Diffeomorphic Statistical Deformation Models

June 25, 2007

Abstract

In this paper we present a new method for constructing diffeomorphic statistical deformation models in arbitrary dimensional images with a nonlinear generative model and a linear parameter space.

Our deformation model is a modified version of the diffeomorphic model by Cootes et al. The modifications ensure that no boundary restriction has to be enforced on the parameter space to prevent folds or tears in the deformation field.

For straightforward statistical analysis, principal component analysis and sparse methods, we assume that the parameters for a class of deformations lie on a linear manifold and that the distance between two deformations are given by the metric introduced by the L_2-norm in the parameter space. The chosen L_2-norm is shown to have a clear and intuitive interpretation on the usual nonlinear manifold.

Our model is validated on a set of MR images of corpus callosum with ground truth in form of manual expert annotations.

We anticipate applications in unconstrained diffeomorphic synthesis of images, e.g. for tracking, segmentation, registration or classification purposes.

1 Introduction

Registration is the problem of establishing correspondence between points in different images. It has been used for building models of variation in groups of images for several years. Cootes et al. proposed the very successful active appearance models in 1998 [3], which, once trained, can establish correspondence between points in the model and the images using a piecewise affine mapping. Rueckert et al. presented a statistical deformation model based on registrations of an atlas to the images of the group [9]. Joshi et al. demonstrate how to construct an unbiased atlas from a population [5], and Cootes et al. presented a guaranteed diffeomorphic shape model [2] by using smooth kernels for interpolating a warp field and putting restrictions on the variation of the parameters. Vester-Christensen et al. have presented an accelerated version of this algorithm [10], which is based on the inverse compositional method by Baker et al., which we have also made extensive use of in the presented work [1].

2 Methods

We define image registration as the identification of correspondence between positions in images. In the current work we address problems where the correspondences can be represented by a diffeomorphic function $f \in \mathcal{H}$, where $\mathcal{H}$ denotes the infinite dimensional group of diffeomorphisms on $\mathbb{R}^N$. The mapping from one image to the other is differentiable and the inverse exists and is also differentiable. Popular speaking this limits the problem of registration to the problem of finding smooth warps without folds or tears.

In the statistical analysis of the warp functions we are interested in estimating an unbiased atlas of the structures we are registering. We identify such an atlas as the groupwise maximizer of similarity between the atlas and the deformed images, while minimizing the deformation field.

$$[\phi_i,\ \hat{R}] = \min_{\phi_i, \hat{R}} \sum_i \mathcal{S}[\hat{R}, T \circ \phi] + \alpha \mathcal{D}(\phi_i)^2 \ . \qquad (1)$$

where $\mathcal{S}$ denotes the similarity measure and $\mathcal{D}(\phi)$ denotes the regularization term, introduced to regularize the warp ϕ further than just restricting it to the space of the parameters, and α is the regularization parameter.

2.1 Parameterized diffeomorphisms

Fletcher et al. have investigated geodesic curves on the nonlinear manifolds of the parameters of the M-reps parameterization [4]. Most of the current statistical analysis, however, is based on the assumption that the data is located on a linear manifold with the Euclidian metric, e.g. principal component analysis (PCA) and independent component analysis (ICA), which have nice properties as analytical tools. This is our motivation for introducing a function $\mathcal{G}$ which identifies $\mathbb{R}^M$ with a (hopefully interesting) subset of diffeomorphisms.

Let $\mathcal{H}(\mathbb{R}^N)$ denote the set of diffeomorphisms ($f : \mathbb{R}^N \to \mathbb{R}^N$). Now let $\mathcal{G}$ be a bijective mapping:

$$\mathcal{G} : \mathbb{R}^M \to \mathcal{H}_t \ . \tag{2}$$

where $\mathcal{H}_t = \mathcal{G}(\mathbb{R}^M) \subset \mathcal{H}$. We let $\mathcal{H}_t$ inherit the Euclidian metric from the parameter space $\mathbb{R}^M$

$$\begin{aligned} d(\mathcal{G}(t_1), \mathcal{G}(t_2)) \equiv d(t_1, t_2) = \|t_1 - t_2\|_2 \ \ , \\ t_1, t_2 \in \mathbb{R}^M \text{ and } \mathcal{G}(t_1), \mathcal{G}(t_2) \in \mathcal{H}_t \ , \end{aligned} \tag{3}$$

from which we conclude that $\mathcal{G}$ is a homeomorphism, and that the spaces $\mathcal{H}_t = \mathcal{G}(\mathbb{R}^M)$ and $\mathbb{R}^M$ are topologically equivalent. To conclude it can be observed that the defined metric on the space of parameterized warps is the L_2 norm on $\mathbb{R}^M$ as intended.

