale κ for DWI. However, we shall optimize over a global similarity measure as we work with registration. To accommodate this, we let $\alpha \to \infty$ such that W becomes constant over the image domain and the ROI spans the entire image. Discounting the locality scale W, we now work with a *globally* orderless image, though it is still only a special case of the LOI being a PW as shown in [Darkner and Sporring, 2013]. The LOI, that we use in eq. (4.1), now simplifies to

$$h_{\beta,\sigma,\kappa}(i|\boldsymbol{x}) = \int_{\Omega\times S^2} P_\beta(\boldsymbol{I}_{\sigma,\kappa}(\boldsymbol{x},\boldsymbol{v}) - i)d\boldsymbol{x}\times d\boldsymbol{v} \tag{4.3}$$

To optimize the similarity in image registration, we combine two such LOIs into a joint histogram, of a target image $\boldsymbol{J}$ and a moving image $\boldsymbol{I}$, under a spatial and directional transformation ϕ and ψ

$$\begin{aligned} h_{\beta\alpha\sigma\kappa}(i,j|\tilde{\Phi},\boldsymbol{x}) = \\ \int_{\Omega\times S^2} P_\beta(\boldsymbol{I}_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - i)P_\beta(J_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) - j)d\boldsymbol{x}\times d\boldsymbol{v} \end{aligned} \tag{4.4}$$

The entries in the joint histogram are now only pair-wise orderless, in the sense that the images are aligned and the values of the LOI is a combination of any two intensities. We would not be able to ignore the order of voxels in the two images separately, i.e. random sampling must be pair-wise. While all of this means, that we do not strictly adhere to the generalized definition of the LOI, it provides a mathematically solid scale-space framework for our own spatio-directional scale-space model.

4.2 Choice of Scale-Space Kernels

The LOR-DWI framework in our paper is presented as the theoretically sound extension of the LOR framework using infinite support Gaussians models and other generalizations. We go through our choice of computationally feasible, discrete scale-space kernels.

4.2.1 Scale-Space Kernels: Image and Orientation

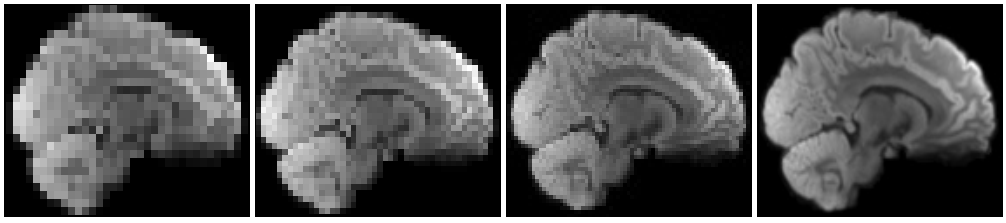
The interpolation of DWI data is both spatial and directional. In the LOR-DWI framework, the spatial and directional operations are commutative, and we can compute them in any of the following order

$$\boldsymbol{I}_{interp} = \Gamma_\kappa \circ K_\sigma \circ \boldsymbol{I} \circ \Phi = K_\sigma \circ \Gamma_\kappa \circ \boldsymbol{I} \circ \Phi \tag{4.5}$$

where Γ is the directional interpolation or orientation scale, B the spatial interpolation or image scale, Φ the deformation model, and $\boldsymbol{I}_{interp}$ is equivalent to eq. (4.2) from the LOI definition.

The Image Scale: Cubic B-Splines

There are two common ways of controlling the scale of an image. One is to smooth the full resolution image to produce different scales, and another is to use a smooth kernel to downsample the image, as shown in Figure 4.1. In our experimental setup of [Jensen et al., 2015], we use the first approach, but change it to the second and quicker version in the extended nonrigid version, as we return to inChapter 5.



(a) spacing= 4 **(b)** spacing= 3 **(c)** spacing= 2 **(d)** spacing= 1

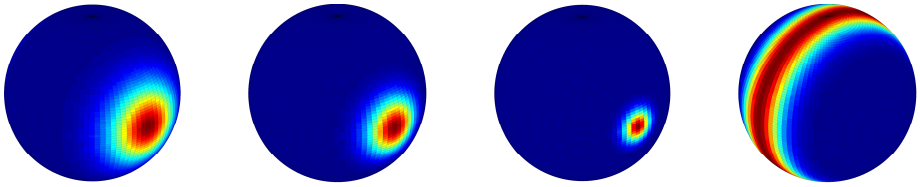
Figure 4.1: Sagittal view of the mean diffusivity of non-normalized HCP subject `103818` (same as used in Section 2.2). From left to right, cubic B-spline interpolation at every 4th voxel, every 3rd, 2nd, and finally the full resolution image.

We discretize the kernel K_σ by approximating it with a uniform cubic B-Spline, as a full Gaussian kernel is not tractable and a truncated Gaussian has lower accuracy than a third-order B-Spline [Bouma et al., 2007]. The cubic B-spline model is a popular and well-known kernel with local support, that uses a linear combination of basis function to create a smooth curve over an area, based on the surrounding control points, defined by the kernel. It has the following attractive properties:

- Stable with local and symmetric support.
- Partition of unity[1].
- Piecewise polynomial and non-negative.
- Smooth with trivial analytical derivatives.

For a deeper introduction, such as the definition of basis functions and their derivatives, we again refer to Chapter 5, which constitutes the nonrigid part of this framework.

[1]The weights sum to one and does not change the density of the image.



(a) Watson $\kappa = 5$ **(b)** Watson $\kappa = 10$ **(c)** Watson $\kappa = 30$ **(d)** Watson $\kappa = -10$

Figure 4.2: The Watson distribution with four different concentration parameters. Note that the kernel is antipodal symmetric and would be mirrored on the hidden side of the sphere for (**a-c**). The center of (**d**) is the same as for (**a-c**).

The Orientation Scale: The Watson Density Function

We use the Watson density function Γ_κ to represent the orientational scale and model the ODF [Jupp and Mardia, 1989]. The choice of kernel is similar to the reasoning for using B-Splines. We do not perform interpolation based on peaks in the diffusion representing fiber orientations. We use a more data-independent approach, where we convolve the orientation distribution with a smooth antipodal symmetric kernel - adhering to the symmetric nature of diffusion scans. In other words, we take a step back from trying to fit a function to a certain distribution on the sphere, which is otherwise a common use for models such as the Watson distribution combined with EM-approaches [Bijral et al., 2007].

When it comes to DWI and modeling the raw HARDI signal or the ODF, there are two predominant ways to perform interpolation on the sphere. The first is through a combination of periodical orthonormal basis functions that are fitted to the ODF as often seen with spherical harmonics (SH) [Tuch, 2004]. The second is to use an exponential density function that also has a convenient approximation of the tomographic inversion (FRT), such as the von Mises-Fisher (vMF) or Watson distributions [Jupp and Mardia, 1989]. For the Watson density function, we get the FRT from using a negative concentration parameter. The function is defined as

$$\Gamma(u, v)_\kappa = C \cdot \sum_i e^{\kappa \cdot (\langle \boldsymbol{v}_i, \boldsymbol{u} \rangle)^2} \cdot a_i \tag{4.6}$$

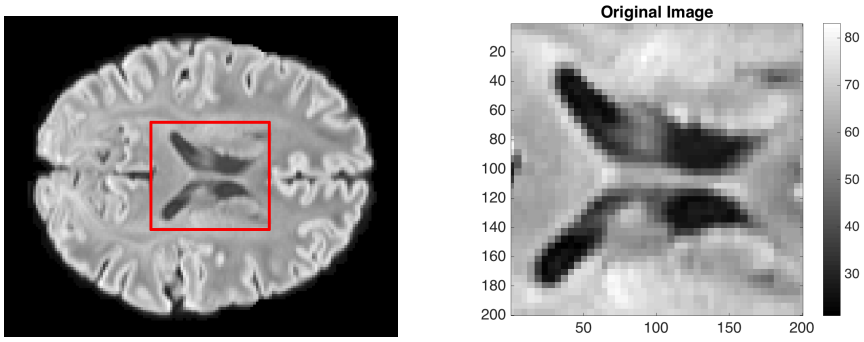
where a_i is the magnitude of each vector on the sphere and C is a normalization constant, that in our discrete case takes the form $C = \frac{1}{\sum_i e^{\kappa \cdot (\langle \boldsymbol{v}_i, \boldsymbol{u} \rangle)^2}}$. Furthermore, κ is a concentration parameter around the mean vector $\boldsymbol{u}$ for $\kappa > 0$, the average magnitude of all vectors $\boldsymbol{v}_i$ if $\kappa = 0$ (i.e. mean diffusivity), and the perpendicular great circle if $\kappa < 0$. Figure 4.2 shows the spherical

support of different κ-values. Unlike the vMF kernel, the negative concentration parameter gives a closed-form solution to the Funk-Radon transform. It is a kernel function that is suitable to represent the ODF due to its relatively low computational complexity and a nice mathematical justification [Rathi et al., 2009].

4.2.2 The Intensity Scale: The Parzen-Window (PW)

The final scale is the intensity scale or the topology scale as described in [Koenderink and Van Doorn, 1999]. The ability to optimize over and manipulate the isocurves between two DWI scans is perhaps the greatest strength of our registration framework, next to the density-based formulation that allows us to use information-theoretic similarity measures. Both of these are a direct effect of the Parzen-Window (PW) formulation.

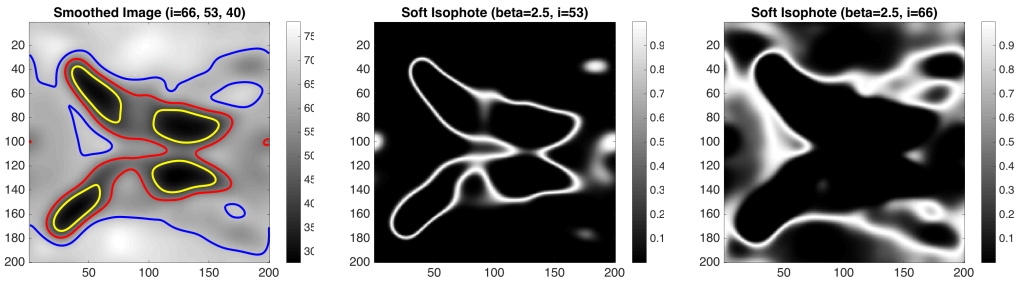
Since there is no easy and intuitive way to illustrate smooth isocurves of the 4D structure of DWI, the isocurves are perhaps better illustrated on a 2D slice of a 3D MR image, similar to the random images presented in [Darkner and Sporring, 2013]. Figure 4.3 shows a slice and, within this, the ROI around the ventricles that we use for this example.



(a) **(b)**

Figure 4.3: Axial or from above view of the brain. The images show the selection around the ventricles (the dark area in **(b)**) used to illustrate the isocurves, or isophotes, in Figure 4.4.

Three smooth isocurves also referred to as soft isophote lines are shown in Figure 4.4a. They are the soft bins, or densities, from the PW-generated histogram. The isophotes are inversely proportional and perpendicular to the gradients of the image, which allow us to smooth *along* the isocontours instead of just smoothing the spatial scales, known to result in partial volume effects.



(a) **(b)** **(c)**

Figure 4.4: **(a)** shows 3 isophote lines around the area of the ventricles, shown as a spatially smooth version of Figure 4.3. **(b-c)** are 2 of the red and blue isophotes extracted from the PW, where β is the scale of the PW, reused for purely illustrative purposes from [Darkner and Sporring, 2013].

The PW density estimation [Parzen, 1962] is a widespread data-interpolation technique, where a kernel function is superposed at each discrete observation x_i to estimate the probability $P(x_i)$ within the window of support. PW methods have been widely used in image registration to allow for gradient-based optimization by explicit differentiation of the otherwise discrete joint histogram, required for information-theoretic similarity measures [Darkner and Sporring, 2013]. For a Gaussian PW kernel P_β of standard deviation β, we use the definition of the interpolated volume $\boldsymbol{I}_{\sigma,\kappa}$ from eq. (4.2) to write the LOI of eq. (4.1) and the joint LOI of eq. (4.4) as the normalized density estimates

$$p_{\beta\alpha\sigma}(i|\boldsymbol{x}) \simeq \frac{h_{\beta\alpha\sigma}(i|\boldsymbol{x})}{\int_\Lambda h_{\beta\alpha\sigma}(k|\boldsymbol{x})dk} \tag{4.7}$$

$$p_{\beta\alpha\sigma}(i,j|\boldsymbol{x}) \simeq \frac{h_{\beta\alpha\sigma}(i,j|\boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma}(k,l|\boldsymbol{x})dk\,dl}\,. \tag{4.8}$$

In [Jensen et al., 2015], we use a 2D uniform cubic B-Spline as the choice of kernel for an efficient approximation to a Gaussian Parzen-window. This is equivalent to an intensity scale parameter β at a Gaussian variance of around 0.6 for each smooth entry into the histogram [Darkner and Sporring, 2013], see Chapter 5 for details. Furthermore, it has some important properties, also described with the image scale, such as the *partition of unity*, which makes normalization of the densities trivial, as each kernel entry will sum to one and the normalization factor becomes one over the number of entries - independent of the transformation. This can, for instance, be found in [Thévenaz and Unser, 1998] who also uses a Parzen-Window based on cubic B-splines. It will also ensure that the marginal probabilities $p_{\boldsymbol{J}}$ of the station-

ary reference image $\boldsymbol{J}$ remain independent of the transformation parameters, as we can move the normalization factor outside the kernel. This, in turn, ensures that we do not need to compute the derivatives with respect to p_J [Thevenaz and Unser, 1997]. There is also the local support, which spreads the contribution over several bins, allowing for the recovery of more information than a traditional histogram.

The intensity values of both images are linearly scaled to fall within a given range of bins. The choice of the number of bins is a well-known issue, and it is a trade-off between removing details and a high sparsity which may result in noisy estimates. In [Jensen et al., 2015], we measure the effects of the different scales during optimization. In Chapter 5, we further investigate the effect of hierarchically changing the size of histogram during registration in the nonrigid extension.

4.3 Choice of Similarity Measure

As introduced in Section 2.1, the cost function $\mathcal{M}$ defines how well-aligned two images are in image registration:

$$\mathcal{M}(\boldsymbol{I}, \boldsymbol{J}, \Phi) = \mathcal{F}(\boldsymbol{I} \circ \Phi, \boldsymbol{J}) + \mathcal{S}(\Phi) \; . \tag{4.9}$$

Our choice of similarity measure $\mathcal{F}$ in [Jensen et al., 2015] is the information-theoretic similarity measure Mutual Information (MI). However, the use of the smooth differentiable PW formulation, combined with spatial and directional derivatives, gives us the potential to swap the similarity measure with another measure, that can be defined over a histogram or normalized density estimate. This is trivial for linear measure, though somewhat more complicated for non-linear measures. We briefly discuss our decision to use the non-linear MI as the similarity measure.

4.3.1 Similar Similarity Measures

It is worth noting that only a handful of similarity measures in publicly available frameworks include explicit reorientation, as part of the cost function and the similarity measure. As we investigated in Chapter 3, the only two both popular and recognized frameworks were `DTI-TK` [Zhang et al., 2006] and `DT-REFinD` [Yeo et al., 2009]. As also noted, both frameworks have since been expanded but are primarily based on the finite strain models, and none of the expansions use non-linear measures such as information-theoretic measures, which makes it more unclear how matching across b-values and different scanner parameters should be performed.

4.3.2 A Non-linear Similarity Measure for DWI

Information-theoretic and correlation-based similarity measures are the most popular measures for comparing images that have a complex relationship not easily matched or normalized [Rogelj and Kovacic, 2001]. Such multi-modal images, or images of a complex nature like the bi-exponential DWI, often have a non-linear relationship and require a measure that can model the statistical relationship between them. The two most popular similarity measures that fit this criterion are the Correlation Ratio (CR) and Mutual Information (MI) [Roche et al., 1999]. CR assumes that some function exists that can approximate one image with the other. MI is similar but more statistical than functional, as it is theoretically more robust to variations from an ideal functional relationship. With few assumptions about the nature of the relationship between image intensities, MI is particularly attractive to use for the complex and noisy DWI images, that rely on b-values as a single product over multiple non-trivial parameters, such as acquisition time, magnetic field strength, etc. MI has not been extended beyond scalar-based registration in DWI, but it is already being used extensively used for distortion correction [Treiber et al., 2016] and motion correction [Rohde et al., 2004].

Mutual Information in the LOR framework

Once the probabilistic density between two DWI images has been calculated in the shape of a normalized histogram, formulating the similarity measure is identical to the scalar-based LOR framework on which it was based [Darkner and Sporring, 2013]. The density formulation gives access to a set of generalized linear and non-linear similarity measures for DWI registration as an extension of the LOR framework, defined as

$$\mathcal{F}_{lin} = \int_{\Lambda^2} f(i,j)p(i,j)di\,dj \qquad \mathcal{F}_{non-lin} = \int_{\Lambda^2} f(p(i,j))di\,dj \tag{4.10}$$

where p is the normalized joint histogram and f is the similarity. The linear version is a position-independent loss function, where only p is influenced by the registration parameters and f can easily be changed. It is more complicated with the position-dependent non-linear version, for which [Darkner and Sporring, 2013] defines MI as

$$\mathcal{F}_{\text{MI}} = \mathcal{H}_I + \mathcal{H}_R - \mathcal{H}_{I,R} \tag{4.11}$$

where

$$\mathcal{H}_I = -\int_\Gamma p_I(i)\log p_I(i)\,di, \qquad \mathcal{H}_R = -\int_\Gamma p_R(j)\log p_R(j)\,dj,$$
$$\mathcal{H}_{I,R} = -\int_{\Gamma^2} p_{I,R}(i,j)\log p_{I,R}(i,j)\,didj, \tag{4.12}$$

For a list of both linear and non-linear loss functions, following the above notation, we refer to the technical report [Sporring and Darkner, 2011], which also defines CR and more.

This ends the extended background introduction to [Jensen et al., 2015] presented at MICCAI 2015, which continues on the following page.

Locally Orderless Registration for Diffusion Weighted Images

Henrik G. Jensen, Francois Lauze, Mads Nielsen, and Sune Darkner

Department of Computer Science, University of Copenhagen
Universitetsparken 1, DK-2100 Copenhagen, Denmark
{henne,madsn,francois,darkner}@di.ku.dk

Abstract. Registration of Diffusion Weighted Images (DWI) is challenging as the data, in contrast to scalar-valued images, is a composition of both directional and intensity information. The DWI signal is known to be influenced by noise and a wide range of artifacts. Therefore, it is attractive to use similarity measures with invariance properties, such as Mutual Information. However, density estimation from DWI is complicated by directional information. We address this problem by extending Locally Orderless Registration (LOR), a density estimation framework for image similarity, to include directional information. We construct a spatio-directional scale-space formulation of marginal and joint density distributions between two DWI, that takes the projective nature of the directional information into account. This accounts for orientation and magnitude and enables us to use a wide range of similarity measures from the LOR framework. Using Mutual Information, we examine the properties of the scale-space induced by the choice of kernels and illustrate the approach by affine registration.

1 Introduction

The registration of Diffusion Weighted Images (DWI) is interesting as it contains information about the fibrous micro-architecture otherwise invisible to structural MRI. Registration of these structures enables us to compare connectivity within and across subjects. However, registration is challenging due to the inherent geometry of DWI; notably high-angular resolution diffusion imaging (HARDI) which models more complex displacement profiles. We extend the Locally Orderless Registration (LOR) [2] density estimation framework for image similarity from scalar-valued images to DWI. The LOR is a scale-space framework for image density estimation that allows us to employ a wide range of similarity measures for registration, including MI. By introducing a spatio-directional kernel, thus including the space of gradient directions, we model the relationship between direction and measurements as histograms. The histograms are mapped to probability density estimates by normalization and marginalization over the deformed space.

Our contribution is a full LOR scale-space formulation for DWI, offering explicitly control of orientation, image, intensity and integration scale. We examine

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 305–312, 2015.
DOI: 10.1007/978-3-319-24571-3_37

the effects of the scale-space and illustrate the application of the density estimate by affine image registration of DWI data using Mutual Information.

2 Previous Work and Background

Locally Orderless Images (LOI) [6] is a scale-space representation of intensity distributions in images modeling three inherent scales: the image scale (i.e. image smoothing), the integration scale (local histogram), and the intensity scale (soft bin width). The first mention of LOI in the context of image registration was in by Hermosillo et al. [4] where a variational approach to image registration was presented. The LOR framework [2], an extension of [1], generalized a range of similarity measures as linear and non-linear functions of density estimates for scalar-valued images. One such non-linear similarity measure is Mutual Information (MI) [13]. MI is one of the most frequently used similarity measures in image registration and was introduced as a multi-modal similarity measure. MI is frequently used in MRI due to its invariance properties with respect to intensity values and is associated with scalar-valued images. It is used in the context of DWI for distortion-correction [9] on e.g. b_0 or individual DWI directions. Van Hecke et al. [12] used MI for non-rigid registration of DWI. Under an assumption of alignment, each gradient direction was evaluated separately as well as in a pooled fashion to form a joint density distribution. Interpolation of directional information in DWI was introduced by Tao and Miller [10] for affine registration using SSD and extended by Duarte-Carvajalino et al. [3] to non-rigid B-spline registration. The angular interpolation was extended with a Watson distribution by Rathi et al. [8]. Raffelt et al. [7] used SSD after spherical deconvolution for fiber modeling (FODs), while others, like Yap et al. [14], compared the coefficients of the spherical harmonics.

Image registration is the process of spatially aligning (two) images (I and J) under some transformation Φ given some regularity condition $\mathcal{S}(\Phi)$ and similarity $\mathcal{F}(I \circ \Phi, J)$ such that $\mathcal{M}(I, J, \Phi)$ is minimized

$$\mathcal{M}(I, J, \Phi) = \mathcal{F}(I \circ \Phi, J) + \mathcal{S}(\Phi) \tag{1}$$

In this paper we address the estimation of $\mathcal{F}$ of single shell DWI as an extension of LOR with application to MI. DWI MR attenuation signals at location $\boldsymbol{x}$, for a gradient direction $\boldsymbol{v}$, are modeled by $S(\boldsymbol{x}, \boldsymbol{v}) = S_0(\boldsymbol{x})e^{-bI(\boldsymbol{x},\boldsymbol{v})}$ [10] and apparent diffusion coefficients volumes are given by $I(\boldsymbol{x}, \boldsymbol{v}) = -\frac{1}{b}\log\frac{S(\boldsymbol{x},\boldsymbol{v})}{S_0(\boldsymbol{x})}$. Gradient directions $\boldsymbol{v}$ are taken on the unit sphere $\mathbb{S}^2$ although diffusion are orientation-free and have $I(\boldsymbol{x}, \boldsymbol{v}) \approx I(\boldsymbol{x}, -\boldsymbol{v})$, i.e., with antipodal symmetry. Such a symmetric function $I(\boldsymbol{x}, -)$ on the sphere can be represented by a function on the projective space $\mathbb{P}^2$ of directions of $\mathbb{R}^3$, $\mathbb{P}^2 \simeq \mathbb{S}^2/\{\pm 1\}$.

We start by defining the type of transformation considered for the LOR density estimates for single shell DWI presented in this paper. For any transformation ϕ of a point $\boldsymbol{x}$, we consider only diffeomorphic mappings $\phi(\boldsymbol{x})\colon \mathbb{R}^3 \to \mathbb{R}^3$. Under this assumption, ϕ is invertible and its differential, or Jacobian $d_{\boldsymbol{x}}\phi$ at

$\boldsymbol{x}$, gives naturally rise to a projective transformation on $\mathbb{P}^2$: $t\boldsymbol{v} \mapsto td_{\boldsymbol{x}}\phi(\boldsymbol{v}), t \in \mathbb{R}\backslash\{0\}$. We drop $\boldsymbol{x}$ and simply write $d\phi$. Its representation over $\mathbb{S}^2$ is $\boldsymbol{v} \in \mathbb{S}^2 \mapsto \pm\frac{d\phi(\boldsymbol{v})}{|d\phi(\boldsymbol{v})|}$. This term appears within a spherical kernel Γ_κ with antipodal symmetry, making the sign irrelevant. Setting $\psi(\boldsymbol{v}) = \frac{d\phi(\boldsymbol{v})}{|d\phi(\boldsymbol{v})|}$, it corresponds to the transformation proposed by [10]. We therefore extend our transformation to $\tilde{\boldsymbol{\Phi}} : (\boldsymbol{x}, \boldsymbol{v}) \mapsto \big(\phi(\boldsymbol{x}), \psi(\boldsymbol{v})\big)$. This type of transformation is also argued in [7], although neither [10] nor [7] did consider its projective nature. We proceed to describe the LOR.

The LOR framework defines the similarity over three scales: The image scale σ, the intensity scale β, and the integration scale α. In registration, for a transformation ϕ, we get

$$h_{\beta\alpha\sigma}(i,j|\phi,\boldsymbol{x}) = \int_\Omega P_\beta(I_\sigma(\phi(\boldsymbol{x})) - i)P_\beta(J_\sigma(\boldsymbol{x}) - j)W_\alpha(\boldsymbol{\tau} - \boldsymbol{x})d\boldsymbol{\tau} \tag{2}$$

$$p_{\beta\alpha\sigma}(i,j|\phi,\boldsymbol{x}) \simeq \frac{h_{\beta\alpha\sigma}(i,j|\phi,\boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma}(k,l|\phi,\boldsymbol{x})dk\,dl} \tag{3}$$

where $i,j \in [a_1,a_2]$ are values in the image intensity range, $I_\sigma(\boldsymbol{\phi}(\boldsymbol{x})) = (I * K_\sigma)(\boldsymbol{\phi}(\boldsymbol{x}))$ and $J_\sigma(\boldsymbol{x}) = (J * K_\sigma)(\boldsymbol{x})$ are images convolved with the kernel K_σ with standard deviation σ, P_β is a Parzen-window of scale β, and W_α is a Gaussian integration window of scale α. The marginals are trivial and obtained by integration over the appropriate variable. The LOR-approach to similarity lets us use a set of generalized similarity measures, the linear and non-linear

$$\mathcal{F}_{lin} = \int_{\Lambda^2} f(i,j)p(i,j)di\,dj \qquad \mathcal{F}_{non-lin} = \int_{\Lambda^2} f(p(i,j))di\,dj \tag{4}$$

where the linear measure $f(i,j)$ includes e.g. sum of squared differences and Huber, and the non-linear $f(p(i,j))$ includes e.g. MI, NMI, see [2] for details.

3 Locally Orderless DWI

To extend the density estimates of LOR to include directional information, we introducing a kernel on the sphere to account for directional smoothing. With that in mind, we extend spatial smoothing to be spatio-directional, where the directional smoothing preserves this symmetry, and thus the projective structure, via a symmetric kernel $\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v})$ on $\mathbb{S}^2$. We define the smoothed signal $I_{\sigma,\kappa}$ at scales (σ,κ) by

$$I_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) = \int_{S^2}\left(\int_\Omega I(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\boldsymbol{x})d\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v})d\boldsymbol{\nu} = (I * (K_\sigma \otimes \Gamma_\kappa))(\boldsymbol{x},\boldsymbol{v}) \tag{5}$$

where $K_\sigma(\boldsymbol{x})$ is a Gaussian kernel with σ standard deviation. We use a symmetric Watson distribution [5] as $\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v})$ for directional smoothing on $\mathbb{S}^2$, given by

$$\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v}) = Ce^{\kappa(\langle\boldsymbol{\nu},\boldsymbol{v}\rangle)^2}, \qquad C = M(\frac{1}{2},\frac{1}{d},\kappa) = \sum_{i=0\ldots\infty}\frac{\kappa^n}{(\frac{3}{2})^n n!} \tag{6}$$

where M is the Kummer function for a d-dimensional unit vector $\boldsymbol{\nu}$, (in this case $d = 3$), $\pm\boldsymbol{v}$ the center of the distribution, and κ the concentration parameter, which is roughly inverse proportional to the variance on the sphere. As one alternative, a symmetrized von Mises-Fisher [5] distribution could be considered.

In order to use the similarity measures provided by the LOR framework, we extend the LOR formulation from scalar-valued images to DWI. The joint histogram, that is, the contribution to $h(i,j)\colon \Lambda^2 \to \mathbb{R}_+$ of the joint histogram and normalization can be written as

$$h_{\beta\alpha\sigma\kappa}(i,j|\boldsymbol{x}) = \int_{\Omega\times S^2} P_\beta(I_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) - i)P_\beta(J_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) - j)W_\alpha(\boldsymbol{\tau}-\boldsymbol{x})d\boldsymbol{x}\times d\boldsymbol{v} \tag{7}$$

$$p_{\beta\alpha\sigma\kappa}(i,j|\boldsymbol{x}) = \frac{h_{\beta\alpha\sigma\kappa}(i,j|\boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma\kappa}(k,l|\boldsymbol{x})dk\,dl} \tag{8}$$

where $I_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v})$ and $J_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v})$ are defined as in Equation (5), P is a Gaussian Parzen-window with standard deviation β, and W a Gaussian window of integration around $\boldsymbol{x}$ with standard deviation α. The marginals are trivial and obtained by integration over the appropriate variable. The joint and marginal probability densities allow us to apply the generalized similarity measures in Equation (4). In this paper, we use the non-linear MI.

