CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

www.csgb.dk

RESEARCH REPORT

2019

Henrik G. Jensen, François Lauze and Sune Darkner

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

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

Henrik G. Jensen, François Lauze and Sune Darkner

Department of Computer Science, University of Copenhagen, Denmark

Abstract

We present an information-theoretic approach to registration of DWI with explicit optimization over the orientational scale, with an additional focus on normalized mutual information as a robust information-theoretic similarity measure for DWI. The framework is an extension of the LOR-DWI density-based hierarchical scale-space model, that varies and optimizes over the integration, spatial, directional, and intensity scales. We extend the model to non-rigid deformations and show that the formulation provides intrinsic regularization through the orientational information. Our experiments illustrate that the proposed model deforms ODFs correctly and is capable of handling the classic complex challenges in DWI-registrations, such as the registration of fiber-crossings along with kissing, fanning and interleaving fibers. Our results clearly illustrate a novel promising regularizing effect, that comes from the nonlinear orientation-based cost function. We illustrate the properties of the different image scales, and show that including orientational information in our model make the model better at retrieving deformations compared to standard scalar-based registration.

Keywords: 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 geometry of DWI makes it a challenge in image registration, a key tool for comparing and analyzing medical image data. In addition, DWI acquired on different scanners or with different protocols have a non-linear relationship and defining the similarity between two DWI is inherently difficult. [1].

Our contribution is a scale-space formulation of density estimation for image similarity for DWI in a nonrigid registration setting. The model encompasses explicit

reorientation providing the computational framework for estimating non-linear similarity measures well-suited for DWI. The model allows for image similarities such as normalized mutual information (NMI) by extending the Locally Orderless Registration for Diffusion-Weighted Images (LOR-DWI) [2] to nonrigid registration. LOR-DWI includes orientational information in the image similarity, which allows explicit optimization over the reorientation of diffusion gradients. This extends from DTI to the raw high-angular resolution scans (HARDI) or the topographically inverted Orientation Density Functions (ODF).

We demonstrate the properties of the framework, such as the effect of the reorientation on the ODFs, the effect of the scale-space on optimization and regularization on simulated DWI data and on a synthetic deformation of data from a subject of the Human Connectome Project (HCP) [3]. We show that LOR-DWI formulation preserves the ODFs and produce excellent mappings for crossing, kissing and curving fibers, while providing an inherent regularizing effect.

The density formulation also allows us to optimize over the isocurves of the DWI signal $I : \mathbb{R}^3 \times \mathbb{P}^2 \to \mathbb{R}$. In principle, this framework could be extended to other non-trivial geometry of other image modalities where scale-space structures can be defined. In this work, we focus on DWI data where we use a classical scale-space structure on this non-linear geometry.

2 Related Work

This work addresses 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 measures, such as the fractional anisotropy (FA) or the mean diffusivity (MD) [4]. However, as such methods disregard most of the directional information in DWI, 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 after changing the deformation field. Some of the most popular methods can be found in [5, 6, 7, 8]. However, registration frameworks have also 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 frameworks have been shown to generally outperform scalar-based frameworks for DWI [12, 13, 14, 15].

For scalar based approaches to image similarity, registration is straightforward as any popular scalar-based measure can be used, e.g. the sum-of-squares difference (SSD) [10] or mutual information (MI) [16, 15]. Explicit reorientation strategies inherently define the similarity over both position and orientation and, 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] through an extension of the LOR framework in [17]. As argued in [2], the invariant and statistical properties of MI makes it a logical choice for DWI, where multiple factors result in

a more statistical relationship, such as variations in b-values, non-monoexponential behavior in biological tissue, and inter-scanner variability [1]. MI is often used in the standard registration of complex modalities and, as such, in scalar-based registration of DWI [18] and in the pre-processing of DWI [19]. MI and normalized MI (NMI) [20] provides a non-linear statistical measure for DWI. However, it is also likely that more functional measures could be well-suited, such as cross-correlation (CC) and normalized CC (NCC), both of which can be defined from the LOR density-formulation [21, 22]. The density-based DWI comes from the generalized way of estimating image similarity measures based on Locally Orderless Images (LOI) [23]. The first mention of LOI in the context of image registration was in [24] where a variational approach to image registration was presented. The LOR framework [25, 22] 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 $\boldsymbol{I} : \Omega \to \mathbb{R}^3$. 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 be 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 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

$$h_{\beta\alpha\sigma}(i, j|\phi, \boldsymbol{x}) = \int_{\mathbb{R}^3} P_\beta(\boldsymbol{I}_\sigma(\phi(\boldsymbol{x})) - i)P_\beta(\boldsymbol{J}_\sigma(\boldsymbol{x}) - j)W_\alpha(\boldsymbol{\tau} - \boldsymbol{x})d\boldsymbol{\tau} \qquad (3.1)$$

$$p_{\beta\alpha\sigma}(i, j|\phi, \boldsymbol{x}) \simeq \frac{h_{\beta\alpha\sigma}(i, j|\phi, \boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma}(k, l|\phi, \boldsymbol{x})dk \, dl} \qquad (3.2)$$

where $i, j \in [a_1, a_2] =: \Lambda$ are values in the image intensity range, $\boldsymbol{I}_\sigma(\boldsymbol{\phi}(\boldsymbol{x})) = (\boldsymbol{I} * K_\sigma)(\boldsymbol{\phi}(\boldsymbol{x}))$ and $J_\sigma(\boldsymbol{x}) = (\boldsymbol{J} * K_\sigma)(\boldsymbol{x})$ 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

$$\mathcal{F}_{\mathrm{lin}} = \int_{\Lambda^2} f(i,j) p(i,j) di\, dj \tag{3.3}$$

$$\mathcal{F}_{\mathrm{non\text{-}lin}} = \int_{\Lambda^2} f(p(i,j)) di\, dj \tag{3.4}$$

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 [22] 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 this 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 $S(\boldsymbol{x},\boldsymbol{v}) = S_0(\boldsymbol{x}) e^{-b\boldsymbol{I}(\boldsymbol{x},\boldsymbol{v})}$ [26] 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 the diffusion is orientation-free such that $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{v}) \approx \boldsymbol{I}(\boldsymbol{x},-\boldsymbol{v})$ which 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 (3.1) and eq. (3.2) 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 the projective structure, via a symmetric kernel $\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v})$ on $\mathbb{S}^2$. We define the smoothed signal $\boldsymbol{I}_{\sigma,\kappa}$ at scales (σ,κ) by

$$\boldsymbol{I}_{\sigma\kappa}(\boldsymbol{x},\boldsymbol{v}) = \int_{S^2} \left(\int_{\mathbb{R}^3} \boldsymbol{I}(\boldsymbol{\tau},\boldsymbol{\nu}) K_\sigma(\boldsymbol{\tau}-\boldsymbol{x}) d\boldsymbol{\tau} \right) \Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v}) d\boldsymbol{\nu}$$
$$= (\boldsymbol{I} * (K_\sigma \otimes \Gamma_\kappa))(\boldsymbol{x},\boldsymbol{v}) \tag{3.5}$$

where $K_\sigma(\boldsymbol{x})$ is a Gaussian kernel with σ standard deviation. We employ a symmetric Watson distribution [27] as $\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v})$ for directional smoothing on $\mathbb{S}^2$, given by

$$\Gamma_\kappa(\boldsymbol{\nu},\boldsymbol{v}) = C e^{\kappa(\boldsymbol{\nu}^\top \boldsymbol{v})^2} \tag{3.6}$$

$$C = M(\tfrac{1}{2}, \tfrac{1}{d}, \kappa) \tag{3.7}$$

where $M(\tfrac{1}{2}, \tfrac{d}{2}, \kappa)$ the confluent hypergeometric function also called the Kummer function ($d = 3$ in our case) [28], $\pm\boldsymbol{v}$ 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 [27] distribution or a symmetrized heat kernel could be considered.