2.1.1 Composition of warps

The composition of more diffeomorphisms is diffeomorphic, which is a very important property of diffeomorphisms in the present context.

$$\begin{aligned} f_i \in \mathcal{H} \ , \ i \in \{1, 2, \ldots, n\} \\ \phi = f_n \circ f_{n-1} \circ \ldots \circ f_1 \ \Rightarrow \ \phi \in \mathcal{H} \end{aligned} \tag{4}$$

This allows for the construction of diffeomorphisms of higher complexity by the composition of several simpler warps. We shall assume we are dealing with parameterized warp functions, and our statistical analysis of warps can be reduced to the analysis of the warp parameters, in line with (3). For all images in our set the warp parameters shall warp from our *reference*, R, into the current *target*, I. In order to be able to compare parameters from different warp compositions it is evident that all our parameters exist in the same space. This is achieved by ensuring that all warps f_i in a composition warp from the reference coordinate system[2].

2.1.2 Grid based diffeomorphisms

Several grid based representations of diffeomorphisms have been presented and they are commonly used at different levels of detail and composed succeedingly [2, 7, 9]. A general trait of the grid methods is that they manipulate the parameters of the functions describing the diffeomorphism, and that the functions have a local support in the image, either as points defined in the image or as basis functions with support around a control point. Often this parameterization of the grid is linear in the parameters and this obviously imposes some restrictions on the parameters to produce diffeomorphic warps. Cootes et al. specify a cut-off at displacements larger than $\frac{1}{\pi}$ of the cosine based kernel [2] and Lee et al. find a threshold bound on the B-spline parameters to secure that the B-spline based warp function is diffeomorphic [6].

2.1.3 A proposed $\mathcal{G}$

Let $\mathcal{F}$ be the function mapping from a real parameter space $\mathbb{R}^M$ into the space of functions from $\mathbb{R}^N$ to $\mathbb{R}^N$, e.g. in case of the B-spline warps, $\mathcal{F}$ maps from the parameter space into the space of N-dimensional B-spline functions $f : \mathbb{R}^N$ to $\mathbb{R}^N$, the image of $\mathcal{F}$, $\mathcal{K}$ can be shown to contain functions that are not diffeomorphic.

As discussed in the previous section there can for some parameterized warps be specified a threshold such that $\mathcal{P} =]-\tau_1, \tau_1[\ \times \ \cdots \times \ ]-\tau_M, \tau_M[$ and $\mathcal{F} : \mathcal{P} \to \mathcal{H}_t$, where $\mathcal{H}_t \equiv \mathcal{F}(\mathcal{P}) \subset \mathcal{H}$. In the current study we have investigated the use of a function $g : \mathbb{R}^M \to \mathcal{P}$, that is a bounded monotonous injec-

tive function into the space of thresholded displacement parameters. Constructing $\mathcal{G} = \mathcal{F} \circ g$, where $\mathcal{G} : \mathbb{R}^M \to \mathcal{H}_t$ gives us the desired function $\mathcal{G}$, namely a homeomorphic mapping from the parameter space $\mathbb{R}^M$ into the space of diffeomorphisms. As an example of the function g we have chosen a set of hyperbolic tangent function, because the range where it is close to linear in large. The composed mapping $\mathcal{G}$ and the different ranges are illustrated in Figure 1.

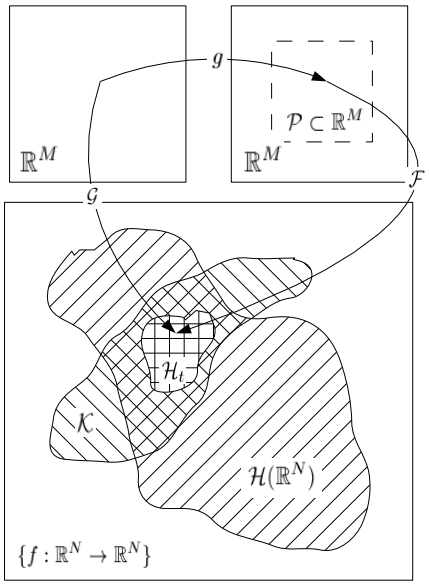


Figure 1: Illustration of the mapping $\mathcal{G}$ from $\mathbb{R}^N$ to $\mathcal{G}$, along with our proposed composed mapping $\mathcal{G} = \mathcal{F} \circ g$

.