4 Image Registration

We write the joint histogram and density for similarity in image registration as

$$h_{\beta\alpha\sigma\kappa}(i,j|\tilde{\Phi},\boldsymbol{x}) = \int_{\Omega\times S^2} P_\beta(I_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - i)P_\beta(J_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) - j)W_\alpha(\boldsymbol{\tau}-\boldsymbol{x})d\boldsymbol{\tau}\times d\boldsymbol{v} \tag{9}$$

$$p_{\beta\alpha\sigma\kappa}(i,j|\tilde{\Phi},\boldsymbol{x}) = \frac{h_{\beta\alpha\sigma\kappa}(i,j|\tilde{\Phi},\boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma\kappa}(i,j|\tilde{\Phi},\boldsymbol{x})dl\,dk} \tag{10}$$

Most similarity measures are global measures, including MI. To make the density estimate global, we let $\alpha \to \infty$ such that W becomes constant. The first-order structure of the similarity (1) is derived following the approach of [2], denoting differentials as $dg = Dg(\boldsymbol{x})d\boldsymbol{x}$, where D is the partial derivative operator and $d\boldsymbol{x}$ a vector of differentials. We seek $d\mathcal{M}$, the derivative of (1), ignoring the regularization term and omitting irrelevant parameters in the notation. The derivative of MI with respect to $h(i,j)$ is found in [2]. Thus, we seek $dh(i,j)$

$$dh(i,j) = \int_{\Omega\times S^2} dP_\beta(I_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - i)P_\beta(J_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) - j)W_\alpha(\boldsymbol{\tau}-\boldsymbol{x})d\boldsymbol{\tau}\times d\boldsymbol{v} \tag{11}$$

$$\text{with} \qquad dP_\beta(I_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - i) = DP_\beta(I_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - i)dI_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))$$

$DP_\beta(I_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)$ can be found in [1,2]. Inserting (5), we get

$$dI_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) = d\int_{S^2}\left(\int_\Omega I(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))d\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu}. \quad (12)$$

and using the Leibniz integration rule and the product rule, we get

$$dI_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) = \int_{S^2}\left(d\int_\Omega I(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))d\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu}+$$
$$\int_{S^2}\left(\int_\Omega I(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))\boldsymbol{\tau}\right)d\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} \quad (13)$$

We consider each of the terms on the sum separately. Using Leibniz integration rule on the first term of the sum, we get

$$\int_{S^2}\left(\int_\Omega I(\boldsymbol{\tau},\boldsymbol{\nu})dK_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} \quad (14)$$

where $dK_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x})) = DK_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))d\phi(\boldsymbol{x})$ which is trivial in the context of registration. From the second term we get $d\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v})) = D\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\psi(\boldsymbol{v})$ and specifying $\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))$ as a Watson distribution gives

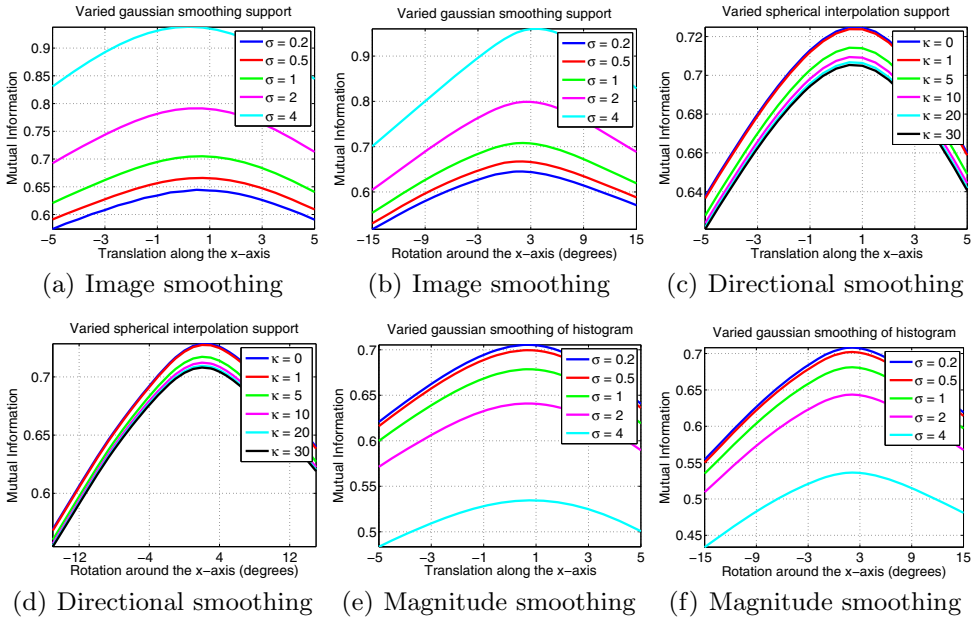
$$D\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v})) = Ce^{\kappa(\langle\boldsymbol{\nu},\psi(\boldsymbol{v})\rangle)^2}2\kappa\langle\boldsymbol{\nu},\psi(\boldsymbol{v})\rangle d\psi(\boldsymbol{v}) \quad (15)$$

which leaves $d\phi(\boldsymbol{x})$ and $d\psi(\boldsymbol{v})$. The first term $d\phi$ is the Jacobian of ϕ and classical in registration literature. The first-order information on spherical reorientation $d\psi(\boldsymbol{v})$ is more complicated, as with our definition of $\psi(\boldsymbol{v})$ as $\frac{d\phi(\boldsymbol{v}}{|d\phi(\boldsymbol{v})|}$, this leads to second-order information of ϕ, which is complex but trivial.

5 Experiments and Results

A series of experiments was conducted to illustrate the scales introduced (spatial, intensity, and directional) with respect to MI. We computed the MI between two subjects and plotted the MI as a function of global rotation and translation (Figure 1) as well as local rotation of three random patches of $10\times10\times10$ voxels (Figure 2). In addition, we performed a few affine registrations of DWI data using the proposed extension of LOR to DWI and MI. We used data from the Human Connectome Project (HCP) database, release Q3, structurally aligned to the MNI-152 template [11], with 90 gradient directions and a b-value of 3000.

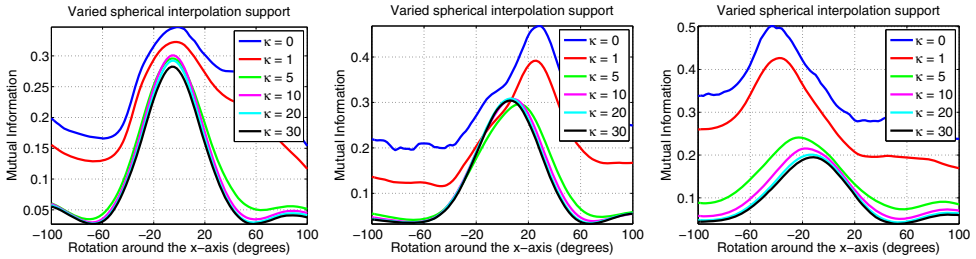
The locally orderless structure introduces four explicit scales on DWI: Image K_σ, intensity P_β, and integration W_α, as well as the extension to orientation Γ_κ. To examine the effect of the scales (ignoring integration scale W_α, i.e. $\alpha\to\infty$), we use Mutual Information, which in itself is a complicated measure. Mutual Information of two observations A, B can be interpreted as the capability of A to encode B. We use this notion of MI to examine the properties of the proposed extension of density estimation to DWI. As a first observation from Figure 1



(a) Image smoothing (b) Image smoothing (c) Directional smoothing

(d) Directional smoothing (e) Magnitude smoothing (f) Magnitude smoothing

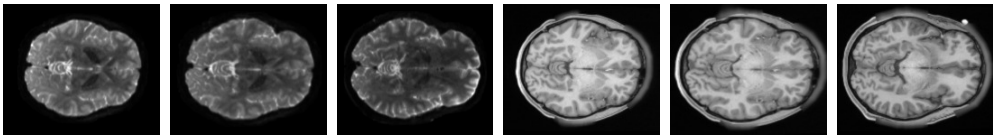
Fig. 1. MI as a function of translation and rotation at different scales. As shown, smoothing in $\mathbb{R}^3$ (a & b) moves the optima, while the change in the angular or diffusion scale (c & d) preserves the MI, despite increased the angular information. This is a good indication of a substantial information in the directions. Smoothing of diffusion magnitudes (e & f) has a similar effect to that observed for scalar-valued images.



(a) Patch 1 (b) Patch 2 (c) Patch 3

Fig. 2. We computed the MI between to DWI volumes within three random patches of $10 \times 10 \times 10$ voxels as a function of rotation and directional smoothing. This clearly illustrates the change in optima as a function of the directional scale. As the images are reasonably well-aligned, this is a strong indication that directional information is required for proper local alignment

it is clear that the DWI optima does not correspond to the structural optima of the registration (to MNI) provided by the HCP. This is illustrated by the fact that the maxima of Figure 1 are not at 0.



(a) Original (b) Affine (c) Target (d) Original (e) Affine (f) Target

Fig. 3. Two DWI images registered using affine transformation and MI for DWI. (left) b=0 gradient images. (Right) T1-weighted images.

The Image Scale influences the MI significantly. Smoothing in the individual directions increases the MI (Figures 1(a) and 1(b)). This increase is not surprising as this smoothing of the intensities will transform the distribution of observed intensities towards the mean of the image. **The Intensity Scale** (i.e. Parzen-window) behaves as reported in [2] where the optima displaces with increased kernel size (Figures 1(e) and 1(f)). Increasing the size of the Parzen-window corresponds to reducing the number of bins. **The Orientation Scale** has an effect similar to image smoothing (Figures 1(c) and 1(d)). We observe that smoothing results in increased Mutual Information as the diffusion measurements of all 90 directions converge towards the rotation-invariant *mean diffusivity* for $\kappa \to 0$. Note that the corresponding curves of MI using small kernels, i.e. higher angular resolution, only results in a small decrease in the MI. Figures 1(c) and 1(d) shows preservation of the slope of MI towards the optima is observed, revealing a well-defined optima. Locally, Figure 2, we observe a dramatic shift in optima from mean diffusivity $\kappa = 0$ to high directional resolution $\kappa = 30$. As illustrated, the local optima shifts 30-40 degrees as a function of scale, which justify the need for our proper scale-space formulation for similarity of DWI. **To Illustrate** the LOR for DWI with MI, we performed a few affine registrations using MI, $\kappa = 30$, a cubic B-spline Parzen-window with 200 bins, and B-Spline image interpolation. A registration can be seen in Figure 3.

6 Discussion and Conclusion

The LOR for DWI includes directional information and so first-order information of the deformation is required. We therefore restrict the deformation model to diffeomorphisms to ensure well-defined derivatives. For gradient-based optimization, this implies that the second-order information of the deformation is required, which severely complicates any implementation. We have chosen the Watson distribution for its simplicity compared to e.g. a symmetrized von Mises-Fisher kernel or symmetrized geodesic distances.

We have presented an extension of the Locally Orderless Registration for DWI by introducing a scale-space which accounts for the projective nature of DWI in a theoretically sound manner. Our experiments show that directional resolution is important in order to obtain proper local alignment in registration. Our formulation allows us to directly control the scales of the information from which we

estimate the similarity. By extending the LOR framework, we can easily apply a wide range of similarity measures. We provided the first-order information of the densities, briefly reviewed the effects of the scales, and illustrated the approach by affine registration of DWI using Mutual Information.

References

1. Darkner, S., Sporring, J.: Generalized partial volume: An inferior density estimator to parzen windows for normalized mutual information. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 436–447. Springer, Heidelberg (2011)
2. Darkner, S., Sporring, J.: Locally Orderless Registration. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(6), 1437–1450 (2013)
3. Duarte-Carvajalino, J.M., Sapiro, G., Harel, N., Lenglet, C.: A framework for linear and non-linear registration of diffusion-weighted mris using angular interpolation. Frontiers in Neuroscience 7 (2013)
4. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision 50(3), 329–343 (2002)
5. Jupp, P.E., Mardia, K.: A unified view of the theory of directional statistics, 1975-1988. International Statistical Review, 261–294 (1989)
6. Koenderink, J., Van Doorn, A.: The structure of locally orderless images. International Journal of Computer Vision 31(2), 159–168 (1999)
7. Raffelt, D., Tournier, J., Fripp, J., Crozier, S., Connelly, A., Salvado, O., et al.: Symmetric diffeomorphic registration of fibre orientation distributions. NeuroImage 56(3), 1171–1180 (2011)
8. Rathi, Y., Michailovich, O., Shenton, M.E., Bouix, S.: Directional functions for orientation distribution estimation. Medical Image Analysis 13(3), 432–444 (2009)
9. Rohde, G., Barnett, A., Basser, P., Marenco, S., Pierpaoli, C.: Comprehensive approach for correction of motion and distortion in diffusion-weighted mri. Magnetic Resonance in Medicine 51(1), 103–114 (2004)
10. Tao, X., Miller, J.V.: A method for registering diffusion weighted magnetic resonance images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 594–602. Springer, Heidelberg (2006)
11. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K.: The wu-minn human connectome project: an overview. Neuroimage 80, 62–79 (2013)
12. Van Hecke, W., Leemans, A., D'Agostino, E., De Backer, S., Vandervliet, E., Parizel, P.M., Sijbers, J.: Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information. IEEE Transactions on Medical Imaging 26(11), 1598–1612 (2007)
13. Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1(1), 35–51 (1996)
14. Yap, P.T., Chen, Y., An, H., Yang, Y., Gilmore, J.H., Lin, W., Shen, D.: Sphere: Spherical harmonic elastic registration of hardi data. NeuroImage 55(2), 545–556 (2011)

5 Density-based Nonrigid Registration with Explicit Reorientation

In this chapter, we extend the scale-space transformation model for global registration [Jensen et al., 2015] to a nonrigid model. As with any registration framework built on the explicit reorientation of diffusion data, the analytical solution quickly grows in complexity for the nonrigid model and requires more attention to both the analytical formulation and the implementation itself. The solution and experiments represent the primary scientific contribution, intended for the journal IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). However, we also consider the implemented framework as a contribution, since it is a significant challenge to design a computationally feasible DWI registration framework with explicit reorientation, that can transform high-resolution DWI on a high-end laptop.

In the first part of this chapter, we introduce the TPAMI manuscript that is the journal extension of [Jensen et al., 2015]. An earlier version of the nonrigid framework was presented as an abstract at ISMRM in [Jensen et al., 2017]. It showcases initial non-hierarchical results, and discusses our view on improved registration through full DWI profiles and avoiding warped space reconstructions for quantitative evaluation when possible. The 3-page conference submission can be found in Appendix B.

In the second part, we follow up on our journal manuscript by presenting the framework from a computational perspective in terms of implementation, memory requirements, speed, schematic overviews, and pseudocode. We discuss how the gradients can be calculated by taking a concept from computer science called *program slicing*, and how a parallel solution is essential for real-world applications of the framework.

5.1 The Nonrigid Registration Framework

The analytical solution is the result of a long chain rule, partially described in [Jensen et al., 2015]. In the nonrigid solution, we explicitly define the gradients of the deformation parameters with respect to the similarity, in order to apply quasi-Newton optimization methods. The most notable changes are the local free-form deformation (FFD) model [Rueckert et al., 1999], and the similarity measure Normalized Mutual Information (NMI). The FFD model is a well-defined nonrigid deformation model well-suited for complex image modalities, and the similarity measure is more robust than related popular measures, such as correlation coefficient [Rohde et al., 2004]. In the nonrigid framework, we additionally use the normalized version of MI from [Studholme et al., 1999], as it has been argued to also be invariant to the changes in the overlap region between images through the process of registration.

We follow up the manuscript, in Section 5.3, with additional algorithmic details. The reason for this is that the complexity of calculating the gradients of the new nonrigid model is such that a series of extra algorithmic steps have to be taken to make the framework feasible for two DWI scans. A large part of the complexity comes from the reorientation, which itself depends on the Jacobian of the spatial deformation.

Finally, the manuscript is intended for an expert audience. Though well-described, a cost function that involves explicit DWI reorientation is not trivial. As such, we have attached an additional step-by-step walk-through in Appendix C. It consists of two overall sections that follow Figure 5.1. The first section describes every part the deformation model, while the second section goes through the equivalent derivatives. With this appendix, it is our hope that this will give a clearer understanding of the nonrigid solution, and help others replicate and improve upon our work.

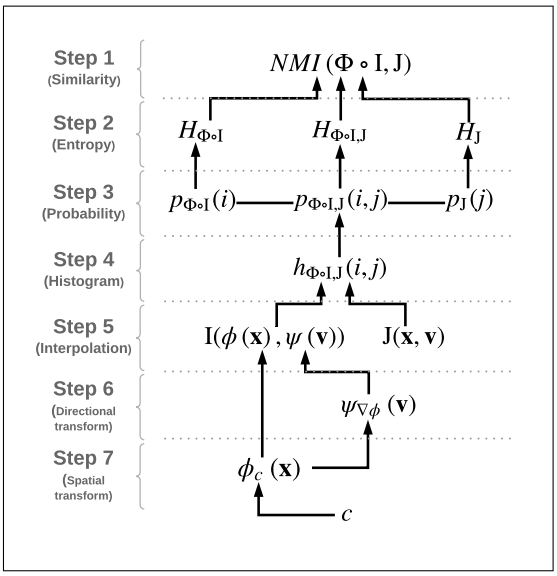


Figure 5.1: LOR-DWI dependency graph.

This ends the brief introduction to the nonrigid LOR-DWI manuscript, which continues on the following page. Implementation details follows the manuscript.

Information-Theoretic Registration with Explicit Reorientation of Diffusion-Weighted Images

Henrik G. Jensen*, Francois Lauze*, and Sune Darkner*

Abstract—Diffusion-weighted MRI (DWI) is the top non-invasive image modality used to study the microstructure of the human brain. The ability to infer and digitally reconstruct neuroanatomy from diffusing molecules has become one of the grand challenges of the 21st century. With a significant growth in sources of publicly available DWI data, groupwise analysis and registration is bound to play a central role in our future understanding of the brain. However, the complex and nonlinear nature DWI requires robust similarity measures on par with multi-modal image analysis, and registration algorithms that takes the directional nature of DWI into account. In this paper, we present a voxel-based nonrigid registration framework for DWI with explicit optimization over the orientational scale, and normalized mutual information as a robust information-theoretic similarity measure for DWI. The framework is a density-based hierarchical scale-space model, that varies and optimizes over both the spatial, directional, and intensity scale. Our results show a promising regularizing effect, that comes inherently from the nonlinear cost function and the increased structural information in DWI data.

Index Terms—Registration, Diffusion-Weighted MRI, Normalized Mutual Information, Explicit Reorientation, Analytical Gradients, Density-Based Formulation, Watson Density Function, Free-Form Deformation, Cubic B-Spline.

1 Introduction

DIFFUSION Weighted Images (DWI) is a non-invasive MRI protocol that can be used to infer cytoarchitecture by tracking the movement of water molecules, otherwise invisible in structural MRI. However, the directional geometry of DWI makes it a challenge in image registration, which is a key tool for comparing and segmenting medical data. In addition, different DWI also often have a non-linear relationship which makes defining the similarity between two DWI difficult [1].

In this paper, we extend the Locally Orderless Registration for Diffusion-Weighted Images (LOR-DWI) [2] framework to a nonrigid formulation. The LOR-DWI is a scale-space framework for DWI density estimation that allows for a wide range of linear and non-linear similarity measures. The directional information is part of the objective function, which results in an explicit optimization over the reorientation of diffusion gradients, both for raw high-angular resolution scans (HARDI) or the topographically inverted Orientation Density Functions (ODF). The density formulation also allows us to optimize over the isoparametric curves, as the 4D image structures are defined in a joint histogram.

Our contribution is a scale-space formulation for DWI that gives a nonrigid registration with explicit reorientation and non-linear similarity measures well-suited for DWI, such as normalized mutual information (NMI). We apply the framework to simulated DWI data and a synthetic deformation for data from a subject of the Human Connectome Project (HCP) [3].

Corresponding author: henrikgjensen@gmail.com
**Department of Computer Science, University of Copenhagen, Denmark*
Manuscript, 2018.

2 Related Work

This work tackles two major challenges in voxel-based registration of DWI: The reorientation of DWI in image registration, and the non-linear similarity between DWI.

Image registration refers to a process that transforms data into a shared coordinate system. For DWI, the common way to register two images is to use scalar-based methods on quantitative measure, such as the fractional anisotropy (FA) or the mean diffusivity (MD) [4]. However, as such methods disregard most of the directional information in DWI, a methods have been developed to also account for the reorientation of the diffusion profile. Most of these are created on top of scalar-based methods and iteratively reorients the gradients based on the deformation field, of which some of the most popular can be found in [5], [6], [7], [8]. However, registration frameworks have been designed with an objective function that explicitly optimizes over the reorientation of the gradients, such as `DTI-TK` [9], `DT-REFineD` [10], and the more recent `DR-TAMAS` [11]. These remain popular frameworks and have been shown to generally outperform scalar-based frameworks [12], [13], [14], [15]. Calculating the analytical gradients required for explicit reorientation, and in a computationally feasible way, is not an easy task.

The image similarity is straight forward for scalar-based registration as any popular scalar-based measure such as the sum-of-squares difference (SSD) [10] or mutual information (MI) [15] can be used. Explicit reorientation strategies must define the similarity over both position and orientation, which can be a daunting analytical task. However, once the full diffusion profile is part of the similarity, such measures can be defined in a well-suited way for the non-linear relationship between DWI. Both [9] and [10] used variations on SSD in the objective function, while MI was used in

[2] which was made possible through the density-based formulation. As argued in [2], the invariant and statistical properties MI makes it a logical choice for DWI where multiple factors results in a more statistical relationship, such as variations in b-values, non-monoexponential behaviour in biological tissue, and inter-scanner variability [1]. MI is often used in standard registration of complex modalities and by extension scalar-based registration of DWI [16] and pre-processing of DWI [17]. MI and normalized MI (NMI) provides a non-linear statistical measure, but it is also likely more functional measures would be well-suited, such as cross-correlation (CC) and normalized CC (NCC). These can be defined from the density-formulation found in [18]. The density-based DWI comes from the generalized way of estimating image similarity measures based on Locally Orderless Images (LOI) [19]. The first mention of LOI in the context of image registration was in [20] where a variational approach to image registration was presented. The LOR framework [18] generalized a range of similarity measures as linear and non-linear functions of density estimates for scalar-valued images.

3 Locally Orderless DWI

3.1 Notations

We start by introducing a few necessary notations. $\Omega \subset \mathbb{R}^3$ is the spatial domain of the images under consideration. A scalar image is a function $I : \Omega \to \mathbb{R}^3$. We assume that we can extend it on the whole $\mathbb{R}^3$, for instance by extending it with 0. The projective space of directions of $\mathbb{R}^3$ is denoted by $\mathbb{P}^2$, and the unit sphere of $\mathbb{R}^3$ by $\mathbb{S}^2$. We will encounter *spatio-directional* images $\boldsymbol{I} : \Omega \times \mathbb{P}^2 \to \mathbb{R}$, which we similarly assume too bee extendable to $\mathbb{R}^3 \times \mathbb{P}^2$. This is necessary in both cases in order to define their spatial smoothing via convolution. We will use the following elementary property: As $\mathbb{P}^2$ can naturally be identified as the quotient $\mathbb{S}^2/\{\pm 1\}$ by the antipodal symmetry, a function $f : \mathbb{P}^2 \to \mathbb{R}$ can be lifted to an antipodal symmetric function $\tilde{f} : \mathbb{S}^2 \to \mathbb{R}$. Conversely, any antipodal symmetric function $g : \mathbb{S}^2 \to \mathbb{R}$ factors through $\mathbb{P}^2$. A spatio-directional image can (and will) be lifted to an antipodal symmetric image $\tilde{\boldsymbol{I}} : \Omega \times \mathbb{S}^2$: $\tilde{\boldsymbol{I}}(\boldsymbol{x}, -\boldsymbol{v}) = \tilde{\boldsymbol{I}}(\boldsymbol{x}, \boldsymbol{v})$. We will denote by $\boldsymbol{I}$ both the spatio-directional image and its antipodal symmetric lifting in the following.

3.2 Recall on the LOR Framework

The LOR framework defines the density estimates over three scales: The image scale σ, the intensity scale β, and the integration scale α. In the context of scalar registration, for a transformation $\phi : \mathbb{R}^3 \to \mathbb{R}^3$, the estimated histogram h and the corresponding density p is computed as

$$h_{\beta\alpha\sigma}(i, j|\phi, \boldsymbol{x}) = \int_{\mathbb{R}^3} P_\beta(I_\sigma(\phi(\boldsymbol{x})) - i) P_\beta(J_\sigma(\boldsymbol{x}) - j) W_\alpha(\boldsymbol{\tau} - \boldsymbol{x}) d\boldsymbol{\tau} \quad (1)$$

$$p_{\beta\alpha\sigma}(i, j|\phi, \boldsymbol{x}) \simeq \frac{h_{\beta\alpha\sigma}(i, j|\phi, \boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma}(k, l|\phi, \boldsymbol{x}) dk\, dl} \quad (2)$$

where $i, j \in [a_1, a_2] =: \Lambda$ are values in the image intensity range, $I_\sigma(\boldsymbol{\phi}(\boldsymbol{x})) = (I * K_\sigma)(\boldsymbol{\phi}(\boldsymbol{x}))$ and $J_\sigma(\boldsymbol{x}) = (J * K_\sigma)(\boldsymbol{x})$ are images convolved with the kernel K_σ with standard deviation σ, P_β is a Parzen-window of scale β, and W_α is a Gaussian integration window of scale α. The marginals are trivially obtained by integration over the appropriate variable. The LOR-approach to similarity lets us use a set of generalized similarity measures, the linear and non-linear

$$\mathcal{F}_{lin} = \int_{\Lambda^2} f(i, j) p(i, j) di\, dj \quad (3)$$

$$\mathcal{F}_{non-lin} = \int_{\Lambda^2} f(p(i, j)) di\, dj \quad (4)$$

where the linear measure $f(i, j)$ includes e.g. sum of squared differences and Huber, and the non-linear $f(p(i, j))$ includes e.g. MI, normalized MI (NMI), see [21] for details.

3.3 The LOR-DWI framework

This work addresses the estimation of the image similarity $\mathcal{F}$ of DWI in the context of nonrigid registration an extension of our previous work [2]. In our context, $\boldsymbol{I}$ and $\boldsymbol{J}$ are spatio-directional signals. Specifically DWI MR attenuation signals at location $\boldsymbol{x}$, for a gradient direction $\boldsymbol{v}$, are modeled by $S(\boldsymbol{x}, \boldsymbol{v}) = S_0(\boldsymbol{x}) e^{-b\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{v})}$ [22] and apparent diffusion coefficients volumes are given by $\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{v}) = -\frac{1}{b} \log \frac{S(\boldsymbol{x}, \boldsymbol{v})}{S_0(\boldsymbol{x})}$. Gradient directions $\boldsymbol{v}$ belong to $\mathbb{S}^2$ but diffusion are orientation-free: $\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{v}) \approx \boldsymbol{I}(\boldsymbol{x}, -\boldsymbol{v})$ and this naturally defines a spatio-directional image $\Omega \times \mathbb{P}^2 \to \mathbb{R}$. In order to apply LOR-DWI, the histogram and density estimates Equation (1) and eq. (2) must be extended to spatio-directional data, and the action of the spatial transformation on the directions must be defined.

We introduce a kernel on the sphere as an extension to the density estimates of LOR to include directional information. This kernel accounts for directional smoothing and defines our LOR-DWI framework. Thus, we extend spatial smoothing to be spatio-directional, such that the directional smoothing preserves this symmetry, and thus the projective structure, via a symmetric kernel $\Gamma_\kappa(\boldsymbol{\nu}, \boldsymbol{v})$ on $\mathbb{S}^2$. We define the smoothed signal $I_{\sigma,\kappa}$ at scales (σ, κ) by

$$\begin{aligned} \boldsymbol{I}_{\sigma\kappa}(\boldsymbol{x}, \boldsymbol{v}) &= \int_{S^2} \left(\int_{\mathbb{R}^3} \boldsymbol{I}(\boldsymbol{\tau}, \boldsymbol{\nu}) K_\sigma(\boldsymbol{\tau} - \boldsymbol{x}) d\boldsymbol{\tau} \right) \Gamma_\kappa(\boldsymbol{\nu}, \boldsymbol{v}) d\boldsymbol{\nu} \\ &= (\boldsymbol{I} * (K_\sigma \otimes \Gamma_\kappa))(\boldsymbol{x}, \boldsymbol{v}) \end{aligned} \quad (5)$$

where $K_\sigma(\boldsymbol{x})$ is a Gaussian kernel with σ standard deviation. We employ a symmetric Watson distribution [23] as $\Gamma_\kappa(\boldsymbol{\nu}, \boldsymbol{v})$ for directional smoothing on $\mathbb{S}^2$, given by

$$\Gamma_\kappa(\boldsymbol{\nu}, \boldsymbol{v}) = C e^{\kappa(\boldsymbol{\nu}^\top \boldsymbol{v})^2} \quad (6)$$

$$C = M(\frac{1}{2}, \frac{1}{d}, \kappa) \quad (7)$$

where $M(\frac{1}{2}, \frac{d}{2}, \kappa)$ the confluent hypergeometric function also called Kummer function ($d = 3$ in our case) [24], $\pm\boldsymbol{v}$ the center of the distribution, and κ the concentration parameter, which is roughly inverse proportional to the variance on the sphere. Because of the symmetry property of the Watson distribution and the antipodal symmetry of $\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{v})$, it is clear that $\boldsymbol{I}_{\sigma\kappa}(\boldsymbol{x}, \boldsymbol{v})$ is antipodal symmetric too. As one alternative, a symmetrized von Mises-Fisher [23] distribution or a symmetrized heat kernel could be considered. W

3.3.1 Action on Orientation

A transformation $\phi : \Omega \to \Omega$ acts on directions via its differential, or Jacobian: at $\boldsymbol{x} \in \Omega$, $J_x\phi$ sends $\mathbb{R}\boldsymbol{v}$ to $\mathbb{R}J_{\boldsymbol{x}}\phi\boldsymbol{v}$ which is well defined as soon as $\det J_{\boldsymbol{x}}\phi \neq 0$ which is the case if we assume that ϕ is diffeomorphic. Via the representation $\mathbb{P}^2 \simeq \mathbb{S}^2/\{\pm 1\}$, we can write $\mathbb{P}_{\boldsymbol{x}}\phi : \{\pm v\} \mapsto \{\pm\frac{J_{\boldsymbol{x}}\boldsymbol{v}}{|J_{\boldsymbol{x}}\boldsymbol{v}|}\}$. We denote by ψ the mapping $(\boldsymbol{x}, \boldsymbol{v}) \mapsto \frac{J_{\boldsymbol{x}}\boldsymbol{v}}{|J_{\boldsymbol{x}}\boldsymbol{v}|}$. Because it will only appear inside an antipodal symmetric kernel, $\psi_{\boldsymbol{x}} = \psi(\boldsymbol{x}, \cdot)$ represents $\mathbb{P}_{\boldsymbol{x}}\phi$ *without ambiguity*. Because of spatial convolutions, we assume that ϕ can be extented to $\mathbb{R}^3$, assuming that it is the identity out of a compact set D containing Ω.

3.3.2 Density, orientation and transformation

We will in the sequel, often omit the parameter $\boldsymbol{x}$. We set $\Phi = (\phi, \psi)$ and we write the joint histogram and density for similarity in image registration as

$$h_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x}) = \int_{\mathbb{R}^3\times S^2} P_\beta(\boldsymbol{I}_{\sigma\kappa}(\phi(\boldsymbol{x}), \psi(\boldsymbol{v})) - i)P_\beta(\boldsymbol{J}_{\sigma\kappa}(\boldsymbol{x}, \boldsymbol{v}) - j) W_\alpha(\boldsymbol{\tau} - \boldsymbol{x})d\boldsymbol{\tau} \times d\boldsymbol{v} \tag{8}$$

$$p_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x}) = \frac{h_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x})dl\ dk} \tag{9}$$

Assume that W is a gaussian kernel with standard deviation α, if we left $\alpha \to \infty$ we obtain full spatial integration and can thus ignore the integration scale α.

3.4 LOR-DWI and Free Form Deformation

In [2], we assumed that ϕ was a global affine transformation, with $\psi_x = \psi$ the projectivization of its linear part. In the present work, we instead assume ϕ to be a nonrigid transformation, and we use Rueckert *et al.* framework [25] where the transformation ϕ is given as a hierarchical spline representation *linearly* parameterized by a spatial grid of control points $\boldsymbol{c}$. In the next subsection, we use its vectorized form, so that $\phi = \boldsymbol{B}\boldsymbol{c}$, $\Phi_{\boldsymbol{c}} = (\phi_{\boldsymbol{c}}, \psi_{\boldsymbol{c}})$. We seek the transformation $\Phi^*_{\boldsymbol{c}}$ which maximizes the regularized NMI

$$\Phi^*_c = \arg\max_c \mathcal{M}(\Phi_c) = \arg\max_c \mathcal{F}(\boldsymbol{I} \circ \Phi_c, \boldsymbol{J}) + \mathcal{S}(\Phi_c) \tag{10}$$

The dependency of $\mathcal{M}$ in $\boldsymbol{c}$ is complex. The following dependency diagram (Fig. 1) for the mutual information term illustrates it.