4

3.3.1 Action on Orientation

A transformation $\phi : \Omega \to \Omega$ acts on directions via its 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

In the following, we will 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

$$h_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x}) = \int_{\mathbb{R}^3 \times S^2} P_\beta(\boldsymbol{I}_{\sigma\kappa}(\phi(\boldsymbol{x}), \psi(\boldsymbol{v})) - i)P_\beta(\boldsymbol{J}_{\sigma\kappa}(\boldsymbol{x}, \boldsymbol{v}) - j)$$
$$W_\alpha(\boldsymbol{\tau} - \boldsymbol{x})d\boldsymbol{\tau} \times d\boldsymbol{v}$$
$$p_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x}) = \frac{h_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x})}{\int_{\Lambda^2} h_{\beta\alpha\sigma\kappa}(i, j|\Phi, \boldsymbol{x})dl \, dk}. \tag{3.8}$$

Assume that W is a Gaussian kernel with standard deviation α. If we left $\alpha \to \infty$, we would obtain full spatial integration and can 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 $\psi_x = \psi$ being the projectivization of its linear part. In the present work, we assume ϕ to be a nonrigid transformation, and we use Rueckert *et al.* framework [29] 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 $\phi = \boldsymbol{Bc}$, $\Phi_{\boldsymbol{c}} = (\phi_{\boldsymbol{c}}, \psi_{\boldsymbol{c}})$. We seek the transformation $\Phi_{\boldsymbol{c}}^*$ which maximizes the regularized NMI

$$\Phi_{\boldsymbol{c}}^* = \arg\max_c \mathcal{M}(\Phi_c) = \arg\max_c \mathcal{F}(\boldsymbol{I} \circ \Phi_c, \boldsymbol{J}) + \mathcal{S}(\Phi_c) \tag{3.9}$$

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 (3.9), in particular L-BFGS from [30], and thus need to compute the gradient of the similarity with respect to control points vector $\boldsymbol{c}$. Most of the required calculations are provided in [29, 22, 2], they correspond to the red frame in Figure 1, a full formula for the gradient is long and not informative for the present work. We therefore only describe how spatio-directional smoothing contributes to the image similarity, i.e., the part of Figure 1 enclosed in the blue frame. Thanks to the LOR-DWI representation, the present extension only appears within the spatio-directional smoothing
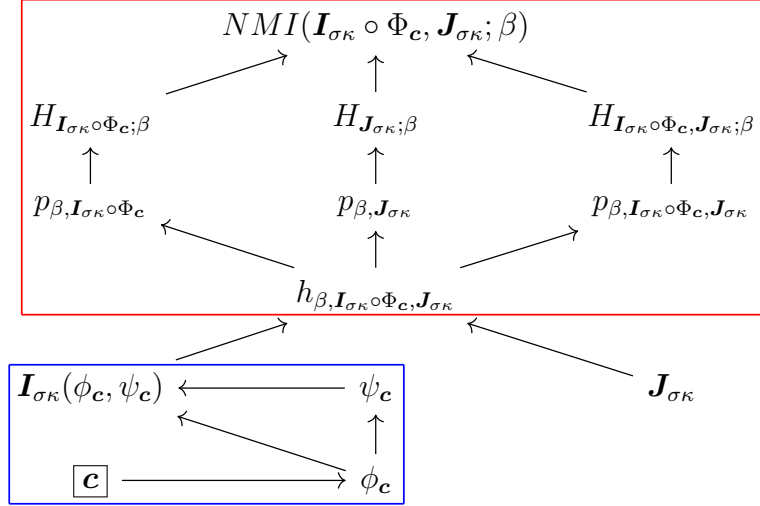
Figure 1: Dependency graph of the nonrigid DWI registration between the moving image $\boldsymbol{I}$ and the target image $\boldsymbol{J}$, with normalized mutual information (NMI) as the similarity measure. The deformation is parameterized by $\boldsymbol{c}$ so that any change in $\boldsymbol{c}$ will eventually affect the total similarity between the two images. The red frame contains MI registration elements which are classical, while the blue frame contains the part which extends the LOR framework to the LOR-DWI.

of $\boldsymbol{I}$. One complication originates from choices in our implementation, where the Kummer function $M(\frac{1}{2}, \frac{3}{2}, \kappa)$ as the normalization constant in the Watson kernel is replaced by an estimate of 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 constant. We rewrite the discrete spatio-directional smoothing as

$$\boldsymbol{I}_{\sigma\kappa}(\phi_{\boldsymbol{c}}(\boldsymbol{x}), \psi_{\boldsymbol{c}}(\boldsymbol{v})) = \sum_{n=1}^{N} \int_{\mathbb{R}^3} \boldsymbol{I}(\tau, \nu_n) K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau) \bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v})) \, d\tau \qquad (3.10)$$

where we have set

$$\bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v})) = \frac{e^{\kappa(\nu_n^\top \psi_{\boldsymbol{c}}(\boldsymbol{v}))^2}}{\sum_{i=1}^{N} e^{\kappa(\nu_i^\top \psi_{\boldsymbol{c}}(\boldsymbol{v}))^2}}. \qquad (3.11)$$

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

$$\begin{aligned}
J_{\boldsymbol{c}} & \boldsymbol{I}_{\sigma\kappa}(\phi_{\boldsymbol{c}}(\boldsymbol{x}), \psi_{\boldsymbol{c}}(\boldsymbol{v})) \\
&= \sum_{n=1}^{N} \Big(\int_{\mathbb{R}^3} \boldsymbol{I}(\tau, \nu_n) J_{\boldsymbol{c}} K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau) \, d\tau \Big) \bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v}) \\
&\quad + \sum_{n=1}^{N} \Big(\int_{\mathbb{R}^3} \boldsymbol{I}(\tau, \nu_n) K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau) \, d\tau \Big) J_{\boldsymbol{c}} \bar{\Gamma}_\kappa(\nu_n, \psi_{\boldsymbol{c}}(\boldsymbol{v})).
\end{aligned} \qquad (3.12)$$

From eq. (3.12), we have

$$J_{\boldsymbol{c}} K_\sigma(\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau) = -\frac{K_\sigma(\boldsymbol{\phi}_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)}{\sigma^2} (\phi_{\boldsymbol{c}}(\boldsymbol{x}) - \tau)^\top \boldsymbol{B} \qquad (3.13)$$

because K_σ is a Gaussian kernel with variance σ^2 and $\phi_c = Bc$, its Jacobian $J_c\phi_c$ is simply B. The part involving directional kernel is somewhat more complex, as it involves the Jacobian $J_x\phi_c$ and the normalizing factor in Equation (3.11). Set $f(x) = e^{\kappa x^2}$ and define $f_i = f(\nu_i \top \psi_c(v))$. Since $f'(x) = 2\kappa x f(x)$, we have

$$J_c f(\nu^\top \psi_c v) = 2\kappa f(\nu^\top \psi_c(v)) \psi_c(v)^\top \nu \nu^\top J_c \psi_c(v). \qquad (3.14)$$

By a straightforward and careful calculation, we obtain

$$J_c \bar{\Gamma}_\kappa(\nu_n, \psi_c(v)) \qquad (3.15)$$

$$= -2 \frac{\kappa \bar{\Gamma}_\kappa(\nu_n, \psi_c(v))}{\sum_{i=1}^N f_i} \psi_c(v)^\top \left(\sum_{i \neq n} f_i \nu_i \nu_i^\top \right) J_c \psi_c(v). \qquad (3.16)$$