We define g coordinate-wise by

$$\boldsymbol{g} = \{g_1, ..., g_M\} \text{ where } g_i : R \to]-\tau_i, \tau_i[$$

$$s_i = g_i(t_i) = \tau_i \tanh a_i t_i \text{ , for } \quad i \in \{1, ..., M\} \tag{5}$$

where τ_i are the threshold parameters reducing the displacement parameter space of the warp to $\mathcal{P} \subset \mathbb{R}^M$, $\boldsymbol{s} = \{s_1, ..., s_M\} \in \mathcal{P}$ are the displacement parameters and a_i are constants ensuring that the impact of each t_i is of the same order of magnitude.

2.1.4 Properties of the g mapping

Before we continue with an empirical validation of our proposed mapping we will make some theoretical considerations over the choice of homeomorphic mapping g. For small values $t \in \mathbb{R}^M$ the L_2 norm in $\mathbb{R}^M$ is equivalent to a scaled L_2 norm in $g(\mathbb{R}^M)$ to a first order. In other words, relating this to diffeomorphic warps, for small deformations the defined norm is equivalent to the usual metric applied in analysis of the warp fields [2, 9].

2.1.5 The parameter distribution

We believe that the distribution of the parameters is well described by normal distribution and we will show what distribution this describes in the displacement parameter space of the warp function. Let f_{t_i} be the marginal distribution of the parameter t_i and f_{g_i} be the marginal distribution of the warp parameter $s_i = g_i(t_i)$ then

$$f_i(t_i) = \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{t_i^2}{2\sigma_i}} \tag{6}$$

$$f_{g_i}(s_i) = \frac{1}{2a_i \cdot \tau_i \sqrt{2\pi\sigma_i}} \left(e^{-\frac{g^{-1}(s_i)^2}{2\sigma_i}} + \frac{e^{\frac{\mu_i^2}{2\sigma_i}}}{2} e^{-\frac{(g^{-1}(s_i)-\mu_i)^2}{2\sigma_i}} + \frac{e^{\frac{\mu_i^2}{2\sigma_i}}}{2} e^{-\frac{(g^{-1}(s_i)+\mu_i)^2}{2\sigma_i}} \right) \tag{7}$$

where $\mu_i = \frac{\sigma_i a_i}{2}$ and this distribution is seen to be the composition of three Gaussian distributions scaled by g^{-1}. For small μ_i this is approaching the Gaussian distribution which is often the distribution for the warp parameters in the small deformation domain and for μ_i big the two μ_i displaced distributions dominate, and we observe a high concentration of parameters around the threshold τ_i. In the presence of strong deformations this also what we expect when imposing a threshold on the warp deformation parameters. Based on these considerations we expect an $M - dimensional$ normal distribution of our parameters to be well suited for modelling the distributions of the observed deformations.

2.1.6 Statistical deformation model

In the previous section we argued that the expected distribution of warps could be modelled as an $M-$ *dimensional* normal distribution. If this is the case PCA is the shown to be the optimal choice of analysis tool for creating a compact model of the observations, and

3 Implementation

To validate our approach for construction of diffeomorphic deformation model we have adapted the grid based diffeomorphisms by Cootes [2] with our g mapping. These diffeomorphisms can be viewed as an extension to standard linear interpolation, where the interpolation coefficients are transformed by a suitable kernel $k(r)$ which ensures smoothness across the grid boundaries. The displacement of a 2D point $\boldsymbol{x} \in \mathbb{R}^2$ is given by

$$\begin{aligned} \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{d}) &= \sum_{m=0}^{1}\sum_{n=0}^{1} k_n(v)k_m(w)\boldsymbol{d}_{i+n,j+m} \\ &= \sum_{m=0}^{1}\sum_{n=0}^{1} a_{i+m,j+n}(\boldsymbol{x})\boldsymbol{d}_{i+n,j+m} \quad (8) \\ &= \begin{bmatrix} \boldsymbol{a}(\boldsymbol{x})^\top & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{a}(\boldsymbol{x})^\top \end{bmatrix} \boldsymbol{d} \quad (9) \end{aligned}$$