3.5 Estimation and optimization of similarity

We use quasi-Newton methods to compute an optimum of Equation (10), in particular L-BFGS from [26], and we need to compute its gradient with respect to control points vector $\boldsymbol{c}$. Most of the needed calculations have been provided in [25], [21], [2]. A full formula for the gradient is very long and not very informative. Therefore we only describe how spatio-directional smoothing contributes to it. Thanks to the LOR-DWI representation, it appears only within the spatio-directional smoothing of $\boldsymbol{I}$. One complication comes from our implementation, where, instead of using the Kummer function $M(\frac{1}{2}, \frac{3}{2}, \kappa)$ as normalization constant in the Watson

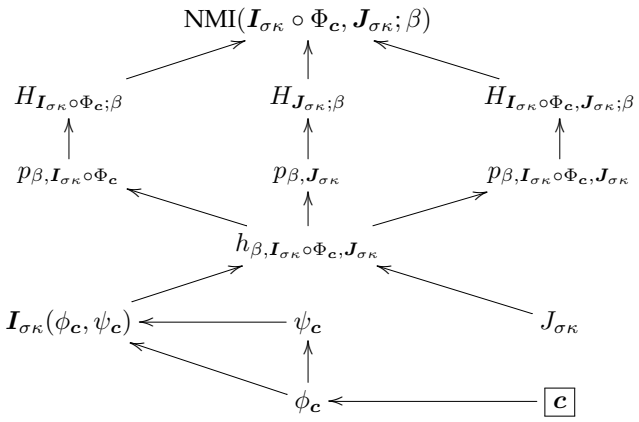


Fig. 1. Dependency graph of the nonrigid DWI registration between the moving image $\boldsymbol{I}$ and the target image $\boldsymbol{J}$, with normalized mutual information (NMI) as the similarity measure. The deformation is parameterized by $\boldsymbol{c}$ so that any change in $\boldsymbol{c}$ will eventually affect the total similarity between the two images.

kernel, we estimate this factor from a discrete set of N directions $\nu_1, \ldots, \nu_N$ at each voxel $\boldsymbol{x}$. It can therefore no longer be considered as constant. We rewrite the discrete spatio-directional smoothing as

$$\boldsymbol{I}_{\sigma\kappa}(\phi_{\boldsymbol{c}}(\boldsymbol{x}), \psi_{\boldsymbol{c}}(\boldsymbol{v})) = \sum_{n=1}^{N} \int_{\mathbb{R}^3} \boldsymbol{I}(\tau, \nu_n) K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)\bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v}))\, d\tau \tag{11}$$

where we have set

$$\bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v}) = \frac{e^{\kappa(\nu_n^\top \psi_{\boldsymbol{c}}(\boldsymbol{v}))^2}}{\sum_{i=1}^{N} e^{\kappa(\nu_i^\top \psi_{\boldsymbol{c}}(\boldsymbol{v}))^2}}. \tag{12}$$

We compute the Jacobian of the spatio-directional smoothing with respect to the control point parameter $\boldsymbol{c}$:

$$J_{\boldsymbol{c}}\boldsymbol{I}_{\sigma\kappa}(\phi_{\boldsymbol{c}}(\boldsymbol{x}), \psi_{\boldsymbol{c}}(\boldsymbol{v})) = \sum_{n=1}^{N} \left(\int_{\mathbb{R}^3} \boldsymbol{I}(\tau, \nu_n) J_{\boldsymbol{c}} K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)\, d\tau \right) \bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v}) + \sum_{n=1}^{N} \left(\int_{\mathbb{R}^3} \boldsymbol{I}(\tau, \nu_n) K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)\, d\tau \right) J_{\boldsymbol{c}}\bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v}) \tag{13}$$

From eq. (13), we have

$$J_{\boldsymbol{c}} K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau) = -\frac{K_\sigma(\boldsymbol{\phi}_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)}{\sigma^2} (\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)^\top \boldsymbol{B}, \tag{14}$$

because K_σ is a Gaussian kernel with variance σ^2 and $\phi_{\boldsymbol{c}} = \boldsymbol{B}\boldsymbol{c}$, its Jacobian $J_{\boldsymbol{c}}\phi_{\boldsymbol{c}}$ is simply $\boldsymbol{B}$. The part involving directional kernel is somewhat more complex, as it involves the Jacobian $J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}$ and the normalizing factor in Equation (12). Set $f(x) = e^{\kappa x^2}$, $f'(x) = 2\kappa x f(x)$, and define $f_i = f(\nu_i \top \psi_c(\boldsymbol{v}))$.

$$J_{\boldsymbol{c}} f(\nu^\top \psi_{\boldsymbol{c}}\boldsymbol{v}) = 2\kappa f(\nu^\top \psi_{\boldsymbol{c}}(\boldsymbol{v}))\psi_{\boldsymbol{c}}(\boldsymbol{v})^\top \nu\nu^\top J_{\boldsymbol{c}}\psi_{\boldsymbol{c}}(\boldsymbol{v}). \tag{15}$$

By a straightforward and careful calculation we obtain

$$J_{\boldsymbol{c}}\bar{\Gamma}_\kappa(\nu_n,\psi_{\boldsymbol{c}}(\boldsymbol{v}))= \tag{16}$$

$$-2\frac{\kappa\bar{\Gamma}_\kappa(\nu_n,\psi_{\boldsymbol{c}}(\boldsymbol{v}))}{\sum_{i=1}^N f_i}\psi_{\boldsymbol{c}}(\boldsymbol{v})^\top\left(\sum_{i\neq n}f_i\nu_i\nu_i^\top\right)J_{\boldsymbol{c}}\psi_{\boldsymbol{c}}(\boldsymbol{v}) \tag{17}$$

To understand $J_{\boldsymbol{c}}\psi_{_c}$, we need to deal with *tensors*. The matrix $\boldsymbol{B}$ is built of cubic B-splines. With k control points, $\boldsymbol{c}$ has dimension $3k$ and $\boldsymbol{B}(\boldsymbol{x})\in\mathbb{R}^{3\times 3k}$. Its spatial Jacobian is a 3D-*tensor* made of quadratic B-splines, $J_{\boldsymbol{x}}\boldsymbol{B}\in\mathbb{R}^{3\times 3k\times 3}$. $J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}=J_{\boldsymbol{x}}\boldsymbol{B}\cdot\boldsymbol{c}$ where the '$\cdot$' operator represents the *contraction* $\mathbb{R}^{3\times 3k\times 3}\times\mathbb{R}^{3k}\to\mathbb{R}^{3\times 3}$, $(r_{uvw},s_v)\mapsto\left(\sum_v r_{uvw}s_v\right)_{uw}$. One has $J_{\boldsymbol{c}}J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\in\mathbb{R}^{3\times 3\times 3k}\neq J_{\boldsymbol{x}}J_{\boldsymbol{c}}(\boldsymbol{B}\boldsymbol{c})=J_{\boldsymbol{x}}\boldsymbol{B}\in\mathbb{R}^{3\times 3k\times 3}$, though the difference is a matter of swapping indices. $J_{\boldsymbol{c}}\psi_{_c}$ is a 3D tensor of the same order as $J_{\boldsymbol{c}}J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}$. Another contraction comes from $J_{\boldsymbol{c}}J_{\boldsymbol{x}}\phi_{_c}(\boldsymbol{v})$. This is a matrix in $\mathbb{R}^{3\times 3k}$ and one can write $J_{\boldsymbol{c}}J_{\boldsymbol{x}}\phi_{_c}(\boldsymbol{v})=J_{\boldsymbol{x}}\boldsymbol{B}\bullet\boldsymbol{v}$ where $\bullet$ is the contraction $\mathbb{R}^{3\times 3k\times 3}\times\mathbb{R}^3\to\mathbb{R}^{3\times 3k}$, $(r_{uvw},t_w)\mapsto\left(\sum_w r_{uvw}t_w\right)_{uv}$. The differentiation of the inner product $\langle J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v},J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v}\rangle$ is given by

$$J_{\boldsymbol{c}}\langle J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v},J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v}\rangle=\left(J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v}\right)^\top\left(J_{\boldsymbol{x}}\boldsymbol{B}\bullet\boldsymbol{v}\right) \tag{18}$$

Denoting by $\boldsymbol{V}$ the vector $J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v}$, so that $\psi_{\boldsymbol{c}}(\boldsymbol{v})=\frac{\boldsymbol{V}}{|\boldsymbol{V}|}$, the Jacobian $J_{\boldsymbol{c}}\psi_{_c}\boldsymbol{v}$ is

$$J_{\boldsymbol{c}}\psi_{_c}\boldsymbol{v}=\left(\mathbf{I}_3-\frac{\boldsymbol{V}\boldsymbol{V}^\top}{|\boldsymbol{V}|^2}\right)\frac{J_{\boldsymbol{x}}\boldsymbol{B}\bullet\boldsymbol{v}}{|\boldsymbol{V}|} \tag{19}$$

where $\mathbf{I}_3$ is the identity of $\mathbb{R}^3$. Note also that $\boldsymbol{V}=J_{\boldsymbol{x}}\phi_{\boldsymbol{c}}\boldsymbol{v}=\left(J_{\boldsymbol{x}}\boldsymbol{B}\cdot c\right)\boldsymbol{v}$ and that $\mathbf{I}_3-\frac{\boldsymbol{V}\boldsymbol{V}^\top}{|\boldsymbol{V}|^2}$ is the orthogonal projection $\boldsymbol{\pi}_{\boldsymbol{V}^\perp}$ onto $\boldsymbol{V}^\perp$, the plane orthogonal to $\boldsymbol{V}$. Putting things together, the Jacobian with respect to the control point parameter $\boldsymbol{c}$ of the spatio-directional smoothing is given by

$$J_{\boldsymbol{c}}\boldsymbol{I}_{\sigma\kappa}(\phi_{\boldsymbol{c}}(\boldsymbol{x}),\psi_{\boldsymbol{c}}(\boldsymbol{v}))=$$
$$-\sum_{n=1}^N\int_{\mathbb{R}^3}\boldsymbol{I}(\tau,\nu_n)K_\sigma(\phi_c(\boldsymbol{x})-\tau)\bar{\Gamma}_\kappa(\nu_n,\psi_{\boldsymbol{c}}(\boldsymbol{v}))\times$$
$$\left\{\frac{(\phi_{\boldsymbol{c}}(\boldsymbol{x})-\tau)^\top\boldsymbol{B}}{\sigma^2}+\frac{2\kappa\boldsymbol{V}^\top\left(\sum_{i\neq n}\nu_i\nu_i^\top\right)}{|\boldsymbol{V}|\sum_{i=1}^N f_i}\boldsymbol{\pi}_{\boldsymbol{V}^\perp}\left(\frac{J_{\boldsymbol{x}}\boldsymbol{B}\bullet\boldsymbol{v}}{|\boldsymbol{V}|}\right)\right\}d\tau. \tag{20}$$

3.6 Regularization

So far, we have not specified the form of the regularizer $\mathcal{S}(\Phi_{\boldsymbol{c}})$ in Equation (10). The regularization has received little attention due to the inherent regularization from the smooth kernels and the additional directional structure. However, we found that the last steps in the hierarchical transformation model, the high resolution of the deformation field, required some regularization to keep the deformation stable. We used a simple regularization that penalizes a non-uniform grid by the squared difference between a point $\boldsymbol{c}_i$ and its direct neighbours. The control points are organized as a family of R grids, from coarse to fine resolution, and the regularizer $\mathcal{S}(\Phi_{\boldsymbol{c}})$ is the sum $\sum_{r=1}^R\mathcal{S}^r(\Phi_{\boldsymbol{c}^r})$ at each resolution, with $\boldsymbol{c}^r=(c_1^r,\dots c_{p_r}^r)^T$ the grid of control points at resolution level r and we denote by $N^r(i)$ the set of indices j such that control point c_j^r is neighbor to control point c_j^r and by $|N^r(i)|$ its cardinal . We set

$$\mathcal{S}^r(\Phi_{\boldsymbol{c}})=-\frac{\lambda_r}{2}\sum_{i=1}^{p_r}\|c_i^r-\frac{1}{|N^r(i)|}\sum_{j\in N^r(i)}c_j^r\|^2. \tag{21}$$

λ_r is a strictly positive parameter controlling the degree of smoothness. The negative sign in Equation (21) comes from the fact that we *maximize* the objective function. In order to compute the gradient of $\mathcal{S}^r(\Phi_{\boldsymbol{c}^r})$, we define series of linear mappings $T_i^r:\boldsymbol{c}\mapsto|N^r(i)|c_i^r-\sum_{j\in N^r(i)}c_j^r$. The regularizer in Equation (21) can be rewritten as

$$\mathcal{S}^r(\Phi_{\boldsymbol{c}})=-\frac{\lambda_r}{2}\sum_{i=1}^{p_r}\frac{1}{|N^r(i)|^2}\|T_i^r\boldsymbol{c}^r\|^2 \tag{22}$$

and by classical manipulation we obtain that

$$\nabla_{\boldsymbol{c}}\mathcal{S}^r(\Phi_{\boldsymbol{c}})=-\lambda_r\sum_{i=1}^{p_r}\frac{1}{|N^r(i)|^2}T_i^{r*}T_i^r\boldsymbol{c}^r. \tag{23}$$

T_i^{r*} is the adjoint of T_i^r. Operator $-\sum_{i=1}^{p_r}\frac{1}{|N^r(i)|^2}T_i^{r*}T_i^r$ is a discrete Laplacian. Our sought gradient is

$$\nabla_{\boldsymbol{c}}\mathcal{S}(\Phi_{\boldsymbol{c}})=-\sum_{r=1}^R\lambda_r\sum_{i=1}^{p_r}\frac{1}{|N^r(i)|^2}T_i^{r*}T_i^r\boldsymbol{c}^r. \tag{24}$$

In this paper, we chose $\lambda_1=\dots=\lambda_R$.

4 Experiments

To illustrate the properties of the proposed method, we conduct a series of experiments on simulated data and artificially warped real data.

4.1 Simulated Examples

The first set of experiments are based on artificially generated distributions of HARDI shells and corresponding ODFs, each representing different fiber constellations. These setups are a good way to visually inspect and validate a framework in a controlled environment. While such artificial experiments can be found in most DWI registration papers, the data is rarely shared. To our knowledge there are no popular open sources of simulated DWI data for comparing registration frameworks, although we noted that the `DIPY` project appears to be a good source for generating simulations [27]. Working in Matlab and C++, we created our own simulated DWI data, which will be free available[1]

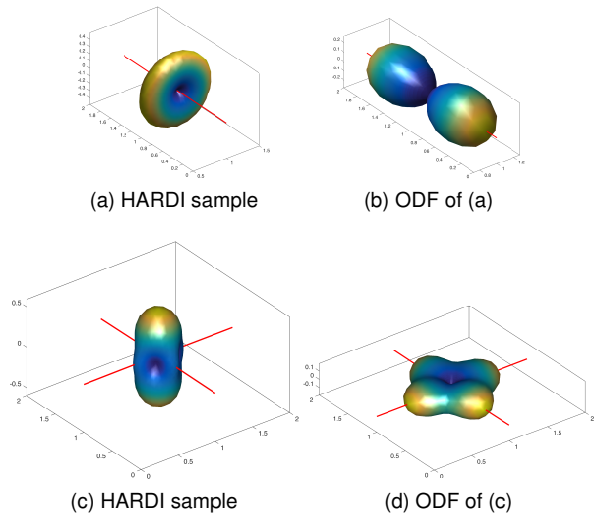
4.1.1 Simulating DWI data

We created all HARDI samples by deforming a unit sphere of equally distributed directions[2] to a certain HARDI or ODF shape. Figure 2 shows two simulated HARDI samples and their corresponding ODFs, where the first is a single fiber ODF, and the second a crossing fiber ODF. DWI samples are antipodal symmetric and every ODF from the simplest to the most complex can be constructed through

1. Contact us on henrikgjensen@gmail.com for the code or examples.

2. The code was provided by [28] who wrote a small toolbox for generating between 9 and 1001 uniformly distributed directions by minimizing the Reisz s-energy configuration of N equal charges confined to the surface of the unit sphere.

a combination of single fiber ODFs. The simulated data is visualized using the regularized QBI algorithm, which uses a linear combination of real spherical harmonics to represent either the direct QBI sample or transformed ODF[29].



(a) HARDI sample (b) ODF of (a)

(c) HARDI sample (d) ODF of (c)

Fig. 2. Simulated DWI samples. The left column shows the raw DWI signal. The right column shows the reconstructed diffusion ODFs that follow anisotropic diffusion. The red lines indicate fiber orientations.

These models can be combined to form simulated DWI tracts in various DWI shapes, such as crossing fiber tracts Figure 3. A 20×20 grid is used through out this section to create blueprints of fiber tract constellations and perform registrations. The images are coloured according to the generalized FA (GFA) value where dark blue regions represent free isotropic diffusion. To simulate a more realistic DWI scenario random uniform noise has been added to the samples.Figure 3.

The simulated voxels with unit density are rescaled for the free diffusion to have a low density or mean diffusivity to resemble real data that has been b_0 normalized,.

4.1.2 Parametric Setup

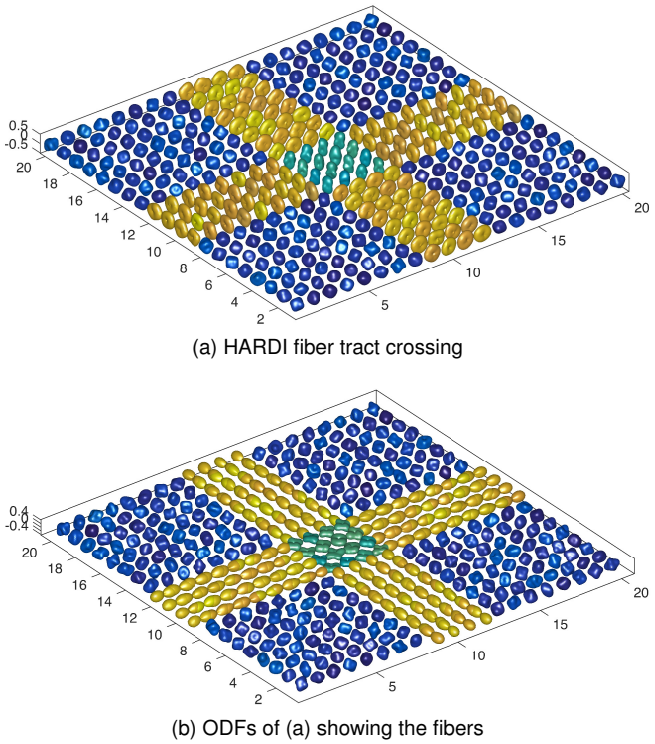
For consistency the same setup is used for all experiments on the simulated data, unless specifically stated otherwise.

Hierarchical mesh resolution. The spacing between the control points is decreased in order to iteratively increase the degrees of freedom in the registration. We use $\Phi_{\text{local}} = \Phi_{\delta=4} + \Phi_{\delta=3.5} + \Phi_{\delta=3} + \Phi_{\delta=2}$. 10 iterations is used for all resolutions except the last, which terminates based on the optimal tolerance of $\epsilon = 1\text{e}{-6}$, or 90 iterations.

Watson concentration, Directional resolution. The concentration parameter is set to $\kappa = 15$, which is sufficiently smooth to represent the 100 uniform directions used.

Spatial resolution. A full spatial resolution is used with a B-spline smoothing at a near-Gaussian variance around $\sigma = 0.6$.

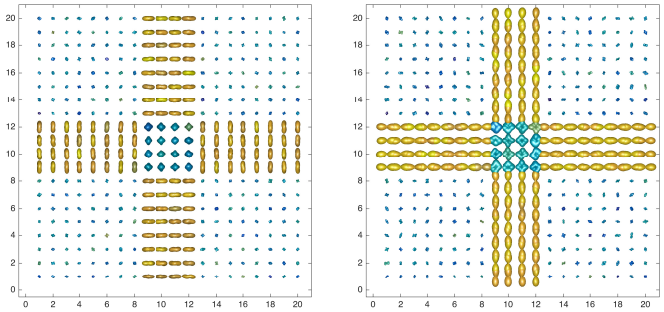
Histogram size. Due to the relatively few fibers being deformed, a small histogram size is used (i.e. intensity resolution) of 20×20 bins (unrelated to the grid size). This allows for some larger, stable but less refined deformations[21].



(a) HARDI fiber tract crossing

(b) ODFs of (a) showing the fibers

Fig. 3. A simulated DWI fiber tract crossing with uniform random noise.



(a) HARDI fiber tract crossing (b) ODFs of (a) showing the fibers

Fig. 4. Simulated DWI fiber tract crossing shown form above. The isotropic diffusion have now been normalized to have a lower mean diffusivity.

HARDI registration, ODF visualization. Note that we will be registering the raw HARDI models, but will visualize the ODFs of the warped data, based on the Funk-Radon transformed (FRT). This is to illustrate that the warped raw data is correctly reoriented and *do not* suffer visibly from affine shearing.

Regularization. We use regularization by a factor of $\lambda = 1\text{e}{-4}$, where 1 is the maximum. The regularization could be omitted in the simulated experiments at the cost of employing multiple levels of resolution of he deformations (large to small), as the simulated data is very structured which may hamper convergence of the optimisation. As an aside, every example below could be highly improved by a situation-specific choice of parameters, but for consistency the same settings were

use for all experiments.

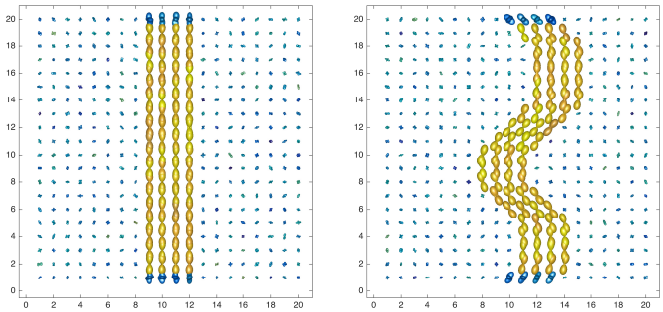
4.2 Experiment 1: Single Fiber Tracts

The first set of experiments maps a straight fiber tract to a curved tract of the same width. It offers an opportunity to discuss some of the differences between a good reconstruction and a correct mapping.

4.2.1 Straight and curved

Three experiments illustrating the different scenarios where straight fibers are registered to wavy fibers. These experiments clearly demonstrate the regularizing effects of using the full diffusion profile for registration.

In the first experiment, the length of the fibers is fixed by adding intersecting fibers at the borders as a shared feature that defines the end and beginning of a segment. In a successful spatial mapping, a correct registration should expand the straight fiber tract to fit the increased length of the curved fiber between the borders as no other features are present. Figure 5 shows the simulated moving (straight) and target (wavy) images.



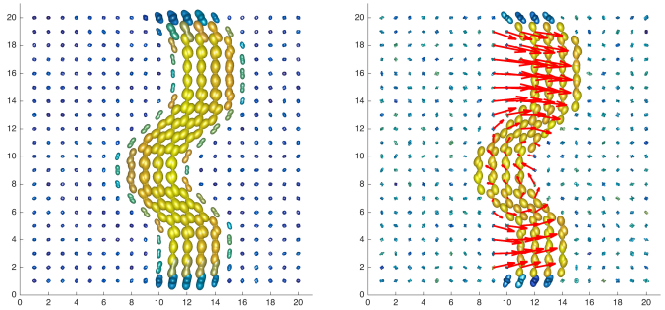
(a) Moving Image (b) Target Image

Fig. 5. Experiment 1: Simulated fiber tract images with fixed boundaries, as illustrated in **??**.

A similar "bounded" DWI example will be used in the all other experiments as well. The results of the first experiment are shown in Figure 6, where 6a is the reconstructed warp, and 6b shows the final spatial mapping from the moving image, overlaid on the original target image. As the figures illustrate the straight fiber is expanding with the curvature as expected, and the reconstructed (smoothed) ODFs from the HARDI-based registration are rotated correctly. Some voxels appear to be a bit off in the spatial mapping, this is due to the lack of surrounding information and is not an issue for a more dense example with more structure throughout the image.

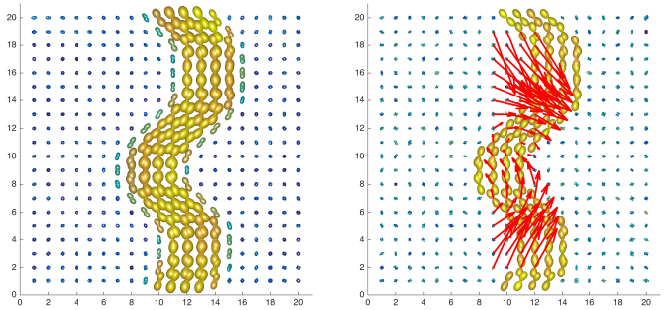
The second experiment is performed without shared features on the boundaries. The results are shown in Figure 7, where we observe that the length is preserved as the straight fiber tract is mapped to a sub-part of the curved tract. Intuitively a correct registration should preserve the length of the straight fiber in this reconstruction.
The fact that this happens with no strong outside regularization, illustrate an inherent regularization effect of the cost function which is defined on both spatial and directional scales.

In the third experiment, the proposed similarity is compared with the equivalent scalar-based registration by performing a mean diffusivity registration ($\kappa = 0$) of the tracts



(a) Reconstructed warped image (b) Spatial mapping

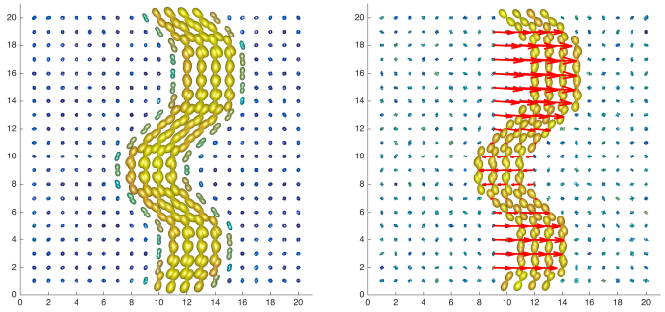
Fig. 6. Experiment 1: Registration of single tract images of varying shape and (given the boundary) length.



(a) Reconstructed warped image (b) Spatial mapping

Fig. 7. Experiment 1: Registration of without boundary fibers. The spatial mapping indicates a nice preservation of length of the straight tract mapped to a subsection of the curved tract.

with no signal on the boundaries. The mean diffusivity carries no directional information. The result can be seen in the Figure 8.



(a) Reconstructed warped image (b) Spatial mapping

Fig. 8. Experiment 1: Scalar-based registration without boundary fibers and no directional directional information in the cost function during optimization.

This pure scalar-based registration is driven by the edges of the simulated tracts only. The reconstruction appears to be corect, but the final spatial mapping indicates a lack of regularization as the fibers are stretched unevenly, the length is not preserved and is from our perspective an inferior mapping.

4.3 Experiment 2: Crossing Fiber Tracts

The second set of experiments are designed to test the registration of crossing fiber tracts.

4.3.1 Straight and shifted

The first experiment examines the framework's ability to register two crossing tracts with a horizontal and vertical shift Figures 9a and 9b. Circular fibers have been added as fixed points in the image to illustrate the local shift of the crossing tracts. The result of the registration is shown in Figures 9c and 9d. The final spatial mapping, shown with arrows, is accurate including the reconstruction, which also shows the added effect of the smoothing.

4.3.2 Two degrees of shearing

The second experiment involving crossing fibers show three fiber tracts crossing under a varying amount of shear with a fixed horizontal tract (Figures 10a and 10b). The purpose is to investigate a relatively large deformation combined with a change to the complex crossing at the center. The results are shown in Figures 10c and 10d. Given that reconstruction has a degree of smoothing, the structure of the complex center-crossing closely matches the orientations of 45 degree crossing fibers. We remind the reader that the registration was not performed on the ODFs but directly on the simulated HARDI models and then subsequently reconstructed.

4.4 Experiment 3: Fanning Fiber Tracts

The kissing fiber experiments involves a high degree of complexity.

4.4.1 Fanning and kissing fiber tracts in a crossing

The first experiment consist of two DWI images, that simulates both fanning (dispersing) and kissing (interleaving) fiber tracts. The moving image in Figures 11a and 11b is a crossing with a few fibers fanning in and out along the vertical line. The target image are two curved tracts fanning in and out, and merging at the central crossing. The results are shown in Figures 11c and 11d, where both the reconstructed warp and spatial mapping show a registration that follows the lines of the original target image. The resulting deformation even manages to move, shrink and turn the center-crossing to fit.

4.4.2 Kissing fiber tracts

The second experiment involves two straight fiber tracts that are registered to two curving and interleaving tracts (Figures 12a and 12b). The result shown in Figures 12c and 12d, displays a large and difficult deformation that by all accounts appears to be a successful registration.

4.5 Synthetic Deformations of Real Data

In this set of experiments, the registration framework is evaluated on real data obtained from the HCP [3]. By introducing a random synthetic warp on a subset of the brain data, a ground truth is obtained where the objective is to register the warped image back to the original image. These types of experiments are useful for evaluating the parameters of the model in a realistic scenario, and for exploring the framework in a guaranteed diffeomorphic scenario. Note that the warped image does not replace real data, and it cannot be compared directly to other frameworks, due to its bias toward the deformation model.

4.5.1 DWI example data

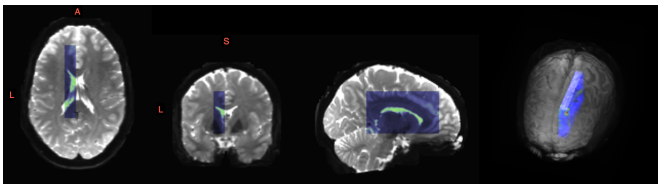


Fig. 13. Selected region of interest for synthetic warp experiments (blue) overlaid on the b_0 image of HCP subject `103818`.

The DWI data for this experiment is shown in Figure 13, where the region of interest (ROI) is the blue overlay on the subjects b_0 image. An ROI of the brain is used ease the visualization of the results. The ROI was chosen at the edge of the corpus callosum (CC) in the left hemisphere due to the characteristic C-shape of the CC and the intersection with other well-known structures e.g. the cingulum. Furthermore, the ROI is in an area with crossing fiber tracts and is near the cortex. A $b = 1000$ DWI volume is used with the ROI being $11 \times 71 \times 41$ at 1.25mm uniform voxel spacing with 90 directions. Only the central sagittal slice along with corresponding ODFs is visualised (Figure 14a), while the deformation is applied to the whole slice. The deformation field for the central slice is shown in Figure 14b.

4.5.2 Studying the effects of various parameters

The ground truth is simply the unwarped image. However, since the similarity measure, NMI, does not reflect a unique point-wise correspondence, three measures for evaluating the resulting registration are introduced: (1) mean squared error (MSE) between coordinates, (2) curl of the deformation field, and (3) divergence of the deformation field. The first defines the average point-wise distance between the target image $\mathbf{J}$ and the moving image $\mathbf{I}$.

