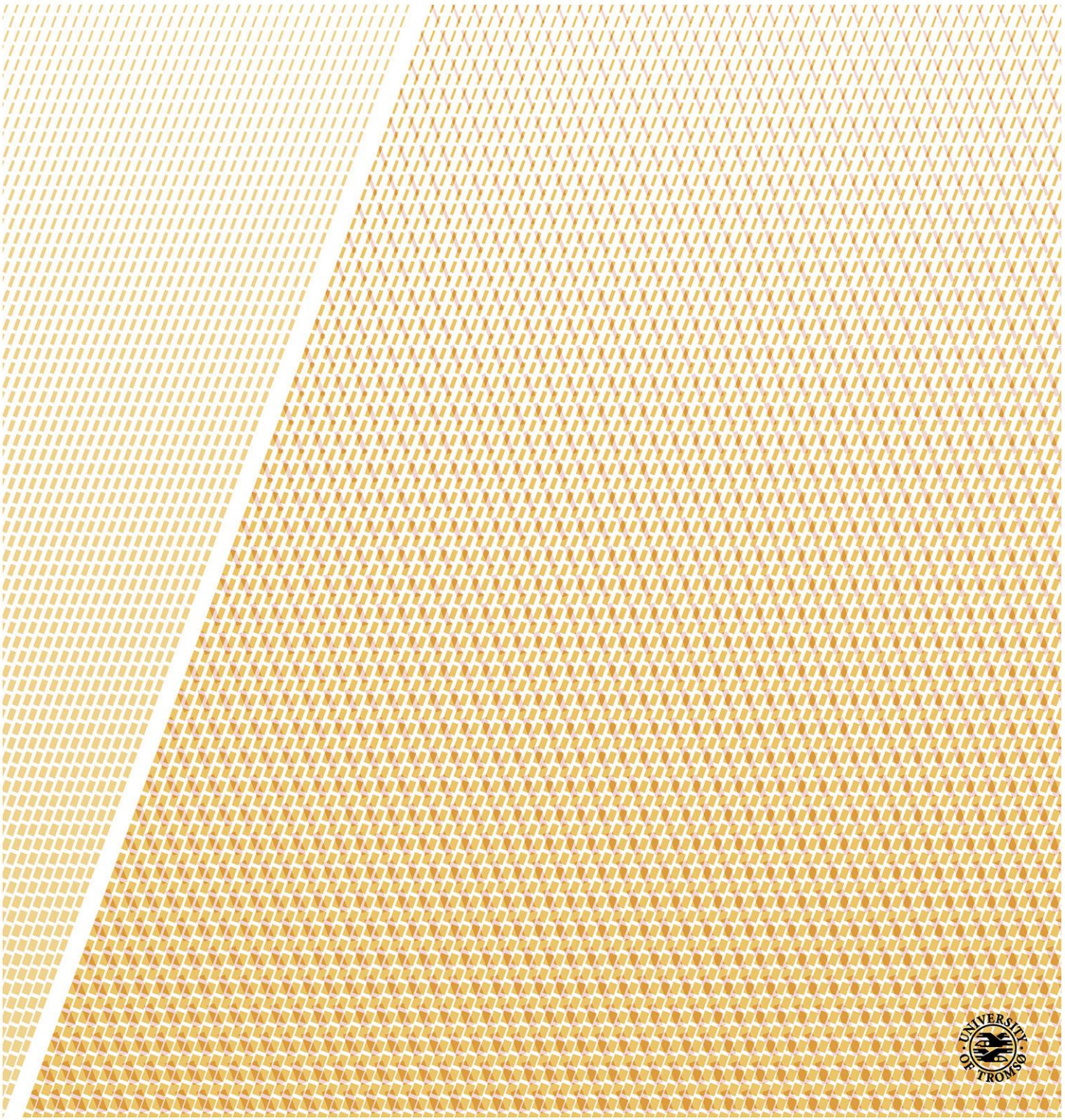
UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Science and Technology, Department of Mathematics and Statistics

Statistical Analysis of Medical Shapes and Directional Data

Jörn Schulz

A dissertation for the degree of Philosophiae Doctor – October 2013



ABSTRACT

The use of statistical shape analysis in medical settings has increased during the last decades. This thesis presents contributions to three major topics of statistical shape analysis. These topics are: the modeling of shape by a geometrical model, the study of rotational shape deformations and the comparison of shapes between populations. All methods that are developed in this thesis are applied to medical problems.

Paper I presents a semiautomatic method for prostate segmentation in radiotherapy treatment planning. To facilitate the manual delineation of the prostate in medical images, an intuitive approach is developed for 3D modeling of the prostate by slice-wise best fitting ellipses in each image. The focus is to estimate a mean shape from a set of training data parametrized by the geometrical model. The proposed estimate is initialized by the definition of a few control points in a new patient. All results are compared to manual delineations by physicians with an average percent-volume overlap of 90%. An additional sample study suggests possible time saving effects by the method and illustrates thereby its potential.

Paper II studies rotational deformations of objects. The rotational deformations such as twisting or bending have been observed as the major variations in some medical applications, where the features of the deformed 3D objects are directional data. Models and estimators are proposed for one or the composition of two simple deformations based on the directional features. The proposed method uses a generalized small circle fitting on the unit sphere. Two analyses of 3D object data are presented in detail: one using skeletal representations in medical image analysis as well as one from biomechanical gait analysis of the knee joint.

Paper III investigates shape differences between two populations of medical objects such as the hippocampi of first episode schizophrenics and controls. Shape differences can be quantified by hypothesis tests, as for instance by a permutation test. Several aspects of a sensitive hypothesis test are elaborated, e.g., geometrical model properties that support accurate statistics of populations. A permutation test is developed to test mean differences of two populations. The proposed method is designed for data whose representations include both Euclidean and non-Euclidean elements. By supporting non-Euclidean components such as directions, the proposed hypothesis test is novel in the study of morphological shape differences. Both global and local analyses showed statistically significant differences between the first episode schizophrenics and controls and demonstrated the sensitivity of the method.

ACKNOWLEDGMENTS

The work presented in this thesis was carried out at the Department of Mathematics and Statistics at the Arctic University of Norway (UiT). Financial support was provided by the Norwegian Research Council through eVita program grant no 176872/V30. The research was performed as part of Tromsø Telemedicine Laboratory, funded by the Norwegian Research Council 2007-2014, grant no 174934.

First, I like to thank my main supervisor Professor Fred Godtliebsen at UiT who gave me the opportunity to do a PhD at such a charming place as Tromsø with impressive nature. Thanks for all the support during the years, the helpful discussions and your never-ending optimism. You also supported my two research visits in the USA in February-April 2010 and 2011 which resulted in Paper II and III of this thesis.

Second, I like to thank my co-supervisors, Professor Stephen M. Pizer and Professor Steve Marron at the University of North Caronlina at Chapel Hill (UNC). I am sincerely grateful for all the help, discussion and tutoring. Stephen widened my knowledge about shape modeling and introduced me in the world of skeletal models. Steve discussed with me several interesting problems in the Euclidean and non-Euclidean world and was always available to give me statistical advices. I also gratefully thank Assistant Professor Sungkyu Jung at the University of Pittsburgh. The collaboration with him was always excellent, helpful and resulted in Paper II. In addition, I wish to thank Jared Vicory, Mark Foskey, Gregg Tracton, Julian G. Rosenman and Martin Styner at the UNC and Juan Carlos Prieto at CREATIS - INSA (France). Jared gave me several advices on running Pablo and fitting skeletal models. I also met Professor Stephan Huckeman at the Georg-August-University of Göttingen during my stay at UNC. Thanks for sharing your impressive knowledge in statistics on manifold data.

I would also like to thank my co-supervisors Stein Olav Skrøvseth at the Norwegian Centre for Integrated Care and Telemedicine (NST) and Professor Håvard Rue at the Norwegian University of Science and Technology in Trondheim. Stein Olav was always there for helpful advices and discussions. Håvard was available for discussions during regular visits in Tromsø. Special thanks goes to Veronika Kristine Tømmerås, Kirsten Marienhagen, Jorunn Andrea Skjelvareid and Rune Sylvarnes at the Department of Oncology at University Hospital of North Norway. They provided me with data and adviced me in radiotherapy planning. Furthermore, I would like to express my gratiude to Jan Terje Kvaløy and the whole department of electrical engineering and computer science at the University of Stavanger to welcome me in their friendly and relaxed enviroment after I moved to Stavanger in January 2013.

Also, I would like to thank my colleagues Marc, Kevin, Kristian, Kajsa and Maciel and many others at TTL,UiT and UNC for an excellent research and social enviroment during the PhD. I will

miss the coffee breaks with random discussions or just to sit on the sofa in the lunch room with an incredible view to the mountains.

Last but not least, I wish to send my gratitude to my family in Germany for the support and the patience to keep the fingers crossed about the years. I wish of course also to thank Izumi and all my friends throughout the world for support, to brighten up and enrich my everyday life during the PhD.

Tromsø, October 2013
Jörn Schulz

LIST OF PUBLICATIONS

Paper I A semiautomatic tool for prostate segmentation in radiotherapy treatment planning, **Schulz, Jörn, Skrøvseth, Stein Olav, Tømmerås, Veronika Kristine, Marienhagen, Kirsten and Godtliebsen, Fred**, submitted to *BMC Medical Imaging*

Paper II Analysis of rotational deformations from directional data, **Schulz, Jörn, Jung, Sungkyu, Huckemann, Stephan, Pierrynowski, Michael, Marron, J.S. and Pizer, Stephen M.**, revised and re-submitted to *Journal of Computational and Graphical Statistics*

Paper III Nonlinear hypothesis testing of geometrical object properties of shapes applied to hippocampi, **Schulz, Jörn, Pizer, Stephen M., Marron, J.S. and Godtliebsen, F.**, preprint

CONTENTS

Part I

INTRODUCTION PART

LIST OF ABBREVIATIONS

AAM	Active appearance model
ASM	Active shape model
BFE	Best fitting ellipse
CPNS	Composite principal nested spheres
CT	Computed tomography
GOP	Geometrical object property
MCMC	Markov chain Monte Carlo
MR	Magnetic resonance
PCA	Principal component analysis
PGA	Principal geodesic analysis
PDM	Point distribution model
PNS	Principal nested spheres
RT	Radiotherapy treatment
RTF	Rotation twisting folding
S-rep	Skeletal representation

1

INTRODUCTION

Statistical shape analysis has become an important field in medicine. The use of medical imaging techniques has increased during the last decades and so has the workload of physicians to manually segment target objects in patient data, e.g., in surgery and radiotherapy treatment planning. Manual image interpretation is a time-consuming and error-prone task (Davies et al., 2008). Segmentation of objects is one application of statistical shape analysis. In shape analysis, a *shape distribution* is often obtained from a *training set*, i.e., a representative set of examples from an object class. The shape distribution can be represented by a mean and the main modes of variation and used in a Bayesian framework as a prior distribution. This allows segmentation of an object in data with low image information. Cootes et al. (2001) proposed a segmentation approach, *active appearance model* (AAM) that combines shape distribution and image intensities. In addition to segmentation of objects, other applications of shape analysis are: the study of morphological differences of human organs and body structures (Albertson et al., 2003; Styner et al., 2004; Ferrarini et al., 2006), the study of drug effects in epidemiology (McClure et al., 2013) and the study of deformations of anatomical shapes (Joshi et al., 2002; Rivest et al., 2008; Oualkacha and Rivest, 2012).