where $k_0(r) = k(r)$, $k_1(r) = 1 - k(r)$, i and j is the local indices of the neighboring grid points, v and w are relative positions of x in the neighborhood and $\boldsymbol{d}$ and $\boldsymbol{d}_{i,j}$ are all the displacements and the displacement of the (i,j)-node, respectively. By substituting the displacements $\boldsymbol{d}$ with the g mapping with a suitable threshold τ this deformation model will no longer be able to generate non-diffeomorphisms.

For notational simplicity the displacement in the ith direction will represented by

$$u_i(\boldsymbol{x}, \boldsymbol{t}_i) = \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{g}_\tau(\boldsymbol{t}_i), \quad (10)$$

and the warp function is written in the form

$$\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{t}). \quad (11)$$

3.1 Image registration

To drive the registration between a reference image R and a target image I we apply the sum-of-squared-differences (SSD) as our similarity measure and the regularization term is given by $\mathcal{D}(\phi) = d(e, \phi) = \|t\|_2$, where e is the identity map corresponding to $\boldsymbol{t} = 0$. The SSD comparison leads us to calculate the reference image as the arithmetic mean of the warped target images, as this is the optimum SSD solution to (1) [5].

$$\begin{aligned} F(\boldsymbol{t}) &= \frac{1}{2}\sum_{\boldsymbol{x}} (R(\boldsymbol{x}) - I(\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}))^2 + \alpha\|\boldsymbol{t}\|_2^2 \quad (12) \\ &= \frac{1}{2}\sum_{\boldsymbol{x}} E^2(\boldsymbol{x}, \boldsymbol{t}) + \alpha\|\boldsymbol{t}\|_2^2. \quad (13) \end{aligned}$$

To achieve a fast optimization we apply the inverse compositional optimization approach by Baker et al. [1] to the cost function. Thus, we obtain a minimum by iteratively minimizing

$$\begin{aligned} F_{ic}(\boldsymbol{t}) &= \frac{1}{2}\sum_{\boldsymbol{x}} (R(\boldsymbol{\varphi}(\boldsymbol{x}, \Delta\boldsymbol{t})) - I(\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}))^2 \\ &\quad + \alpha\|\boldsymbol{t} - \frac{\partial \boldsymbol{t}'}{\partial \Delta\boldsymbol{t}}\Delta\boldsymbol{t}\|^2 \quad (14) \end{aligned}$$

with respect to $\Delta\boldsymbol{t}$ and updating $\boldsymbol{t}$ according to

$$\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}') \leftarrow \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}) \circ \boldsymbol{\varphi}^{-1}(\boldsymbol{x}, \Delta\boldsymbol{t}). \quad (15)$$

In Appendix B it is shown how $\boldsymbol{t}'$ is derived from (15).

By performing a first-order Taylor-expansion on $R(\boldsymbol{\varphi}(\boldsymbol{x}, \Delta\boldsymbol{t})$ around $\boldsymbol{x}$ in (14), taking the derivatives wrt. $\Delta\boldsymbol{t}$ and setting them equal to zero we get

$$\Delta\boldsymbol{t} = \boldsymbol{H}^-1 \left[\sum_{\boldsymbol{x}} \boldsymbol{SD}(\boldsymbol{x})^\top E(\boldsymbol{x}, \boldsymbol{t}) + \alpha\frac{\partial \boldsymbol{t}'}{\partial \Delta\boldsymbol{t}}^\top \boldsymbol{t}\right] \quad (16)$$

where

$$\boldsymbol{SD}(\boldsymbol{x}) = \nabla R(\boldsymbol{x})\frac{\partial \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{0})}{\partial \boldsymbol{t}} \quad (17)$$

and

$$\boldsymbol{H} = \sum_{\boldsymbol{x}} \boldsymbol{SD}(\boldsymbol{x})^\top \boldsymbol{SD}(\boldsymbol{x}) + \alpha\left[\frac{\partial \boldsymbol{t}'}{\partial \Delta\boldsymbol{t}}\right]^\top \left[\frac{\partial \boldsymbol{t}'}{\partial \Delta\boldsymbol{t}}\right]. \quad (18)$$

The advantages with this inverse compositional approach is that $\boldsymbol{SD}(\boldsymbol{x})$ can be pre-computed as it is not dependent on $\boldsymbol{t}$.