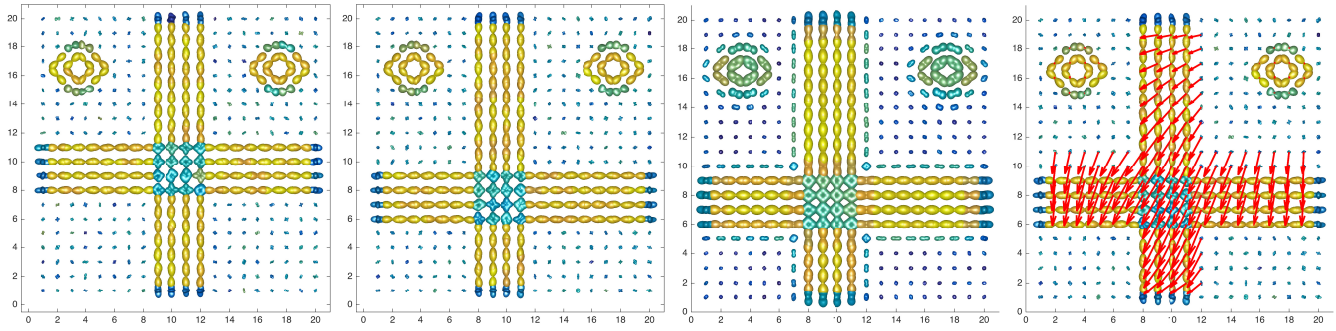
$$MSE(\mathbf{I}(\boldsymbol{x}), \mathbf{J}(\boldsymbol{y})) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{y}_i)^2 \tag{25}$$

The second measure, curl, quantify the amount of orientational change in the deformation field, also referred to as rotation or vorticity. A higher magnitude of the curl is expected for the results of a scalar-based registration compared to a registration over the full diffusion profile. The curl vector is calculated from the Jacobian of the spatial transformation at each voxel in the image

$$\begin{aligned} Curl(\mathbf{I}(\boldsymbol{x})) = &\left(\frac{\partial \boldsymbol{x}_z}{\partial y} - \frac{\partial \boldsymbol{x}_y}{\partial z}\right)\boldsymbol{e}_i + \left(\frac{\partial \boldsymbol{x}_x}{\partial z} - \frac{\partial \boldsymbol{x}_z}{\partial x}\right)\boldsymbol{e}_j + \\ &\left(\frac{\partial \boldsymbol{x}_y}{\partial x} - \frac{\partial \boldsymbol{x}_x}{\partial y}\right)\boldsymbol{e}_k \end{aligned} \tag{26}$$

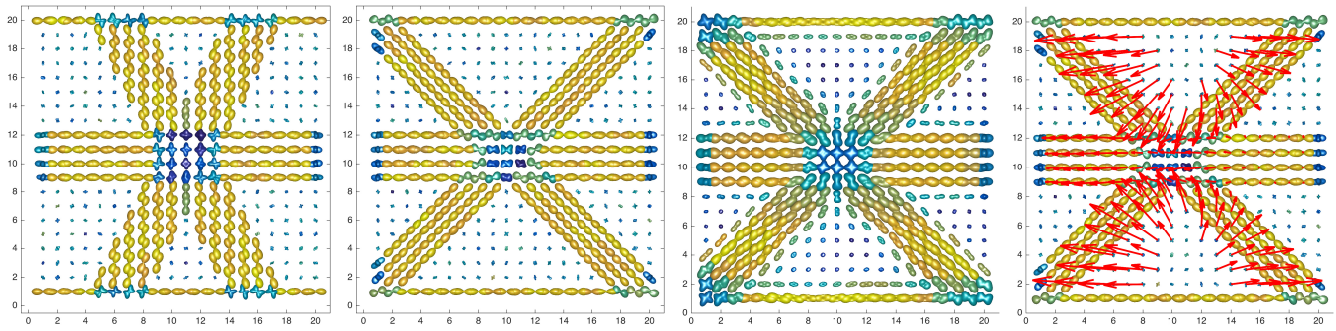
The L_2-norm of the curl vector is used, i.e. the speed of rotation. The third measure, divergence, quantifies the density of the outward flux of a vector field which can be either positive or negative, indicating an expansion or a contraction at a given point. It is a scalar given by the trace of the Jacobian, and we use the L_1 norm of the divergence.

$$Divergence(\mathbf{I}(\boldsymbol{x})) = \frac{\partial \boldsymbol{x}_x}{\partial x} + \frac{\partial \boldsymbol{x}_y}{\partial y} + \frac{\partial \boldsymbol{x}_z}{\partial z} \tag{27}$$



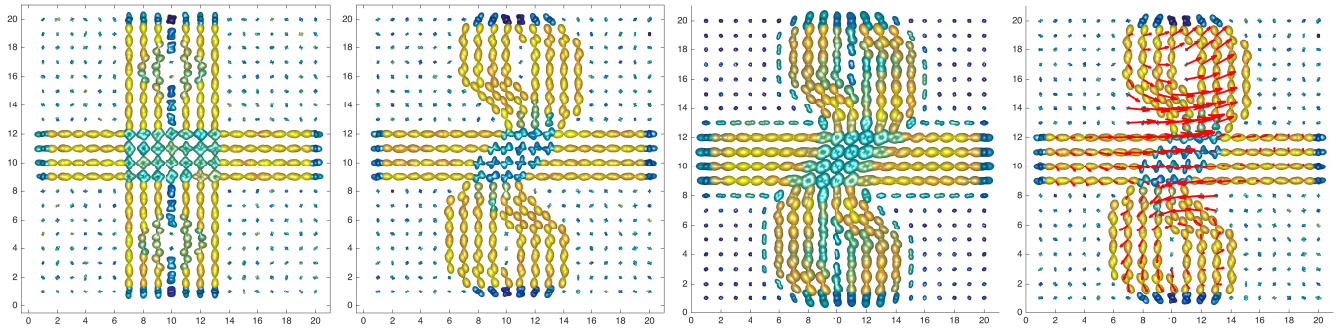
(a) Moving Image (b) Target Image (c) Reconstructed warped image (d) Spatial mapping

Fig. 9. Experiment 2: Simulated crossing tracts (a) with a vertical shift (b). The registered result of the crossing under a both vertical and horizontal shift is reconstructed in (c), and shown with the final spatial mapping from the original position of the moving image in (d).



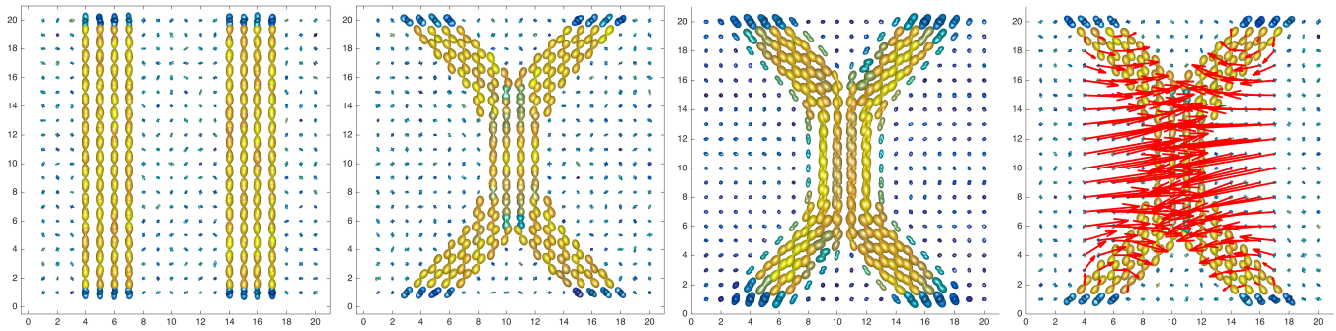
(a) Moving Image (b) Target Image (c) Reconstructed warped image (d) Spatial mapping

Fig. 10. Experiment 2: Simulated crossing tracts with ∼30 degree (a) to 45 degree shearing (b). The reconstruction of the registered result is shown in (c), and the corresponding spatial mapping in (d).



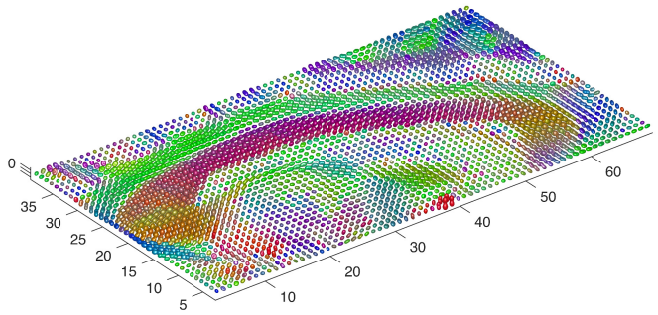
(a) Moving Image (b) Target Image (c) Reconstructed warped image (d) Spatial mapping

Fig. 11. Experiment 3: (a) Simulated fanned opening in the moving image, and (b) an interleaving curved target image. The reconstruction is shown in (c), and the spatial mapping in (d).

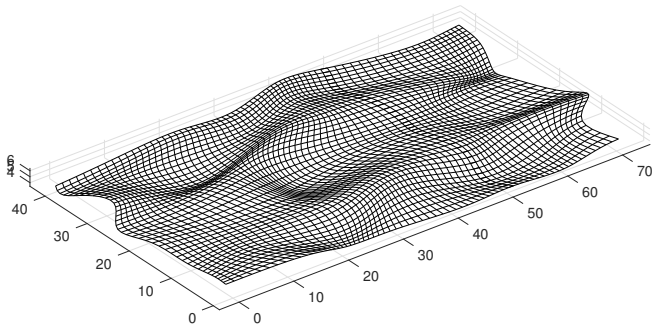


(a) Moving Image (b) Target Image (c) Reconstructed warped image (d) Spatial mapping

Fig. 12. Experiment 3: Simulated straight (a) and kissing (b) fiber tracts. The reconstruction is shown in (c), and the spatial mapping in (d).



(a) Sagittal slice at ROI center.



(b) Deformation field.

Fig. 14. Original central ROI slice with ODFs (a), and the deformation to be applied (b).

These three measures provide information about the magnitude of the voxel-wise distance and the state of the final deformation field. For an example of both curl and divergence as a (regularization) part of a nonrigid registration frame also based on B-splines and NMI, we refer to the scalar-based example by [30].

In the following, experiments are performed over the intensity scale, the orientation scale, and the spatial scale. It is the first two that sets this registration framework apart form most other frameworks is investigated:

1. Isoparametric curves. The density-based formulation allows us to smooth the image inverse proportionally to the image gradients, such as the borders near the CSF, which otherwise can be strongly affected by partial volume effects. When we iteratively decrease the control point-spacing in the FFD model to get an increasingly localized registration, it stands to reason that the intensity range should be increased accordingly, to optimize transformations over relatively isotropic (flat) regions. This is done by starting with a small histogram with a fixed degree of smoothing according to the number of bins and increasing the number of bins after each termination of the optimization.

2. Explicit reorientation. The directional information is expected to provide a more stable and regularized transformation. To investigate this hypothesis, different levels of directional smoothing are examined all the way to a scalar-based mean diffusivity registration at the concentration parameter $\kappa = 0$. The exsperiments on simulated data revealed how the directional information resulted in improved regularized solutions, and these experiments seek to uncover if these observations hold on real data.

Regarding the **spatial scale**, performing multi-resolution registration is a key element in most high resolution registration frameworks, e.g. in the FFD model [25], `ANTs` [5], `Elastix` [31], `FSL` [7], etc. It provides stability while allowing for large and small deformations and reduces the computational complexity. While experiments regarding the spatial part will be performed, its impact on registration quality is fairly well known - in contrast to isocurves and explicit reorientation.

4.5.3 Parametric Setup

Hierarchical mesh resolution. As in the simulated experiments, the spacing between the control points is decreased . We use the composition $\Phi_{\text{local}} = \Phi_{\delta=10} \circ \Phi_{\delta=5} \circ \Phi_{\delta=3.5} \circ \Phi_{\delta=3}$, where the control point spacing δ is scaled down for the smaller sides if the image is not equilateral. The high initial spacing permits the generation of a large random valid synthetic warp at around half of the control point spacing ($\sim 0.4\delta$) [32]. The quasi-Newton L-BFGS optimizer runs for 50 iterations for all resolutions, unless the optimality condition of $\epsilon = 1\text{e}{-6}$ is fulfilled. Importantly, we refer to the result after optimizer termination as *a step*.

Accumulated curl and divergence. While the MSE is easy to calculate at each step, the curl and divergence depend on the first-order derivatives of the spatial deformation, which are accumulated over each step. Thus the final curl and divergence would be defined by product of Jacobians for each step $Curl(\Phi_{\delta=3}) = \boldsymbol{J}(\Phi_{\delta=10})\boldsymbol{J}(\Phi_{\delta=5})\boldsymbol{J}(\Phi_{\delta=3.5})\boldsymbol{J}(\Phi_{\delta=3})$ for all voxels.

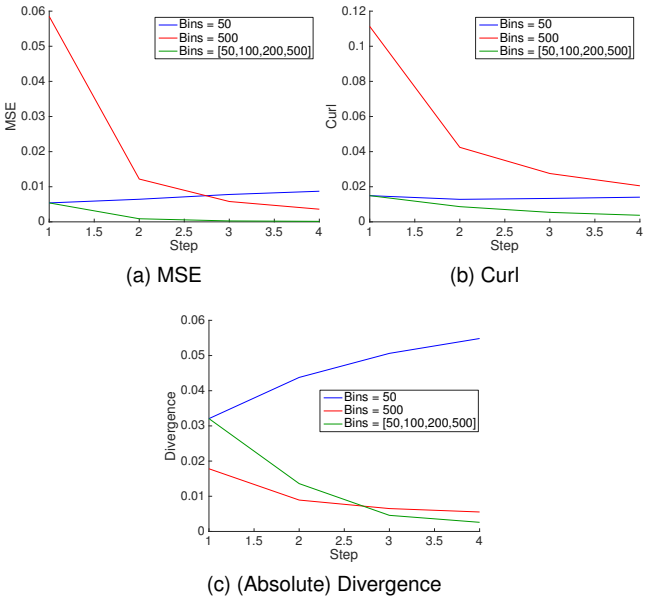
Changes in resolution. Changing the size of the histogram between steps simply requires a rescaling of the intensity range of the images being registered. Changing spatial resolution is more complicated but can be solved by composition.

Other fixed parameters. The regularization is fixed to $\lambda = 1\text{e}{-4}$, and we interpolate and optimize over 30 interpolated orientations of the 90 in the HCP data.

All the following experiments will show results of four accumulating steps. All error measures are reported for the entire ROI and not just the slice visualized.

4.5.4 The Quantitative Effect of Isoparametric Curves

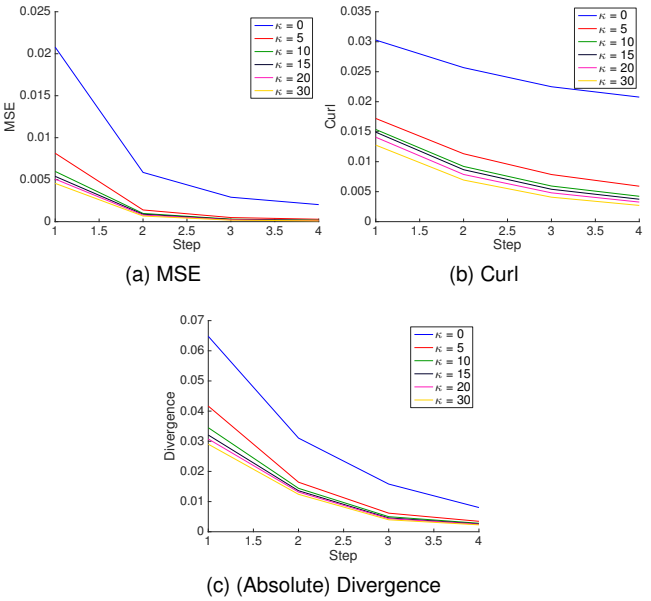
In this experiment, the effect of changing the size of the smoothed joint histogram is examined. Three tests are performed with (i) a fixed histogram of 50×50 bins, (ii) a histogram with 500×500 bins, and (iii) gradually increasing the histogram size in 50, 100, 200 and 500 bins. Figure 15 shows the results in terms of MSE, curl and divergence as mean value over all points along. It is evident how a small histogram with wide isocurves results in an initially faster convergence as higher gradients are required to displace a region of Figure 15 (blue line). However, wide isocurves result in flat regions with small gradients and little structure which causes the result to deteriorate as the degrees of freedom in the transformation is increased with each step. In contrast, starting with a high-resolution of the histogram with thin isocurves has the opposite effect, generating too small gradients in the system with fewer degrees of freedom



(a) MSE (b) Curl

(c) (Absolute) Divergence

Fig. 15. Results from testing the effect of the number of bins through the four step registration. The lines are the mean over all points. The best results are (a): green at $1.42\text{e}{-4}$, (b): green at $3.74\text{e}{-3}$, and (c): green at $2.60\text{e}{-3}$.



(a) MSE (b) Curl

(c) (Absolute) Divergence

Fig. 16. Results from testing the effect of smoothness of the directional interpolation through the four step registration. The lines are the mean over all points. The best results are the yellow line at $\kappa = 30$ with (a): $9.17\text{e}{-5}$, (b): $2.71\text{e}{-3}$, and (c): $2.29\text{e}{-3}$. However, the overall difference between $\kappa = 10$ and $\kappa = 30$ is not more than around $1\text{e}{-5}$.

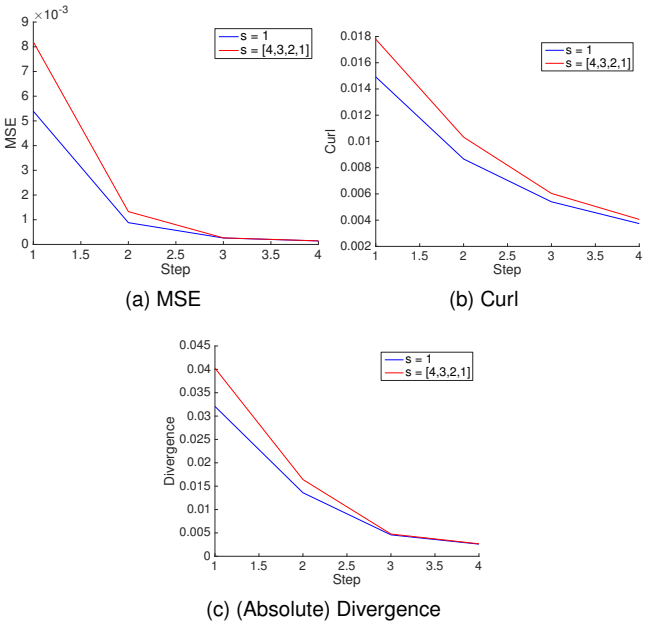
in the initial steps (red line). By iteratively refining the joint histogram, we allow for a wide to thin movement in the image, which provides superior result (green line).

4.5.5 The Quantitative Effect of the Orientation Scale

In the second experiment, we investigate if directional information increases the stability and improves the registration. The size progression of the histogram from the previous experiment (i.e. [50,100,200,500]) is reused, and the concentration parameter κ is varied from mean diffusivity at $\kappa = 0$ to sharp angular features at $\kappa = 30$. The results are shown in Figure 16. It shows how the directional information results in better registration, with significantly less curl than the scalar registration at $\kappa = 0$. Using the best value $\kappa = 30$ is stable in high directional resolution data such as the HCP. As an aside we suggest to use $\kappa = 15$ for low resolution data as this is should suffice.

4.5.6 The Quantitative Effect of the Spatial Resolution

In the last experiment, we use $\kappa = 15$, set the bins to $[50, 100, 200, 500]$, and investigate the effects of changing the spatial scale. The spatial resolution is set to $s = [4, 3, 2, 1]$, which equivalent of smoothly interpolating every 4th point, followed by every 3rd, etc. As with the control point spacing of the deformation field, we scale this to fit the image if the image is not equilateral. For instance, if the image has the spatial dimension $100 \times 150 \times 50$ then space between spatial interpolations for $s = 3$ will be $[2, 3, 1]$ with a bound on no less than 1 (i.e. full resolution). The results are shown in Figure 17. The results are very similar in the final step. However, hierarchical resolution approach compared to the full resolution gave a speedup of a factor 2.4.



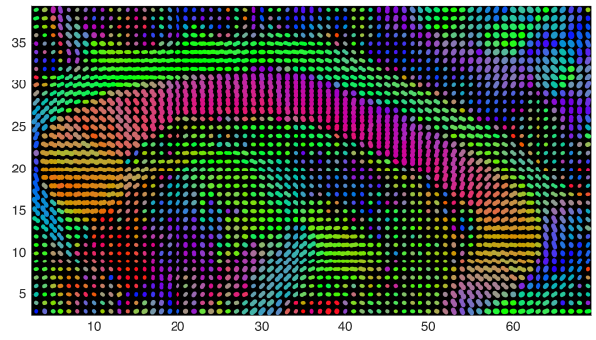
(a) MSE (b) Curl

(c) (Absolute) Divergence

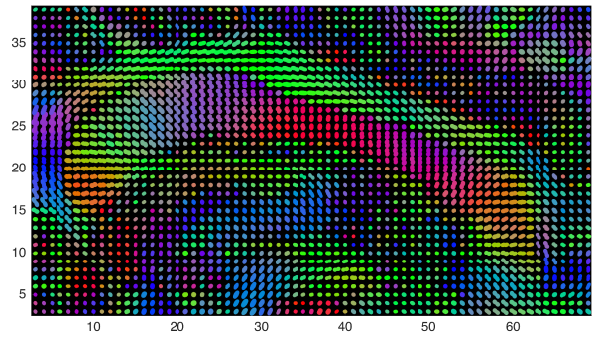
Fig. 17. Results from testing the effect of changing the spatial resolution from low to full. The lines are the mean over all points. For the full (blue) and low-to-full (red) resolution, the difference in results are (a): blue $1.42\text{e}{-4}$ vs red $1.58\text{e}{-4}$, (b): blue $3.74\text{e}{-3}$ vs red $4.21\text{e}{-3}$, and (c): blue $2.60\text{e}{-3}$ vs red $2.70\text{e}{-3}$.

4.5.7 A Qualitative Example of the Results

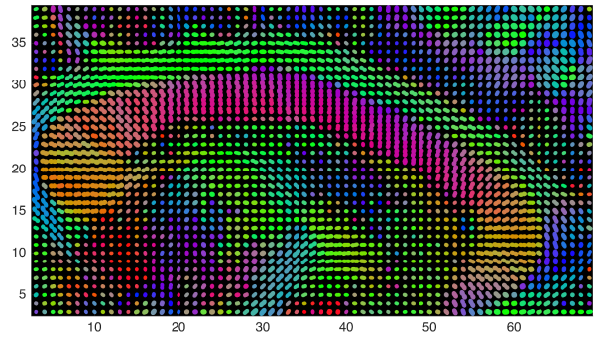
Finally, we visualize the registration to perform a qualitative evaluation of the warp, and the shape and orientation of the individual ODFs. Similar to the simulated experiments, all registrations are based on the raw, noise-corrected HARDI data, while we show the tomographic inversion (FRT) in-



(a) Ground truth

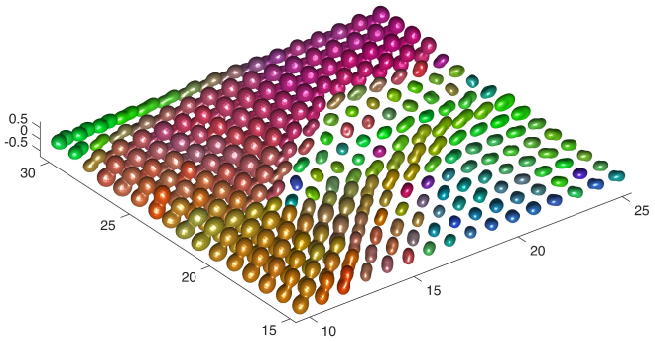


(b) Warped image

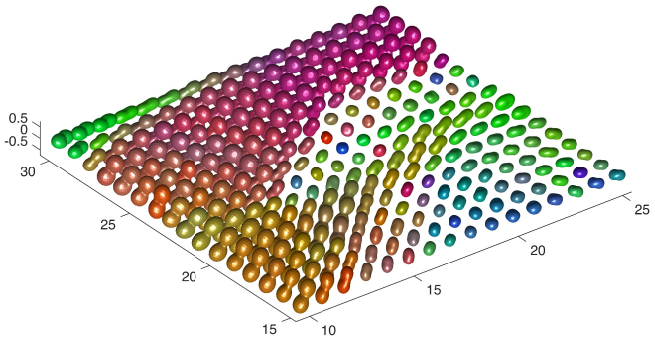


(c) Registered image reconstruction

Fig. 18. 2D visualization of Figure 14a, and the reconstructed warped image after applying the deformation field to the original image. We will be registering (b) back to (a).



(a) Ground truth



(b) Registered image reconstruction

Fig. 19. Same as Figure 18, zoomed in on the left side (anterior) of the figure (around the Genu). The images are not a 100 percent the same, but the difference is hard to notice, and the registration more than adequate.

dicating the direction of the diffusion and likely fiber tract orientations. Unlike the simulated experiments, we fit a B-spline to the image prior to deforming and visualizing the result, which effectively removes most of the smoothing effect applied during registration and spatial interpolation. The results of using an increasing control point resolution with $\kappa = 15$, bins $= [50, 100, 200, 500]$, and spatial resolution $= [4, 3, 2, 1]$ is shown for the central ROI slice in Figure 18, along with a zoomed in version in Figure 19.

5 Conclusion

We have presented a scale-space formulation of density estimation that extends LOR to spatio-directional data, including registration of DWI data with explicit reorientation of the full diffusion profile. We have provided empirical evidence that the underlying structure of the data is preserved during registration, while providing excellent registration results through a number of classical artificial examples, for which registration is known to be difficult. In addition the formulation of the similarity itself provides regularization through the additional information provided by the orientational dimension of the data which which is illustrated clearly in some of the artificial examples. We have investigated the different scales provided by the framework and shown how the different parameters influences the registration results. LOR-DWI provides a smooth but well matched deformation, and the final registration results are improved by integrating the orientational information in the objective function.

References

[1] H. Johansen-Berg and T. E. Behrens, *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press, 2013.

[2] H. G. Jensen, F. Lauze, M. Nielsen, and S. Darkner, "Locally orderless registration for diffusion weighted images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 305–312.

[3] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium *et al.*, "The wu-minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[4] L. J. O'Donnell, A. Daducci, D. Wassermann, and C. Lenglet, "Advances in computational and statistical diffusion mri," *NMR in Biomedicine*, 2017.

[5] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ants)," *Insight j*, vol. 2, pp. 1–35, 2009.

[6] J. Tournier, F. Calamante, A. Connelly *et al.*, "Mrtrix: diffusion tractography in crossing fiber regions," *International Journal of Imaging Systems and Technology*, vol. 22, no. 1, pp. 53–66, 2012.

[7] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.
[8] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.
[9] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, "Deformable registration of diffusion tensor mr images with explicit orientation optimization," *Medical image analysis*, vol. 10, no. 5, pp. 764–785, 2006.
[10] B. T. Yeo, T. Vercauteren, P. Fillard, J.-M. Peyrat, X. Pennec, P. Golland, N. Ayache, and O. Clatz, "Dt-refind: Diffusion tensor registration with exact finite-strain differential," *IEEE transactions on medical imaging*, vol. 28, no. 12, pp. 1914–1928, 2009.
[11] M. O. Irfanoglu, A. Nayak, J. Jenkins, E. B. Hutchinson, N. Sadeghi, C. P. Thomas, and C. Pierpaoli, "Dr-tamas: Diffeomorphic registration for tensor accurate alignment of anatomical structures," *NeuroImage*, vol. 132, pp. 439–454, 2016.
[12] Y. Wang, A. Gupta, Z. Liu, H. Zhang, M. L. Escolar, J. H. Gilmore, S. Gouttard, P. Fillard, E. Maltbie, G. Gerig *et al.*, "Dti registration in atlas based fiber analysis of infantile krabbe disease," *Neuroimage*, vol. 55, no. 4, pp. 1577–1586, 2011.
[13] P. Zhang, M. Niethammer, D. Shen, and P.-T. Yap, "Large deformation diffeomorphic registration of diffusion-weighted imaging data," *Medical image analysis*, vol. 18, no. 8, pp. 1290–1298, 2014.
[14] Y. Wang, Q. Yu, Z. Liu, T. Lei, Z. Guo, M. Qi, and Y. Fan, "Evaluation on diffusion tensor image registration algorithms," *Multimedia Tools and Applications*, vol. 75, no. 13, pp. 8105–8122, 2016.
[15] Y. Wang, Y. Shen, D. Liu, G. Li, Z. Guo, Y. Fan, and Y. Niu, "Evaluations of diffusion tensor image registration based on fiber tractography," *Biomedical engineering online*, vol. 16, no. 1, p. 9, 2017.
[16] W. Van Hecke, A. Leemans, E. D'Agostino, S. De Backer, E. Vandervliet, P. M. Parizel, and J. Sijbers, "Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information," *IEEE transactions on medical imaging*, vol. 26, no. 11, pp. 1598–1612, 2007.
[17] J. M. Treiber, N. S. White, T. C. Steed, H. Bartsch, D. Holland, N. Farid, C. R. McDonald, B. S. Carter, A. M. Dale, and C. C. Chen, "Characterization and correction of geometric distortions in 814 diffusion weighted images," *PloS one*, vol. 11, no. 3, p. e0152472, 2016.
[18] S. Darkner and J. Sporring, "Locally orderless registration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1437–1450, 2013.
[19] J. J. Koenderink and A. J. Van Doorn, "The structure of locally orderless images," *International Journal of Computer Vision*, vol. 31, no. 2, pp. 159–168, 1999.
[20] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras, "Variational methods for multimodal image matching," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 329–343, 2002.
[21] S. Darkner and J. Sporring, "Locally Orderless Registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1437–1450, 2013.
[22] X. Tao and J. V. Miller, "A method for registering diffusion weighted magnetic resonance images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 594–602.
[23] P. Jupp and K. Mardia, "A unified view of the theory of directional statistics, 1975-1988," *International Statistical Review/Revue Internationale de Statistique*, pp. 261–294, 1989.
[24] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,*. Dover, 1974.
[25] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 1999.
[26] M. Schmidt, "minfunc: unconstrained differentiable multivariate optimization in matlab," *Software available at https://www.cs.ubc.ca/ schmidtm/Software/minFunc.html*, 2005.
[27] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, and I. Nimmo-Smith, "Dipy, a library for the analysis of diffusion mri data," *Frontiers in neuroinformatics*, vol. 8, p. 8, 2014.
[28] A. Semechko, "Suite of functions to perform uniform sampling of a sphere," *Software available at https://www.mathworks.com/matlabcentral/fileexchange/37004-suite-of-functions-to-perform-uniform-sampling-of-a-sphere, accessed 2017.*, 2012.
[29] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, "Regularized, fast, and robust analytical q-ball imaging," *Magnetic resonance in medicine*, vol. 58, no. 3, pp. 497–510, 2007.
[30] S. Riyahi-Alam, M. Peroni, G. Baroni, and M. Riboldi, "Regularization in deformable registration of biomedical images based on divergence and curl operators," *Methods of information in medicine*, vol. 53, no. 01, pp. 21–28, 2014.
[31] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 196–205, 2010.
[32] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers, "Diffeomorphic registration using b-splines," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2006, pp. 702–709.

Henrik G. Jensen received his Masters Degree in Computer Science from the University of Copenhagen in 2014. Shortly after, he was awarded an open grant and started as a PhD fellow in the same department, studying medical image analysis and registration of diffusion-weighted MRI. His research interests also include machine learning and data analysis, and before finishing his Bachelor's Degree, he had already defended his first international conference paper on automatic diagnosis of rare diseases. He has also worked as a research assistant analyzing aerial drone footage with the purpose of automatic weed detection to reduce the use of agricultural herbicides.

Francois Lauze studied Mathematics in France, University of Nice Sophia Antipolis where received a PhD in Algebraic Geometry in 1994. He spent some years in Burkina Faso, West Africa, where he taught Mathematics at the University of Ouagadougou. He then moved to Denmark and engaged in yet another PhD, at the IT University of Copenhagen, which he received in 2004. He worked on Variational Methods for Motion Compensated Inpainting and motion recovery among other.He currently holds a position as Associate Professor in Mathematical Image Analysis at the department of Computer Science, University of Copenhagen. He has mainly worked with variational methods, and geometry for Image Analysis - mainly Riemannian but also some metric geometry. More recently with inverse problems in photometric stereo and tomographic imaging.

Sune Darkner received his Masters Degree in Applied Mathematics in 2004. In collaboration with Oticon A/S, he received his Industrial Ph.D. degree in Shape and Deformation Analysis of the Human Ear Canal in 2009, from the Department of Informatics and Mathematical Modelling, at the Technical University of Denmark (DTU). After holding a position at an energy company as data analyst, he rejoined the Department of Informatics and Mathematical Modelling at DTU in 2009 as a post doc. He currently holds a position as Associate Professor in image analysis at the Department of Computer Science, University of Copenhagen. Research interests include image registration and classification, optimization and regularization, and computational physics.

5.3 The Framework from a Computational Perspective

The primary contribution of the nonrigid registration framework lies in the formulation of the analytical solution in the previous section. However, the implementation of such a framework is a significant challenge in itself, due to high memory requirements, speed, and numerical precision of the analytical gradients. The current version of the framework can run on a modern laptop, processing HARDI-based DWI data. We add this as an addendum to the TPAMI manuscript and it should provide a better overview of the method through pseudocode and additional performance measurements.

In the first part, we define a few more steps of the pipeline required to build a complete registration framework. These include the optimization function and a hierarchical scheme that progressively increases the resolution of both the scale-spaces and the nonrigid deformation field.

In the second part, we describe the computational challenges. These are how the implementation deals with heavy memory requirements using program slicing, how the calculation time is decreased through parallelization, and how the numerical precision of the analytical solution is evaluated.

We reused the step-wise and less compact notation from appendix C, which provided an extended description of the analytical solution and was meant to clarify implementational details.