Shape analysis of an object $\Omega \subset \mathbb{R}^d$ requires a definition of the term 'shape', a *geometric model* $\mathcal{M}$ that parametrizes Ω and suitable statistics which deal with all components of the geometric model. This thesis is focused on three-dimensional objects, i.e., $d = 3$.

Often, the size, location and rotation of an object are not connected to the underlying shape variation. This thesis will follow Kendall (1984), who defines a shape as the geometrical information that remains when the group of similarity transforms (location, scale and rotation) are removed. However, scaling can be an important feature if the underlying research question is to find shape differences between two populations. Dryden and Mardia (1998) introduced the term *size-and-shape* which defines the geometrical information that remains after removing location and rotation. It follows that shape comparison is affected by subjectivity.

With increased use of imaging techniques, the number of geometric models has also grown steadily. The most well-known examples are landmark models (Kendall, 1984; Bookstein, 1986), boundary point distribution models (PDMs) (Cootes et al., 1992; Kurtek et al., 2011) and the family of skeletal models (Blum and Nagel, 1978; Siddiqi and Pizer, 2008; Pizer et al., 2013) as discussed in Section 1.1.

The geometric model $\mathcal{M}$ can live in a product of Euclidean space (e.g., $\mathbb{R}^{3n}$), a product of non-Euclidean spaces (e.g., S^{2m}), or a product of Euclidean and non-Euclidean spaces (e.g.,

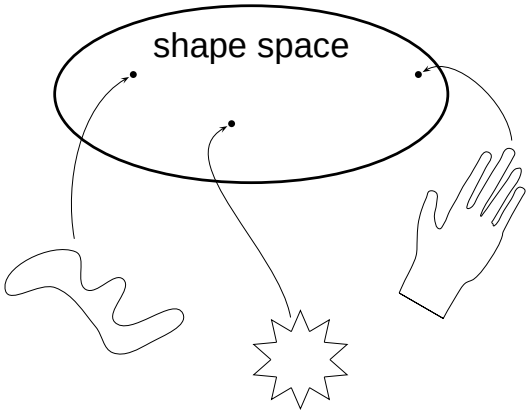


Figure 1: Idea of a shape space. Each object is represented by a point in the shape space.

$\mathbb{R}^{3n} \times S^{2m}$) where n and m depend on the model; $S^d := \{x \in \mathbb{R}^{d+1} : \|x\| = 1\}$ denotes the d-dimensional unit hyper-sphere where $\|x\| := \sqrt{x'x}$ is the Euclidean norm of a column vector $x \in \mathbb{R}^{d+1}$ with transpose x'. The model representation in these coordinate systems reflects a parametrization of an underlying object and is not yet location, rotation and scaling invariant. The *shape space* $\mathcal{S}$ is defined as the space of all possible geometrical configurations of $\mathcal{M}$ which are invariant under global translation, rotation and scaling. Another important notion is the *pre-shape space*, the space of all possible configurations of $\mathcal{M}$ which are invariant under translation and scaling. Dryden and Mardia (1998) have shown that models originally defined in $\mathbb{R}^{3n}$, live on a $(3n-4)$-dimensional unit hyper-sphere S^{3n-4} in the pre-shape space. Thus, given a model $\mathcal{M}$, each shape corresponds to a point in the high-dimensional non-linear shape space $\mathcal{S}$ as depicted in Figure 1. The non-Euclidean data structure requires suitable non-linear statistics.

Multivariate statistics for Euclidean data structures such as *principal component analysis* (PCA) have been intensively developed during the last century. PCA reduces the dimensionality of data and describes the distribution by a selection of eigenvectors, also known as modes of variation. Methods for Euclidean data are not easily applicable to non-Euclidean data. Often, a linear approximation of the non-Euclidean data by an embedding $J : \mathcal{S} \to \mathbb{R}^N$ is applied in order to use classical statistical methods. Huckemann et al. (2010) pointed out that current methods of PCA by linear Euclidean approximation are unsuitable if data in non-linear spaces fall into regions of high curvature, or if they have a large spread. Thus, a generalization of the methods to non-Euclidean data structures is necessary.

The focus of this thesis is the construction and application of geometric models, the use of suitable statistical analyses and the development of novel methods to analyze shapes of objects. In Section 1.1, the most common shape representations will be introduced. Two of these models are used extensively in Paper I-III. In Section 1.2, important definitions and statistical methods will be given that are used particularly in Paper II and Paper III of this thesis. In Section 2, all

three papers are summarized and discussed. The main part of this thesis is Part II which contains the manuscripts of Paper I, II and III, and the corresponding supplementary materials.

1.1 SHAPE REPRESENTATION

First, a geometric model $\mathcal{M}$ is required to parametrize a geometric representation $\Omega \subset \mathbb{R}^3$ where $\mathcal{M}$ typically lies on a manifold. There exists a rich collection of geometrical models in the literature. A selection of important geometrical models are categorized in this thesis as follows:

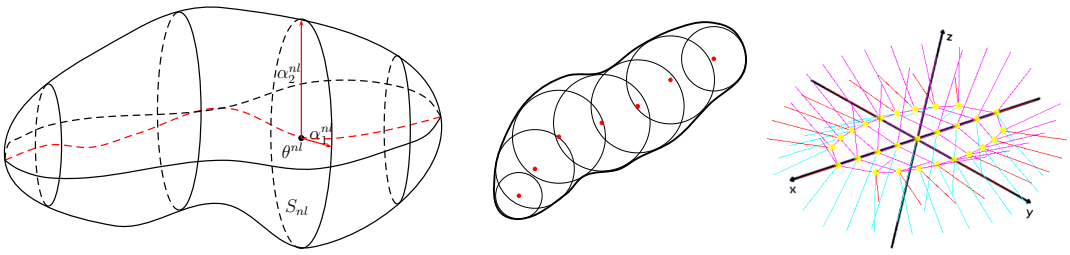
LANDMARK MODELS A solid object is modeled by the positions of a finite number of mathematical or anatomical landmarks represented by a configuration matrix $\mathbf{X} = (X_1, \ldots, X_K)'$ with K landmarks and $X_i \in \mathbb{R}^3$. The landmark model was introduced by Kendall (1984); Bookstein (1986) who did pioneering work in the field of shape statistics. A problem which arises, is the definition of precise landmarks in order to obtain good correspondence between objects. The definition of landmarks is reasonably intuitive in two-dimensional objects but becomes challenging in three-dimensional objects such as the prostate.

PDM A solid object is defined by the positions of the sampled surface points (Cootes et al., 1992; Kurtek et al., 2011). PDMs are also called pseudo-landmarks and the representation is also given by a configuration matrix $\mathbf{X} \in \mathbb{R}^{3K}$ where K is the number of boundary points. A large number of shape models is based on PDMs.

GEOMETRICAL TEMPLATES The boundary of an object is modeled by a set of parametric geometrical components, such as line segments, triangles, circles and ellipses (Dryden and Mardia, 1998). As a result, an object can be described by a list of geometrical parameters. This approach is used in Paper I of this thesis in order to model the prostate by best-fitting-ellipses (BFEs). The geometrical model is described by parameters $(\theta, \alpha, \phi)' \in (\mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\pi, \pi])^L$ where $\theta^l \in \mathbb{R}^2$ defines the position of the l-ellipse, α^l_1, α^l_2 the length of the first and second principal axis and ϕ the rotation angle as depicted in Figure 2a.

DEFORMATION-OF-ATLAS MODELS The shape changes of an object in an image are modeled by the deformations of a template image provided by an atlas (Pennec, 2008; Zhang et al., 2013). The atlas presents a reference of population of shapes obtained from a set of training shapes. Each template is labeled by a collection of representative features and geometric attributes, e.g., a set of landmarks and parametric geometrical components as previously discussed.

SKELETAL MODELS The idea of skeletal models is based on the *medial locus* of an object, first introduced by Blum and Nagel (1978). The medial locus describes the inherent sym-



(a) best fitting ellipses (b) medial locus (c) s-rep

Figure 2: Illustration of geometrical models. A geometrical template model of the prostate by slicewise BFEs is visualized in (a). Each ellipse is defined in a slice $l = 1,\ldots,L$ by a position θ^l, the length α_1^l, α_2^l of the 1st and 2nd principal axis, and a rotation angle ϕ. (b) Visualization of points from the medial locus of a two-dimensional object defined by the centers of maximal circles. (c) Discrete skeletal representation (s-rep) of a flat ellipsoid. The skeletal positions are depicted as small spheres. On each side of the sheet, there is a spoke, a vector with direction and length on the top and on the bottom connecting the skeletal sheet to the boundary. Also, for each edge grid point there is an additional spoke vector connecting the skeletal sheet folded to the crest of the slab.

metry of an object. There exist different possibilities to define a medial locus of an object Ω. A formal definition is given by

$$MA_\Omega := \{q \in \mathbb{R}^n \mid |\{p \in \partial\Omega \mid ||q-p|| = d(q,\Omega)\}| > 1\},$$

where $d(q,\Omega) := \min_{p\in\partial\Omega} ||q-p||$ defines the distance of a point $q \in \mathbb{R}^n$ to the object boundary $\partial\Omega$. Figure 2b visualizes points of the medial locus of a two-dimensional object. In 3D, the medial locus is given by the centers of all maximal spheres. A sphere $S \subseteq \Omega$ is called maximal if there exists no larger sphere $\widetilde{S} \subseteq \Omega$ with $S \subset \widetilde{S}$. In Section 3 of Siddiqi and Pizer (2008) it is shown that the medial locus can be described by an inward "grassfire" that starts at the boundary and dies out at a folded version of the medial locus called M_Ω. Given a folded medial locus M_Ω, the medial representation of an object Ω is determined by a set of spoke directions from points of M_Ω to the corresponding points of tangency on the boundary $\partial\Omega$.