4 Validation: corpus callosum model

To demonstrate our approach we have created a deformation model of the Corpus Callosum from 62 two dimensional MR images of the mid-sagittal cross-section of the corpus callosum brain structure. This data set is part of the LADIS (Leukoaraiosis and DISability) study [8], a pan-European study involving 12 hospitals and more than 700 patients. Furthermore, each corpus callosum have manually been annotated with 72 landmark by a clinician, which we will later use for validation.

Prior to the non-rigid registration a rigid registration was performed to filter out non-anatomical variation. This was achieved by performing Procrustes analysis on the sets of annotation. After the rigid registration an initial reference was created by computing a mean image of the rigid registered images. All corpus callosum images were then non-rigidly registered to the reference, and a new reference was computed by averaging. This was done multiple times until the reference stabilized. For the non-rigid registration the cosine kernel $k(r) = 0.5(1 + cos(\pi r))$ was applied [2]. The non-rigid warps were modelled by composing three grid based diffeomorphisms in a fine-to-coarse maner. The dimension of the applied grids were 5×4, 10×8 and 20×16. The non-rigid registrations were carried out in coarse to find order. After each level φ_i of the warp was estimated the target image was updated by warping the target image back into the reference coordinate frame by $T_{n+1}(x) = T_n(\varphi(x))$. This was done to ensure that different parameters from different warps could be compared [2]. a_i of the g mapping was set proportional to the inverse of the squared grid node distance because the grid was 2 dimensional. The image registration was validated using the Dice measure, which is twice the intersecting area between the ground truth shape outline of the warped image and the outline of the reference shape divided by the total area inside the two outlines. The ground truth was obtained from the expert annotations. The Dice measure was 0.884 ± 0.048. In Fig. 2 we show an example of a typical registration of an image. In Fig. 3 the cumulative overlap of the aligned corpus callosum shapes before and after a rigid registration is illustrated, showing a clear improvement in correspondences between the shapes.

To create a compact deformation model, PCA was applied to the parameters after the groupwise registration of the images. 13 modes of variation could describe 95 % of the observed variation in the population as observed in Figure 4, and the first three modes are illustrated in Fig. 5. The first mode of variation is seen to be related to a vertical stretch and in particular to the size of the septum pellucidum (the dark area between the bright corpus callosum and the bright Fornix), the second mode is related to the kink of the corpus callosum and the thickness of the structure and the same goes for the third mode but with a different bending of the Fornix. Rueckert et al. have also analyzed the corpus callosum and they found modes quite similar to the ones found in the current study [9]

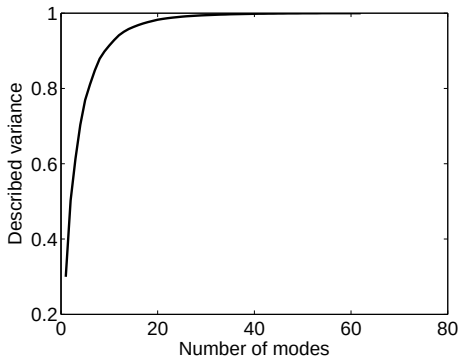
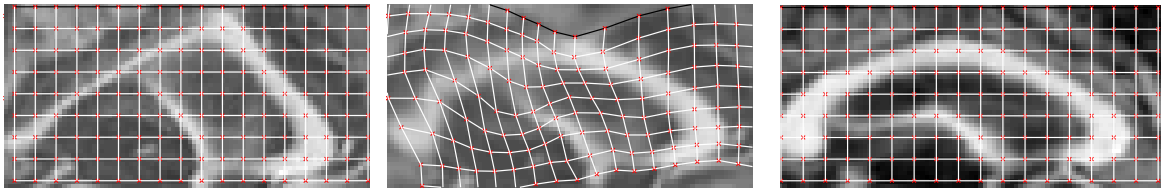


Figure 4: Plot relating described variance with number of modes included in the model.