5.3.1 The Registration Pipeline

In Algorithm 1, we describe the overall nonrigid registration algorithm in a similar format as the pseudocode in [Rueckert et al., 1999]. The initial global affine alignment and increase in control point resolution is described in the following paragraphs. The target image is re-interpolated spatially and directionally to achieve the same degree of smoothing as the moving image. The `while`-loop represents the optimization function, where λ is the step length and ϵ is the stopping criterion. The loop can be replaced by any (cost) function that takes Φ_c as input and outputs the regularized similarity $\mathcal{M}(\Phi_c)$ and its analytical gradients $d_c\mathcal{M}(\Phi_c)$. In our current implementation, this loop is replaced with the L-BFGS algorithm in the `minFunc` framework.

Global alignment

Global registration is the first step of any nonrigid registration algorithm, as a detailed local mapping requires a good initial alignment of the scale, rotation, position, and stretch. Our transformation model is thus a combined

Algorithm 1 The overall steps in the registration of two DWI volumes

Input

I : original (moving) image.
J : reference (target) image.

Output

Φ_c : control points of the deformation field.

Algorithm

normalize intensities of both images to fall within histogram bin range.
interpolate the target image with kernels Γ_κ and K_σ.
calculate the optimal global affine alignment Φ_{global}.
initialize the mesh of control points $\Phi_c = \Phi_{\text{global}}$ at spacing δ_c.
repeat
  calculate gradients of the cost function with respect control points $d_c\mathcal{M}(\Phi_c)$.
  while $||d_c\mathcal{M}(\Phi_c)|| < \epsilon$ **do**
    recalculate the control points $\Phi_c = \Phi_c + \lambda \frac{d_c\mathcal{M}(\Phi_c)}{||d_c\mathcal{M}(\Phi_c)||}$.
    recalculate the analytical gradients $d_c\mathcal{M}(\Phi_c)$.
  end while
  increase control point resolution by decreasing mesh spacing δ_c.
  increase spatial resolution by interpolating on a finer grid.
  increase intensity resolution by increasing the size of the histogram.
until finest control point mesh resolution is reached.

transformation

$$\Phi(\boldsymbol{x}, \boldsymbol{v}) = \Phi_{\text{global}}(\boldsymbol{x}, \boldsymbol{v}) \circ \Phi_{\text{local}}(\boldsymbol{x}, \boldsymbol{v}) \ . \tag{5.1}$$

To increase the robustness and speed, we have split the global transformation into a translation, followed by an affine transformation

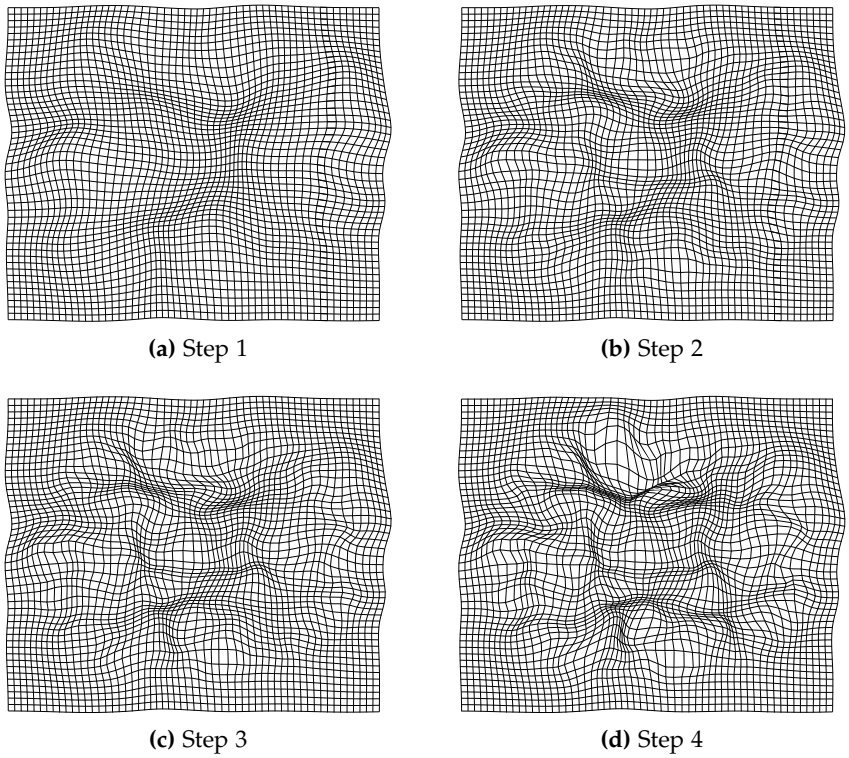
$$\Phi_{\text{global}}(\boldsymbol{x}, \boldsymbol{v}) = \Phi_{\text{translation}}(\boldsymbol{x}, \boldsymbol{v}) \circ \Phi_{\text{affine}}(\boldsymbol{x}, \boldsymbol{v}) \ . \tag{5.2}$$

This ensures that the affine transformation is calculated based on overlapping images with approximately the same center of mass. We only require the global alignment to be an approximate alignment, which means this part can be calculated using common simple methods, like `FSL FLIRT` on the FA or MD images. We use the same optimizer as for the nonrigid part, `minFunc` by [Schmidt, 2005], where we provide a cost function that applies the global transformation and returns the NMI score under the same interpolation as the nonrigid model. We let `minFunc` optimize the 12 parameters of the affine transformation by forward-difference numerical differentiation.

Optimizer

We use the popular quasi-newton L-BFGS[1] optimization algorithm from the efficient optimization framework `minFunc` by [Schmidt, 2005]. The algorithm approximates the Hessian with a low-rank approximation. Since it has a linear to super-linear convergence and linear memory requirement, L-BFGS is particularly well-suited for optimization problems with a large number of degrees of freedom, such as the FFD model used. A similar prominent example with FFD and a bounded version of L-BFGS can be found in [Mattes et al., 2003]. Brief attempts to use another popular built-in optimization scheme, the Stochastic Conjugate Gradient method, resulted in similar results, however with a much lower convergence rate.

Hierarchical Registration



(a) Step 1 **(b)** Step 2

(c) Step 3 **(d)** Step 4

Figure 5.2: Axial views of a hierarchical nonrigid 4-step deformation from a real HCP subject registration. Each step is the result of an optimized alignment for a set of parameters. Each of these steps are calculated on top of the accumulated effect of previous lower resolution alignment steps.

[1] Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS).

In order to avoid local minima, we use a hierarchical multi-resolution optimization scheme. Unlike [Rueckert et al., 1999] who use a subdivision scheme for increasing the resolution of control points, we take the more direct approach where we combine a set of FFD optimization runs, or steps, at an increasing granularity. Repeating the notation of eq. (5.1), let Φ^t_{local} denote a mesh resolution at knot-spacing δ_t where we have left out $\boldsymbol{x}$ and $\boldsymbol{v}$, and let $\Phi^{t+1}_{\text{local}}$ be the next step, which is a higher resolution and lower knot-spacing such that $\delta_t > \delta_{t+1}$. The composition of transformations then defines a final local transformation given by

$$\Phi^n_{\text{local}} = \Phi^{t+n}_{\text{local}}(\Phi^{t+n-1}_{\text{local}}(\cdots(\Phi^{t+1}_{\text{local}}(\Phi^t_{\text{local}})))\cdots) \tag{5.3}$$

where n is the highest resolution of control points. A multi-level scheme reduces the need for regularization as the optimization is more gradual and avoids local minima. However, every time the resolution of the control mesh increases, it raises the requirements of regularization to avoid local overfitting or creating a non-invertible deformation[2]. Our final deformation model can be written out as

$$\Phi(\boldsymbol{x},\boldsymbol{v}) = \Big(\Phi_{\text{translation}}(\boldsymbol{x},\boldsymbol{v}) \circ \Phi_{\text{affine}}(\boldsymbol{x},\boldsymbol{v})\Big) \circ \Big(\Phi^n_{\text{local}}(\boldsymbol{x},\boldsymbol{v})\Big)\,. \tag{5.4}$$

Additionally, we iteratively increase the spatial resolution and the size of the histogram. Iteratively increasing the resolution on multiple fronts avoids local minima, and it decreases the computation time of the initial alignment.

5.3.2 Implementational and Practical Challenges

Implementing the nonrigid registration model, with an analytical solution and explicit reorientation, has been a major challenge in terms of computational memory, speed, and precision. It required the introduction of additional steps in the framework to make it computationally feasible. We go through these three challenges, followed by a schematic and informal overview of the implementation. In short:

- To solve the memory challenge of storing the derivatives for the analytical solution, we use a technique from computer science called *program slicing*, in which we split the chain rule calculations into two runs.
- To solve the speed challenge of registrations taking days, we parallelized each step in the dependency graph from the paper manuscript.
- To solve the precision challenge of the derived gradients suffering from high numerical errors, we located critical sources of imprecision and increased floating point precision at these locations.

[2] At a very local scale the difference between images can be high, causing steep gradients and an unstable optimization.

Challenge: Memory

It is easy to illustrate the memory problem if we take two single-shell HARDI brain scans from the HCP. The dimensions of the brains are 145x174x145 voxels at 1.25 mm with 90 gradient directions which means 182x218x182 voxel at 1 mm re-interpolated resolution. With a cubic B-Splines FFD model, we have 64 control point vectors influencing each point $\boldsymbol{x}$ and direction $\boldsymbol{v}$. For illustrative purposes, we interpolate the images to a 1 mm spatial resolution, yet choose to interpolate them in only 30 uniformly distributed directions to save some space (cutting away about 2/3 of the original HCP scan). We ignore smaller memory requirements such as the joint histogram. Next, say we made a quick naive solution in which we stored all values needed to calculate any set of derivatives with respect to the deformation. Focusing on a rough estimate of the most memory consuming values and ignoring index book-keeping, this puts us at around 950 GB as shown in Table 5.1.

Variables	**# Double-Precision Values**	**Memory in GB**
$d\phi_c(\boldsymbol{x})$	$64 \cdot 182 \cdot 218 \cdot 182$	3.44
$d\psi_c(\boldsymbol{v})$	$3 \cdot 3 \cdot 64 \cdot 182 \cdot 218 \cdot 182 \cdot 30$	929.67
$dK(\phi(\boldsymbol{x})) \circ \mathbf{I}$	$3 \cdot 182 \cdot 218 \cdot 182 \cdot 30$	4.84
$d\Gamma(\boldsymbol{\nu}, \psi(\boldsymbol{v})) \circ \mathbf{I}$	$3 \cdot 182 \cdot 218 \cdot 182 \cdot 30$	4.84
Volume **I** & **J**	$2 \cdot 145 \cdot 174 \cdot 145 \cdot 90$	4.91
Interpolated **I** & **J**	$2 \cdot 182 \cdot 218 \cdot 182 \cdot 30$	3.22
Total	-	950.92

Table 5.1: An estimate of the memory costs of storing the derivatives with respect to all degrees of freedom (or partial derivatives) in major steps in the framework for two high-resolution DWI scans.

Clearly the derivatives of the Jacobian in $d\psi_c(\boldsymbol{v})$ are infeasible to store in short or even long-term memory. Additionally, due to the normalization in the Watson kernel and the histogram, we are unable to update the control points with respect to a single point and direction at a time. Inspired by *program slicing*[3] in computer science, we defined our own version of *algorithmic slicing*, to make the problem computationally feasible.

Program slicing refers to identifying a set of program statements — the program slices — that may affect the values of the program at some point. These can be precomputed. In our algorithm, this corresponds to precomputing the tail of the chain rule, such that for every partial derivative of the transformation $d_c\phi(\boldsymbol{x})$ and $d_c\psi(\boldsymbol{v})$, the product of derivatives has already been

[3]See `https://en.wikipedia.org/wiki/Program_slicing`

precomputed between the transformed image gradients and the similarity measure. In essence, we perform two slicing operations which save memory at the cost of adding extra computations:

Slice 1 Calculate $\Phi_c(\boldsymbol{x}, \boldsymbol{v})$ to transform all points $\boldsymbol{x}$ and directions $\boldsymbol{v}$ but *do not* calculate the derivatives $d\Phi_c(\boldsymbol{x}, \boldsymbol{v})$. Instead, calculate the derivatives $d_{\boldsymbol{I}\circ\Phi}NMI$, i.e. the gradients of the transformed image with respect to the similarity, which is Step 1 to Step 5 in Appendix C.

Slice 2 Repeat the calculation of $\Phi_c(\boldsymbol{x}, \boldsymbol{v})$ one voxel at a time, this time calculating the derivatives $d\Phi_c(\boldsymbol{x}, \boldsymbol{v})$, multiplying them with the relevant $d_{\boldsymbol{I}\circ\Phi}NMI$ from Slice 1, and adding this product to the 64 $d\boldsymbol{c}$ parameters affecting the neighbourhood of the voxel.

The additional computational cost, that comes from calculating the transformation of the moving image twice, is not a heavy operation. Even for an HCP subject, this only takes a few seconds on a high-end laptop. To get a better overview of the slicing, see the implementation diagram in Figure 5.3 and pseudocode in Algorithm 2 further below, or go to Appendix A.1 for a more detailed overview.

Challenge: Speed

Parallelization is essential to achieve an acceptable computation time, given the number of calculations required in DWI registration. Fortunately, each step in the framework is trivial to parallelize individually as each point, direction, or value is updated independently of its neighbors.

The CPU parallelization of each step is based on the C++11 Thread Pool implementation [Jakob and Vclav, 2012]. A pre-initialized pool of re-assignable threads is preferable, as opposed to creating and deleting new processes asynchronously which is time-consuming. Each thread is assigned a chunk of work[4] from a list of `future`-class objects[5] allowing for asynchronous retrieval of the results, which means that the limited amount of threads do multiple chunks of work per thread before retrieving the results. Table 5.2 shows how this leads to a significant decrease in computation time. The speed-up ratio in Table 5.2 is 3 - 5 times faster when going from 1 to 5 threads, and an 11 - 24 times improvement between 1 and 40 threads. While it can be run on a high-end laptop, a full resolution HCP subject is computationally

[4]The amount of work per thread is set on to be either an equal number-of-threads-dependent fragment of the total amount of work or some specified amount, e.g. 500 iterations per thread. This was set by a quick bit of trial and error, and can likely be improved even further in the future.

[5]`http://en.cppreference.com/w/cpp/thread/future`

Concurrency	Φ	dK	$d\Gamma$	dP	$dNMI$	$d\Phi$
Threads: 1	0.1089	2.4342	4.0974	0.0416	0.1279	4.1613
Threads: 5	0.0228	0.4378	0.9205	0.0137	0.0371	0.9819
Threads: 10	0.0154	0.2453	0.4884	0.0079	0.0224	0.5248
Threads: 15	0.0105	0.1782	0.3485	0.0050	0.0161	0.3923
Threads: 20	0.0088	0.1380	0.3104	0.0041	0.0146	0.3522
Threads: 25	0.0074	0.1254	0.2508	0.0035	0.0128	0.3112
Threads: 30	0.0072	0.1142	0.2401	0.0027	0.0100	0.2818
Threads: 35	0.0070	0.1102	0.2311	0.0026	0.0089	0.2796
Threads: 40	0.0069	0.1038	0.2221	0.0024	0.0078	0.2650

Table 5.2: Tests run on a department cluster node supporting up to 40 concurrent threads on an HCP subject at full resolution. The results are measured in milliseconds, and represents a single update of the FFD parameters. Each column corresponds to *STEP 1-6* in the detailed diagram of Appendix A.1, where 1-5 represent the first algorithmic slice and step 6 represent the second slice. In other terms: deformation, spatial interpolation, directional interpolation, density estimation, similarity, and derivative deformation.

very demanding and due to the lack of available hardware, we have only been able to report on few comparative results on parallelism. This is only a single parameter update out of the many runs performed by L-BFGS optimizer. Running without parallelism would likely take days instead of a few hours.

GPU parallelization has become the gold standard of many *embarrassingly parallel* problems. However, memory transfer to and from the GPU is often an issue of large datasets. An experiment was performed to evaluate the potential of using GPUs in the individual steps. The experiment was done on the histogram, as it is tricky to parallelize efficiently in terms of coalesced memory access. However, even with latency hiding and using multiple streams, the memory transfer dominated the computations and, under the current structure of the framework, we did not see any significant advantages in using the GPU. For details on these experiments, see Appendix A.2.

Challenge: Precision

In this complex system, where the gradient of each single control point depends on the sum over thousands of value-pairs all part of a long chain of derivatives, and with kernels fitted to relatively unknown inputs of data, we cannot avoid numerical instability. As such, we need to evaluate the precision of our analytical derivatives, as well as the correctness of our implementation. One common way of doing so is by numerically checking the gradients using finite difference approximation. If we want to minimize our cost function $\mathcal{M}$ by changing the transformation parameters $\boldsymbol{c}$, we would expect that the

analytical gradient is equal to the numerical gradient such that

$$d_c\mathcal{M} = \lim_{\epsilon \to 0} \frac{\mathcal{M}(\boldsymbol{c} + \epsilon) - \mathcal{M}(\boldsymbol{c})}{\epsilon} \tag{5.5}$$

where ϵ in practice is set to some sufficiently small value that does not cause more round-off errors. In this case, we set it to $\epsilon = 10^{-6}$, which often means that the left- and right-hand sides of eq. (5.5) should agree on at least six significant digits. For this level of precision, we write the gradient error for a change in the model parameters $\boldsymbol{c}$ as

$$\mathcal{E}(d_c\mathcal{M}) \approx \left| d_c\mathcal{M} - \frac{\mathcal{M}(\boldsymbol{c} + \epsilon) - \mathcal{M}(\boldsymbol{c})}{\epsilon} \right| \tag{5.6}$$

So far, we have looked at the whole system, where a change in a parameter of the deformation field will lead to a change in the similarity of two DWI volumes being registered. However, this analysis was also performed at each individual step of the chain rule in the analytical solution. The analysis found that the gradients of the spherical interpolation did not satisfy the criteria of the error being smaller than six significant digits. In fact, the derivatives of the Watson kernel was off by around $\mathcal{E}(d_\psi \Gamma(\boldsymbol{\nu}, \psi(\boldsymbol{v}))) \approx 10^{-3}$ for a change in $\psi(\boldsymbol{v})$. This is not at all surprising when considering the derivatives of the Watson kernel. The sum over exponentials on the sphere will result in round-off errors when we evaluate the support of vectors $\boldsymbol{\nu}$ far from the center of the distribution in $\psi(\boldsymbol{v})$. While there might be a numerical trick for solving this, we have reduced the gradient error to acceptable limits by increasing the precision from the C++ floating-point data type `double` to `long double` for the Watson interpolation. The total precision of all steps in the analytical solution, for data inputs such as HCP or ADNI subjects with standard parameters[6], is around

$$\mathcal{E}(d_c\mathcal{M}(\boldsymbol{I}, \boldsymbol{J}, \Phi)) \approx 10^{-8} \tag{5.7}$$

for the gradients of the parameters in transformation model with respect to the cost function.

Implementation: Diagram

Figure 5.3 shows a diagram of the optimization over the FFD deformation field. The two slicing operations (yellow and purple) are described in much more detailed schematic diagrams in Appendix A.1.

[6]This number varies depending on the data and parameters of the model.

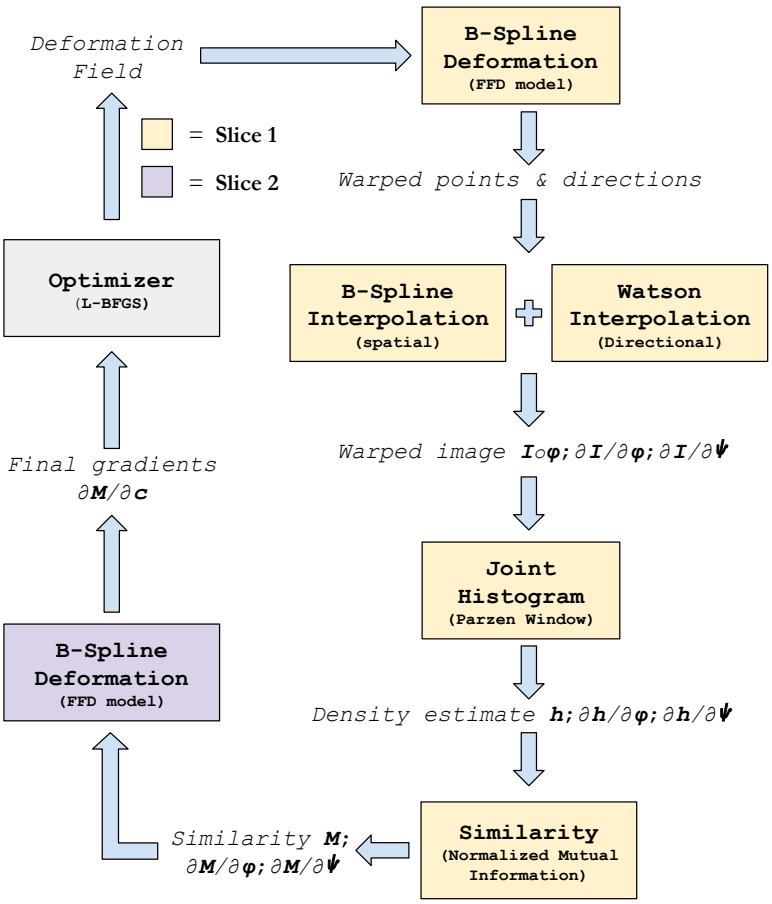


Figure 5.3: Sketched overview of the registration framework. Each step is implemented in C++ and parallellized, and the Watson Interpolation is implemented with high numerical precision.

Implementation: Pseudocode

In Algorithm 2, we have summarized the implementation which calculates the similarity measure and analytical gradients, corresponding to the input to L-BFGS that replaces the optimization `while`-loop in Algorithm 1 (disregarding the regularization).There are a few things to be aware of when reading the pseudocode.

First, the parameters (κ, number of bins in $\mathbf{h}$, control point and image resolution spacing, stopping criteria, etc.) are not included in the input.

Second, the function $dF(\cdot)$ is a bit compact, as it calculates the gradients over four levels in the dependency graph Step 1 (similarity) to Step 4 (histogram). The reason is simply that to calculate the effect of a Parzen-window entry into

the histogram with respect to the final similarity, we have to precompute a series of steps and then reconstruct the histogram while storing the gradients.

Third, we use the gradient notation $\nabla_a B$ to indicate "all the partial derivatives of B with respect to the parameter a".

Algorithm 2 The cost function and analytical derivatives using slicing.

Input

Φ_c : $\lceil \frac{X2}{\delta_x} \rceil \times \lceil \frac{Y2}{\delta_y} \rceil \times \lceil \frac{Z2}{\delta_z} \rceil \times 3$, deformation field at spacing δ.

$\mathbf{X}_I$: $\lceil \frac{X2}{\gamma_x} \rceil \times \lceil \frac{Y2}{\gamma_y} \rceil \times \lceil \frac{Z2}{\gamma_z} \rceil \times 3$, spatial coordinates at image resolution γ.

$\mathbf{V}_I$: $\lceil \frac{X2}{\gamma_x} \rceil \times \lceil \frac{Y2}{\gamma_y} \rceil \times \lceil \frac{Z2}{\gamma_z} \rceil \times N1 \times 3$, directional vectors at image resolution γ.

$\mathbf{I}$: $X1 \times Y1 \times Z1 \times N1$, original (moving) image.

$\mathbf{J}$: $X2 \times Y2 \times Z2 \times N2$, reinterpolated (target) image.

$\mathbf{w}_J$: $N2 \times 3$, directional interpolation coordinates.

Output

$\nabla_{\mathcal{F}}\Phi_c$: $\lceil \frac{X2}{\delta_x} \rceil \times \lceil \frac{Y2}{\delta_y} \rceil \times \lceil \frac{Z2}{\delta_z} \rceil \times 3$, gradients of Φ_c w.r.t. $\mathcal{F}$.

$\mathcal{F}$: The NMI similarity measure between $\mathbf{I} \circ \Phi_c$ and $\mathbf{J}$.

Algorithm - Slice 1

calculate the transformed coordinates $(\mathbf{X}^*, \mathbf{V}^*) = \Phi_c(\mathbf{X}, \mathbf{V})$.

interpolate spatially $(\mathbf{I}^*, \nabla_{\phi(\boldsymbol{x})}\mathbf{I}) = K(\mathbf{I}, \mathbf{X}^*, \mathbf{V})$
where $\mathbf{I}^* : X2 \times Y2 \times Z2 \times N1$ and $\nabla_{\phi(\boldsymbol{x})}\mathbf{I} : X2 \times Y2 \times Z2 \times N1 \times 3$.

interpolate directionally, $(\mathbf{I}^{**}, \nabla_{\psi(\phi(\boldsymbol{x}))}\mathbf{I}, \nabla_{\psi(\phi(\boldsymbol{v}))}\mathbf{I}) = \Gamma_\kappa(\mathbf{I}^*, \mathbf{V}^*, \nabla_{\phi(\boldsymbol{x})}\mathbf{I}, \mathbf{w}_J)$
where $\nabla_{\psi(\phi(\boldsymbol{x}))}$ & $\nabla_{\psi(\phi(\boldsymbol{v}))} : X2 \times Y2 \times Z2 \times N2 \times 3$,
and $\mathbf{I}^{**} : X2 \times Y2 \times Z2 \times N2$.

calculate the joint histogram $\mathbf{h}_{IJ} = P(\mathbf{I}^{**}, \mathbf{J})$.

calculate the joint probability $\mathbf{p}_{IJ} = \frac{\mathbf{h}_{i,j}}{\sum_{IJ}(i,j)\mathbf{h}_{IJ}}$
and the marginals $\mathbf{p}_I = \sum_i \mathbf{p}_{IJ}(i,j)$ and $\mathbf{p}_J = \sum_j \mathbf{p}_{IJ}(i,j)$.

calculate the entropy $\mathbf{H}_{IJ} = \sum_{i,j} \mathbf{p}_{IJ}(i,j) log(\mathbf{p}_{IJ}(i,j))$, $\mathbf{H}_I = \sum_i \mathbf{p}_I(i) log(\mathbf{p}_I(i))$
and $\mathbf{H}_J = \sum_j \mathbf{p}_J(j) log(\mathbf{p}_J(j))$. Save $\mathbf{p}^*_{IJ} = \log(\mathbf{p}_{IJ})$ and $\mathbf{p}^*_I = \log(\mathbf{p}_I)$

calculate the similarity measure $\mathcal{F} = -\frac{\mathbf{H}_I + \mathbf{H}_J}{\mathbf{H}_{IJ}}$.

calculate the remaining gradients $(\nabla_{\mathcal{F}(H(p(h(\psi(\phi(\boldsymbol{x}))))))}\mathbf{I}, \nabla_{\mathcal{F}(H(p(h\psi(\phi(\boldsymbol{v}))))))}\mathbf{I})$
$= dF(\mathbf{I}^{**}, \mathbf{J}, \mathbf{H}_I, \mathbf{H}_J, \mathbf{H}_{IJ}, \mathbf{p}^*_{IJ}, \mathbf{p}^*_I, \nabla_{\psi(\phi(\boldsymbol{x}))}\mathbf{I}, \nabla_{\psi(\phi(\boldsymbol{v}))}\mathbf{I})$,
where each spatial gradient is also summed over all $N2$ directions.

Algorithm - Slice 2

for all $\boldsymbol{x}_i \in \mathbf{X}_I$ **do**
  for l = 1..64 **do**
    update $\nabla_{\mathcal{F}}\Phi_{c_l} = \nabla_{\mathcal{F}}\Phi_{c_l} + (\nabla_{\phi(\boldsymbol{x}_i)}\boldsymbol{c}_i \cdot \nabla_{\mathcal{F}(H(p(h(\psi(\phi(\boldsymbol{x}_i))))))}\mathbf{I})$
  end for
  for j = 1..$N2$ **do**
    for l = 1..64 **do**
      update $\nabla_{\mathcal{F}}\Phi_{c_l} = \nabla_{\mathcal{F}}\Phi_{c_l} + (\nabla_{\psi(\boldsymbol{v}_{i,j})}\boldsymbol{c}_i \cdot \nabla_{\mathcal{F}(H(p(h(\psi(\phi(\boldsymbol{v}_{i,j}))))))}\mathbf{I})$
    end for
  end for
end for

6 Conclusions

This dissertation proposes a new formulation for density-based DWI similarity, along with a framework for nonrigid DWI registration that minimizes the objective function based on the full diffusion profile.

6.1 Summary

Chapter 3 set the foundation of this work by presenting a critical review of the current state of methods for registration and validation of DWI for voxel- and tract-based approaches, focused on the human brain. Under a significant lack of comparative studies of DWI registration, we reviewed popular approaches to comparing different DWI scans. We found that most studies rely on locally trained experts for in-house segmentations and that there is an immediate requirement for public evaluation protocols and community challenges, as DWI is becoming an issue of Big Data.

Chapter 4 presented the first density-based cost function between two full DWI scans. The objective function is based on a scale-space model, that allows us to perform optimization over the spatial, orientation, and intensity scales. The integration over the orientational scale takes our method from 3D to DWI, and the intensity kernel offers a unique isocurve optimization by integrating over the joint density of the DWI scan. This is exemplified in the entropy-based measure Mutual Information that was used as a non-linear similarity for global affine registration. We showed that it is important to take the full diffusion profile into account for a proper spatial alignment.

Chapter 5 presented a full, nonrigid DWI registration framework based on the previously defined density formulation. It was presented as a theoretical contribution with an additional focus on the implementation itself. The framework exploits the smooth scale-space formulation to perform registration of full DWI profiles with explicit reorientation of the gradient vectors

as part of the analytical gradients. Both raw HARDI data and ODFs can be used as input, and the results demonstrate a capability to correctly reorient the ODFs even if the registration is based on the HARDI data. We also investigated the parameters of the different scales and their influence on the results. Interestingly, simulated results indicate that the cost function has an inherent regularizing effect and that the underlying DWI structure is preserved.

6.2 Discussion & Future Work

6.2.1 Validation Issues in DWI

DWI is a hot topic used with very different goals in mind, such as studying regional changes in diffusion or the brain connectivity. Consequently, DWI is also a highly volatile topic in terms of the various types of raw DWI data and how the data is modeled. The diversity, rapid growth, and cytoarchitectonic complexity have resulted in a lack of comparative surveys and quantitative validation. Results are often qualitatively validated by visual inspection or by help from trained medical experts. In our critical review, we concluded that this will be increasingly problematic as DWI data grows in resolution and public availability, and that computational evaluation will be required to handle microscopic details and the sheer amount of data. However, while we believe that group-wise validation is a critical issue, it is worth pointing out that the DWI community[1] seems to be aware of the growing confusion and lack of cross-study evaluation. Much is actively being done to validate state-of-the-art methods on the scale of individual DWI.

For current on-going challenges, the focus is often on improving fiber tract identification and tractography as in the scan-rescan subject in [TraCED, 2017], and the phantom, tracer-injected monkey and high-resolution ex-vivo monkey in [VoTEM, 2018]. Voxel-based validation has been less profiled but that might also be changing with [CDMRI, 2017, MUSHAC, 2018], where data is combined from multiple scanners aimed at predicting high-resolution scans from low-resolution clinical samples. With the focus on DWI as an issue of Big Data [O'Donnell et al., 2017, Smith and Nichols, 2018, Yeatman et al., 2018] and with similar problems being tackled in related fields in the Continuous Registration Challenge [CRC, 2018], we predict that interesting challenges on large datasets of DWI will soon appear. Already, there are 76TB 1200 subjects from the HCP that can be found on the Amazon S3 for effective, large-scale cloud computing [WU-Minn, 2017].

[1] Conferences such as MICCAI, ISBI, ISMRM, and similar which endorse such DWI challenges and workshops.