Strictly medial representations are limited by the fact that every protruding boundary kink results in additional medial branches. Thus, two versions of the same object with small noise can have drastically different medial representations. Skeletal models achieve additional stability by relaxing the medial constraint. Figure 2c visualizes a discrete version of the skeletal object representation in $\mathbb{R}^3$ composed of a *skeletal sheet* and *spokes* which emanate from a *skeletal position* p_i on the skeletal sheet to the surface. Each spoke is composed of a normal direction u_i with length r_i. The discrete skeletal representation (s-rep) is described by a feature vector $\boldsymbol{s} = (p_1,\ldots,p_{n_a},r_1,\ldots,r_{n_s},u_1,\ldots,u_{n_s}) \in \mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s} \times S^{2n_s}$

with the number n_a of skeletal positions and n_s of skeletal directions. This model and its properties will be discussed and used extensively in Paper III of this thesis. Furthermore, the model is used as an application in Paper II for the study of rotational deformations.

Additional representations are for example implicit models (Leibe et al., 2006) or models using spherical harmonics (Styner et al., 2006).

The calculation of statistics on a set of shapes requires quantification of shape differences. A central difficulty is the identification of correspondence between the parameterized shapes. Hufnagel et al. (2009) proposed the use of correspondence probabilities for the case of unstructured point sets such as PDMs. Alternatively, Cates et al. (2007) introduced correspondence by the optimization of a dynamic particle system by starting with few initial points on the surface (called particle) and splitting the particles in each iteration step. The optimal configuration is defined as one in which the entropy of each shape is balanced against the entropy of the ensemble shapes. Skeletal models introduce correspondence by a stable branching structure of the skeletal sheet, in addition to modeling the skeleton as medial as possible (Pizer et al., 2013). This is discussed in detail in Paper III of this thesis. The geometrical template model used in Paper I introduces correspondence by the templates themself in addition to constraints between the geometrical parameters.

Landmark models and PDMs are often analyzed by PCA although Dryden and Mardia (1998) have shown that these models, originally defined in $\mathbb{R}^{3n}$, live on the unit hyper-sphere S^{3n-4} after removing location and scaling. Other models, such as geometrical template and skeletal models live by definition on a product of Euclidean and non-Euclidean spaces. Therefore, statistics should incorporate all elements of these models.

1.2 SHAPE STATISTICS

1.2.1 *General definitions*

The shape space $\mathcal{S}$ of a model $\mathcal{M}$ can often be understood as a *manifold*. A space is called a d-dimensional manifold if every point of $\mathcal{S}$ has a neighborhood to an open subset of $\mathbb{R}^d$, i.e., for every point p of the manifold exists a map $J : U_p \to \mathbb{R}^d$ for an open neighborhood U_p of p (Bronstein et al., 2008). For example, the tangent space $T_c S^d$ at $c \in S^d$, $d \geq 2$ can be parametrized by $\mathbb{R}^d$. Without loss of generality, let c be the $(d+1)$-dimensional vector $c = (0, \ldots, 0, 1)'$. The exponential map $\mathrm{Exp}_c : \mathbb{R}^d \to S^d$ is defined for $v \in \mathbb{R}^d$ by

$$\mathrm{Exp}_c(v) = \left(\frac{v}{\|v\|} \sin \|v\|, \cos \|v\| \right),$$

with a convention of $\mathrm{Exp}_c(0) = c$. The exponential map has an inverse, called the log map, and is denoted by $\mathrm{Log}_c : S^d \to T_c S^d$. Thereby, S^d is a manifold.

In general, there are two views to manifolds. The *extrinsic view* understands the manifold embedded in a Euclidean space. Statistics are calculated in the tangent space on an extrinsic mean. The *intrinsic view* understands the manifold as a topological space by itself. Shape spaces are often described by a product of a real space $\mathbb{R}^n$ and a d-dimensional unit hyper-sphere S^d as shown before (e.g., Section 1.1). Elements from $\mathbb{R}_+$ can be transformed by the logarithm into $\mathbb{R}$. Thus, it is important to understand statistics on S^d.

To work with observations on S^d, a distance measure is required. An intrinsic distance measure is defined by the Riemannian geodesic distance function d_g by the arc length of the shortest great circle segment joining $x, y \in S^d$, and is

$$d_g(x,y) = \arccos(x'y). \tag{1}$$

For a random element X whose domain is S^d, a sensible notion of the mean $\mu(X)$ is defined by a minimizer of the mean squared distance,

$$\mu_g(X) = \operatorname*{argmin}_{x \in S^d} \mathbb{E}\{d_g^2(x,X)\}, \tag{2}$$

often called the geodesic or Fréchet mean (Fréchet, 1948; Karcher, 1977). A useful measure of dispersion is the geodesic variance which is defined as

$$\mathrm{Var}_g(X) = \mathbb{E}\{d_g^2(\mu_g(X),X)\} = \min_{x \in S^d} \mathbb{E}\{d_g^2(x,X)\} \tag{3}$$

provided that $\mu(X)$ exists. Papers II and III use the intrinsic Fréchet mean to analyze components on S^d. The spherical information in Paper I (the rotation angle of the 1st principal axis of an ellipse) is analyzed by an extrinsic mean. In all three papers, the understanding of a rotational motion in the three-dimensional space (3-space) is important.

The axis–angle pair $(c, \phi) \in S^2 \times [0, 2\pi)$ represents a rotation in 3-space, where a vector $x \in \mathbb{R}^3$ is rotated by (c, ϕ) by applying $x \mapsto R(c, \phi)x$ with the matrix

$$R(c,\phi) = I_3 + \sin\phi\,[c]_\times + (1 - \cos\phi)(cc' - I_3). \tag{4}$$

Here, I_3 is the the three-dimensional unit matrix, and the 3×3 matrix $[c]_\times$ is the cross product matrix satisfying $[c]_\times v = c \times v$ for any $v \in \mathbb{R}^3$. The group of rotations is denoted by $SO(3)$ with $R'R = I_3$ for $R \in SO(3)$.

Thereby, a configuration matrix $\mathbf{X} = (X_1, \ldots, X_K)'$ with $X_i \in \mathbb{R}^3$, $i = 1, \ldots, K$ can be aligned by $\mathbf{X} \to sR(c,\phi)(\mathbf{X} - t)$ with a translation vector $t \in \mathbb{R}^3$, a scaling factor $s \in \mathbb{R}_+$ and a rotation

matrix $R \in SO(3)$. The trajectory of a rotation of a direction vector $x \in S^2$ about an axis $c \in S^2$ is a circle, centered at c with radius $r = \arccos(x'c)$, and is denoted by

$$\delta(c,r) = \{x \in S^2 : x'c = \cos(r)\} \subset S^2. \tag{5}$$

Since $\delta(c,r) = \delta(-c,\pi - r)$, we may assume that $r \leq \pi/2$. A circle $\delta(c,r)$ is called a *great circle* or *geodesic* if $r = \pi/2$, and a *small circle* if $r < \pi/2$. These observations are important for the analysis of rotational motion in Paper II, such as global rotation, rotational bending and rotational twisting.

1.2.2 *Estimation of shape distributions*

An important aspect of shape analysis is to describe the variability of a set of objects which are parametrized by a model $\mathcal{M}$. Another aspect is to find a subspace that best represents the variability of the high-dimensional data (dimension reduction). The geometrical parameters of the BFEs in Paper I are modeled by Gaussian distributions as described in Dryden and Mardia (1998). Another powerful method is PCA for Euclidean data or the generalization of PCA to non-Euclidean data.

PCA describes each data point by a mean and a sum of weighted eigenmodes. It is also a dimension reduction method when retaining only the main modes of variation in a data set. PCA of d-dimensional data can be calculated by a *forward* or *backward* approach corresponding to the order in which the dimensions are considered. In a traditional forward approach the mean is estimated as the 0-dimensional affine space (a point), which is the best fitting subspace to the data. Then, PCA is described by increasing the dimension of the best fitting subspace successively

$$0 \to 1 \to \cdots \to d-1.$$

Starting with fitting the best subspace of dimension $d-1$ to the data, PCA can also be carried out in the reverse order by

$$0 \leftarrow 1 \leftarrow \cdots \leftarrow d-1.$$

In non-Euclidean spaces, such as the d-dimensional unit sphere S^d, generalizations of PCA depend on this forward or backward direction. Damon and Marron (2013) have studied this across a variety of contexts, and have shown that backwards subspace fitting is generally more amenable to analysis, because it is equivalent to a simple adding of constraints. Generalizations of PCA for data on S^d are discussed in the following.