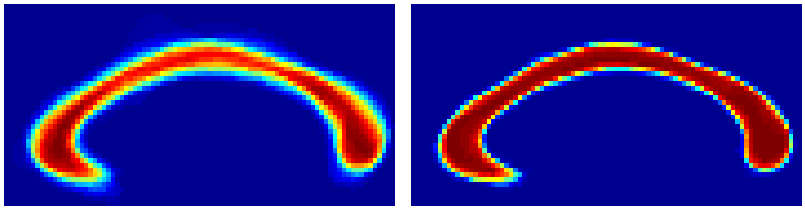
5 Discussion

We have shown how a parametric function can be defined on the unbounded linear space $\mathbb{R}^M$ and still produce diffeomorphic warps. When this is accomplished by first mapping $\mathbb{R}^M$ into an open bounded subset of $\mathbb{R}^M$, which inevitably leads to an asymptotic behavior at the closure of the bounded set. In our im-



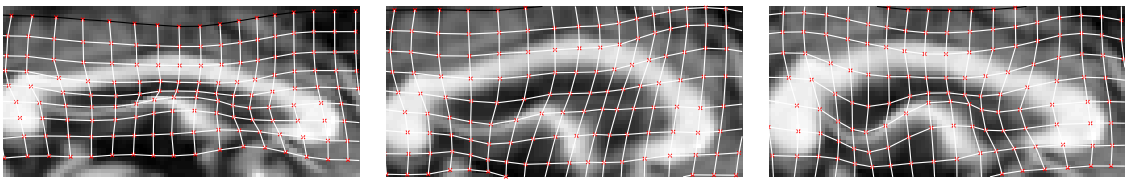
(a) Template image (b) Warped template (c) Reference image

Figure 2: Registration of an image to the reference.s

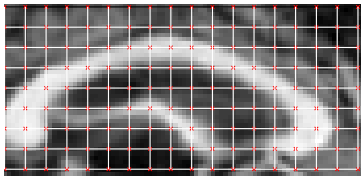


(a) Before nonrigid registration (b) After nonrigid registration

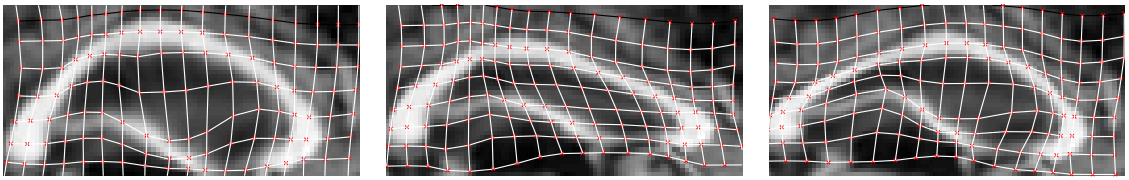
Figure 3: Cumulative overlap of the aligned corpus callosum shapes before and after a rigid registration



(a) 1st mode (b) 2nd mode (c) 3rd mode



(d) Reference



(e) 1 st mode (f) 2nd mode (g) 3rd mode

Figure 5: First three modes of the PCA model of the corpus callosum, shown as the reference warped $\pm$ 3 std. deviations

plemented example the parameters of the model by Cootes et al. asymptotically approach $\frac{1}{\pi}$ where singularities in the warp *may* occur. We believe that our distance measure is very reasonable when we are indeed approaching a singularity, as a small change in the displacement parameters of the warp will cause a huge impact on curvature of the warp function. In Fig. 6, where -6 std. deviations of the first mode is shown. We see that a singularity start to form in the contracting area but this is highly unlikely as predicted by our model and metric.

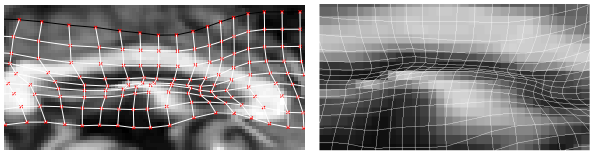


Figure 6: -6 Stnd. deviations of the first mode, normal view and a zoomed view on the beginning singularity.

With the choice of tanh function, the asymptotic behavior is assumed to be exponential, which may not always be the case. There are obviously an infinite variety of monotonous bounded functions, e.g. arcus tangent, and we will be investigating the choice of function in more detail.

A problem, we believe, that may occur with the proposed method is that we cannot be sure that the threshold does actually mark a singularity. A simple translation would for instance be asymptotic as well, which is why initial rigid alignment is very important indeed. Currently we investigate more involved parameter restrictions than the simple threshold to circumvent this possible problem.