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

$$J_c \langle J_x \phi_c v, J_x \phi_c v \rangle = (J_x \phi_c v)^\top (J_x B \bullet v). \qquad (3.17)$$

Denoting by V the vector $J_x\phi_c v$, so that $\psi_c(v) = \frac{V}{|V|}$, the Jacobian $J_c\psi_c v$ is

$$J_c \psi_c v = \left(I_3 - \frac{VV^\top}{|V|^2} \right) \frac{J_x B \bullet v}{|V|} \qquad (3.18)$$

where I_3 is the identity of $\mathbb{R}^3$. Note also that $V = J_x\phi_c v = (J_x B \cdot c) v$ and that $I_3 - \frac{VV^\top}{|V|^2}$ is the orthogonal projection $\pi_{V^\perp}$ onto $V^\perp$, the plane orthogonal to V. Putting things together, the Jacobian with respect to the control point parameter c of the spatio-directional smoothing is given by

$$J_c I_{\sigma\kappa}(\phi_c(x), \psi_c(v))$$

$$= -\sum_{n=1}^N \int_{\mathbb{R}^3} I(\tau, \nu_n) K_\sigma(\phi_c(x) - \tau) \bar{\Gamma}_\kappa(\nu_n, \psi_c(v)) \qquad (3.19)$$

$$\times \left\{ \frac{(\phi_c(x) - \tau)^\top B}{\sigma^2} + \frac{2\kappa V^\top \left(\sum_{i \neq n} \nu_i \nu_i^\top \right)}{|V| \sum_{i=1}^N f_i} \pi_{V^\perp} \left(\frac{J_x B \bullet v}{|V|} \right) \right\} d\tau.$$

3.6 Regularization

So far, we have not specified the form of the regularizer $\mathcal{S}(\Phi_c)$ in Equation (3.9). 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 required some regularization to keep the deformation stable, due to the high resolution of the deformation field. We used a simple regularization that penalizes a non-uniform grid by the squared difference between a point 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 $\mathcal{S}(\Phi_c)$ is the sum $\sum_{r=1}^{R} \mathcal{S}^r(\Phi_{c^r})$ at each resolution, where $c^r = (c_1^r, \ldots c_{p_r}^r)^T$ is 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

$$\mathcal{S}^r(\Phi_c) = -\frac{\lambda_r}{2} \sum_{i=1}^{p_r} \|c_i^r - \frac{1}{|N^r(i)|} \sum_{j \in N^r(i)} c_j^r\|^2. \tag{3.20}$$

λ_r is a strictly positive parameter controlling the degree of smoothness. The negative sign in Equation (3.20) comes from the fact that we *maximize* the objective function. In order to compute the gradient of $\mathcal{S}^r(\Phi_{c^r})$, we define series of linear mappings $T_i^r : c \mapsto |N^r(i)|c_i^r - \sum_{j \in N^r(i)} c_j^r$. The regularizer in Equation (3.20) can be rewritten as

$$\mathcal{S}^r(\Phi_c) = -\frac{\lambda_r}{2} \sum_{i=1}^{p_r} \frac{1}{|N^r(i)|^2} \|T_i^r c^r\|^2 \tag{3.21}$$

and by classical manipulation, we obtain that

$$\nabla_c \mathcal{S}^r(\Phi_c) = -\lambda_r \sum_{i=1}^{p_r} \frac{1}{|N^r(i)|^2} T_i^{r*} T_i^r c^r. \tag{3.22}$$

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

$$\nabla_c \mathcal{S}(\Phi_c) = -\sum_{r=1}^{R} \lambda_r \sum_{i=1}^{p_r} \frac{1}{|N^r(i)|^2} T_i^{r*} T_i^r c^r. \tag{3.23}$$

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

4   Experiments

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

4.1   Simulated Examples

Artificially generated samples enables visual inspection and validating of the framework in a highly controlled environment. While artificial experiments can be found in most DWI registration papers, the artificial data is rarely available. To our knowledge, there are no open sources of simulated DWI data for comparing registration frameworks, although the `DIPY` project may be a good source for generating simulations [31]. To this end, we created our own simulated DWI data, which is freely available. The generation of the artificial data is described in the following.[1]

---

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

4.1.1 Simulating DWI data

The simulated DWI (HARDI) samples were created by deforming a unit sphere of equally distributed directions [32] 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 is a crossing fiber ODF. DWI samples are antipodal symmetric and every ODF from the simplest to the most complex can be constructed through 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 [33]. These
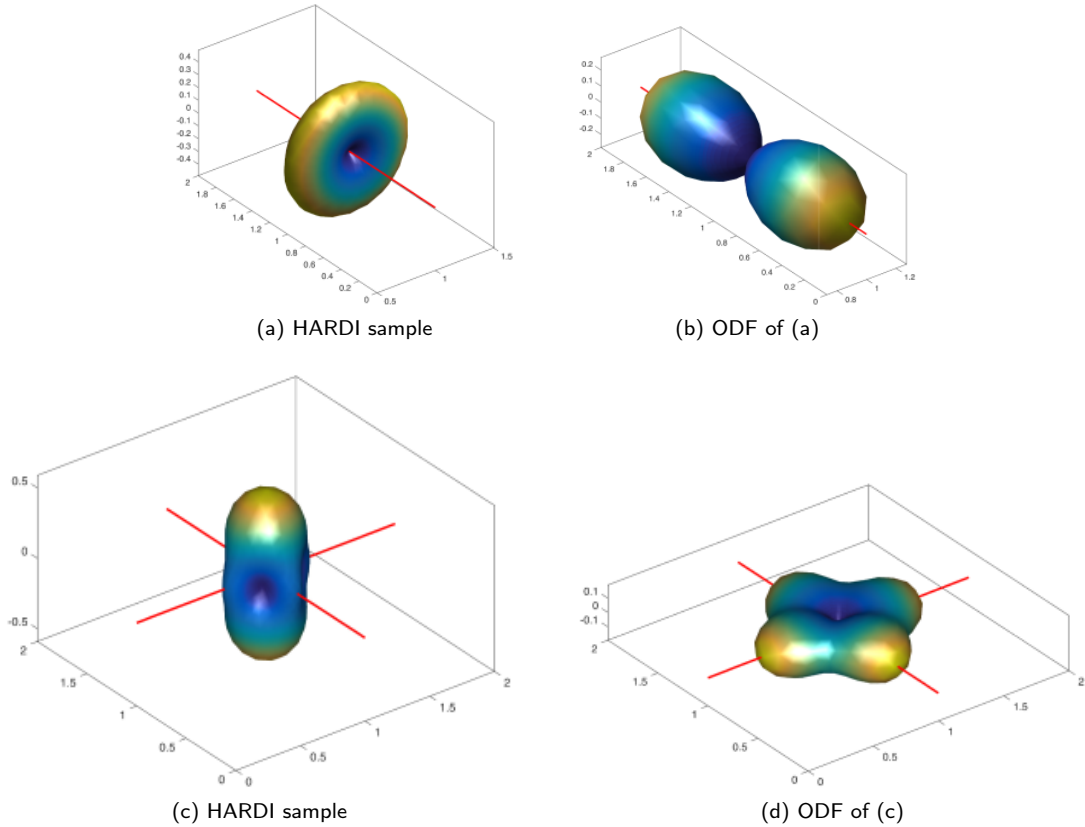

(a) HARDI sample   (b) ODF of (a)

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

Figure 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.

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 throughout this section to create blueprints of fiber tract constellations with the purpose of performing LOR-DWI 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 in 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 b_0 normalized data.