Fletcher et al. (2004) proposed a forward generalization of PCA to manifolds called *principal geodesic analysis* (PGA). The first step of PGA is to find a center point of the data by a forward Fréchet mean. Given the Fréchet mean, the next step is to find the best fitting geodesic passing

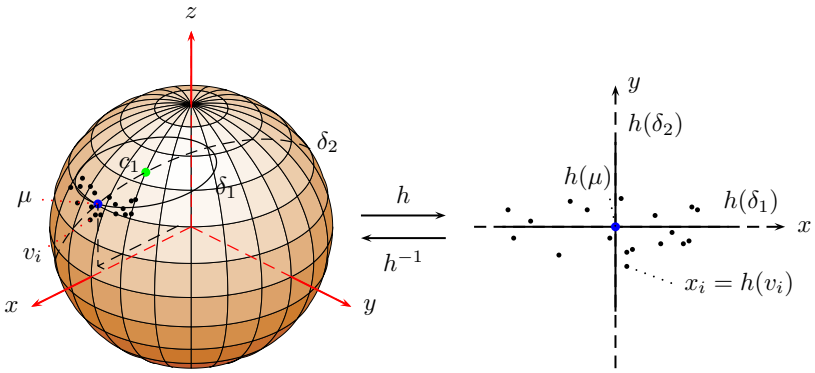


Figure 3: Illustration of point mapping from S^2 to $\mathbb{R}^2$ with PAA. The coordinates $v_i \in S^2$ are mapped to $x_i \in \mathbb{R}^2, i \in \{1,\ldots,n\}$ by a projection h. The coordinate system is defined by the principal circles (δ_1, δ_2). The first principal circle δ_1 is defined by the center c_1 and the radius $r_1 = d_g(c,\mu)$ between the center and the principal circle mean μ. The second principal circle δ_2 is defined by the geodesic passing through the coordinates $(c_1, \mu)^T$. The second principle circle is orthogonal to the first principle circle.

through the mean and afterwards the best fitting geodesic orthogonal to this geodesic. The approach leads to undesirable results if the data are uniformly distributed on a geodesic (e.g., the equator), because the mean would be located at the north or south pole. Huckemann et al. (2010) proposed an important modification to solve this problem by a backward step during the first dimension reduction. Instead of finding the best fitting 0-space (a point), Huckemann proposed to find the best fitting geodesic in the first step. Afterwards, the mean is found on this geodesic, followed by the forward approach of finding the best fitting geodesics through the mean, orthogonal to each other for all dimension $d > 2$. Neither of the methods are suitable for data which are distributed along small circles. This motivates *principal arc analysis* (PAA) by Jung et al. (2011), a generalization of PCA on S^2 by finding the best fit of any circle (small or great) to the data. This approach is illustrated in Figure 3. The first approach discussed above is a classical forward approach, whereas the last two methods contain a backward view to the data in the first step followed by the classical forward view.

Principal nested sphere (PNS) is a fully backward approach that fits the best lower dimensional subsphere in each dimension starting with S^d (Jung et al., 2012). The subspheres can be great or small and thereby, the method includes variation along small and great circles. PNS is described in more detail in Paper III where it is used extensively in order to produce suitable means of shape populations.

2

RESULTS AND DISCUSSION

All three papers in this thesis are motivated by medical applications. Paper I presents a semiautomatic method for prostate segmentation in radiotherapy treatment planning. Paper II presents a novel framework to analyze the biomechanical gait analysis of the knee joint. Paper III presents a novel approach to analyze anatomical shape differences between hippocampi of first episode schizophrenics and controls. Thus, the work in this thesis has demonstrated that statistical shape analysis is an essential approach in medical applications.

2.1 PAPER I - PROSTATE SEGMENTATION

This paper is focused on the development of an intuitively appealing method for the 3D modeling and segmentation of the prostate in radiotherapy treatment (RT) planning of prostate cancer. Prostate cancer is the second most diagnosed cancer worldwide (Jemal et al., 2011). The delineation of the target volume is a time-consuming task that is often performed manually by a physician, slice by slice in a set of computed tomography (CT) or magnetic resonance (MR) images. The proposed semiautomatic method for prostate delineation reduces the workload of physicians and offers comparable accuracy. The method is based on the statistical shape analysis of a training set of prostates that can be understood as the prior information in a Bayesian framework.

The developed geometric model $\mathcal{M}$ of the prostate follows the object parametrization by geometrical templates as proposed in Dryden and Mardia (1998). The prostate has an ellipsoid-like and smooth shape. This motivates modeling the prostate by slicewise BFEs. The BFEs are found by a least-square minimizer as described in Ahn et al. (2001). Suppose the training set consists of N data sets and the prostate is manually outlined in L_n slices in each data set $n = 1,\dots,N$; the stacked ellipses are parametrized by

$$BFE_{nl} := (\theta^{nl}, \alpha^{nl}, \phi^{nl})^T \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\frac{\pi}{2}, \frac{\pi}{2}], l = 1,\dots,L_n, n = 1,\dots,N \tag{6}$$

with positions θ^{nl}, lengths $\alpha_1^{nl}, \alpha_2^{nl}$ of the first and second principal axes, and rotation angles ϕ^{nl} as depicted in Figure 2a. The number of slices L_n usually varies between the data sets $n = 1\dots,N$. Therefore, either an interpolation of the model parameters or an interpolation of the image intensities is necessary. Because the image information of the prostate is low, an interpolation of the image intensities should be avoided. The implemented method assumes to set

manual control points on the boundary which are used to make the BFEs comparable by the transformation of the parametrized ellipses model to a common position, scale and orientation. Control points are assumed in the first, last and center slice. Moreover, the first and second principal axes are reordered, the rotation parameter ϕ^{nl} is relaxed in case of circularity and smoothing is performed between neighboring slices to avoid large forward and backwards rotations. Thereby, correspondence is introduced in the proposed geometric model in Paper I. The normalized parameters result in a mean shape and shape distribution based on the training set. The mean shape is deformed and registered in a new data set based on the control points.

Several approaches are proposed in the literature to segment the prostate. A collection of methods available for prostate segmentation is reviewed by Ghose et al. (2012). The works of Saroul et al. (2008) and Mahdavi et al. (2011) are related to the parametrization method by BFEs as used in Paper I. In contrast to the work of those authors, the approach of slice-wise BFEs introduces more flexibility into the model. Moreover, the method has similarities to a tubular medial representation (Saboo et al., 2009). The proposed approach is very intuitive for physicians; the registration of the mean shape has resulted in fairly accurate estimates in a set of test data with an average volume overlap of 90% compared to the manual segmentations. Moreover, a time saving effect of 30% could be achieved. Thus, the proposed method can provide the physicians with a good initialization of the prostate contour in RT planning. The model fitting procedure for a training data set and the calculation of a mean shape is easy to understand compared to more sophisticated geometric models such as s-reps (Pizer et al., 2013) which support more accurate statistics.

In addition to the deformation and registration of the obtained mean shape in a new data set, the prior was used in a Bayesian framework in order to estimate the posterior shape model in a new data set using a *Markov chain Monte Carlo* (MCMC) method. Only small improvements in the volume overlap could be achieved with large computational costs. This demonstrated the high performance of the mean shape model based on stacked ellipses.

Future work

The small improvements of the segmentation results using MCMC motivates the implementation of an alternative approach, e.g., an *active shape model* (ASM) or an AAM (Cootes et al., 2001; Davies et al., 2008; Heimann, 2008). ASM and AAM model the image intensity variation in addition to the shape variation derived from a training data set. Both shape and image intensity variation are combined in a single statistical model. The image intensity is sampled near the model edges in ASM whereas all image intensity information which is covered by the target object is used in AAM. Both approaches find the final shape model parameters by a minimization procedure. The implementation of these methods may also reduce manual interaction such as the definition of control points.

Further improvements may be achieved by the incorporation of additional constraint and regularization terms such as the surface curvature. Around the center slice of the prostate, a low surface curvature can be expected and smaller changes of model parameters between neighbor slices. Towards the outer slices of the prostate, larger surface curvature can be expected and also larger changes of model parameters between neighbor slices. Thus, the surface curvature could be used as a weight parameter for the implemented regularization terms.

A larger study would be an important step for the implementation of the proposed method into the clinical praxis. This includes a full integration of the method into the work-flow of physicians during the RT planning.

2.2 PAPER II - ESTIMATION OF ROTATIONAL DEFORMATIONS

Segmentation of 3D objects in medical imaging is one application in statistical shape analysis, another is the study of deformations. The deformations of objects such as organs and anatomical structures can appear in several variants, simple and complex types. An important class is rotational deformations including rigid rotation, bending and twisting of an object. The analysis and understanding of such rotational deformations can be important for an accurate medical treatment and diagnosis. For example, the analysis of biomechanical deformations are crucial for the construction of biomechanical models and the treatment of orthopedic problems (Rivest et al., 2008; Ball and Greiner, 2012). Paper II of this thesis studies the rotational deformations and proposes models and estimators based on the directional features $x \in S^2$ of the deformed objects.

If a K-tuple of $K \geq 2$ direction vectors $\mathbf{x} = (x_1, \ldots, x_K) \in (S^2)^K$ is rotated together about a common axis c, then each of the rotated direction vectors is on a circle $\delta(c, r_j)$ as defined in (5) with common center c but with different radii $r_j = \arccos(c' x_j)$, $j = 1, \ldots, K$. As a results, a collection of concentric circles with a common center c and radii tuple $\mathbf{r} = (r_1, \ldots, r_K) \in [0, \pi/2] \times [0, \pi]^{K-1}$ is obtained by

$$\delta(c, \mathbf{r}) = \{(x_1, \ldots, x_K) \in (S^2)^K : x_j' c = \cos(r_j),\ j = 1, \ldots, K\}. \tag{7}$$

The proposed method in Paper II uses the concentric small circles to analyze the rotational deformations of an object. Based on this idea, a rotation model is proposed by

$$X_j = R(c, \theta_j)\mu_j \oplus \varepsilon_j \quad (j = 1, \ldots, K) \tag{8}$$

for a K-tuple of random direction vectors $\mathbf{X} = (X_1, \ldots, X_K)$, unknown constants $c, \mu_j \in S^2$ and latent random variables $\theta_j \in [-\pi/2, \pi/2)$. The error terms ε_j are assumed to be independently distributed, and the $\oplus$ sign defines a suitable action of the error distribution for observations

on S^2. Estimators are developed for $c, \mathbf{r}, \mu_j$ and the latent variable θ_j. Moreover, this model is extended to hierarchical rotations to model and estimate rotational deformations which consist of two independent deformations, e.g., bending and twisting. The numerical performance of the proposed estimator is demonstrated by a simulation study for different situations. Furthermore, the methods are applied to rotationally deformed ellipsoids represented by s-reps as discussed in Section 1.1. This is of interest because several organs have an ellipsoid-like shape such as the prostate. Finally, an analysis of biomechanical data of the knee joint is presented and demonstrates the quality of the method.

Future work

The method introduced in Paper II is an initial work in the analysis of rotational deformations from directional data and opens several potential work directions. The extension of the proposed models to time series models can be investigated in the future. Time series models such as autoregressive-moving-average models could model more accurate rotational deformations observed over time. Furthermore, an improvement of the prediction of θ_j is of interest. The estimations of rotational deformations from s-reps have indicated that it is necessary to incorporate the surface curvature into the model in order to obtain accurate predictions. The Gaussian curvature, mean curvature and shape operator may be useful descriptors (Gray, 1998; Kühnel, 2006). Other research topics would be the extension of the hierarchical deformations to more than two rotational deformations in addition to the incorporation of locational information to the analysis of directional data.