Our validation on corpus callosum data showed that we were able to learn the important modes of variation, similar to previous obtained results, while the relatively high Dice coefficient illustrated that our warp representation was able to capture the large variations in the data set. We believe it is an advantage that all configurations in our parameter space are valid diffeomorphism, such that all gradients and derivatives during the optimization are well defined.

6 Conclusions

This paper proposed a new warp representation which allows statistical analysis on unrestricted linear parameter space. Furthermore, we have shown that L_2-norm the parameter space introduces a reasonable metric in the actual space of modelled diffeomorphisms.

Acknowledgements

The authors extend their gratitude to the LADIS work group for supplying the corpus callosum data. In particular we acknowledge the annotation effort of Charlotte Ryberg and Egill Rostrup from the Danish Research Center for Magnetic Resonance, Copenhagen University Hospital, Hvidovre, Denmark. We also wish to thank Vagn L. Hansen and Niels V. Christensen for fruitful discussion during this work.

A Warp inversion

Theorem A.1. *Consider the function* $\boldsymbol{\varphi} : \mathbb{R}^N \times \mathbb{R}^M \mapsto \mathbb{R}^N$ *of type* $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{t})$ *and let* $\boldsymbol{\varphi_t}(\boldsymbol{x}) = \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t})$ *be a* $\mathcal{C}^1$*-diffeomorphism. If* $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{0}) = \boldsymbol{0}$ *and* $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{t}) = -\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{t})$*,* $\boldsymbol{\varphi}(\boldsymbol{x}, -\boldsymbol{t})$ *converges with second-order to* $\boldsymbol{\varphi}^{-1}(\boldsymbol{x}, \boldsymbol{t})$*.*

Proof.

$$
\begin{aligned}
& |\xi_i(h\boldsymbol{t})| = |\varphi_i(\boldsymbol{\varphi}(\boldsymbol{x}, h\boldsymbol{t}), -h\boldsymbol{t}) - x_i| \\
= \; & |x_i + u_i(\boldsymbol{x}, h\boldsymbol{t}) - u_i(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, h\boldsymbol{t}), h\boldsymbol{t}) - x_i| \\
< \; & |u_i(\boldsymbol{x}, h\boldsymbol{t}) - u_i(\boldsymbol{x}, h\boldsymbol{t}) + \frac{\partial u_i}{\partial \boldsymbol{x}}(\boldsymbol{x}, h\boldsymbol{t})\boldsymbol{u}(\boldsymbol{x}, h\boldsymbol{t})| \\
< \; & |h\boldsymbol{t}^\top \frac{\partial^2 u_i}{\partial \boldsymbol{x} \partial h\boldsymbol{t}}(\boldsymbol{x}, \boldsymbol{0}) \frac{\partial \boldsymbol{u}}{\partial h\boldsymbol{t}}(\boldsymbol{x}, \boldsymbol{0}) h\boldsymbol{t}| \\
< \; & |c| \cdot |h^2| \qquad (19)
\end{aligned}
$$

□

B Derivation of update function

In general, it is unlikely that $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}) \circ \boldsymbol{\varphi}^{-1}(\boldsymbol{x}, \Delta\boldsymbol{t})$ can be parameterized with $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}')$, and thus it has to be

approximated.
In Appendix A, it was shown that $\boldsymbol{\varphi}(\boldsymbol{x}, -\boldsymbol{t})$ is a first-order approximation to $\boldsymbol{\varphi}^{-1}(\boldsymbol{x}, \boldsymbol{t})$ as the error converges with second-order to zero. The composition in Eq. 15 is approximated with the parameters $\boldsymbol{t}'$ which minimizes the SSD between the true compositional warp and the warp $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}')$

$$\sum_x \Delta\boldsymbol{\varphi}(\boldsymbol{x})^\top \Delta\boldsymbol{\varphi}(\boldsymbol{x}) \tag{20}$$

where

$$\begin{aligned} \Delta\boldsymbol{\varphi}(\boldsymbol{x}) &= \boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{x}, \Delta\boldsymbol{t}), \boldsymbol{t}) - \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{t}') \\ &= \boldsymbol{a}(\boldsymbol{x})^\top (\boldsymbol{g}_\tau(\Delta\boldsymbol{t}) - \boldsymbol{g}_\tau(\boldsymbol{t}')) \\ &\quad + \boldsymbol{a}(\boldsymbol{\varphi}(\boldsymbol{x}, \Delta\boldsymbol{t}))\boldsymbol{g}_\tau(\boldsymbol{t}). \end{aligned} \tag{21}$$