9
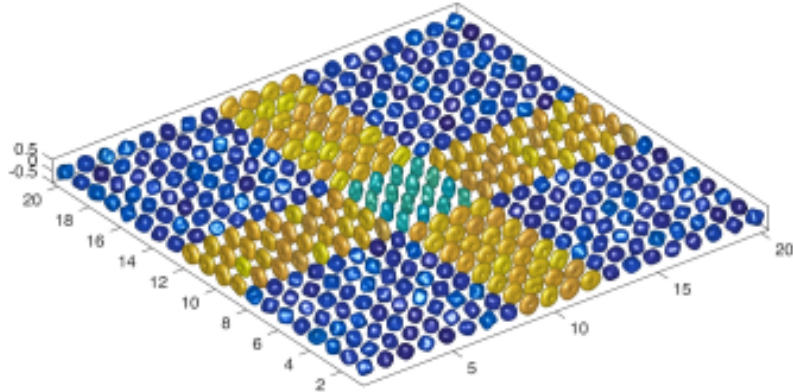
(a) HARDI fiber tract crossing
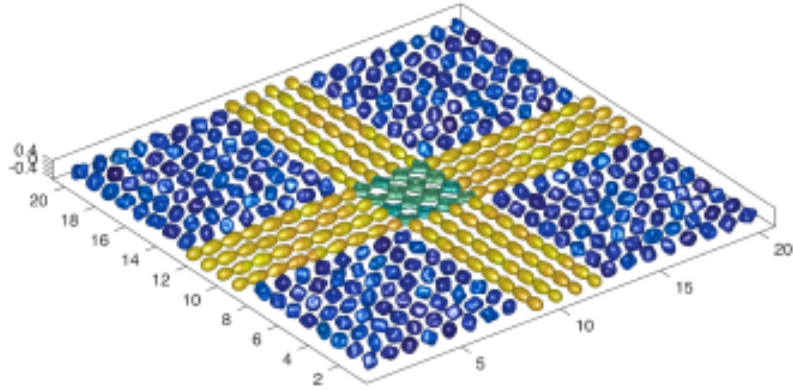
(b) ODFs of (a) showing the fibers

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

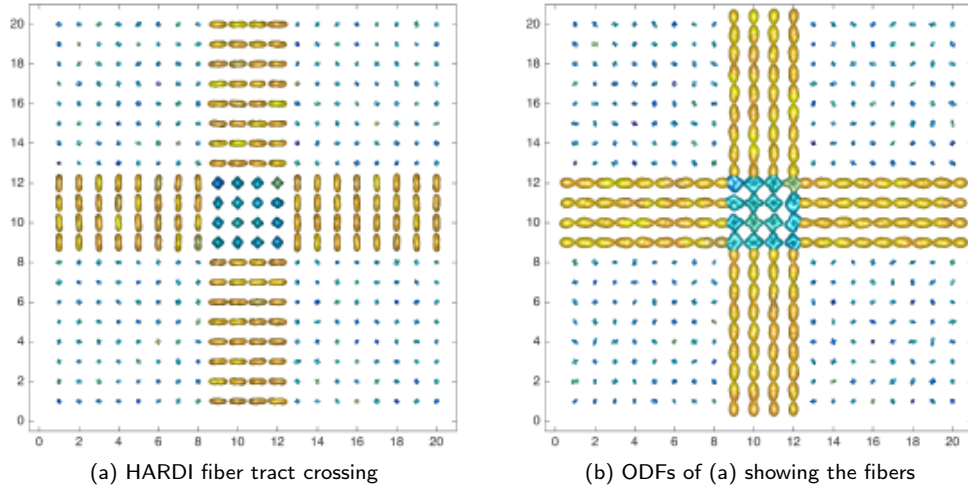(a) HARDI fiber tract crossing

(b) ODFs of (a) showing the fibers

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

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. In the B-spline deformation, 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 $\epsilon = 10^{-6}$, 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 of $\sigma = 0.6$ voxels.

Histogram size. Due to the relatively small data sample, a low number of bins is used for the histogram (i.e. intensity resolution) 20×20. This allows for some larger, stable but less refined deformations [22].

HARDI registration, ODF visualization. We register the raw HARDI models but we visualize the ODFs of the warped data, based on the Funk-Radon transformed (FRT). We do this 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 $\lambda = 10^{-4}$. The regularization could have be omitted in the simulated experiments and substituted by multiple levels of resolution of the deformations (large to small), as the simulated data is very structured. As an aside, every example below could be highly improved by a experiment-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 are based on artificially generated distributions of HARDI shells and corresponding ODFs, each representing different fiber constellations. The experiment 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 illustrate the regularizing effects of using the full diffusion profile for registration. In the first experiment, the position of the end of the fibers is constrained by adding intersecting fibers at the end and beginning. Figure 5 shows the simulated moving (straight) and target (wavy) images. The results of the first experiment is 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 fibers are stretched in correspondence with the curvature, and the reconstructed ODFs from the HARDI-based registration are rotated correctly, although smoothed by the interpolation.
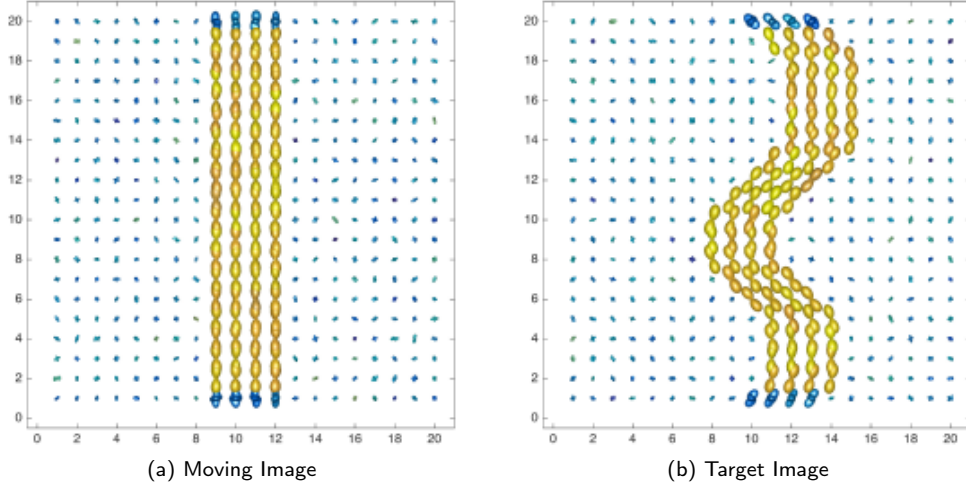
(a) Moving Image

(b) Target Image

Figure 5: Experiment 1: Simulated fiber tract images with intersecting fibers at the boundaries.

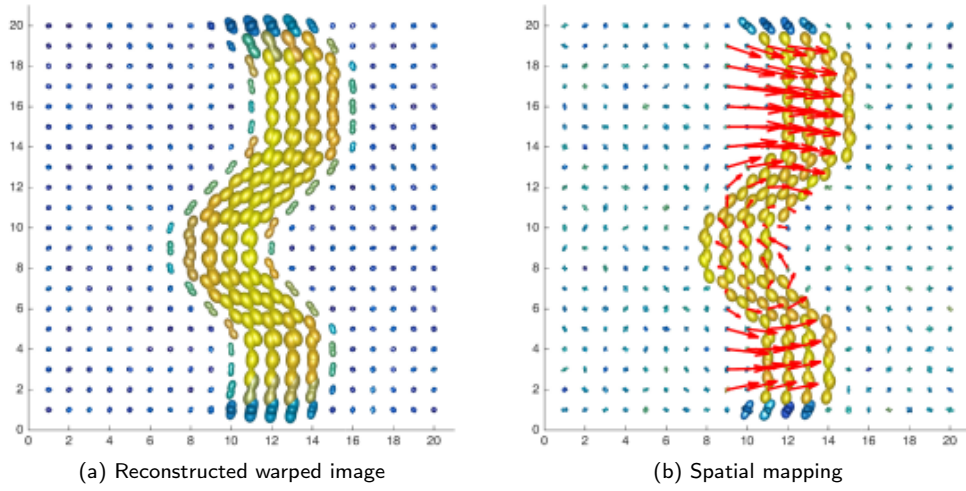(a) Reconstructed warped image

(b) Spatial mapping

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

The second experiment is performed without 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 in such a case should preserve the length of the straight fiber after deformation. The fact that this happens with no strong outside regularization, illustrate an inherent regularization effect of the cost function.