The next two sections present an alternative approach to estimate rotational deformations and an exploratory tool to discriminate rotational deformations. The methodologies presented here, are suggestions for future research.

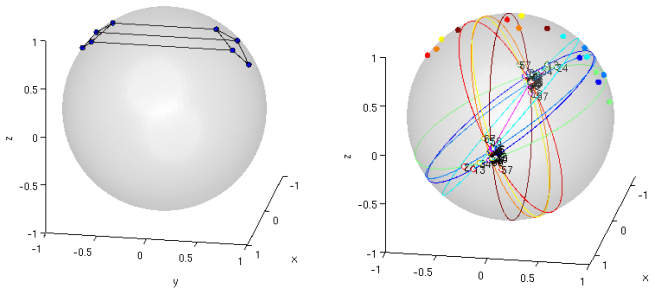
Analysis of rotational deformations by intersecting geodesics

In Paper II, the estimation of a rotation axis $c \in S^2$ is based on the Fréchet mean of the rotation axes of a set of small circles. Instead of using small circles, an alternative estimator could be based on intersecting geodesics (large circles). Suppose, we have K direction vectors $x_{ij} \in S^2$ at time points $i = 1,2$ with $j = 1,\ldots,K$. An estimation of the rotation axis can be obtained from the intersection of geodesics defined by the pairs (v_i, c_i) with

$$v_j = \frac{x_{1j} + x_{2j}}{||x_{1j} + x_{2j}||} \quad \perp \quad c_j = \frac{x_{1j} \times x_{2j}}{||x_{1j} \times x_{2j}||}, \quad j = 1,\ldots,K. \tag{9}$$

Here, v_j is the bisector of the pair (x_{1j}, x_{2j}). Notice, that the intersection points of two geodesics are perpendicular on S^2. Figure 4 visualizes the intersecting geodesics of 8 directions which are

bent and rigidly rotated around the common rotation axis $c = (1,0,0)'$. It can be observed that the intersection points cluster around the true rotation axis.



(a) $i = 1$ (b) $i = 1,2$

Figure 4: Toy example. (a) Visualization of 8 observed direction vectors on S^2. (b) Visualization of the directions before and after the rotational deformation. Each color belongs to a direction $j = 1,\ldots,8$, i.e., each color appears three times (two points and the corresponding geodesic). The intersecting geodesics clustering around the true rotation axis $c = (1,0,0)'$.

This idea is an interesting alternative to the proposed approach in Paper II. However, simulations have shown more accurate and robust results using concentric small circles because in certain situations undesirable side-clusters were observed using intersecting geodesics, e.g., in case of observations close to the underlying rotation axis.

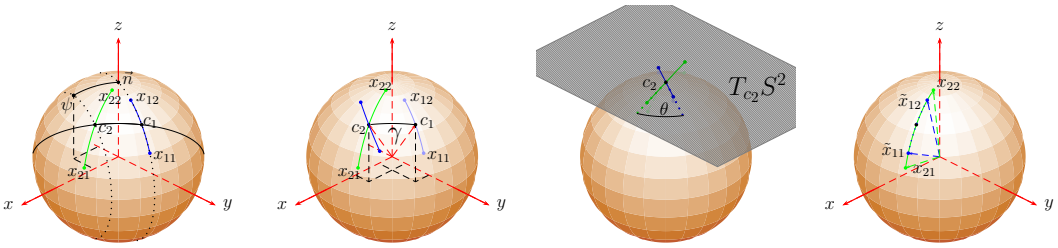
An exploratory discrimination tool for rotational deformations

In the following section, a representation called *rotation twisting folding* (RTF) is proposed. RTF compares pairwise directions between different observations to discriminate rotational deformations in rigid rotation, bending and twisting. The RTF representation is designed to capture rotation, bending and twisting by a set of parameters. Thereby, the representation allows to distinguish between the three different types of rotational motions. In particular, specific patterns in the scatterplots of the RTF parameters can be observed.

Suppose, we observe a tuple of directions before rotation (x_{11}, x_{12}) and after rotation (x_{21}, x_{22}), i.e., directions $x_{ij} \in S^2$ with $j = 1,2$ number of directions at time points $i = 1,2$. The RTF representation of (x_{i1}, x_{i2}) is defined by a triple $(\psi, \theta, \tau) \in S^2 \times S^1 \times \mathbb{R}^+$ where ψ indicates the amount of rigid rotation, θ indicates the amount of twisting and τ the amount of bending as visualized in Figure 5 and explained in more detail in the following.

The bisector between the tuple (x_{i1}, x_{i2}) is given by

$$c_i = \frac{x_{i1} + x_{i2}}{||x_{i1} + x_{i2}||}, \quad i = 1,2.$$



(a) Definition of ψ (b) Definition of γ (c) Definition of θ (d) Definition of τ

Figure 5: Parameter definition in the RTF representation. (a) Definition of $\psi \in S^2$ which defines the rotation direction and rotation angle of bisector c_1 to c_2. (b) Location of the tuple (x_{11}, x_{12}) after rotation corresponding to ψ with rotation angle $\gamma = \arccos(\psi_3)$. (c) Definition of θ by the signed cosinus between the vectors $\mathrm{Log}_{c_2}(\widetilde{R}(c_1,c_2)x_{11})$ and $\mathrm{Log}_{c_2}(x_{21})$ in the tangential plane $T_{c_2}S^2$. (d) Definition of τ by the ratio of the geodesic distances of the tuples (x_{11}, x_{12}) and (x_{21}, x_{22}). Here, $(\tilde{x}_{11}, \tilde{x}_{12})$ is the tuple (x_{11}, x_{12}) after transformation by ψ and θ.

For convenience, denote by $\widetilde{R}(a,b) \in SO(3)$ a rotation matrix that moves $a \in S^2$ to $b \in S^2$, and $\vec{n} = (0,0,1)$ defines the north pole on S^2. Thereby, the direction vector $\psi = (\psi_x, \psi_y, \psi_z) \in S^2$ is defined by

$$\psi := \widetilde{R}(c_1, \vec{n})c_2 \tag{10}$$

which rotates the bisector c_1 to c_2 with rotation angle $\gamma = \arccos(\psi_z)$. The direction $\psi \in S^2$ indicates the amount of rigid rotation. The ratio of the two pairs of directional vectors is defined by

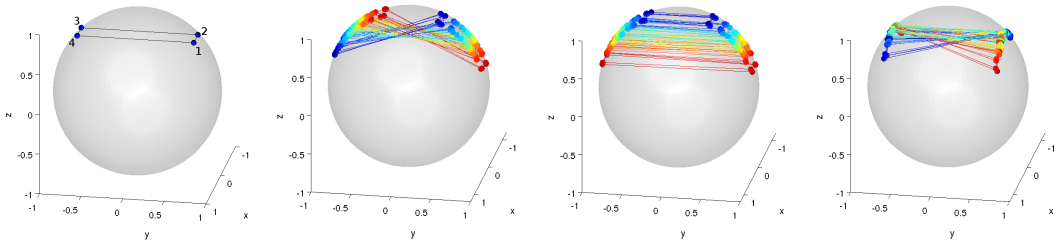
$$\tau := \frac{d_g(x_{21}, x_{22})}{d_g(x_{11}, x_{12})} \tag{11}$$

with the geodesic distance d_g as defined in (1). The ratio τ indicates the amount of bending between the observations with $\tau = 1$ in case of no bending. Finally, x_{11} is mapped to $\kappa_1 \in \mathbb{R}^2$ by $\kappa_1 = \mathrm{Log}_{c_2}(\widetilde{R}(c_1,c_2)x_{11})$; and x_{21} is mapped to $\kappa_2 \in \mathbb{R}^2$ by $\kappa_2 = \mathrm{Log}_{c_2}(x_{21})$. The vectors κ_1 and κ_2 are elements of the tangent space $T_{c_2}S^2$. Thereby, the parameter θ is defined by

$$\theta := sign(\kappa_1 \times \kappa_2) \arccos\left(\frac{\kappa_1' \kappa_2}{||\kappa_1|| ||\kappa_2||}\right), \tag{12}$$

where $sign(\kappa_1 \times \kappa_2)$ is the sign of the cross product of κ_1 and κ_2. The parameter θ indicates the amount of twisting with $\theta = 0$ in case of no twisting.

The decomposition by the RTF representation can be understood by considering a simple example, see Figure 6. Suppose, an object is modeled by four surface points with attached directions $x_{0j}, j = 1, \ldots, 4$. Moreover, consider 30 random observations after rigid rotation, bending



(a) Direction vectors (b) Rigid rotation (c) Bending (d) Twisting

Figure 6: Toy example for the RTF representation. (a) Each direction vector x_{0j} is a point on S^2, $j = 1,\ldots,4$. (b-d) 30 realizations of random rotational deformations. Different colors represent different observations. (b) Rigid rotation about $(1,0,0)'$. (c) Bending reflected by a rotation of the directions on the left and right side of the sphere about the common axis $(1,0,0)'$ by common angle but in opposite directions. (d) Twisting reflected by a rotation of directions on the left and right side about the common axis $(0,1,0)'$ by common angle but in opposite directions.

and twisting of the object as described in Figure 6 where each rotational deformation was generated by

$$X_{ij} = R(\omega,\phi)x_{0j} \oplus \varepsilon \quad \text{or} \quad X_{ij} = R(\omega,-\phi)x_{0j} \oplus \varepsilon$$