6.2.2 New Similarity Measures Explicitly for DWI

The similarity between DWI scans is commonly defined in the context scalar images as a basis for registration. That is, quantitative measures such as the FA are used for the registration and reorientation, if involved, is built on top of this - as we saw with implicit reorientation methods in Chapter 3. This is different from similarities in DWI based on fiber tracts, that are often found to be measures of pair-wise Euclidean distances - possibly including curvature information or quantitative features along the tracts [O'Donnell et al., 2017]. Euclidean distance measures make sense for tract-based methods that already embodies a diffusion model. However, this does not apply to voxel-based methods. Well-known scalar-based algorithms applied to an FA image do not utilize the diffusion profile as an explicit part of the similarity. Moreover, registration algorithms with similarities fully designed for DWI, such as `DTI-TK` [Zhang et al., 2006] and `DT-REFinD` [Yeo et al., 2009], have repeatedly been shown to outperform scalar-based methods [Zhang et al., 2014, Wang et al., 2016, Wang et al., 2017].

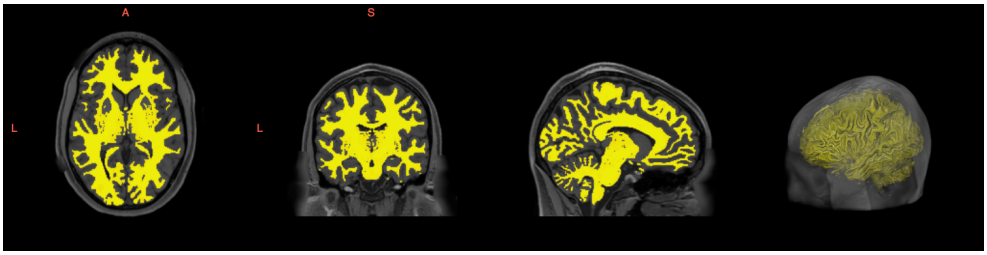
We reason that the popularity of scalar-based DWI registration, with implicit use of the diffusion profile, is due to the computational speed and applications for state-of-the-art DWI in clinical practice. However, it is clear that future methods would benefit from being explicitly designed for DWI, and we hope to contribute to this by introducing a density formulation between DWI. In Chapters 4 and 5, we applied both Mutual Information (MI) and Normalized MI (NMI) as similarity measures between DWI, and we argued that such information-theoretic measures are well-suited for the problem, as also pointed out by others [Van Hecke et al., 2007, Bhushan et al., 2012]. DWI scans clearly have a non-linear and perhaps even purely statistical relationship in terms of b-values, noise, different scanners, etc. [Wiest-Daesslé et al., 2007, Johansen-Berg and Behrens, 2013]. However, given a set of high-resolution temporal or phantom DWI scans with ground truth labels and a one-to-one correspondence, it could be very interesting to compare the effects of different similarity measures. For instance, it would be relevant to compare linear and non-linear measures, as this could help quantify the relationship between different b-values or across-scan noise. Or the functional and statistical relationship could potentially be quantified between Normalized Cross-Correlation (NCC) and NMI, where it might be that complex scan variations can be described as a function with NCC while multi-centric studies with different scanners can be described by the entropy in NMI. In either case, there is a lot of potentials if the proper data is available.

6.2.3 Validating DWI in Whole-Brain Registration

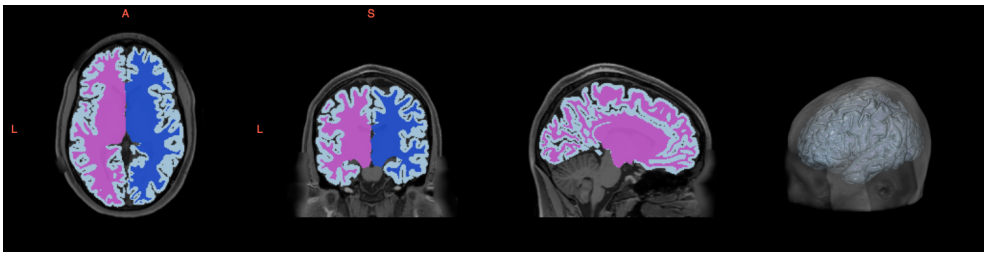
Throughout our work, the density-based registration framework has been applied to a diverse set of problems. Before developing the expertise to simulate DWI data, initial experiments were performed on different high-resolution HCP subjects in [Jensen et al., 2015, Jensen et al., 2017]. Inter-subject registration raises the question of when a spatial alignment is as good as it can be. We know that brains are not diffeomorphic [Mangin et al., 2016], and to our knowledge, no regions or anatomical structures in the brain are guaranteed to have a one-to-one correspondence. Structural landmarks used for validation in group studies often have a trained expert to place them or correct the automatic placement [Mori et al., 2008, Klein et al., 2009].

In [Jensen et al., 2015], we were able to use inter-subject registration since the transformation was global without risk of overfitting, and we were only interested in effects of the scales with respect to the similarity. In [Jensen et al., 2017], we demonstrated that large nonrigid deformations were possible and that the brains seemed aligned by visual inspection in both structure and orientation. However, from the review in Chapter 3, we could conclude that there are no generally accepted ways of validating inter-subject registration, aside from manual labels or the comparison with competing methods. The latter is an option for studies presenting new methods while lacking ground truth data. For instance, segmentations from the HCP could be used as these are generally considered state-of-the-art automatic (not manual) labels. Examples of such segmentations are shown for illustrative purposes in Figure 6.1. However, without expert knowledge of competing frameworks, this approach is inherently flawed and it is easy to fall into the trap of using the default settings of other methods. Alternative approaches to validation were thoroughly discussed in Chapter 3. However, most depend on the data available.

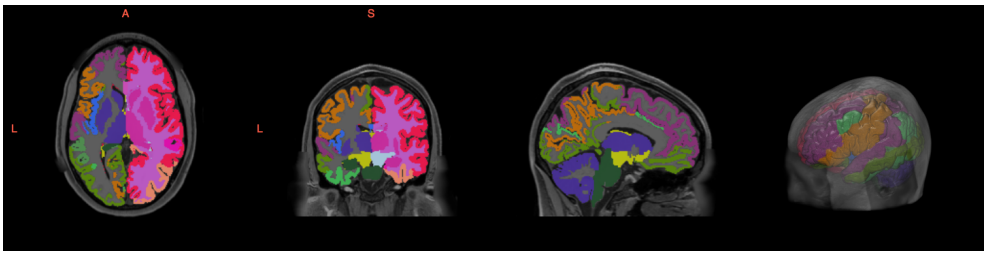
In this work, we initiated several projects towards an openly available way of validating nonrigid DWI registration. One of our attempts was to register the public data form the Alzheimer's Disease Neuroimaging Initiative (ADNI) [Jack et al., 2008]. While DWI is only a recent addition to the ADNI database, we managed to download a suitably large dataset of preprocessed healthy and sick subjects [Nir et al., 2013]. The results are described in Appendix D. The goal was to evaluate the temporal progression in healthy and sick subjects to locate new potential diffusion-related biomarkers, and to validate our framework through existing biomarkers and large intra-subject deformations. However, in the end, it turned out that the resolution of the preprocessed DWI data in ADNI is too poor for proper validation of nonrigid registration.



(a) White matter segmentation.



(b) "Ribbon" segmentation of the cortical grey matter.



(c) Both surface-based and subcortical multi-label segmentation.

Figure 6.1: A few of the different examples of available preprocessed automatic `FreeSurfer` segmentations from the HCP dataset.

We found no suitable datasets for inter-subject whole-brain validation, there are datasets that would potentially be interesting to future investigations. The HCP has test-retest scans available by request [WU-Minn, 2017], which could be used to test intra-subject registration and noise-correction. Additionally, they offer restricted access to a database of related subjects with twin and sibling cases, which might have more suitably matching brain structures than unrelated subjects and could be used for testing inter-subject registration.

On a final note, our framework is not specific in any way to brain data. There are no built-in bias and regularization towards brains, and any experiment involving DWI data would be applicable to the formulation.

A Implementational Details

A.1 Detailed Schematic Overview of the Implementation of the Analytical Gradients

The diagrams in Figures A.1 to A.3 show a detailed version of the steps that calculates the analytical gradients of FFD parameters with respect to the NMI similarity measure. The input and output in the diagrams are colored for navigational purposes (green output in step 1 corresponds to green input in step 2 and so on). Gray represents exterior inputs and do not come from any previous steps. The function names are taken directly from the framework, which follows the algorithmic slicing pseudocode in Algorithm 2.

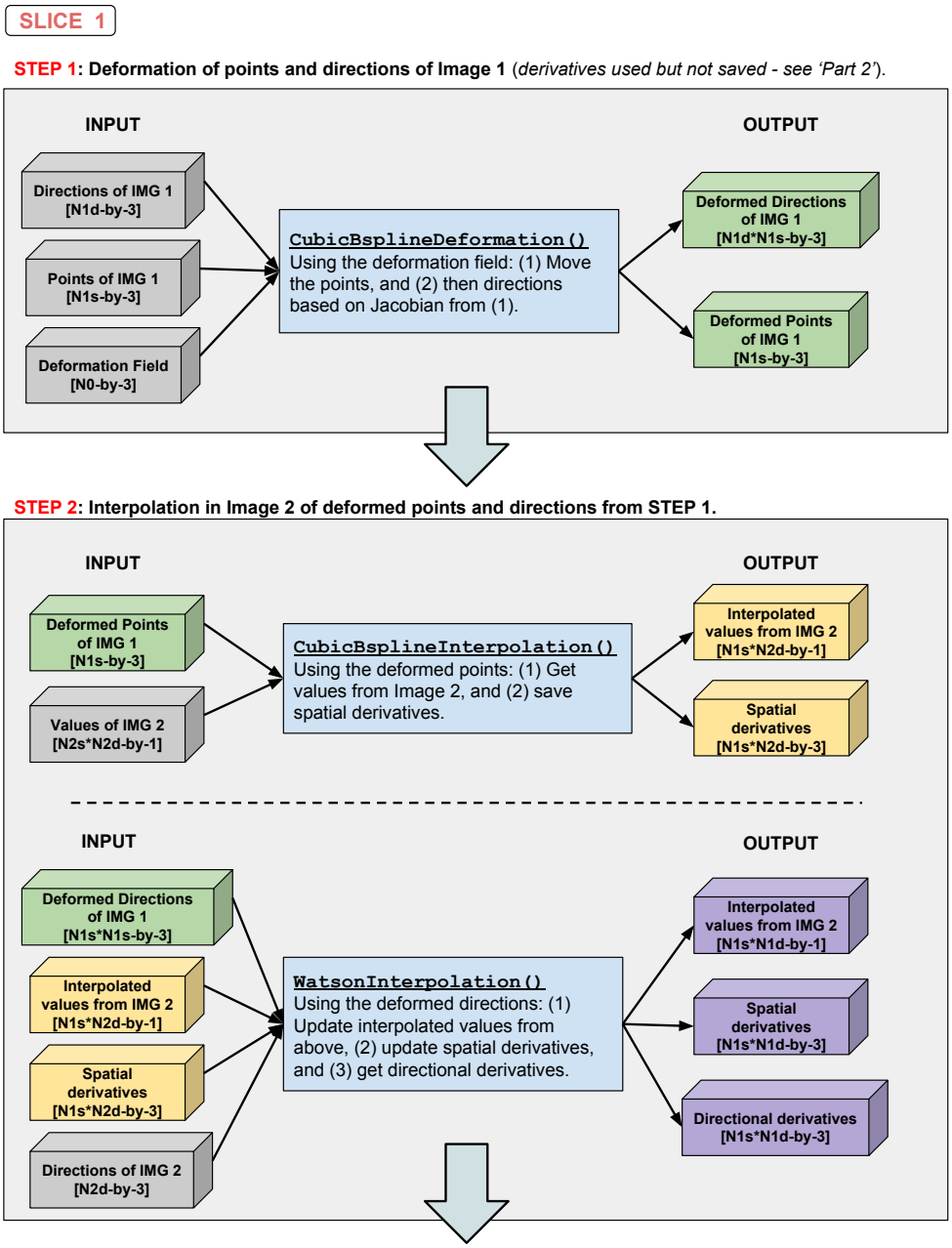


Figure A.1: The first part of the slicing operation. Step 1 is calculating the transformed points and directions without storing the derivatives. Step 2 is the spatial and directional interpolation.

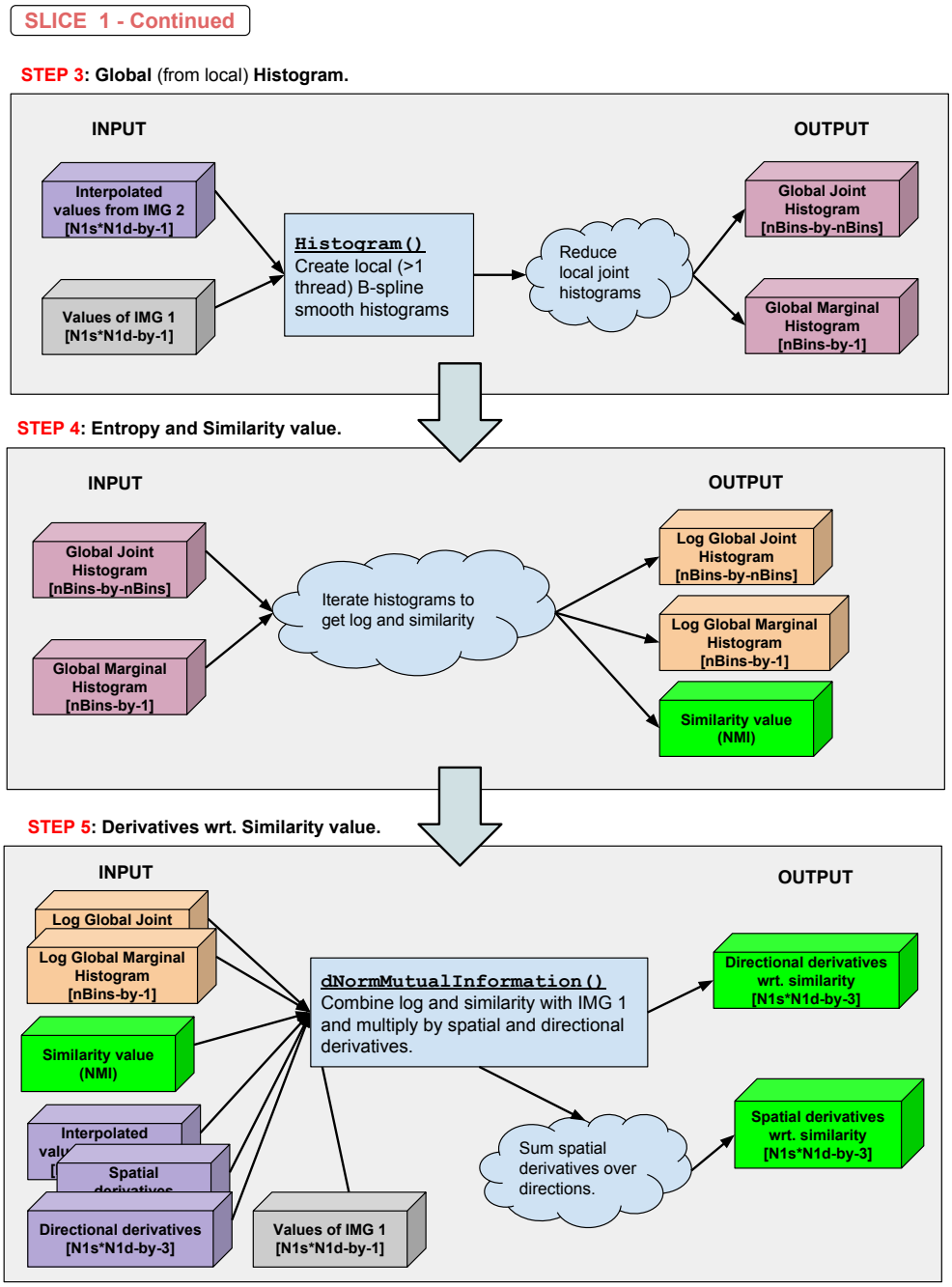


Figure A.2: Continuing the first slice, Step 3 calculates the histogram, Step 4 calculates the entropy (log) and similarity measure, while Step 5 finally calculates the derivatives with respect to the similarity measure. At the end of Slice 1, we now have the derivatives of the transformed image with respect to NMI.

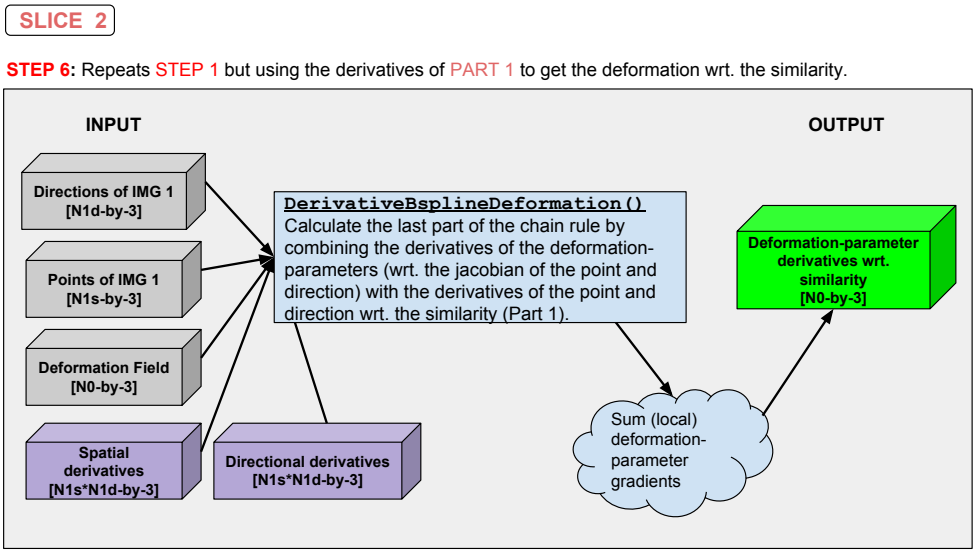


Figure A.3: The second slicing operation, goes through the deformation of the very first step, this time calculating the derivatives of the deformation one at a time, and multiplying on the corresponding gradients from the first slice. This finally gives us the gradients of the control points with respect to the similarity NMI.

A.2 A Brief Experiment in Utilizing the GPU for Parallelization of the Histogram

Given millions of independent calculations doing the DWI registration on a GPU makes sense. However for GPU parallelization to add a significant speedup two criteria should be matched: (1) a large number of independent computations, and (2) a favorable ratio of memory I/O to the amount of work per processing unit. The current implementation clearly favors the first but not the second criteria - the calculations are very quick compared to the memory that needs to be transferred on to the GPU device memory. However, in a collaboration with a three other students[1], we investigated the potential of speeding up the registration by taking the hardest part of the implementation (the histogram) to throw on the GPU and analyze the potential for latency hiding of datasets that are larger than available device memory by using streaming techniques.

For histograms, an efficient parallel implementation yields a number of problems related to memory performance. One of these is the unknown order of indices from the map-phase which prevents us from working with the data in a coalesced manner during the reduce-phase. This was solved by partially sorting (using radix sort) the input data into segments such that the values of each segment fit into small local histograms which fit into shared memory and can be flushed to corresponding parts of the global histogram. Skipping the details on optimizing the hardware parallelism for radix sort, we move on to CUDA streams. They give us the possibility to better utilize the hardware in a fashion shown on Figure A.4. The figure shows copying to the device (HD), histogram computation (K) and copying back the resulting histogram (DH). As seen, while one buffer is filling up,

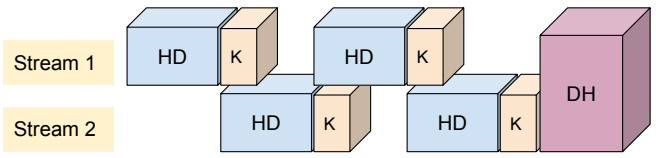


Figure A.4: Using streams to overlap copying with computation.

the other is calculating a histogram and both commit to the same global histogram. However, HD takes up more than 90% of the run time, making

[1] In collaboration with Joachim T. Kristensen, Christian K. Larsen, and Mathias Grymer as part of the course 'Programming Massively Parallel Hardware' at University of Copenhagen, 2016.

it impossible to create enough streams to hide the memory transfer. This is clearly seen in a benchmark tested for a non-parallelized CPU and 1, 2, and 4 streams (1 stream is to check whether memory transfers are the dominating factor) with a data size of 2G elements and 400k bins. As seen in Table A.1 adding more streams does not reduce time, underlining memory transfer as a highly dominant factor. There is no doubt that a significant speedup can be

Method	Time
CPU	9.64 s
1 stream	4.91 s
2 streams	4.84 s
4 streams	4.92 s

Table A.1: Time of streaming to the device.

achieved using streams if more complex and time-consuming computations were performed in parallel. However, the steps in the framework are not independent and while complex bookkeeping and redundant computations might be possible for a worthwhile speedup, it is beyond the scope of this work.

B Abstract from ISMRM 2017

The first non-hierarchical version of the nonrigid LOR-DWI registration framework was accepted as an abstract and presented at International Society for Magnetic Resonance in Medicine (ISMRM) in April 2017.

Titled "Density-Based Nonrigid Registration of Diffusion-Weighted Images", the abstract demonstrates the nonrigid framework on two simulated fiber crossings and between two HCP subjects. The validation is purely visual but demonstrates that a correct registration seems to occur, as random gradient directions match up and large deformations are possible. Moreover, the abstract also offers a brief discussion on the purpose of DWI registration as a spatial mapping and what we believe to be the poor idea of reconstructing DWI in warped space for anything but visual validation.

Density-Based Non-Rigid Registration of Diffusion-Weighted Images

Henrik Grønholt Jensen[1], Francois Lauze[1], Mads Nielsen[1], and Sune Darkner[1]

[1]*Computer Science, University of Copenhagen, Copenhagen, Denmark*

Synopsis

We present a non-rigid registration method for Diffusion-Weighted MRI which uses a density and scale space approach to estimate image similarity. This allows us to employ smooth intensity-invariant similarity measures, such as Mutual Information (MI) in contrast to the model-driven registrations. Using the inherent microstructure of High Angular Resolution Diffusion Imaging (HARDI) scans, we obtain a less regularized and more flexible registration that can be used on either raw diffusion signals or reconstructions of the fiber orientations. We show some promising results on Human Connectome Project (HCP) subjects and an artificial example.

Purpose

We present a non-rigid registration approach which uses the apparent diffusion coefficient (ADC) directly, based on Locally Orderless Registration for DWI1. While model-driven registration methods (such as the ones using Spherical Deconvolution, Spherical Harmonics, or similar) reduce noise or offers smoother solutions by modelling signal frequencies and selecting peaks, they tend to be sensitive to scaling and multi-scanner variance. It is desirable to use information theoretical measures of the observations in DWI data due to their invariance properties. With a scale space approach to density estimation we are able to register diffusion data directly using the ADC between subjects. Our framework allows us to use similarity measures like Mutual Information (MI), previously reserved for registration of scalar-valued images. We illustrate the approach on a subject from the HCP and on artificial fiber-crossings.

Methods

We use a non-rigid registration based on B-splines, and the density estimation scheme described by Jensen et al2, to estimate the joint and marginal distributions over intensity, location, and orientation. The histogram is estimated as

$$h_{\beta\alpha\sigma\kappa}(i,j|\vec{x}) = \int_{\Omega\times S^2} P_\beta(I_{\sigma\kappa}(\vec{x},\vec{v}) - i)P_\beta(J_{\sigma\kappa}(\vec{x},\vec{v}) - j)d\vec{x}\times d\vec{v}$$

where

$$I_{\sigma\kappa}(\vec{x},\vec{v}) = \int_{S^2}\left(\int_\Omega I(\vec{\tau},\vec{\nu})K_\sigma(\vec{\tau}-\vec{x})d\vec{\tau}\right)\Gamma_\kappa(\vec{\nu},\vec{v})d\vec{\nu} = (I*(K_\sigma\otimes\Gamma_\kappa))(\vec{x},\vec{v})$$

and σ is the image scale, κ is the orientation scale, and β is the intensity scale. P, K, and Γ are kernels in intensity, location, and orientation respectively. In our experiments, the location scale is modelled by B-splines and the orientation scale by a bidirectional Watson kernel. The histogram is normalised to generate a density estimate from which MI, normalized MI, and other density-based image similarity measures can be estimated.

Results

We selected a full brain DWI of a single HARDI shell from 2 randomly chosen subjects from the HCP. We used the scans with a b-value of 3000. The resolution is 1mm isotropic voxel with 90 directions. In Figure 1 and 2, the subject is first aligned by a global rigid transformation and then aligned by the non-rigid registration.

We also constructed two artificial DWI samples containing fiber ODFs, one that simulates a fiber-crossing with a 90 degree angle and one that simulates a 45 degree angle crossing. The resolution is set to 1mm isotropic voxels with 90 directions. Figure 3 illustrates the transformation of the 45 degree crossing to the 90 degree crossing.

Discussion

As the results show, the method is able to align both the real images and the artificial example.

Although unlikely to occur in real data, the artificial example illustrates the ability of the method to capture the transformation under such extreme cases. A direct map between the two crossings is not to be expected given the large local deformation needed, but we see that our model generalises well. Smoothing effects on the isotropic ODFs are to be expected, seen in the corners of the transformation.

Our experiments on real data show that the transformation, and the estimated non-rigid deformation, shows an excellent fit between the interpolation of subject 1 and the transformed subject 2 (this also generalises to other gradient directions). The application of the registration using B-splines shows that a very good apparent alignment can be achieved.

It is worth mentioning that the use of a Watson distribution γ allows us to register fiber orientations by using negative κ. However, it appears that no real difference is observed between positive and negative κ.

We believe in taking advantage of the full recorded signal (the whole of the DWI signal) including noisy observation to estimate the transformation in contrast to using the dominant features derived from the signal. The latter may in some cases oversimplify the signal which potentially could lead to additional overfitting.

It is our belief that in the context of registration, computations on deformed DWI data (such as tractography or modelling of fibers) should be avoided as the deformed data in no way represents any anatomy, but merely serves the purpose of estimating a transformation between any two subjects. Computation should be performed in the native space. However, for illustration purposes we have warped the ODFs according to the deformation to show the optimum found using the similarity measure.

Conclusion

We have presented a non-rigid registration similarity measure which directly relies upon the ADC and allows us to register diffusion scans rather than a fiber-model. As the results show, we are able to effectively perform inter-subject registration. Furthermore, we are able to match regions with potential large anatomical variability as the artificial example shows.

Acknowledgements

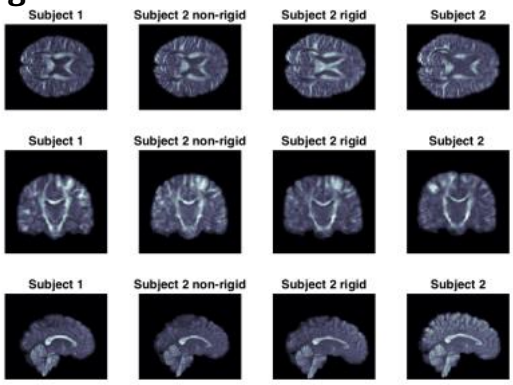
Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

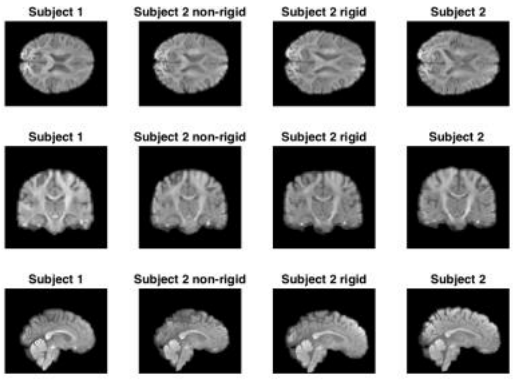
[1] Jensen, Henrik G., et al. "Locally Orderless Registration for Diffusion Weighted Images." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[2] David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, Kamil Ugurbil, for the WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. NeuroImage 80(2013):62-79.

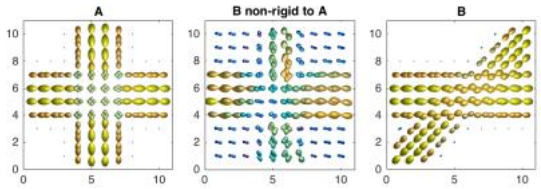
Figures



Non-rigid registration of HCP Subject 2 to Subject 1, shown from a random gradient orientation. The interpolation of the original scans are seen to the left and right. The second column is the non-rigid results that maps Subject 1 to 2, and the third column shows the rigid alignment prior to the non-rigid registration.



Same as Figure 1 but shown is mean diffusivity instead of from a single gradient direction.



A 45 degree crossing registered to a 90 degree crossing with a little vertical overlap.

C Step-by-Step Analytical Formulation of LOR-DWI Registration

The purpose of this appendix is to unfold the dependencies and write up the explicit gradients of the transformation parameters with respect to the similarity measure. The dependency graph is illustrated in Figure C.1, along with the steps that we will go through.

In the first part, we write up the functions and equations that allow us to define the similarity $\mathcal{F}$ as a function of the transformation model Φ_c, specifically its parameters c. We follow the steps in Figure C.1, where we have chosen the non-linear similarity measure Normalized Mutual Information (NMI)

$$\mathcal{F}_{non-lin} = \int_{\Lambda^2} f(p(i,j))di\ dj$$
$$\text{where} \qquad f(p(i,j)) = NMI(\Phi_c, \boldsymbol{I}, \boldsymbol{J})$$

where p is the joint density, $\boldsymbol{I}$ is the moving image, and $\boldsymbol{J}$ target image.

In the second part, we again move through the dependency graph, this time writing up the chain of derivatives. In all parts, we will define both the continuous and discrete representations.

C.1 Unfolding the Dependencies - Part 1

We start from the top, and follow the branches that depends on the transformation Φ_c.

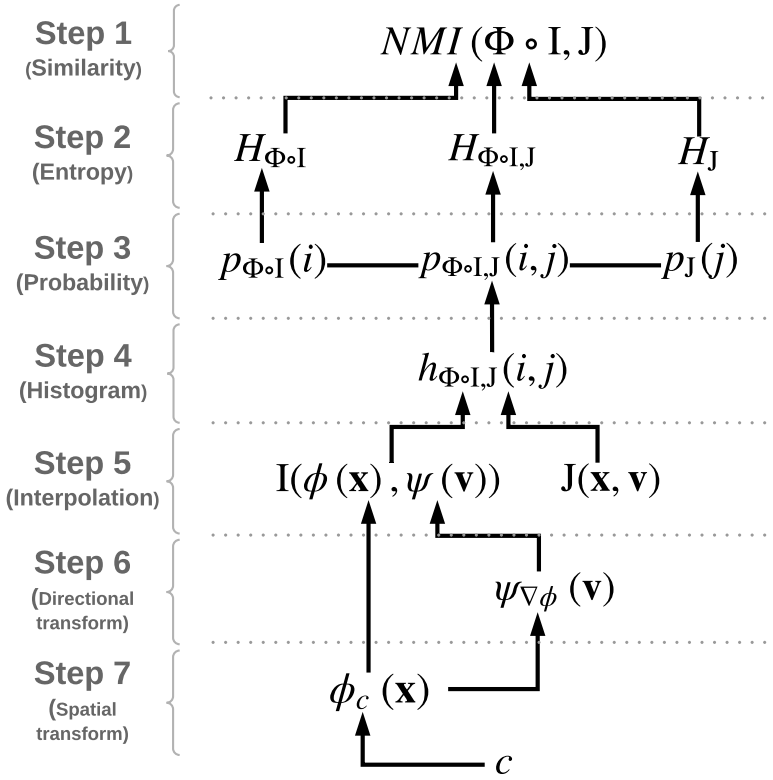


Figure C.1: Dependency graph of the nonrigid DWI registration between the moving image **I** and the target image **J**, with normalized mutual information (NMI) as the similarity measure. The deformation is parameterized by **c** so that any change in **c** will eventually affect the total similarity between the two images.

Step 1 - Similarity Measure

$$NMI(\Phi, I, J) = \frac{H_J + H_{I\circ\Phi}}{H_{I\circ\Phi,J}} \tag{C.1}$$

where H is the marginal and joint entropy.