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 with no signal on the boundaries. The mean diffusivity carries no directional information. The result can be seen in Figure 8. This pure scalar-based registration is driven by the edges of the simulated tracts only. The reconstruction appears to be correct, but the final spatial mapping indicates a lack of regularization as the fibers are stretched unevenly.
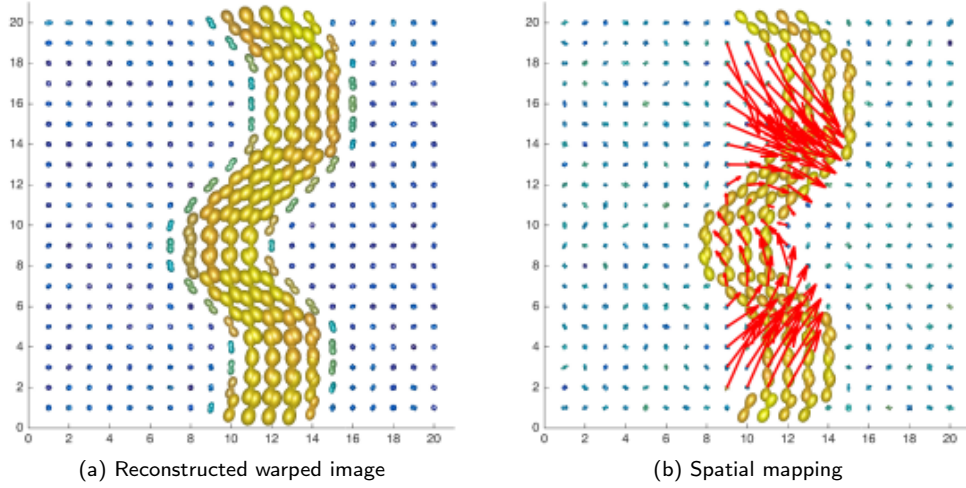
(a) Reconstructed warped image   (b) Spatial mapping

Figure 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.
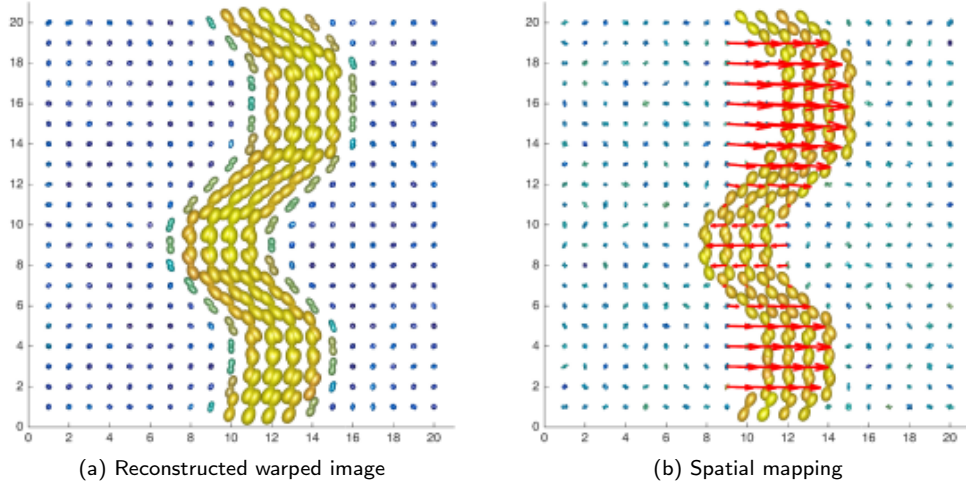

(a) Reconstructed warped image   (b) Spatial mapping

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

4.3 Experiment 2: Crossing Fiber Tracts

The second set of experiments are designed to investigate 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. Note that the reconstruction is subject to the effect of the smoothing in the interpolation.

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. As the figure illustrates 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 subsequently reconstructed.

4.4 Experiment 3: Fanning Fiber Tracts

The third set of experiments examines the most complex configurations namely the registration of kissing and fanning fibers. Note that although the artificial cases constructed for these experiments have a 1-1 correspondence.

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 contains a crossing with a few fibers fanning in and out along the vertical center line. The target image contains two curved tracts fanning in and out, and merging at the central crossing. The results of the registration experiment 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 registration is able to move, shrink and turn the center-crossing to fit to fit the target image.

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.
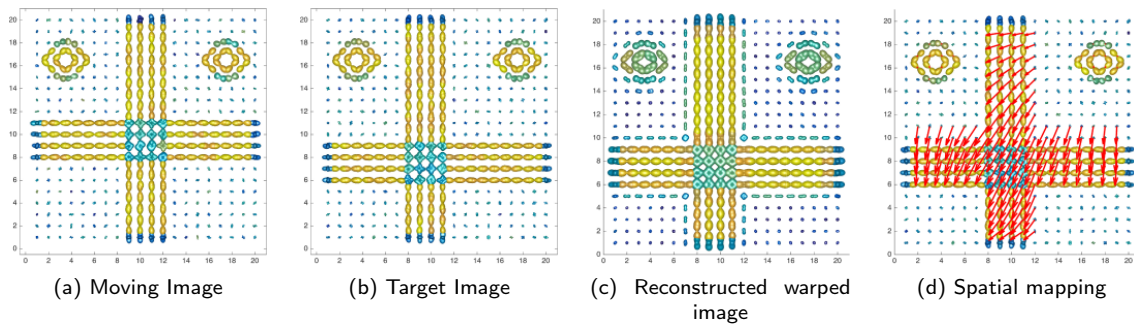

(a) Moving Image   (b) Target Image   (c) Reconstructed warped image   (d) Spatial mapping

Figure 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).
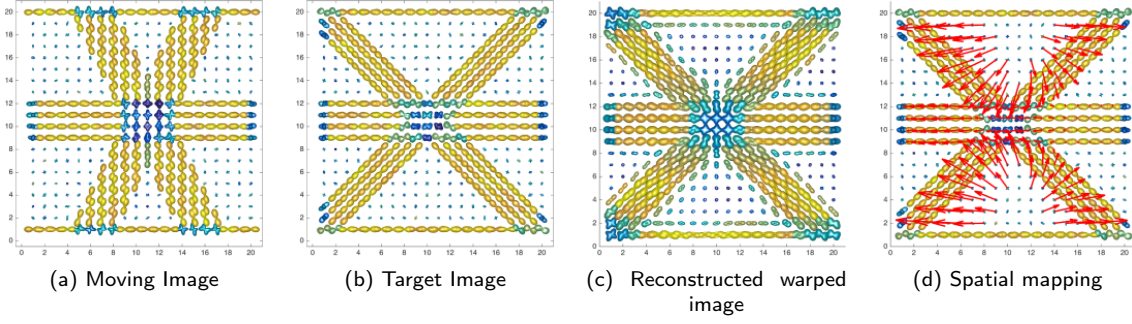
14