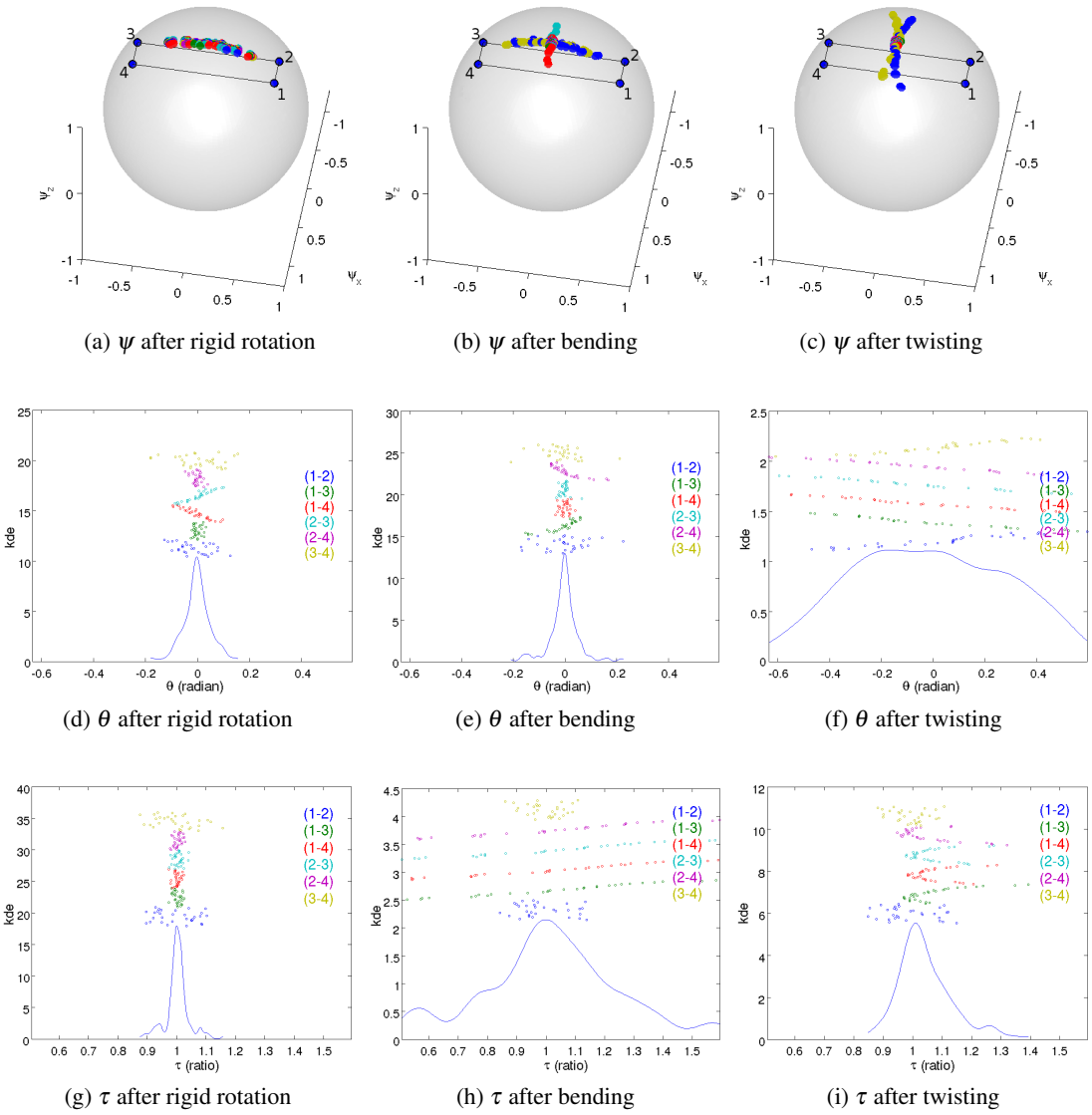
with $i = 1,\ldots,30$, $j = 1,\ldots,4$, $\phi \sim N(0,\eta^2)$ and a rotation axis $\omega \in S^2$. The RTF parametrization is calculated between all pairwise combinations $(x_{0k_1}, x_{0k_2}, X_{ik_1}, X_{ik_2})$ with

$$\{k_1,k_2\} \in \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\},$$

represented by the tuples (X_{ik_1}, X_{ik_2}).

The RTF representation for each rotational deformation is visualized in Figure 7. Observations obtained by rigid rotation (Figure 6b) result in a small spread of the parameters θ and τ (Figures 7d and 7g) but in a wide spread of ψ around the underlying rotation axis (Figure 7a). Observations obtained by a bending deformation (Figure 6c) result in a small spread of the parameter θ but in a wide spread of τ (Figures 7e and 7h). The parameter ψ indicates a rigid rotation for the tuples (X_{i1}, X_{i2}) and (X_{i3}, X_{i4}) about $(1,0,0)'$, no rotation for the tuples (X_{i1}, X_{i4}) and (X_{i2}, X_{i3}), and a minor rotation for the tuples (X_{i1}, X_{i3}) and (X_{i2}, X_{i4}) about $(0,0,1)'$ (Figure 7b). Observations obtained by twisting (Figure 6d) result in a small spread of the parameter τ but in a wide spread of θ (Figures 7f and 7i). The parameter ψ indicates only rigid rotation for the tuples (X_{i1}, X_{i2}) and (X_{i3}, X_{i4}) (Figure 7c).

The presented results reveal the potential of the RTF representation as an exploratory tool for the discrimination analysis of rotational deformations from directional data. However, further simulations and research is required on the basis of more complex objects and real data examples.



(a) ψ after rigid rotation (b) ψ after bending (c) ψ after twisting

(d) θ after rigid rotation (e) θ after bending (f) θ after twisting

(g) τ after rigid rotation (h) τ after bending (i) τ after twisting

Figure 7: RTF representation of rigid rotated, bent and twisted directions. (a-c) Parameter ψ, (d-f) parameter θ (g-i) parameter τ of the RTF representation. In each plot 180 realizations (points) are depicted of the corresponding RTF parameter. Different colors represent different direction tuples, e.g., (1-2) for the tuple (X_{i1}, X_{i2}) with $i = 1, \ldots, 30$.

The proposed method may be useful for a classification of rotational deformations into rigid rotation, bending and twisting. Such a classification is left for future research.

2.3 PAPER III - NONLINEAR HYPOTHESIS TESTING

Another application of statistical shape analysis is presented in Paper III by exploring shape differences between hippocampi of first episode schizophrenics and controls. The study of mor-

phological changes of human organs and body structures is of general interest in medical analysis, e.g., the understanding of neuroanatomical structures of the human brain (Gerig et al., 2001; Styner et al., 2004; Ferrarini et al., 2006).

Paper III proposes a novel method to test mean differences of populations on data whose representations include both Euclidean and non-Euclidean elements. Several aspects of a sensitive hypothesis test are elaborated such as important geometrical model properties which support accurate statistics of populations. The s-rep model (Pizer et al., 2013) is presented as a suitable model that includes Euclidean components and components which live on spheres. A statistical method based on PNS (Jung et al., 2012) is proposed to calculate means and to analyze the Euclidean and non-Euclidean components of the models. Finally, a permutation test is presented based on Pesarin (2001); Terriberry et al. (2005), in order to test for mean differences. A permutation test is a non-parametric approach that uses the data to estimate the sampling distribution of the test statistic under the null-hypothesis of equal distributions. Because of the assumption that populations have equal distributions, it is valid to permute the data between the populations without affecting the distribution of the test statistic. The test statistic is defined by a difference measure. An appropriate difference measure is presented for the analysis of s-reps. The difference measure quantify the mean difference of geometrical object properties (GOP) which are single or a collection of s-rep features. A global and local test based on the GOP differences are proposed taking into account the problem of multiple comparison correction.

The analysis of first episode schizophrenics compared to controls demonstrated the power of the permutation test, both globally as well as locally. Several statistically significant findings could be made. The global test established a significant shape difference of the hippocampi between schizophrenics and controls. The local test showed a significant loss of hippocampal volume for schizophrenics. The significant volume difference was observed in the x and y-directions but not in the z-direction for the aligned hippocampi. Moreover, several spoke directions were found as statistically significant by the local test. These findings may be of high interest in neuroscience. In addition, Paper III is a novel study that analyzes morphological differences using non-Euclidean data components such as directional information using s-reps. Moreover, a bias for the test results was observed depending on the shape distribution used for the model fitting. The choice of an appropriate shape distribution of a population is discussed in more detail in the paper. Furthermore, the impact of a chosen data pre-processing was discovered.

Future work

Several open questions have been raised during the implementation of the proposed nonlinear hypothesis test. A possible extension of the study is to investigate hippocampal changes of schizophrenics and controls in longitudinal data. Studies such as Narr et al. (2004); Mamah et al. (2012); McClure et al. (2013) have indicated the need for a more sensitive hypothesis test

in order to examine shape differences inside a treatment group over time. A detailed power study on the basis of simulated data and investigation of alternative combining functions for the global test are also left open. Moreover, potential improvements can be made in the data modeling. This includes the incorporation of image intensities in addition to morphological features in order to test for population differences. Furthermore, the fitting procedure of s-reps to the data could be improved in the future, e.g., by an automatic parameter choice for the fitting procedure and an adaptive grid size choice of the skeletal sheet.

BIBLIOGRAPHY

Ahn, S. J., Rauh, W., and Warnecke, H.-J. (2001), "Least-Squares Orthogonal Distances Fitting of Circle, Sphere, Ellipse, Hyperbola, and Parabola," *Pattern Recognition*, 34, 2283–2303.

Albertson, R. C., Streelman, J. T., and Kocher, T. D. (2003), "Genetic Basis of Adaptive Shape Differences in the Cichlid Head," *Journal of Heredity*, 94, 291–301.

Ball, K. A. and Greiner, T. M. (2012), "A Procedure to Refine Joint Kinematic Assessments: Functional Alignment," *Computer Methods in Biomechanics and Biomedical Engineering*, 15, 487–500.

Blum, H. and Nagel, R. (1978), "Shape Description Using Weighted Symmetric Axis Features," *Pattern Recognition*, 10, 167–180.

Bookstein, F. L. (1986), "Size and Shape Spaces for Landmark Data in Two Dimensions," *Statistical Science*, 1, 181–242.

Bronstein, A. M., Bronstein, M. M., and Kimmel, P. (2008), *Numerical Geometry of non-Rigid Shapes*, Heidelberg, Germany: Springer.

Cates, J., Fletcher, P., Styner, M., Shenton, M., and Whitaker, R. (2007), "Shape Modeling and Analysis with Entropy-Based Particle Systems," in *Information Processing in Medical Imaging*, vol. 4584 of *Lecture Notes in Computer Science*, pp. 333–345.

Cootes, T. F., Edwards, G., and Taylor, C. (2001), "Active Appearance Models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23, 681–685.

Cootes, T. F., Taylor, C., Cooper, D., and Graham, J. (1992), "Training Models of Shape from Sets of Examples," in *Proc. British Machine Vision Conference*, eds. Hogg, D. and Boyle, R., Berlin. Springer-Verlag, pp. 9–18.