Step 2 - Entropy

$$H_{I\circ\Phi} = -\int_{\Lambda} p_{I\circ\Phi}(i) \cdot log(p_{I\circ\Phi}(i))di \tag{C.2}$$

$$H_{I\circ\Phi,J} = -\int_{\Lambda^2} p_{I\circ\Phi,J}(i,j) \cdot log(p_{I\circ\Phi,J}(i,j))di\, dj \tag{C.3}$$

where p is the marginal and joint probability density function. The discrete version simply replaces the integral with a sum.

Step 3 - Probability Density Function

$$p_{I\circ\Phi}(i) = \int_{\Lambda} p_{I\circ\Phi,J}(i)di \tag{C.4}$$

$$p_{I\circ\Phi,J}(i,j) = \frac{h_{I\circ\Phi,J}(i,j)}{\int_{\Lambda^2} h(k,l)dk\ dl} \tag{C.5}$$

where h is the joint histogram. Again, the discrete version replaces the integrals with sums, as the histogram is normalized by the total number of entries. The straightforward normalization is due to the partition of unity property of the Parzen-window in Step 4.

Step 4 - Histogram

$$h_{I\circ\Phi,J}(i,j) = \int_{\Omega\times S^2} P_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - i)P_\beta(J(\boldsymbol{x},\boldsymbol{v}) - j)d\boldsymbol{x}\times d\boldsymbol{v} \tag{C.6}$$

where $I(\cdot)$ and $J(\cdot)$ are image intensities linearly normalized to the histogram size, and P is a Gaussian Parzen-window with variance β - the first scale-space kernel in our model (intensity scale) used to create a smooth histogram.

In the discrete version, we approximate the Parzen-Window with a cubic B-spline kernel for local and computationally feasible support. This also ensures that each smooth entry into the histogram sums to one (partition of unity), which makes normalization trivial. Let $\mathbf{I}$ and $\mathbf{J}$ be the two discrete images where the intensities are normalized by the histogram size. We write up the well-known 1D cubic B-spline basis functions

$$B_t(u) = \begin{cases} (1-u)^3/6 & \text{if } t = 0 \\ (3u^3 - 6u^2 + 4)/6 & \text{if } t = 1 \\ (-3u^3 + 3u^2 + 3u + 1)/6 & \text{if } t = 2 \\ u^3/6 & \text{if } t = 3 \\ 0 & \text{otherwise} \end{cases} \tag{C.7}$$

where u is a value between 0 and 1. Replacing the Gaussian Parzen-Window in eq. (C.6) with a cubic B-spline definition, we define the size of a bin (i,j) in the discrete histogram $\mathbf{h}$ over the entire image domain as

$$\mathbf{h}(i,j) = \sum_{u\in\mathbf{I},v\in\mathbf{J}} B_{t_i}(u)B_{t_j}(w) \tag{C.8}$$

where u, w are the non-negative fractional part of the image intensity $u = \mathbf{I}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) - \lfloor\mathbf{I}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))\rfloor$, $w = \mathbf{J}(\boldsymbol{x},\boldsymbol{v}) - \lfloor\mathbf{J}(\boldsymbol{x},\boldsymbol{v}))\rfloor$, and $t_i = \lfloor\mathbf{I}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))\rfloor - i + 1$, $t_j = \lfloor\mathbf{J}(\boldsymbol{x},\boldsymbol{v})\rfloor - j + 1$. This definition also accounts

for the non-trivial contributions to $\mathbf{h}(i,j)$ from neighbouring bins, as any update to $\mathbf{h}$ by a pair of intensity values changes 16 bins instead of just 1. Let $\mathbf{h}_{\text{sub}^{(i,j)}}$ be the 4×4 submatrix of $\mathbf{h}$ formed from rows $(i-1, i, i+1, i+2)$ and columns $(j-1, j, j+1, j+2)$. Then, we define an update of the joint histogram $\mathbf{h}$ at (i,j) for some pair (u,w) as

$$\mathbf{h}_{\text{sub}^{(i,j)}} = \mathbf{h}_{\text{sub}^{(i,j)}} + \begin{bmatrix} B_0(u)B_0(w) & B_0(u)B_1(w) & B_0(u)B_2(w) & B_0(u)B_3(w) \\ B_1(u)B_0(w) & B_1(u)B_1(w) & B_1(u)B_2(w) & B_1(u)B_3(w) \\ B_2(u)B_0(w) & B_2(u)B_1(w) & B_2(u)B_2(w) & B_2(u)B_3(w) \\ B_3(u)B_0(w) & B_3(u)B_1(w) & B_3(u)B_2(w) & B_3(u)B_3(w) \end{bmatrix} \tag{C.9}$$

We can further simplify eq. (C.9) by writing eq. (C.7) as a vector $\mathbf{B}(u) = [B_0(u), B_1(u), B_2(u), B_3(u)]^T$ so that the matrix of basis functions is given by the outer (Kronecker) product

$$\mathbf{h}_{\text{sub}^{(i,j)}} = \mathbf{h}_{\text{sub}^{(i,j)}} + \mathbf{B}(u) \otimes \mathbf{B}(w) \tag{C.10}$$

Step 5 - Interpolation

$$I_{\Phi_c}(\phi(\boldsymbol{x}), \psi(\boldsymbol{v})) = \int_{S^2} \left(\int_{\Omega} I(\boldsymbol{\tau}, \boldsymbol{\nu}) K_\sigma(\boldsymbol{\tau} - \phi(\boldsymbol{x})) d\boldsymbol{\tau} \right) \Gamma_\kappa(\boldsymbol{\nu}, \psi(\boldsymbol{v})) d\boldsymbol{\nu}. \tag{C.11}$$

where K is the spatial Gaussian kernel at variance σ, and Γ is the directional Watson kernel at concentration κ. These are the other two kernels in our scale-space model, representing the spatial and orientational smoothness.

For the discrete case, we start with the spatial Gaussian kernel which we approximate with a cubic B-spline (similar to Step 4). Since each interpolated value is independent of its neighbors, we skip the outer sum over the image domain and simply state that the calculation is applied for all points $\boldsymbol{x}$. Let $(\delta_x, \delta_y, \delta_z)$ denote the scale of the interpolated 3D image relative to the original image. We then have

$$\mathbf{I}(\phi(\boldsymbol{x}), \boldsymbol{\nu}) = \mathbf{I} \circ K_\sigma(\phi(\boldsymbol{x})) \tag{C.12}$$

$$= \sum_{t_x=0}^{3} \sum_{t_y=0}^{3} \sum_{t_z=0}^{3} B_{t_x}(u) B_{t_y}(w) B_{t_z}(q)\ a_{\boldsymbol{l}, \boldsymbol{\nu}_n} \tag{C.13}$$

where B is the 1D cubic basis functions from eq. (C.7), $u = \phi(\boldsymbol{x})_x/\delta_x - \lfloor \phi(\boldsymbol{x})_x/\delta_x \rfloor$, $w = \phi(\boldsymbol{x})_y/\delta_y - \lfloor \phi(\boldsymbol{x})_y/\delta_y \rfloor$, $q = \phi(\boldsymbol{x})_z/\delta_z - \lfloor \phi(\boldsymbol{x})_z/\delta_z \rfloor$, and a is a scalar value in the image $\mathbf{I}$ at the spatial coordinate $\boldsymbol{l} = (\lfloor \phi(\boldsymbol{x})_x/\delta_x \rfloor +$

t_x , $\lfloor \phi(\boldsymbol{x})_y/\delta_y \rfloor + t_y$, $\lfloor \phi(\boldsymbol{x})_z/\delta_z \rfloor + t_z) - 1$ in the direction $\boldsymbol{\nu}_n$. Equation (C.13) is repeated every point $\boldsymbol{x}$, and we move on to the discrete directional Watson kernel. Again, the interpolation of every direction $\boldsymbol{v}$ is independent of its neighbors and the following is applied to all $\boldsymbol{v}$

$$\mathbf{I}(\phi(\boldsymbol{x}), \psi(\boldsymbol{v})) = \mathbf{I}(\phi(\boldsymbol{x})) \circ \Gamma_\kappa(\boldsymbol{\nu}, \psi(\boldsymbol{v})) \tag{C.14}$$

$$= C \cdot \sum_n e^{\kappa \cdot (\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v}) \rangle)^2} \mathbf{I}(\phi(\boldsymbol{x}), \boldsymbol{\nu}_n) \tag{C.15}$$

$$= \frac{1}{\sum_n e^{\kappa \cdot (\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v}) \rangle)^2}} \sum_n e^{\kappa \cdot (\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v}) \rangle)^2} \mathbf{I}(\phi(\boldsymbol{x}), \boldsymbol{\nu}_n) \tag{C.16}$$

where n is the number of $\boldsymbol{\nu}$ directions in the original image image, κ is the concentration parameter, the operator $\langle \cdot, \cdot \rangle$ denotes the inner product, and C is the normalization constant. Normally, C would be defined by the Kummer function [Jensen et al., 2015]. However, when the Watson distribution is used as an interpolating function, it does not ensure the partition of unity property, and so the Kummer function was replaced with simple normalization over the discrete set of directions. The normalization becomes a function of the interpolated result, which adds to the complexity.

Step 6 and 7 - Transformation Model

The directional transformation $\psi_{\nabla \phi_{\boldsymbol{c}}}(\boldsymbol{v})$ of Step 6 depends on the first-order derivatives of the spatial transformation $\phi_{\boldsymbol{c}}(\boldsymbol{x})$ in Step 7. To make the notation easier to follow, we start with the spatial transformation model which is similar to eq. (C.13). Let $(\delta_x, \delta_y, \delta_z)$ denote the uniform spacing of the vector field $\boldsymbol{c}$, where $\boldsymbol{c}$ are the parameters controlling the transformation. The transformation applies to all points $\boldsymbol{x}$ and is defined as

$$\phi_c(\boldsymbol{x}_d) = \boldsymbol{x}_d + \mathcal{B}_d(\boldsymbol{x}, \boldsymbol{c}) \tag{C.17}$$

$$= \boldsymbol{x}_d + \sum_{t_x=0}^{3} \sum_{t_y=0}^{3} \sum_{t_z=0}^{3} B_{t_x}(u) B_{t_y}(w) B_{t_z}(q) \boldsymbol{c}_{\boldsymbol{l},d} \tag{C.18}$$

where $u = \boldsymbol{x}_x/\delta_x - \lfloor \boldsymbol{x}_x/\delta_x \rfloor$, $w = \boldsymbol{x}_y/\delta_y - \lfloor \boldsymbol{x}_y/\delta_y \rfloor$, $q = \boldsymbol{x}_z/\delta_z - \lfloor \boldsymbol{x}_z/\delta_z \rfloor$, and d is the xyz-index in the vector $\boldsymbol{c}_{\boldsymbol{l},d}$ located at $\boldsymbol{l} = (\lfloor \boldsymbol{x}_x/\delta_x \rfloor) + t_x - 1$, $\lfloor \boldsymbol{x}_y/\delta_y \rfloor) + t_y - 1$, $\lfloor \boldsymbol{x}_z/\delta_z \rfloor) + t_z - 1)$. This is the spatial transformation model based on the Free-Form Deformation (FFD) by [Rueckert et al., 1999] for nonrigid registration. B-Splines are an attractive way of modeling the transformation as they are computationally efficient, independent of the image modality, popular in the registration community, and the derivatives easy to compute. The parameters $\boldsymbol{c}$ are also referred to as *knots* or *control points* in the B-spline model, and uniform mesh of control points is known

as the *deformation field*. The spacing δ between control points determine the degree of nonrigid deformation where a low mesh resolution models global changes and a high-resolution models small local changes.

The directional transformation for the vector $\boldsymbol{v}$ at the point $(\boldsymbol{x}, \boldsymbol{v})$ is then given by

$$\psi_{\nabla\phi_{\boldsymbol{c}}}(\boldsymbol{v}) = \psi_{\boldsymbol{c}}(\boldsymbol{v}) = \frac{\nabla\phi_{\boldsymbol{c}}(\boldsymbol{v})}{||\nabla\phi_{\boldsymbol{c}}(\boldsymbol{v})||} = \frac{\boldsymbol{J}_\phi\boldsymbol{v}}{||\boldsymbol{J}_\phi\boldsymbol{v}||} = \frac{\boldsymbol{J}_\phi\boldsymbol{v}}{\sqrt{(\boldsymbol{J}_\phi\boldsymbol{v})^T\boldsymbol{J}_\phi\boldsymbol{v}}} \tag{C.19}$$

where $\nabla\phi$ is the gradients of the spatial transformation, i.e. Jaocbian $\boldsymbol{J}$, applied to the directional vector $\boldsymbol{v}$. This projective transformation $\mathbb{P}^2 : t\boldsymbol{v} \mapsto t\nabla\phi_{\boldsymbol{c}}(\boldsymbol{v}), t \in \mathbb{R}\backslash\{0\}$ of the directional vectors can also be found in other works of DWI registration [Tao and Miller, 2006, Yap et al., 2010], and offer an approach to reorientation that is independent of specific DWI model assumptions. We write the Jacobian of the spatial transformation as

$$\nabla\phi_{\boldsymbol{c}}(\boldsymbol{v}) = D\phi_{\boldsymbol{c}}(\boldsymbol{x})\boldsymbol{v} = \boldsymbol{J}_\phi\boldsymbol{v} \tag{C.20}$$

where D is the partial derivative operator. The partial derivatives of the weights $\mathcal{B}$ is defined as

$$D_x\mathcal{B}_x = \sum_{t_x=0}^{3}\sum_{t_y=0}^{3}\sum_{t_z=0}^{3} dB_{t_x}(u)B_{t_y}(w)B_{t_z}(q)\boldsymbol{c}_{\boldsymbol{l},x} \tag{C.21}$$

$$D_y\mathcal{B}_x = \sum_{t_x=0}^{3}\sum_{t_y=0}^{3}\sum_{t_z=0}^{3} B_{t_x}(u)dB_{t_y}(w)B_{t_z}(q)\boldsymbol{c}_{\boldsymbol{l},x} \tag{C.22}$$

$$D_z\mathcal{B}_x = \sum_{t_x=0}^{3}\sum_{t_y=0}^{3}\sum_{t_z=0}^{3} B_{t_x}(u)B_{t_y}(w)dB_{t_z}(q)\boldsymbol{c}_{\boldsymbol{l},x} \tag{C.23}$$

then we can write the Jacobian of eq. (C.18) as

$$D\phi_{\boldsymbol{c}}(\boldsymbol{x}) = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} + \begin{bmatrix}D_x\mathcal{B}_x & D_x\mathcal{B}_y & D_x\mathcal{B}_z\\ D_y\mathcal{B}_x & D_y\mathcal{B}_y & D_y\mathcal{B}_z\\ D_z\mathcal{B}_x & D_z\mathcal{B}_y & D_z\mathcal{B}_z\end{bmatrix} = \boldsymbol{J}_\phi \tag{C.24}$$

The first-order derivatives for the basis functions of eq. (C.7) are

$$dB_t(u) = \begin{cases}(-3u^2 + 6u - 3)/6 & \text{if } t = 0\\ (9u^2 - 12u)/6 & \text{if } t = 1\\ (-9u^2 + 6u + 3)/6 & \text{if } t = 2\\ 3u^2/6 & \text{if } t = 3\\ 0 & \text{otherwise}\end{cases} \tag{C.25}$$

To optimize the objective function in our framework, we use quasi-Newton methods. To do so requires computing the gradient of our objective with respect to parameters $\boldsymbol{c}$. This is a complex task due to the reorientation already defined through the first-order derivatives of the spatial motion.

C.2 Unfolding the Dependencies - Part 2

The previous section has demonstrated the complex dependency of the objective to $\boldsymbol{c}$. To compute the gradient $D_{\boldsymbol{c}}\mathcal{F}$, we use the chain rule together with the differentials of all the unfolding steps, from the top, writing up the derivatives of each node in the path NMI to $\Phi_{\boldsymbol{c}}$ of the dependency graph in Figure C.1. We will be using common rules, such as the quotient and product rule, to write up the derivatives. In particular, we will use the special case of the Leibniz Integral Rule (LIR), which allow us to move the differential operators inside the integral under the condition that integral limits are constant. We use the chain rule in the form $d_f g = Dg(f)df$ where $Dg(f)$ is the differential of g, i.e. the matrix of partial derivatives of g, computed at f, and df is the vector of "variations" of f. The subscript denotes the dependency that we seek if the function g is multivariate. In differential geometric df is a *differential form*.

Step 1 - Derivative of the Similarity Measure

The derivatives of the similarity measure are found by using the quotient rule

$$dNMI(\Phi, I, J) = d\left(\frac{H_J + H_{I\circ\Phi}}{H_{I\circ\Phi,J}}\right) \tag{C.26}$$

$$= \frac{(dH_{I\circ\Phi} + dH_J)H_{I\circ\Phi,J} - dH_{I\circ\Phi,J}(H_{I\circ\Phi} + H_J)}{(H_{I\circ\Phi,J})^2} \tag{C.27}$$

Since we only care about the dependency on Φ_c we have that $dH_J = 0$ and thus

$$d_\Phi NMI(\Phi, I, J) = \frac{dH_{I\circ\Phi}\ H_{I\circ\Phi,J} - dH_{I\circ\Phi,J}(H_{I\circ\Phi} + H_J)}{(H_{I\circ\Phi,J})^2} \tag{C.28}$$

Step 2 - Derivative of the Entropy

We use LIR and the product rule

$$dH_{I\circ\Phi} = -d\int_{\Lambda} p_{I\circ\Phi}(i)\cdot log(p_{I\circ\Phi}(i))di \tag{C.29}$$

$$= -\int_{\Lambda}(1+log(p_{I\circ\Phi}(i)))dp_{I\circ\Phi}di \tag{C.30}$$

$$dH_{I\circ\Phi,J} = -d\int_{\Lambda^2} p_{I\circ\Phi,J}(i,j)\cdot log(p_{I\circ\Phi,J}(i,j))di\ dj \tag{C.31}$$

$$= -\int_{\Lambda^2}(1+log(p_{I\circ\Phi,J}(i,j)))dp_{I\circ\Phi,J}di\ dj \tag{C.32}$$

Step 3 - Derivative of the Probability Density Function

We use LIR and the quotient rule

$$dp_{I\circ\Phi}(i) = d\int_{\Lambda} p_{I\circ\Phi,J}(i)di \tag{C.33}$$

$$= \int_{\Lambda} dp_{I\circ\Phi,J}(i)di \tag{C.34}$$

$$dp_{I\circ\Phi,J}(i,j) = d\frac{h_{I\circ\Phi,J}(i,j)}{\int_{\Lambda^2} h_{I\circ\Phi,J}(k,l)dk\ dl} \tag{C.35}$$

$$= \frac{dh_{I\circ\Phi,J}(i,j)}{\int_{\Lambda^2} h_{I\circ\Phi,J}(l,k)dl\ dk} - \frac{h_{I\circ\Phi,J}(i,j)\int_{\Lambda^2} dh_{I\circ\Phi,J}(l,k)dl\ dk}{(\int_{\Lambda^2} h_{I\circ\Phi,J}(l,k)dl\ dk)^2} \tag{C.36}$$

Step 4 - Derivative of the Histogram

We use LIR and the product rule

$$dh_{I\circ\Phi,J}(i,j) = d\int_{\Omega\times S^2} P_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)P_\beta(J(\boldsymbol{x},\boldsymbol{v})-j)d\boldsymbol{x}\times d\boldsymbol{v} \tag{C.37}$$

$$= \int_{\Omega\times S^2} dP_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)P_\beta(J(\boldsymbol{x},\boldsymbol{v})-j)d\boldsymbol{x}\times d\boldsymbol{v}$$

$$+ \quad P_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)dP_\beta(J(\boldsymbol{x},\boldsymbol{v})-j)d\boldsymbol{x}\times d\boldsymbol{v} \tag{C.38}$$

As with Step 1, we seek only the dependency on Φ_c and so we have that

$$d_\Phi h_{I\circ\Phi,J}(i,j) = \int_{\Omega\times S^2} dP_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)P_\beta(J(\boldsymbol{x},\boldsymbol{v})-j)d\boldsymbol{x}\times d\boldsymbol{v}$$

$$+ P_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)\cdot 0\ d\boldsymbol{x}\times d\boldsymbol{v} \tag{C.39}$$

$$= \int_{\Omega\times S^2} dP_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))-i)P_\beta(J(\boldsymbol{x},\boldsymbol{v})-j)\ d\boldsymbol{x}\times d\boldsymbol{v} \tag{C.40}$$

From here we can use the chain rule on the histogram kernel

$$dP_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v}))) = DP_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})))dI(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) \tag{C.41}$$

Relating this to the discrete B-spline kernel version of the Parzen-window, the total effect of adding an entry from $\mathbf{I}$ to the joint histogram is simply the sum or inner product over eq. (C.9) so that

$$dP_\beta(I(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})))P_\beta(J(\boldsymbol{x},\boldsymbol{v})) = \langle D\mathbf{B}(u),\mathbf{B}(w)\rangle\, dI(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) \tag{C.42}$$

where $D\mathbf{B}(u)$ is defined in eq. (C.25).

Step 5 - Derivative of the Interpolation

We use the product rule and LIR

$$d\mathbf{I}_{\sigma\kappa}(\phi(\boldsymbol{x}),\psi(\boldsymbol{v})) = d\int_{S^2}\left(\int_\Omega \mathbf{I}(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))d\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} \tag{C.43}$$

$$= \int_{S^2}\left(d\int_\Omega \mathbf{I}(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))d\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} +$$

$$\int_{S^2}\left(\int_\Omega \mathbf{I}(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))\boldsymbol{\tau}\right)d\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} \tag{C.44}$$

$$= \int_{S^2}\left(\int_\Omega \mathbf{I}(\boldsymbol{\tau},\boldsymbol{\nu})dK_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))d\boldsymbol{\tau}\right)\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} +$$

$$\int_{S^2}\left(\int_\Omega \mathbf{I}(\boldsymbol{\tau},\boldsymbol{\nu})K_\sigma(\boldsymbol{\tau}-\phi(\boldsymbol{x}))\boldsymbol{\tau}\right)d\Gamma_\kappa(\boldsymbol{\nu},\psi(\boldsymbol{v}))d\boldsymbol{\nu} \tag{C.45}$$

As with the histogram, the partial derivatives of the discrete version of spatial kernel $K(\phi_\sigma(\boldsymbol{x}))$ is straight forward for the B-spline kernel. As the interpolation, with the use of a kernel, is a convolution, we use the shorthand notation of eq. (C.21)-eq. (C.23) and the derivatives can be written as

$$d(\mathbf{I} * K_\sigma(\phi(\boldsymbol{x}))) = \mathbf{I} * dK_\sigma(\phi(\boldsymbol{x})) \tag{C.46}$$

$$dK_\sigma(\phi(\boldsymbol{x})) = D\mathcal{B}_\mathcal{I}(\phi(\boldsymbol{x}))d\phi(\boldsymbol{x}) \tag{C.47}$$

where $\mathcal{B}_\mathcal{I}$ is the B-spline image interpolation kernel. The directional Watson

kernel is somewhat more complex due to the normalization constant

$$\mathbf{I}(\phi(\boldsymbol{x})) \circ d\Gamma_\kappa(\boldsymbol{\nu}, \psi(\boldsymbol{v})) \tag{C.48}$$

$$= D\Big(\frac{1}{\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}} \sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2} \mathbf{I}(\phi(\boldsymbol{x}), \nu_n)\Big) d\psi(\boldsymbol{v}) \tag{C.49}$$

$$= \Bigg(D\Big(\frac{1}{\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}}\Big) \sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2} + \frac{1}{\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}} D\Big(\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}\Big)\Bigg) d\psi(\boldsymbol{v}) \tag{C.50}$$

$$= \Bigg(D\Big(\frac{1}{\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}}\Big) \sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2} + \frac{1}{\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}} \sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}\, 2\kappa D(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)\Bigg) d\psi(\boldsymbol{v}) \tag{C.51}$$

$$= \Bigg(\frac{1}{\big(\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}\big)^2} \sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}\, 2\kappa D(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle) - \frac{1}{\sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}} \sum_n e^{\kappa\cdot(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)^2}\, 2\kappa D(\langle \boldsymbol{\nu}_n, \psi(\boldsymbol{v})\rangle)\Bigg) d\psi(\boldsymbol{v}) \tag{C.52}$$

The product rule was used in eq. (C.50), and the chain rule once in eq. (C.51) and twice in eq. (C.52). The partial derivatives of the inner product of two vectors is trivial.

Step 6 - Derivative of the Directional Transformation

This is a slightly complex part, as we seek to write up the derivatives of the spatial Jacobian applied to the directional vectors with respect to the

transformation parameters

$$d\psi_{\boldsymbol{c}}(\boldsymbol{v}) = D\left(\frac{\boldsymbol{J}_\phi \boldsymbol{v}}{||\boldsymbol{J}_\phi \boldsymbol{v}||}\right) d\boldsymbol{c} \tag{C.53}$$

$$= D\left(\frac{\boldsymbol{J}_\phi \boldsymbol{v}}{\sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}}}\right) d\boldsymbol{c} \tag{C.54}$$

$$= \frac{1}{\langle \boldsymbol{J}_\phi \boldsymbol{v}, \boldsymbol{J}_\phi \boldsymbol{v} \rangle} \left(D\boldsymbol{J}_\phi \boldsymbol{v} \sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}} - \boldsymbol{J}_\phi \boldsymbol{v}\, D\left(\sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}}\right) \right) d\boldsymbol{c} \tag{C.55}$$

$$= \frac{1}{\langle \boldsymbol{J}_\phi \boldsymbol{v}, \boldsymbol{J}_\phi \boldsymbol{v} \rangle} \left(D\boldsymbol{J}_\phi \boldsymbol{v} \sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}} - \frac{\boldsymbol{J}_\phi \boldsymbol{v}}{\sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}}} D\left((\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}\right) \right) d\boldsymbol{c} \tag{C.56}$$

$$= \frac{1}{\langle \boldsymbol{J}_\phi \boldsymbol{v}, \boldsymbol{J}_\phi \boldsymbol{v} \rangle} \left(D\boldsymbol{J}_\phi \boldsymbol{v} \sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}} - \frac{\boldsymbol{J}_\phi \boldsymbol{v}}{\sqrt{(\boldsymbol{J}_\phi \boldsymbol{v})^T \boldsymbol{J}_\phi \boldsymbol{v}}} (D\boldsymbol{J}_\phi \boldsymbol{v})^T D\boldsymbol{J}_\phi \boldsymbol{v} \right) d\boldsymbol{c} \tag{C.57}$$

where we have used the quotient in eq. (C.55), and the chain rule in eq. (C.56). For a B-spline deformation, the derivative of eq. (C.24) with respect to the parameters is a $3 \times 3 \times 3$ matrix of partial derivatives. To see why, take $D_x \mathcal{B}_x$ from eq. (C.21) as example. The partial derivatives with respect to $\boldsymbol{c}$ simply becomes

$$D_{\boldsymbol{c}} D_x \mathcal{B}_x = \sum_{t_x=0}^{3} \sum_{t_y=0}^{3} \sum_{t_z=0}^{3} dB_{t_x}(u) B_{t_y}(w) B_{t_z}(q) \tag{C.58}$$

which essentially boils down to the sum of derivatives of the B-spline basis functions replicated three times since $D_{\boldsymbol{c}_x} D_x \mathcal{B}_x = D_{\boldsymbol{c}_y} D_x \mathcal{B}_x = D_{\boldsymbol{c}_z} D_x \mathcal{B}_x$.

Step 7 - Derivative of the Spatial Transformation

The partial derivatives of the spatial transformation with respect to $\boldsymbol{c}$ is simply the weights of the basis functions

$$d\phi_c(\boldsymbol{x}_d) = D(\boldsymbol{x}_d + \mathcal{B}_d(\boldsymbol{x}, \boldsymbol{c}))d\boldsymbol{c} \tag{C.59}$$

$$= \left(\sum_{t_x=0}^{3} \sum_{t_y=0}^{3} \sum_{t_z=0}^{3} B_{t_x}(u) B_{t_y}(w) B_{t_z}(q) \right) d\boldsymbol{c} \tag{C.60}$$

Which brings us to the end of the chain of dependencies from Step 1 to Step 7. The tricky part is connecting all the partial derivatives bottom-up in a single long chain rule.

Putting it together

As mentioned in the beginning of the previous paragraph, we optimize our objective via a quasi-Newton method and we need need to compute its gradient w.r.t $\boldsymbol{c}$. Similar to gradient back-propagation, this is done by chaining the steps of Part 2 via the chain rule.

$$
\begin{aligned}
&d_c NMI \\
&= D_{H(p(\cdots))} NMI \; dH(p(h(I(\psi(\phi(\boldsymbol{c}))))))\\
&= D_{H(p(\cdots))} NMI \cdot D_{p(h(\cdots))} H \; dp(h(I(\psi(\phi(\boldsymbol{c})))))\\
&= D_{H(p(\cdots))} NMI \cdot D_{p(h(\cdots))} H \cdot D_{h(I(\cdots))} \; dh(I(\psi(\phi(\boldsymbol{c}))))\\
&= D_{H(p(\cdots))} NMI \cdot D_{p(h(\cdots))} H \cdot D_{h(I(\cdots))} \cdot D_{I(\psi(\cdots))} \; dI(\psi(\phi(\boldsymbol{c})))\\
&= D_{H(p(\cdots))} NMI \cdot D_{p(h(\cdots))} H \cdot D_{h(I(\cdots))} \cdot D_{I(\psi(\cdots))} \cdot D_{\psi(\phi(\boldsymbol{c}))} \; d\psi(\phi(\boldsymbol{c}))\\
&= D_{H(p(\cdots))} NMI \cdot D_{p(h(\cdots))} H \cdot D_{h(I(\cdots))} \cdot D_{I(\psi(\cdots))} \cdot D_{\psi(\phi(\boldsymbol{c}))} \cdot D_{\psi(\phi(\boldsymbol{c}))} \; d\phi(\boldsymbol{c})\\
&= D_{H(p(\cdots))} NMI \cdot D_{p(h(\cdots))} H \cdot D_{h(I(\cdots))} \cdot D_{I(\psi(\cdots))} \cdot D_{\psi(\phi(\boldsymbol{c}))} \cdot D_{\psi(\phi(\boldsymbol{c}))} \cdot D_{\phi(\boldsymbol{c})} d\boldsymbol{c}
\end{aligned}
\tag{C.61}
$$

In the previous part, we have established analytic formulas of all the differentials appearing in (C.61), and by computing from right to left, with an actual value for $\boldsymbol{c}$, we obtain the differential of our objective function and thus its gradient. Details of computations were provided in Section 5.3 and Appendix A.

D Longitudinal Registration of Alzheimer's DWI

To validate our registration framework and investigate new diffusion biomarkers, we downloaded a dataset of normal controls (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD) subjects from the public ADNI database [Jack et al., 2008]. The MCI are further divided into early-MCI and late-MCI depending on the level of impairment, but due to the size of the datasets, we have pooled them into MCI in the results. The distribution of dataset can be seen in Table D.1. All of the subjects have a follow-up scan

Population	Females	Males	Age
NC	15	13	73.55 ± 6.14
EMCI	6	17	70.91 ± 8.09
LMCI	9	18	72.70 ± 6.26
AD	9	20	75.97 ± 9.94
Total	39	68	73.42 ± 7.26

Table D.1: Distribution of gender and age in the populations of DWI subjects downloaded from ADNI.

that was taken 1-2 years later. The DTI data corresponds to the preprocessed data made available in [Nir et al., 2013]. The structural T1w images has a voxel size $= 1.2 \times 1.0 \times 1.0$ mm and the DWI are isotropic voxels of size $2.7 \times 2.7 \times 2.7$ mm at $b = 1000$, re-sampled during preprocessing to $2 \times 2 \times 2$ mm with 41 diffusion gradients. We refer to [Nir et al., 2013] for additional details on the data.