If

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}(\boldsymbol{x_1})^\top \\ \vdots \\ \boldsymbol{a}(\boldsymbol{x_n})^\top \end{bmatrix} [], \text{and } \boldsymbol{A}_\varphi = \begin{bmatrix} \boldsymbol{a}(\boldsymbol{\varphi}(\boldsymbol{x}_1, \Delta\boldsymbol{t})^\top \\ \vdots \\ \boldsymbol{a}(\boldsymbol{\varphi}(\boldsymbol{x}_n, \Delta\boldsymbol{t})^\top \end{bmatrix} []$$

the updated warp parameters $\boldsymbol{t}'$ can be found by solving the system

$$\boldsymbol{0} = \boldsymbol{A}(\boldsymbol{g}_\tau(\Delta\boldsymbol{t}_i) - \boldsymbol{g}_\tau(\boldsymbol{t}_i')) + \boldsymbol{A}_\varphi \boldsymbol{g}_\tau(\boldsymbol{t}_i). \tag{22}$$

The least square solution to the system is

$$\boldsymbol{t}_i' = \boldsymbol{g}_\tau^{-1}\left(\boldsymbol{A}^\dagger \boldsymbol{A}_\varphi \boldsymbol{g}_\tau(\boldsymbol{t}_i) + \boldsymbol{g}_\tau(\Delta\boldsymbol{t}_i)\right) \tag{23}$$

where $\boldsymbol{A}^\dagger = \left[\boldsymbol{A}^\top \boldsymbol{A}\right]^{-1} \boldsymbol{A}^\top$.

As A_φ has to be evaluated on warped points it is relatively computational expensive to evaluate. Thus, we perform a first-order Taylor expansion on A_φ and arrive at

$$\boldsymbol{t}_i' = \boldsymbol{k}^{-1}\left(\boldsymbol{A}^\dagger \boldsymbol{A}_{J_i} \boldsymbol{A}\boldsymbol{g}_\tau(\Delta\boldsymbol{t}_i) + \boldsymbol{g}_\tau(\Delta\boldsymbol{t}_i) + \boldsymbol{g}_\tau(\boldsymbol{t}_i)\right), \tag{24}$$

where

$$A_{J_i} = \boldsymbol{I} + \mathrm{diag}(\frac{\partial \boldsymbol{a}(\boldsymbol{x}_j)}{\partial x_i}^\top \boldsymbol{g}_\tau(\boldsymbol{t}_i))_{j=1\ldots n} \tag{25}$$

References

[1] S. Baker and I. Matthews. Lucas-kanade 20 years on A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[2] T. Cootes, C. Twinning, and C. Taylor. Diffeomorphic statistical shape models. *British Machine Vision Conference*, 1:447–456, 2004.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *5th European Conference on Computer Vision*, 2:484–498, 1998.

[4] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Prinicpal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Image Analysis*, 23(8):995–1005, 2004.

[5] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:151–160, 2004.

[6] S. Lee, G. Wolberg, K.-Y. Chwa, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Trans. Visualizat. Comput. Graphics*, 2(4):337Ũ354, 1996.

[7] J. Modersitzki. Numerical methods for image registration. *Oxford Uni. Press*, 2004.

[8] L. Pantoni, A. M. Basile, G. Pracucci, K. Asplund, J. Bogousslavsky, H. Chabriat, T. Erkinjuntti, F. Fazekas, J. M. Ferro, M. Hennerici, J. O'brien, P. Scheltens, M. C. Visser, L. O. Wahlund, G. Waldemar, A. Wallin, and D. Inzitari. Impact of age-related cerebral white matter changes on the transition to disability - the LADIS study: Rationale, design and methodology. *Neuroepidemiology*, 24(1-2):51–62, 2005.

[9] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–25, 2003.

[10] M. Vester-Christensen, S. G. Erbou, S. Darkner, and R. Larsen. Accelerated 3D image registration. In *International Symposium on Medical Imaging 2007, San Diego, {CA}*, feb 2007.