Damon, J. and Marron, J. S. (2013), "Backwards Principal Component Analysis and Principal Nested Relations," *Journal of Mathematical Imaging and Vision*, dOI: 10.1007/s10851-013-0463-2.

Davies, R., Twining, C., and Taylor, C. (2008), *Statistical Models of Shape - Optimisation and Evaluation*, London, UK: Springer.

Dryden, I. L. and Mardia, K. V. (1998), *Statistical Shape Analysis*, Chichester, UK: Wiley.

Ferrarini, L., Palm, W. M., Olofsen, H., van Buchem, M. A., Reiber, J. H., Admiraal-Behloul, F., et al. (2006), "Shape Differences of the Brain Ventricles in Alzheimer's Disease," *NeuroImage*, 32, 1060–1069.

Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. (2004), "Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape," *IEEE Transactions on Medical Imaging*, 23, 995–1005.

Fréchet, M. (1948), "Les Éléments Aléatoires de Nature Quelconque dans un Espace Distancié," *Annales de l'Institut Henri Poincaré*, 10, 215–310.

Gerig, G., Styner, M., Shenton, M. E., and Lieberman, J. A. (2001), "Shape Versus Size: Improved Understanding of the Morphology of Brain Structures," *Medical Image Computing and Computer-Assisted Intervention*, 24–32.

Ghose, S., Oliver, A., Martá, R., Lladó, X., Vilanova, J. C., Freixenet, J., Mitra, J., Sidibé, D., and Meriaudeau, F. (2012), "A Survey of Prostate Segmentation Methodologies in Ultrasound, Magnetic Resonance and Computed Tomography Images," *Computer Methods and Programs in Biomedicine*, 108, 262–287.

Gray, A. (1998), *Modern Differential Geometry of Curves and Surfaces with Mathematica*, Boca Raton, USA: CRC Press, 2nd ed.

Heimann, T. (2008), *Statistical shape models for 3D medical image segmentation*, Saarbrücken, Germany: VDM Verlag Dr. Müller.

Huckemann, S., Hotz, T., and Munk, A. (2010), "Intrinsic Shape Analysis: Geodesic PCA for Riemannian Manifolds modulo Isometric Lie Group Actions," *Statistica Sinica*, 20, 1–58.

Hufnagel, H., Ehrhardt, J., Pennec, X., Ayache, N., and Handels, H. (2009), "Computing of Probabilistic Statistical Shape Models of Organs Optimizing a Global Criterion," *Methods of Information in Medicine*, 48, 314–319.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011), "Global Cancer Statistics," *CA: A Cancer Journal for Clinicians*, 61, 69–90.

Joshi, S., Pizer, S. M., Fletcher, P., Yushkevich, P., Thall, A., and Marron, J. S. (2002), "Multiscale Deformable Model Segmentation and Statistical Shape Analysis Using Medial Descriptions," *IEEE Transactions on Medical Imaging*, 21, 538–550.

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568.

Jung, S., Foskey, M., and Marron, J. S. (2011), "Principal Arc Analysis on Direct Product Manifolds," *Annals of Applied Statistics*, 5, 578–603.

Karcher, H. (1977), "Riemannian Center of Mass and Mollifier Smoothing," *Communications on Pure and Applied Mathematics*, 30, 509–541.

Kendall, D. G. (1984), "Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces," *Bulletin of the London Mathematical Society*, 16, 81–121.

Kühnel, W. (2006), *Differential Geometry*, vol. 16 of *Student Mathematical Library*, Wiesbaden, Germany: Friedr. Vieweg & Sohn Verlag, 2nd ed.

Kurtek, S., Ding, Z., Klassen, E., and Srivastava, A. (2011), "Parameterization-Invariant Shape Statistics and Probabilistic Classification of Anatomical Surfaces," in *Information Processing in Medical Imaging*, vol. 22, pp. 147–158.

Leibe, B., Leonardis, A., and Schiele, B. (2006), "An Implicit Shape Model for Combined Object Categorization and Segmentation," in *Toward Category-Level Object Recognition*, Springer Berlin Heidelberg, Lecture Notes in Computer Science, pp. 508–524.

Mahdavi, S. S., Chng, N., Spadinger, I., Morris, W. J., and Salcudean, S. E. (2011), "Semi-Automatic Segmentation for Prostate Interventions," *Medical Image Analysis*, 15, 226–237.

Mamah, D., Harms, M. P., Barch, D. M., Styner, M. A., Lieberman, J., and Wang, L. (2012), "Hippocampal Shape and Volume Changes with Antipsychotics in Early Stage Psychotic Illness," *Frontiers in Psychiatry*, 3, 1–10.

McClure, R. K., Styner, M., Maltbie, E., Liebermann, J. A., Gouttard, S., Gerig, G., Shi, X., Zhu, H., et al. (2013), "Localized Differences in Caudate and Hippocampal Shape are Associated with Schizophrenia but not Antipsychotic Type," *Psychiatry Research: Neuroimaging*, 211, 1–10.

Narr, K. L., Thompson, P. M., Szeszko, P., Robinson, D., Jang, S., Woods, R. P., Kim, S., Hayashi, K. M., Asunction, D., Toga, A. W., and Bilder, R. M. (2004), "Regional Specificity of Hippocampal Volume Reductions in First-Episode Schizophrenia," *NeuroImage*, 21, 1563–1575.

Oualkacha, K. and Rivest, L.-P. (2012), "On the Estimation of an Average Rigid Body Motion," *Biometrika*, 99, 585–598.

Pennec, X. (2008), "Statistical Computing on Manifolds: from Riemannian Geometry to Computational Anatomy," *Emerging Trends in Visual Computing*, 5416, 347–386.

Pesarin, F. (2001), *Multivariate Permutation Tests with Applications to Biostatistics*, Chichester, UK: John Wiley & Sons.

Pizer, S. M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013), "Nested Sphere Statistics of Skeletal Models," in *Innovations for Shape Analysis: Models and Algorithms*, Springer Lecture Notes in Computer Science., pp. 93–115.

Rivest, L.-P., Baillargeon, S., and Pierrynowski, M. (2008), "A Directional Model for the Estimation of the Rotation Axes of the Ankle Joint," *Journal of the American Statistical Association*, 103, 1060–1069.

Saboo, R., Levy, J., Chaney, E., and Pizer, S. M. (2009), "Medial Models of Populations of Nearly Tubular Objects," in *Proceedings of the MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, pp. 232–243.

Saroul, L., Bernard, O., Vray, D., and Friboulet, D. (2008), "Prostate Segmentation in Echographic Images: A Variational Approach Using Deformable Super-Ellipse and Rayleigh Distribution," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008*, pp. 129–132.

Siddiqi, K. and Pizer, S. (2008), *Medial Representations: Mathematics, Algorithms and Applications*, Computational Imaging and Vision, Vol. 37, Dordrecht, Netherlands: Springer, 1st ed.

Styner, M., Lieberman, J., Pantazis, D., and Gerig, G. (2004), "Boundary and Medial Shape Analysis of the Hippocampus in Schizophrenia," *Medical Image Analysis*, 8, 197–203.

Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Levitt, J., Shenton, M., and Gerig, G. (2006), "Framework for the Statistical Shape Analysis of Brain Structures Using SPHARM-PDM," *The Insight Journal*, 242–250.

Terriberry, T., Joshi, S., and Gerig, G. (2005), "Hypothesis Testing with Nonlinear Shape Models," in *Information Processing in Medical Imaging*, eds. Christensen, G. and Sonka, M., Springer Berlin Heidelberg, vol. 3565 of *Lecture Notes in Computer Science*, pp. 15–26.

Zhang, S., Zhan, Y., Cui, X., Gao, M., Huang, J., and Metaxas, D. (2013), "3D Anatomical Shape Atlas Construction Using Mesh Quality Preserved Deformable Models," *Computer Vision and Image Understanding*, 117, 1061–1071.

Part II

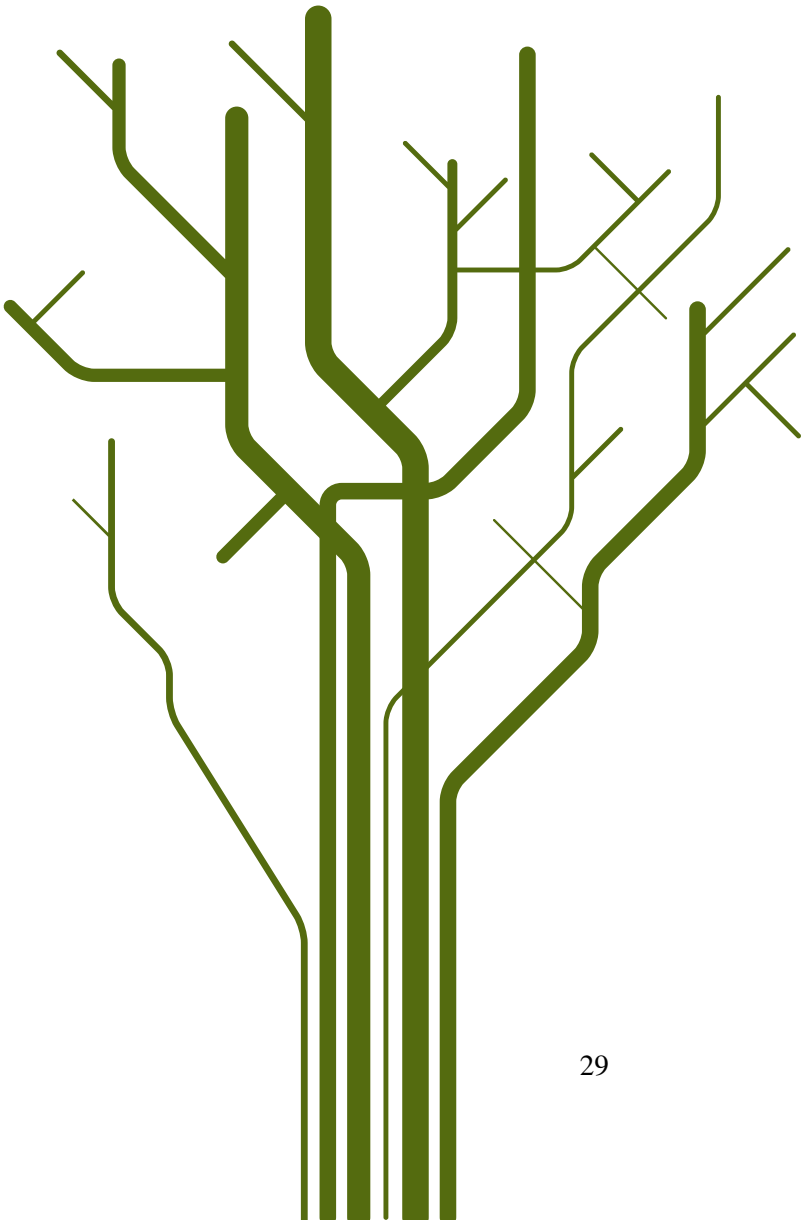
PAPERS

Paper I

A semiautomatic tool for prostate segmentation in radiotherapy treatment planning