To segment the data, we used `VolBrain` on the T1w images. `VolBrain` [Manjón and Coupé, 2016] is an online openly available tool for automatic volumetric brain segmentation and statistical information. We found it to provide excellent and precise segmentations of the T1w scans. We used `FSL FLIRT` to convert the labels to the DWI space through global affine registra-

tion. Finally, we applied the LOR-DWI framework for nonrigid registration between the original scans and their follow-up scans in order to investigate the temporal difference. It should be noted that at the time of the experiment, the registration was only hierarchical in the control point spacing of the deformation field, and not hierarchical in the spatial and intensity scales as described in Chapter 5.

For each subject, we calculated the mean of the Jacobian determinant based on the deformation at each voxel within a given segmentation. We then calculated the mean over each population, which is the values shown in Table D.2. These values represent the net contraction (< 1) or expansion (> 1) of each anatomical structure in the population, which are often used as indications of atrophy or growth.

ROI	$\det \boldsymbol{J}$ (NC)	$\det \boldsymbol{J}$ (MCI)	$\det \boldsymbol{J}$ (AD)
White Matter	0.9713	0.9744	0.9937
Grey Matter	0.9612	0.9617	0.9825
Ventricles	1.0110	1.0501	1.0734
Caudate	1.0190	1.0177	1.0201
Putamen	1.0331	1.0037	0.9957
Thalamus	1.0022	0.9930	1.0124
Globus Pallidus	1.0517	1.0233	1.0101
Hippocampus	0.9913	0.9907	0.9877
Amigdala	0.9697	0.9584	0.8921
Accumbens	1.0517	0.9976	0.9952
Brainstem	0.9866	0.9856	1.0079
Cerebellum	0.9633	0.9658	0.9928

Table D.2: The mean of each population's mean Jacobian determinant for each label category. The red and green colors indicate some of the most prominent incorrect and correct results.

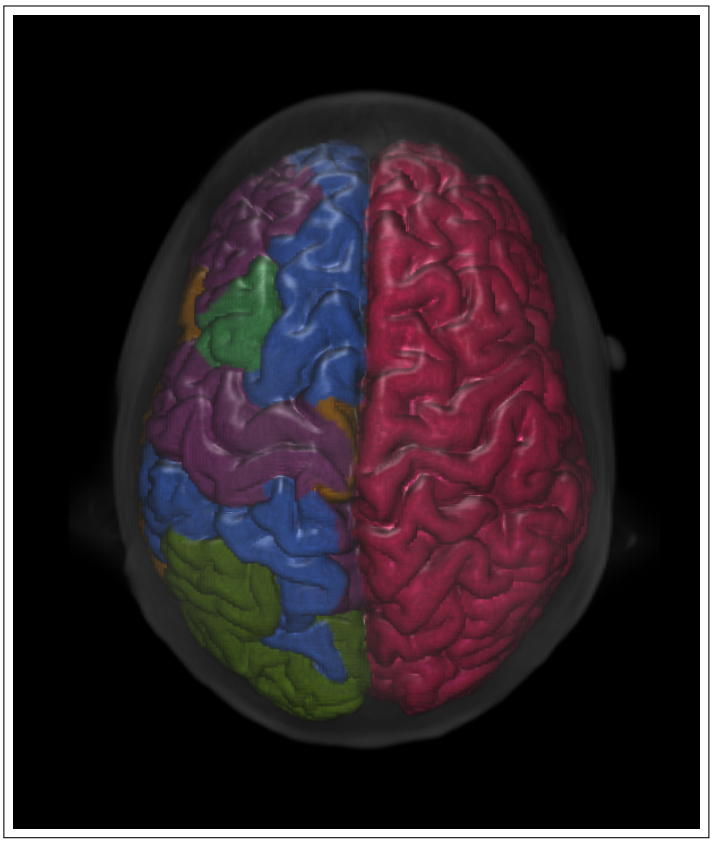
Evidently, the overall registrations were unable to capture the expected deformations, as it is well-known that the there is atrophy in the white and gray matter of AD subjects. Here, indicated in red text, we observe more atrophy in NC than in MCI and AD, which is incorrect [Frisoni et al., 2002, Thompson et al., 2003, Serra et al., 2010]. The growth in ventricles and the atrophy in the hippocampus is as expected. However, given the clear indications of a poor registration, we can not trust the results. The reasons for the poor registrations can likely be found in the preprocessed data.

There are several issues with the dataset. First of all, the image resolution is insufficient for measuring atrophy, as indicated by the results in the gray

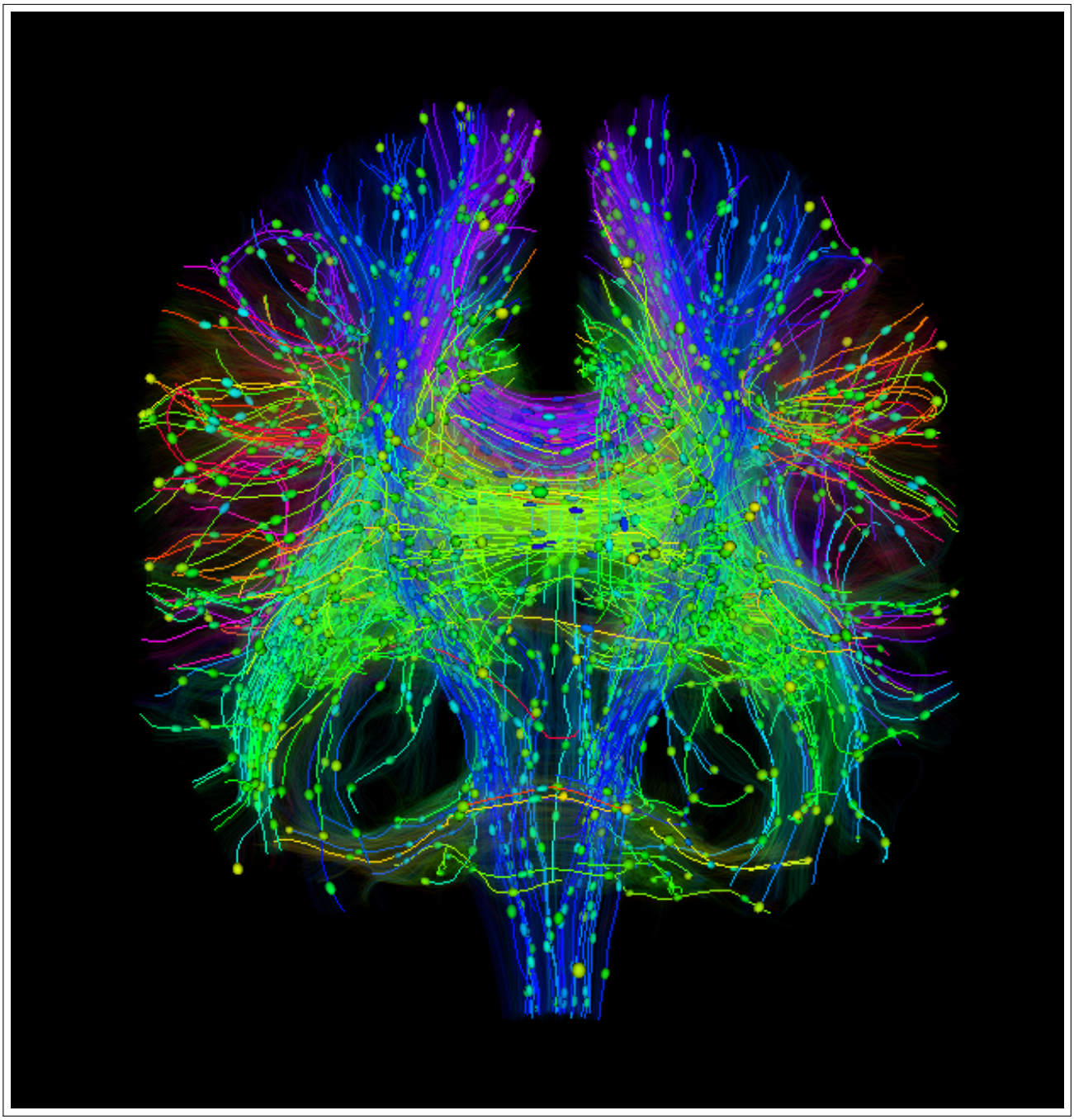
matter where NC has the highest atrophy. The average cerebral cortex has a thickness ranging from 1.5 to 4.5 mm [Narr et al., 2004, Winkler et al., 2010]. Since the original isotropic voxel had a size of 2.7 mm^3, it is highly unlikely that the accuracy is good enough to capture atrophy in the cortex without severe CSF contamination. Additionally, the motion correction performed with eddy correct should cause a change in the diffusion directions, but this does not seem to be the case. Nor was the EPI-correction performed according to a field map, but through a non-linear registration, which has little to do with the actual susceptibility artifacts. As we discovered several of these issues, we looked into performing the preprocessing. However, without a field map from the original scan, there does not seem to be a good way to correct the data.

E Artistic Illustrations of Brains and DWI Data

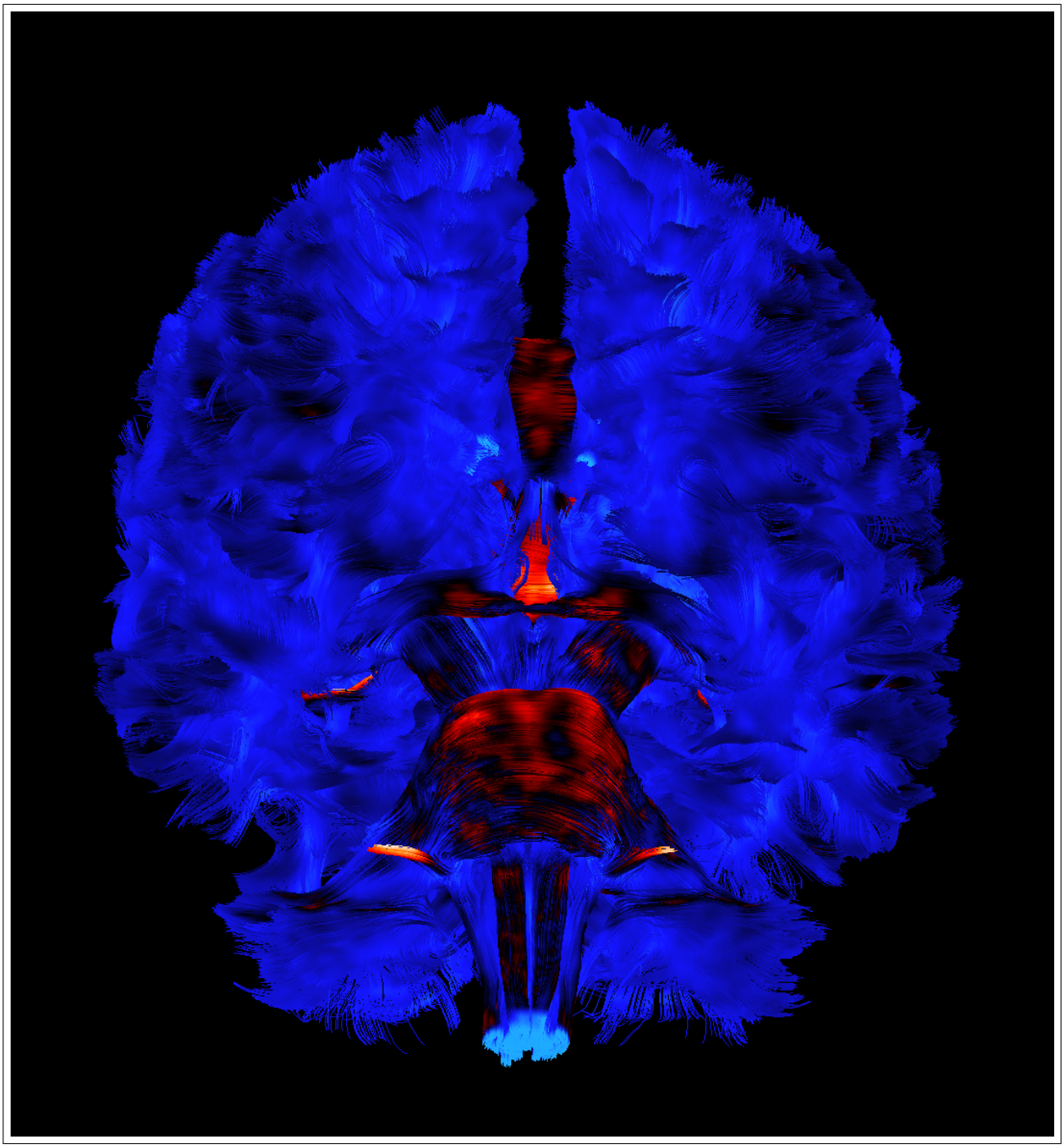
On a personal note, one of the really cool things about MRI and DWI was the visualization of the brain, its subdivisions and connections. The following images are based on real HCP subjects, and, while I can not guarantee anatomical correctness on e.g. the DTI tractographies, I still think it gives a really inspiring view of what DWI can be used to.



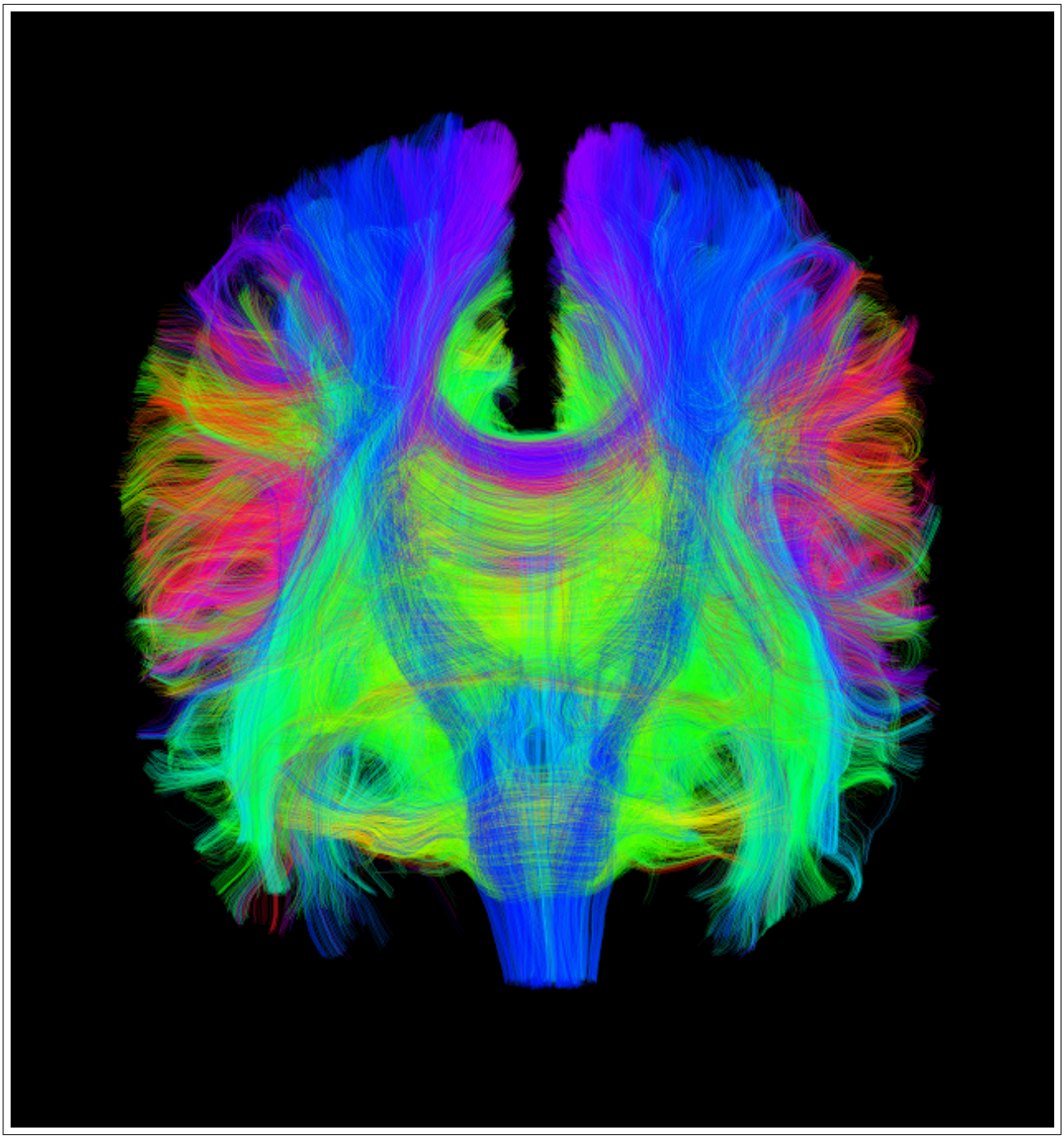
Axial view of sparse whole-brain tractography coloured by the orientation of the principal eigenvector. The ODFs used for streamline-seeding are interspersed.



Coronal view of sparse whole-brain tractography coloured by the orientation of the principal eigenvector. Glyphs indicating ODFs used for streamline-seeding are interspersed.



A dense image of whole-brain tractography showing the streamlines coloured by anisotropy.



Coronal view of a dense whole-brain tractography coloured by the orientation of the principal eigenvector.

Bibliography

[Alexander et al., 2000] Alexander, A. L., Hasan, K., Kindlmann, G., Parker, D. L., and Tsuruda, J. S. (2000). A geometric analysis of diffusion tensor measurements of the human brain. *Magnetic Resonance in Medicine*, 44(2):283–291.

[Assaf and Cohen, 2009] Assaf, Y. and Cohen, Y. (2009). Inferring microstructural information of white matter from diffusion mri. *Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy*, 127.

[Basser et al., 1994] Basser, P. J., Mattiello, J., and LeBihan, D. (1994). Mr diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267.

[Bhushan et al., 2012] Bhushan, C., Haldar, J. P., Joshi, A. A., and Leahy, R. M. (2012). Correcting susceptibility-induced distortion in diffusion-weighted mri using constrained nonrigid registration. *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*.

[Bijral et al., 2007] Bijral, A. S., Breitenbach, M., and Grudic, G. (2007). Mixture of watson distributions: a generative model for hyperspherical embeddings. In *Artificial Intelligence and Statistics*, pages 35–42.

[Bouma et al., 2007] Bouma, H., Vilanova, A., Bescós, J. O., ter Haar Romeny, B. M., and Gerritsen, F. A. (2007). Fast and accurate gaussian derivatives based on b-splines. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 406–417. Springer.

[Brown, 1992] Brown, L. G. (1992). A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376.

[Budde et al., 2007] Budde, M. D., Kim, J. H., Liang, H.-F., Schmidt, R. E., Russell, J. H., Cross, A. H., and Song, S.-K. (2007). Toward accurate diagnosis of white matter pathology using diffusion tensor imaging. *Magnetic resonance in medicine*, 57(4):688–695.

[Callaghan, 1993] Callaghan, P. T. (1993). *Principles of nuclear magnetic resonance microscopy*. Oxford University Press on Demand.

[Carr and Purcell, 1954] Carr, H. Y. and Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical review*, 94(3):630.

[CDMRI, 2017] CDMRI, . (2017). Diffusion mri data harmonisation. `https://projects.iq.harvard.edu/cdmri2017`. Accessed: 2017-12-10.

[CRC, 2018] CRC, W. . (2018). Continuous registration challenge. `https://continuousregistration.grand-challenge.org`. Accessed: 2018-03-09.

[Darkner and Sporring, 2013] Darkner, S. and Sporring, J. (2013). Locally orderless registration. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1437–1450.

[De Santis et al., 2014] De Santis, S., Drakesmith, M., Bells, S., Assaf, Y., and Jones, D. K. (2014). Why diffusion tensor mri does well only some of the time: variance and covariance of white matter tissue microstructure attributes in the living human brain. *Neuroimage*, 89:35–44.

[Descoteaux, 2008] Descoteaux, M. (2008). *High angular resolution diffusion MRI: from local estimation to segmentation and tractography*. PhD thesis, Université Nice Sophia Antipolis.

[Descoteaux, 2010] Descoteaux, M. (2010). *High angular resolution diffusion MRI: from local estimation to segmentation and tractography*. PhD thesis, Max Planck Institute, Germany.

[Einstein, 1905] Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 322(8):549–560.

[Frisoni et al., 2002] Frisoni, G., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., and Laakso, M. (2002). Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6):657–664.

[Hawkins, 2004] Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12.

[Jack et al., 2008] Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C., et al. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of magnetic resonance imaging*, 27(4):685–691.

[Jakob and Vclav, 2012] Jakob, P. and Vclav, Z. (2012). ThreadPool.cpp. `https://github.com/progschj/ThreadPool`. Accessed: 2017-07-31.

[Jensen, 2014] Jensen, H. G. (2014). Master's Thesis. Request: `henrikgjensen@gmail.com`.

[Jensen et al., 2015] Jensen, H. G., Lauze, F., Nielsen, M., and Darkner, S. (2015). Locally orderless registration for diffusion weighted images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 305–312. Springer.

[Jensen et al., 2017] Jensen, H. G., Lauze, F. B., Nielsen, M., and Darkner, S. (2017). Density-based non-rigid registration of diffusion-weighted images. In *25th ISMRM Annual Meeting*.

[Johansen-Berg and Behrens, 2013] Johansen-Berg, H. and Behrens, T. E. (2013). *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press.

[Jones, 2010] Jones, D. K. (2010). Challenges and limitations of quantifying brain connectivity in vivo with diffusion mri. *Imaging in Medicine*, 2(3):341.

[Jupp and Mardia, 1989] Jupp, P. and Mardia, K. (1989). A unified view of the theory of directional statistics, 1975-1988. *International Statistical Review/Revue Internationale de Statistique*, pages 261–294.

[Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802.

[Koenderink and Van Doorn, 1999] Koenderink, J. J. and Van Doorn, A. J. (1999). The structure of locally orderless images. *International Journal of Computer Vision*, 31(2):159–168.

[Le Bihan et al., 1986] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M. (1986). Mr imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407.

[Maes et al., 1997] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198.

[Mangin et al., 2016] Mangin, J.-F., Lebenberg, J., Lefranc, S., Labra, N., Auzias, G., Labit, M., Guevara, M., Mohlberg, H., Roca, P., Guevara, P., et al. (2016). Spatial normalization of brain images and beyond.

[Manjón and Coupé, 2016] Manjón, J. V. and Coupé, P. (2016). volbrain: An online mri brain volumetry system. *Frontiers in neuroinformatics*, 10:30.

[Mattes et al., 2003] Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). Pet-ct image registration in the chest using free-form deformations. *IEEE transactions on medical imaging*, 22(1):120–128.

[Mori et al., 2008] Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A. V., Mahmood, A., Woods, R., et al. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an icbm template. *Neuroimage*, 40(2):570–582.

[MUSHAC, 2018] MUSHAC, . (2018). Multi-shell diffusion mri harmonisation challenge. `https://projects.iq.harvard.edu/cdmri2018`. Accessed: 2018-04-02.

[Narr et al., 2004] Narr, K. L., Bilder, R. M., Toga, A. W., Woods, R. P., Rex, D. E., Szeszko, P. R., Robinson, D., Sevy, S., Gunduz-Bruce, H., Wang, Y.-P., et al. (2004). Mapping cortical thickness and gray matter concentration in first episode schizophrenia. *Cerebral cortex*, 15(6):708–719.

[Nir et al., 2013] Nir, T. M., Jahanshad, N., Villalon-Reina, J. E., Toga, A. W., Jack, C. R., Weiner, M. W., Thompson, P. M., (ADNI, A. D. N. I., et al. (2013). Effectiveness of regional dti measures in distinguishing alzheimer's disease, mci, and normal aging. *NeuroImage: clinical*, 3:180–195.

[Norton et al., 2017] Norton, I., Essayed, W. I., Zhang, F., Pujol, S., Yarmarkovich, A., Golby, A. J., Kindlmann, G., Wasserman, D., Estepar, R. S. J., Rathi, Y., et al. (2017). Slicerdmri: open source diffusion mri software for brain cancer research. *Cancer research*, 77(21):e101–e103.

[O'Donnell et al., 2017] O'Donnell, L. J., Daducci, A., Wassermann, D., and Lenglet, C. (2017). Advances in computational and statistical diffusion mri. *NMR in Biomedicine*.

[O'Donnell and Westin, 2011] O'Donnell, L. J. and Westin, C.-F. (2011). An introduction to diffusion tensor image analysis. *Neurosurgery clinics of North America*, 22(2):185–196.

[Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076.

[Peter Basser and Pierpaoli, 1996] Peter Basser, J. and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *J Magn Reson B*, 111(3):209–219.

[Rathi et al., 2009] Rathi, Y., Michailovich, O., Shenton, M. E., and Bouix, S. (2009). Directional functions for orientation distribution estimation. *Medical image analysis*, 13(3):432–444.

[Roche et al., 1999] Roche, A., Malandain, G., Ayache, N., and Prima, S. (1999). Towards a better comprehension of similarity measures used in medical image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 555–566. Springer.

[Rogelj and Kovacic, 2001] Rogelj, P. and Kovacic, S. (2001). Similarity measures for nonrigid registration. In *Medical Imaging 2001: Image Processing*, volume 4322, pages 569–579. International Society for Optics and Photonics.

[Rohde et al., 2004] Rohde, G., Barnett, A., Basser, P., Marenco, S., and Pierpaoli, C. (2004). Comprehensive approach for correction of motion and distortion in diffusion-weighted mri. *Magnetic resonance in medicine*, 51(1):103–114.

[Rueckert et al., 1999] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721.

[Schlaug et al., 1997] Schlaug, G., Siewert, B., Benfield, A., Edelman, R., and Warach, S. (1997). Time course of the apparent diffusion coefficient (adc) abnormality in human stroke. *Neurology*, 49(1):113–119.

[Schmidt, 2005] Schmidt, M. (2005). minfunc: unconstrained differentiable multivariate optimization in matlab. *Software available at https://www.cs.ubc.ca/ schmidtm/Software/minFunc.html*.

[Serra et al., 2010] Serra, L., Cercignani, M., Lenzi, D., Perri, R., Fadda, L., Caltagirone, C., Macaluso, E., and Bozzali, M. (2010). Grey and white matter changes at different stages of alzheimer's disease. *Journal of Alzheimer's Disease*, 19(1):147–159.

[Smith and Nichols, 2018] Smith, S. M. and Nichols, T. E. (2018). Statistical challenges in "big data" human neuroimaging. *Neuron*, 97(2):263–268.

[Sotiras et al., 2013] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190.

[Sporring and Darkner, 2011] Sporring, J. and Darkner, S. (2011). Jacobians for lebesgue registration for a range of similarity measures. *Department of Computer Science, University of Copenhagen, Tech. Rep*, 4.

[Stejskal and Tanner, 1965] Stejskal, E. and Tanner, J. (1965). Spin diffusion measurements: Spin-echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42:288.

[Studholme et al., 1999] Studholme, C., Hill, D. L., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition*, 32(1):71–86.

[Tao and Miller, 2006] Tao, X. and Miller, J. V. (2006). A method for registering diffusion weighted magnetic resonance images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 594–602. Springer.

[Thevenaz and Unser, 1997] Thevenaz, P. and Unser, M. (1997). Spline pyramids for inter-modal image registration using mutual information. *Proceedings of SPIE: Wavelet Applications in Signal and Image Processing, San Diego, CA*, pages 236–247.

[Thévenaz and Unser, 1998] Thévenaz, P. and Unser, M. (1998). An efficient mutual information optimizer for multiresolution image registration. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, volume 1, pages 833–837. IEEE.

[Thompson et al., 2003] Thompson, P. M., Hayashi, K. M., De Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., et al. (2003). Dynamics of gray matter loss in alzheimer's disease. *Journal of neuroscience*, 23(3):994–1005.

[TraCED, 2017] TraCED, I. . (2017). Tractography-reproducibility challenge with empirical data. `https://my.vanderbilt.edu/ismrmtraced2017`. Accessed: 2018-01-12.

[Treiber et al., 2016] Treiber, J. M., White, N. S., Steed, T. C., Bartsch, H., Holland, D., Farid, N., McDonald, C. R., Carter, B. S., Dale, A. M., and Chen, C. C. (2016). Characterization and correction of geometric distortions in 814 diffusion weighted images. *PloS one*, 11(3):e0152472.

[Tuch, 2004] Tuch, D. S. (2004). Q-ball imaging. *Magnetic resonance in medicine*, 52(6):1358–1372.

[Tuch et al., 1999] Tuch, D. S., Weisskoff, R., Belliveau, J., and Wedeen, V. (1999). High angular resolution diffusion imaging of the human brain. In *Proceedings of the 7th Annual Meeting of ISMRM, Philadelphia*, volume 321.

[Van Essen et al., 2013] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79.

[Van Hecke et al., 2007] Van Hecke, W., Leemans, A., D'Agostino, E., De Backer, S., Vandervliet, E., Parizel, P. M., and Sijbers, J. (2007). Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information. *IEEE transactions on medical imaging*, 26(11):1598–1612.

[Viola and Wells III, 1997] Viola, P. and Wells III, W. M. (1997). Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154.

[VoTEM, 2018] VoTEM, I. . (2018). 3-d validation of tractography with experimental mri. `https://my.vanderbilt.edu/votem/`. Accessed: 2018-01-12.

[Wang et al., 2017] Wang, Y., Shen, Y., Liu, D., Li, G., Guo, Z., Fan, Y., and Niu, Y. (2017). Evaluations of diffusion tensor image registration based on fiber tractography. *Biomedical engineering online*, 16(1):9.

[Wang et al., 2016] Wang, Y., Yu, Q., Liu, Z., Lei, T., Guo, Z., Qi, M., and Fan, Y. (2016). Evaluation on diffusion tensor image registration algorithms. *Multimedia Tools and Applications*, 75(13):8105–8122.

[Wiest-Daesslé et al., 2007] Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S. P., and Barillot, C. (2007). Non-local means variants for denoising of diffusion-weighted and diffusion tensor mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 344–351. Springer.

[Winkler et al., 2010] Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., and Glahn, D. C. (2010). Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *Neuroimage*, 53(3):1135–1146.

[WU-Minn, 2017] WU-Minn, H. (2017). 1200 subjects data release reference manual.

[Yablonskiy and Sukstanskii, 2010] Yablonskiy, D. A. and Sukstanskii, A. L. (2010). Theoretical models of the diffusion weighted mr signal. *NMR in Biomedicine*, 23(7):661–681.

[Yap et al., 2010] Yap, P., Chen, Y., An, H., Gilmore, J., Lin, W., and Shen, D. (2010). Non-parametric deformable registration of high angular resolution diffusion data using diffusion profile statistics. In *ISMRM*, volume 18, page 3968.

[Yeatman et al., 2018] Yeatman, J. D., Richie-Halford, A., Smith, J. K., Keshavan, A., and Rokem, A. (2018). A browser-based tool for visualization and analysis of diffusion mri data. *Nature communications*, 9(1):940.

[Yeo et al., 2009] Yeo, B. T., Vercauteren, T., Fillard, P., Peyrat, J.-M., Pennec, X., Golland, P., Ayache, N., and Clatz, O. (2009). Dt-refind: Diffusion tensor registration with exact finite-strain differential. *IEEE transactions on medical imaging*, 28(12):1914–1928.

[Zhang et al., 2006] Zhang, H., Yushkevich, P. A., Alexander, D. C., and Gee, J. C. (2006). Deformable registration of diffusion tensor mr images with explicit orientation optimization. *Medical image analysis*, 10(5):764–785.

[Zhang et al., 2014] Zhang, P., Niethammer, M., Shen, D., and Yap, P.-T. (2014). Large deformation diffeomorphic registration of diffusion-weighted imaging data. *Medical image analysis*, 18(8):1290–1298.

[Zucchelli et al., 2017] Zucchelli, M., Descoteaux, M., and Menegaz, G. (2017). Noddi-sh: a computational efficient noddi extension for fodf estimation in diffusion mri. *arXiv preprint arXiv:1708.08999*.