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

Figure 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).
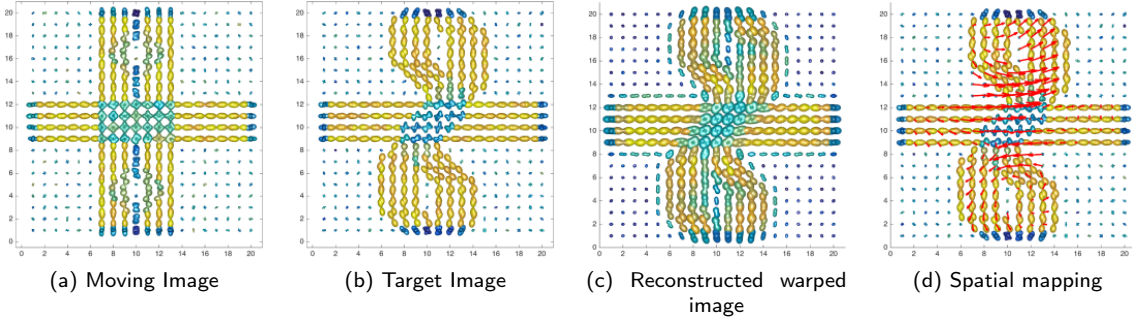

(a) Moving Image   (b) Target Image   (c) Reconstructed warped image   (d) Spatial mapping

Figure 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).
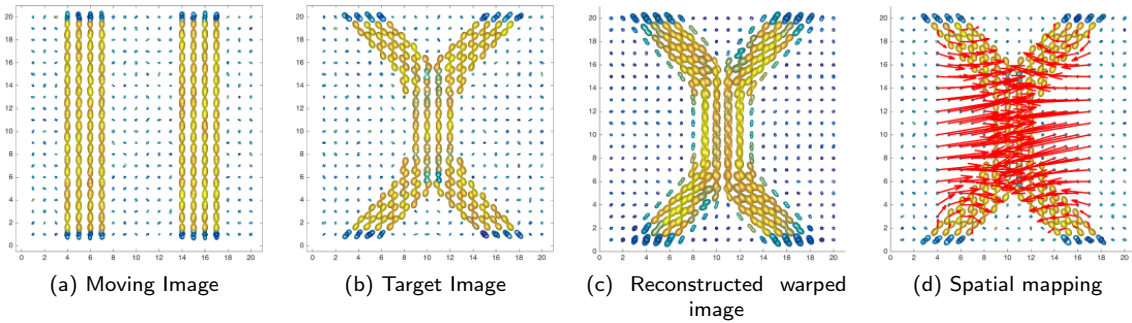

(a) Moving Image   (b) Target Image   (c) Reconstructed warped image   (d) Spatial mapping

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

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

15

brain data, a ground truth is obtained where the objective is to register the warped image back to the original image. These experiments illustrate the effect of the parameters of the model in a realistic scenario in a guaranteed diffeomorphic scenario.
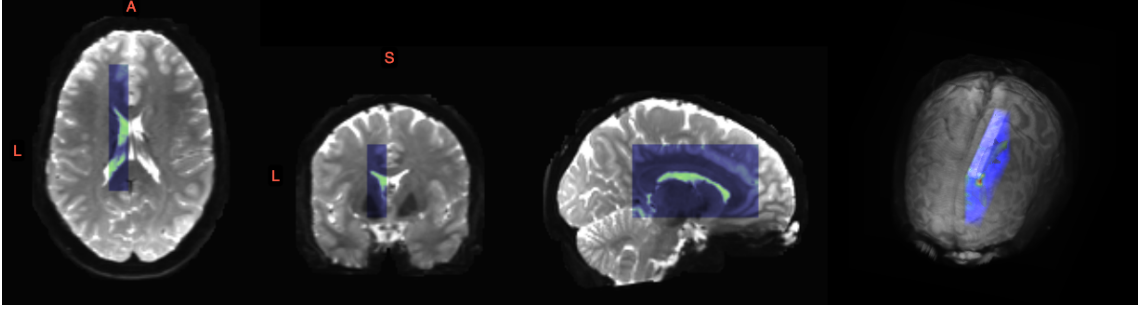
4.5.1 DWI example data

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

The DWI data used in 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 to improve 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 1.25 mm isotropic voxels with 90 directions. Only the central sagittal slice along with corresponding ODFs is visualised (Figure 14a), while the deformation is applied to the whole ROI. The deformation field for the central slice is shown in Figure 14b.

4.5.2 The effects of the parameters

The experimental setup provides a ground truth, which is simply the identity transformation. As the similarity measure, NMI, does not reflect a unique point-wise correspondence, we introduce three measures for evaluating the resulting registration: (1) mean squared error (MSE) between coordinates, (2) curl of the final deformation field, and (3) divergence of the final deformation field [34]. *Curl* quantifies the amount of orientational change in the deformation field, also referred to as rotation or vorticity. We use the L_2-norm of the curl vector, i.e. the velocity of rotation. 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. *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. In combination, these three measures provide information about the magnitude of the voxel-wise distance and the state of the final deformation field. Curl and divergence has previously been used as regularization in a nonrigid registration framework [35].

In the following, the experiments are performed while varying the intensity scale, the orientation scale, and the spatial scale, as it is the intensity and orientational
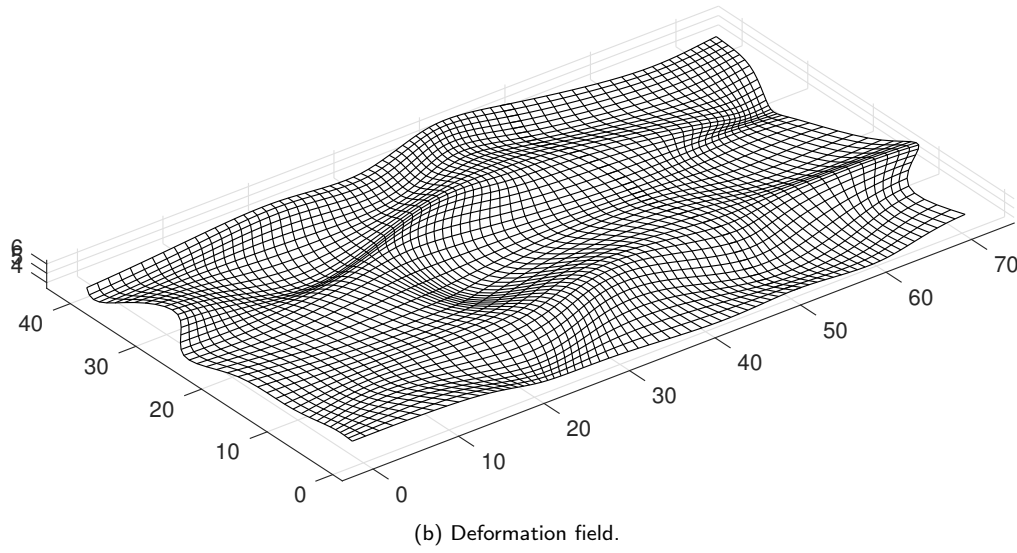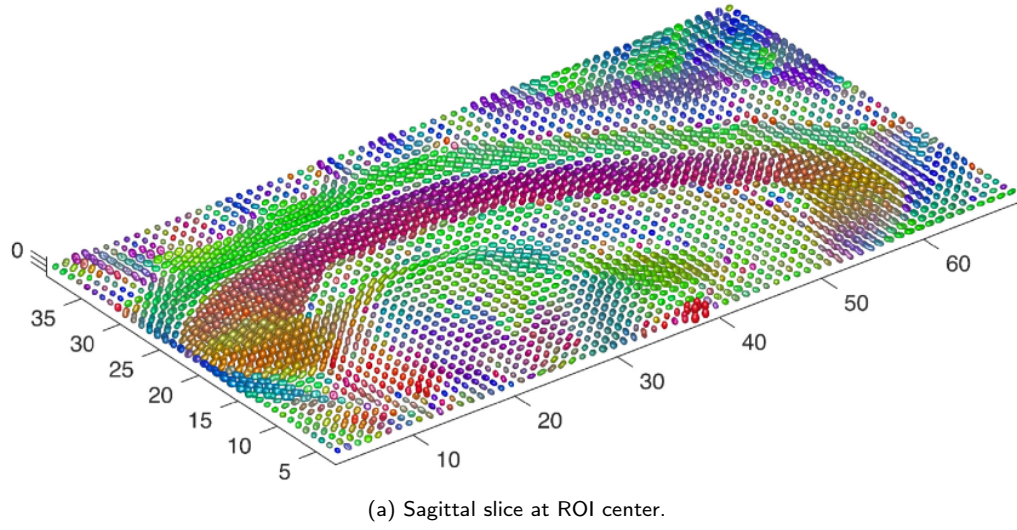
(a) Sagittal slice at ROI center.

(b) Deformation field.

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

scale that sets this registration framework apart form other frameworks. The following is investigated:

1. **Isointensity 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 can be strongly affected by partial volume effects. We iteratively decrease the control point-spacing in the FFD model to get an increasingly localized registration, and increase the intensity range accordingly. This is realized with an initial histogram with relatively few bins and a fixed degree of smoothing and then successively increase the number of bins as the deformation resolution is increased of iterations.

2. **Explicit reorientation.** The directional information is, according to our experiments, expected to provide a more stable and regularized transformation. To investigate this hypothesis, different levels of directional smoothing are examined from $\kappa = 30$ a highly peaked ODF to a scalar-based mean diffusivity registration i.e. setting the concentration parameter κ to 0. The experiments on simulated data indicated that the directional information results in direct regularized solutions through the similarity, and these experiments seek to uncover if these observations hold for real data.

Performing multi-resolution registration (a continuation method) is a key element in most high resolution registration frameworks, e.g. in the FFD model [29], `ANTs` [5], `Elastix` [36], `FSL` [7], etc. It provides stability while allowing for large and small deformations while reducing the computational complexity. While experiments regarding the spatial part will be performed, its impact on registration quality is fairly well-described in contrast to iso-intensity curves and explicit reorientation.

4.5.3 Parametric Setup

Hierarchical mesh resolution. Similar to the simulated experiments, the spacing between the control points is successively decreased. To this end 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 according to the dimensions of the image. The low initial resolution of the deformation permits the generation of a large random and valid synthetic warp at around half of the control point spacing ($\sim 0.4\delta$) [37]. For the optimization we use a quasi-Newton L-BFGS optimizer which runs for 50 iterations for all resolutions, unless the optimality condition of $\epsilon = 10^{-6}$ is fulfilled. Note that we refer to the result after each optimizer termination as *a step*, thus 4 steps in total.

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 is defined by the product of Jacobians for each step $\text{Curl}(\Phi_{\delta=3}) = \boldsymbol{J}(\Phi_{\delta=10})\boldsymbol{J}(\Phi_{\delta=5})\boldsymbol{J}(\Phi_{\delta=3.5})\boldsymbol{J}(\Phi_{\delta=3})$ for all voxels.

Change in resolution. Changing spatial resolution is achieved by convolution.

Other fixed parameters. The regularization is fixed to $\lambda = 10^{-4}$, and we interpolate and optimize over 30 interpolated orientations of the 90 in the HCP data.

All the following experiments will show results of the 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

This experiment illustrates the effect of changing the size of the smoothed joint histogram. Three experiments 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.
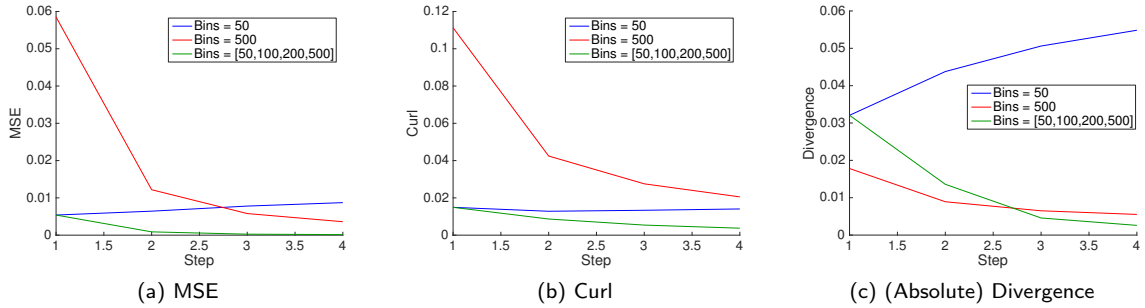

(a) MSE   (b) Curl   (c) (Absolute) Divergence

Figure 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 1.42×10^{-4}, (b): green at 3.74×10^{-3}, and (c): green at 2.60×10^{-3}.

It is evident how a small histogram with wide isocurves results in an initially faster convergence Figure 15 (blue line). However, wide iso-curves result in flat regions with small gradients and little structure which may cause the result to become sub-optimal or degenerate with increasing degrees of freedom in the transformation. In contrast, starting with a high-resolution of the histogram with thin iso-curves has the opposite effect in the initial steps (red line). However, 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 this experiment, we investigate if directional information increases the stability and improves the registration. We employ the size progression of the histogram from the previous experiment (i.e. [50,100,200,500]), and vary the concentration parameter κ from mean diffusivity at $\kappa = 0$ to sharp angular features at $\kappa = 30$. The results are shown in Figure 16. The figure shows how the directional information results in better registration, with significantly less curl than the scalar registration at $\kappa = 0$ and we observe that the best value $\kappa = 30$ is stable in high directional resolution data such as the HCP. We suggest setting $\kappa = 15$ as this is should suffice for both high and low resolution data, such as DTI.
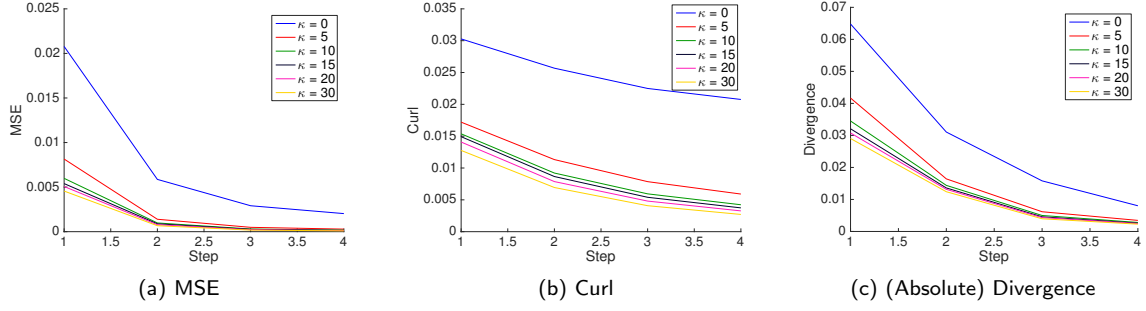
(a) MSE (b) Curl (c) (Absolute) Divergence

Figure 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): 9.17×10^{-5}, (b): 2.71×10^{-3}, and (c): 2.29×10^{-3}. However, the overall difference between $\kappa = 10$ and $\kappa = 30$ is not more than around 10^{-5}.

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.
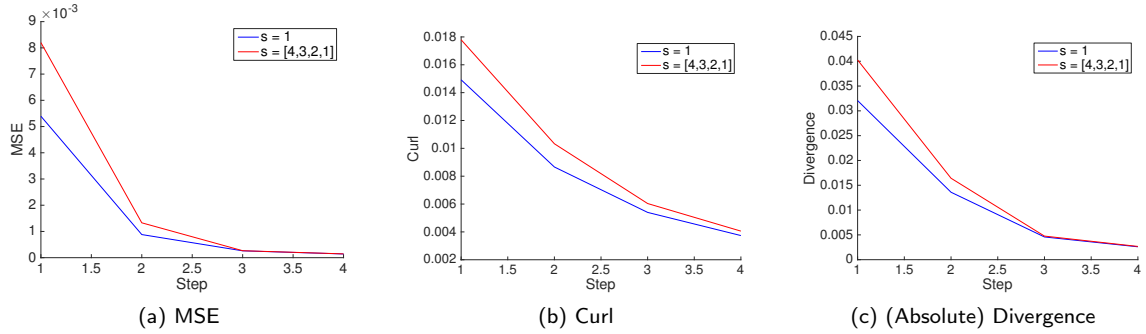

(a) MSE (b) Curl (c) (Absolute) Divergence

Figure 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 1.42×10^{-4} vs red 1.58×10^{-4}, (b): blue 3.74×10^{-3} vs red 4.21×10^{-3}, and (c): blue 2.60×10^{-3} vs red 2.70×10^{-3}.

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.

4.5.7 A Qualitative Example of the Results

Finally, we visualize the registration to perform a qualitative assessment of the warp, shape and orientation of the individual ODFs. All registrations are based on the raw, noise-corrected HARDI data, while we show the tomographic inversion

(FRT) indicating the direction of the diffusion and likely fiber tract orientations. Unlike the simulated experiments, we have fitted a B-spline to the image prior to deforming and visualizing the result, in contrast to the smoothing spline used in the artificial examples. The results where obtained using an increasing control point resolution with $\kappa = 15$, bins $= [50, 100, 200, 500]$, and spatial resolution $= [4, 3, 2, 1]$ and Figure 18 shows the result for the central ROI slice, along with a zoomed in version in Figure 19.

(a) Ground truth   (b) Warped image   (c) Registered image reconstruction
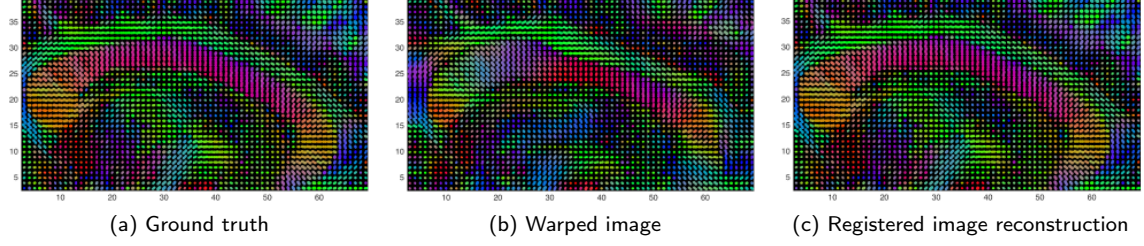
Figure 18: 2D visualization of Figure 14a, and the reconstructed warped image after applying the deformation field to the original image. We will be registering (b) back to (a).
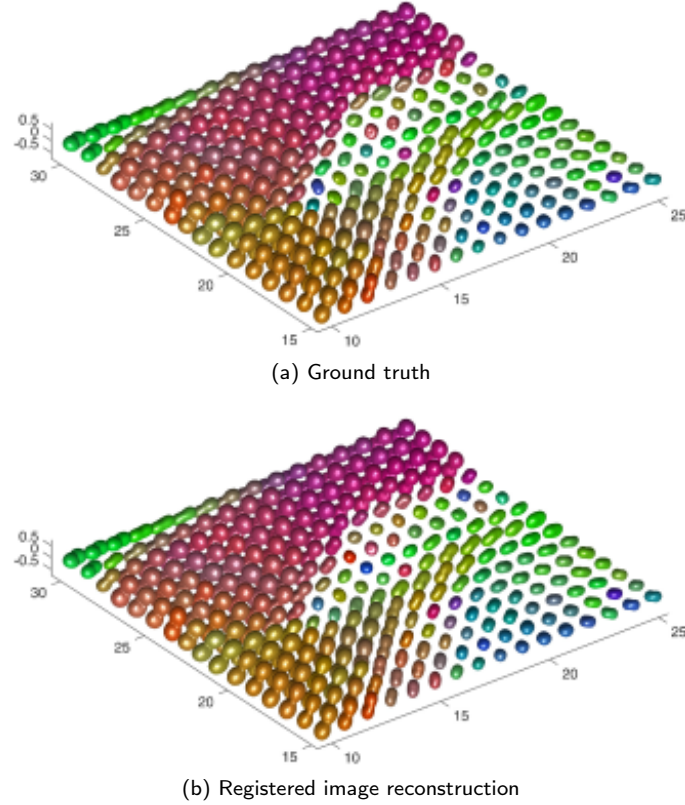

(a) Ground truth

(b) Registered image reconstruction

Figure 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.

5 Conclusion

We have presented a scale-space formulation of density estimation that extends LOR to spatio-directional data, explicitly for the registration of DWI data with explicit reorientation of the full diffusion profile. We have provided empirical evidence showing that the underlying structure of the data is preserved during registration, while providing excellent registration results through a number of classical artificial examples known to be difficult for registration. In addition, we have shown that the formulation of the similarity itself provides regularization through the additional information given by the orientational dimension, illustrated clearly in the artificial examples. We have investigated the different scales provided by the framework and shown how the different parameters influences the registration results. In conclusion, the LOR-DWI provides a smooth and well-matched deformation of DWI data, and the registration results are improved by the induced regularization obtained by integrating the orientational information in the objective function. Finally, we showed that the deformation have very little impact on the shape of the deformed ODF.

Acknowledgments

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. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[17] S. Darkner and J. Sporring, "Locally orderless registration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1437–1450, 2013.

[18] 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.

[19] 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.

[20] C. Studholme, D. L. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[21] J. Sporring and S. Darkner, "Jacobians for lebesgue registration for a range of similarity measures," *Department of Computer Science, University of Copenhagen, Tech. Rep*, vol. 4, 2011.

[22] S. Darkner and J. Sporring, "Locally Orderless Registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1437–1450, 2013.

[23] 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.

[24] 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.

[25] S. Darkner and J. Sporring, "Generalized partial volume: An inferior density estimator to parzen windows for normalized mutual information," in *IPMI*, ser. LNCS, vol. 6801. Springer, 2011, pp. 436–447.

[26] 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.

[27] 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.

[28] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,*. Dover, 1974.

[29] 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.

[30] M. Schmidt, "minfunc: unconstrained differentiable multivariate optimization in matlab," *Software available at https://www.cs.ubc.ca/ schmidtm/Software/minFunc.html*, 2005.

[31] 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.

[32] 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.

[33] 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.

[34] H. Helmholtz, "Uber integrale der hydrodynamischen gleichungen, welcher der wirbelbewegungen entsprechen" (on integrals of the hydrodynamic equations which correspond to vortex motions)," *Journal für die reine und angewandte Mathematik*, vol. 55, no. 8, pp. 25–55, 1858.

[35] 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.

[36] 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.

[37] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers, "Diffeo-morphic registration using b-splines," in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2006, pp. 702–709.