Schulz, Jörn, Skrøvseth, Stein Olav, Tømmerås, Veronika Kristine, Marienhagen, Kirsten and Godtliebsen, Fred

submitted to *BMC Medical Imaging*



A semiautomatic tool for prostate segmentation in radiotherapy treatment planning

Jörn Schulz*[1], Stein Olav Skrøvseth[2] , Veronika Kristine Tømmerås[3] , Kirsten Marienhagen[3] and Fred Godtliebsen[1]

[1]Department of Mathematics and Statistics, University of Tromsø, 9037 Tromsø, Norway
[2]Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, 9038 Tromsø, Norway
[3]Department of Oncology, University Hospital of North Norway, 9038 Tromsø, Norway

Email: Jörn Schulz*- jorn.schulz@uit.no; Stein Olav Skrøvseth - stein.olav.skrovseth@telemed.no; Veronika Kristine Tømmerås - veronika.kristine.tommeras@unn.no; Kirsten Marienhagen - kirsten.marienhagen@unn.no; Fred Godtliebsen - fred.godtliebsen@uit.no;

*Corresponding author

Abstract

Background: Delineation of the target volume is a time-consuming task in radiotherapy treatment planning, yet essential for a successful treatment of cancers such as prostate cancer. To facilitate the delineation procedure, the paper proposes an intuitive approach for 3D modeling of the prostate by slice-wise best fitting ellipses.

Methods: The proposed estimate is initialized by the definition of a few control points in a new patient. The method is not restricted to particular image modalities but assumes a smooth shape with elliptic cross sections of the object. A training data set of 23 patients was used to calculate a prior shape model. The mean shape model was evaluated based on the manual contour of 10 test patients. The patient records of training and test data are based on axial T1-weigthed 3D fast-field echo (FFE) sequences. The manual contours were considered as the reference model. Volume overlap (Vo), accuracy (Ac) (both ratio, range 0-1, optimal value 1) and Hausdorff distance (HD) (mm, optimal value 0) were calculated as evaluation parameters.

Results: The median and median absolute deviation (MAD) between manual delineation and deformed Mean Best Fitting Ellipses (MBFE) was Vo (0.9 ± 0.02), Ac (0.81 ± 0.03) and HD (4.05 ± 1.3)mm and between manual delineation and Best Fitting Ellipses (BFE) was Vo (0.96 ± 0.01), Ac (0.92 ± 0.01) and HD (1.6 ± 0.27)mm. Supplementary Materials are available and show a moderate improvement of the MBFE results after Monte Carlo Markov Chain (MCMC) method.

Conclusions: The results emphasize the potential of the proposed method of modeling the prostate by best fitting ellipses. It shows the robustness and reproducibility of the model. A small sample test on 8 patients suggest possible time saving using the model.

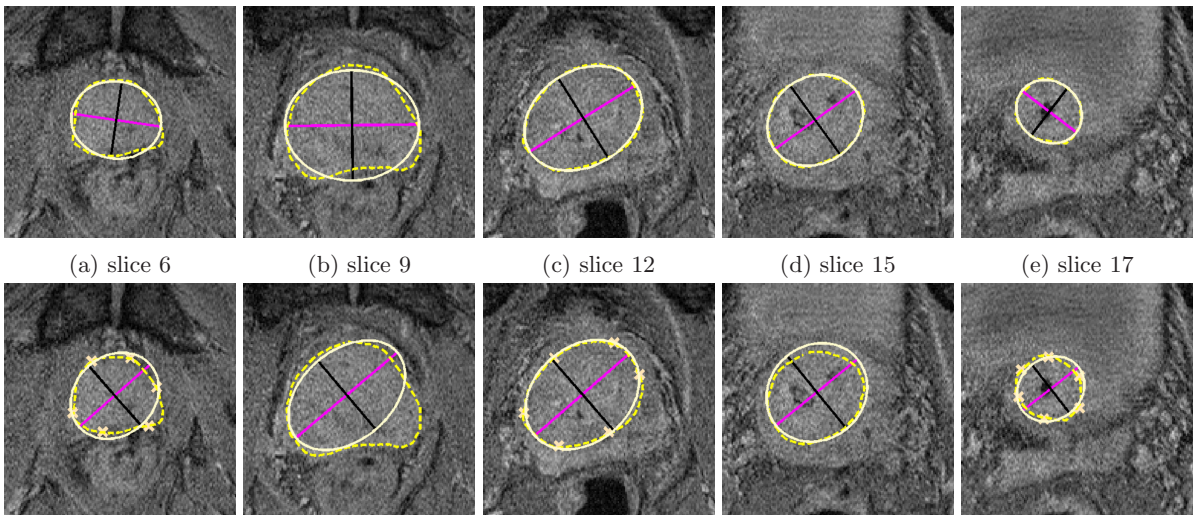
Keywords: Delineation, Ellipse model, Empirical Bayes, Prostate, Radiotherapy treatment planning, Statistical shape analysis

Background

Prostate cancer is the second most diagnosed cancer accounting for 14 percent of all cancers diagnosed worldwide [1]. It is most common in males over the age of 50, and has the highest incidence rate in the developed countries. Aggressive tumors are usually treated with extern radiotherapy or brachytherapy which requires a precise treatment plan for the target volume. In any type of radiotherapy treatment, radiation of healthy tissue should be minimized while maintaining the desired dose to the target volume. Therefore, a successful treatment of prostate cancer relies on an accurate segmentation of the prostate from the surrounding tissue, by image-based description of the shape and location of the target volume. The volume of interest is characterized by a smooth shape, and for this reason an algorithmic description of the volume is feasible.

Transrectal ultrasound (TRUS), magnetic resonance (MR) and computed tomography (CT) images are the three main imaging techniques used in diagnosis, treatment planning and follow-up examination of prostate cancer. Smith et al. [2] investigated the effects of these imaging techniques on the properties of the prostate volume. A collection of methods available for prostate segmentation is reviewed by Ghose et al. [3]. In addition to the methods presented by Ghose et al., alternative approaches are available in the literature, such as the medial or skeleton representation of the prostate [4–8]. The present work proposes a segmentation method which falls into the category of deformable meshes in Ghose et al. [3], but refers to the term geometrical parametrization as described in Dryden and Mardia [9]. The main focus of this paper is the development of a statistical shape model for the prostate. An overview about this type of models in 3D medical image segmentation is presented for example by Davies et al. [10] and Heimann and Meinzer [11].

The works of Saroul et al. [12] and Mahdavi et al. [13] are related to the stacked ellipses parametrization



(a) slice 6 (b) slice 9 (c) slice 12 (d) slice 15 (e) slice 17

(f) slice 6 (g) slice 9 (h) slice 12 (i) slice 15 (j) slice 17

Figure 1: Selected slices of MR data set 3 from the test data set. (a)-(e) Manual delineation of the prostate (dashed line) and best fitting ellipse (solid line). (f)-(j) Manual delineation of the prostate (dashed line), deformed mean shape (solid line) and defined control points in the first, center and last slice.

method used in this paper. In contrast to them, we focus on slice-wise best fitting ellipses which will introduce more flexibility into the model. The approach of slice-wise best fitting ellipses has similarities to a tubular medial representation [14].

Beside the single segmentation of the prostate, several attempts have been tried out for a joint segmentation of neighbor organ and structure to gain improved segmentation results [15–17].

To our knowledge, despite the substantial effort in this area, no widely implemented algorithm exists. In oncology departments this means that the physician has to delineate the prostate slice by slice. This is time-consuming and inefficient. We propose a less ambitious approach in that we use a method that gives a useful starting point for the physician after the definition of few control points. Given the initial estimate of the volume of interest, the physician can adjust the estimate according to their evaluation of the image rather than starting from scratch. By this approach, we obtain the same accuracy with less effort. The main points in our approach are as follows: First, we accept that the algorithm cannot give a fully precise description of the volume. Our main aim is therefore to give a good estimate which can be used as a starting point for the physician. Second, we use a simple ellipse model that is easy to interpret and understand. Our hypothesis is that a more efficient use of physicians in Radiotherapy Treatment Planning (RTP) of patients with prostate cancer can be obtained by an easy-to-interpret semiautomatic tool.

Figure 1 shows an example of the initial estimate we typically obtain for a single image slice. The dashed

line in (a) to (e) describes the manual contour while the solid line shows the best fitting ellipse including the the two principal axes for the observed data of this slice. Note that the fitted model is very much in agreement with the manual line, indicating that the stacked ellipses model gives a good description of the object of interest. The solid lines in (f)-(j) shows the outcome from our model in this situation together with few defined control points. This result shows a typical performance of the method, and that the estimate is close to the best fit we can obtain with the ellipse model. The full processing demands little computational resources, such that the suggested delineation can be presented immediately. The example is discussed further in the Methods and Results and discussion section.

The rest of the paper is organized as follows. In the Method section, we introduce the data sources and the proposed stacked ellipses model, and discuss the shape space and statistics along with constraints and parameters. Results are presented in the Results and discussion section using a test data set to show the potential of the mean shape model, followed by a Conclusion section. Supplementary Materials with further detailed discussion are available online.

Methods

Preliminaries

Each prostate must be described by a shape model in order to calculate statistics, e.g., by stacked ellipses as a parametric shape model. The parameters of a parametric shape model can be estimated from a training set. The training set models also the geometric variability of anatomical structures by a shape probability distribution. The training set contains volume and contour information of segmented prostates from N patients. The volume information describes the image modalities (e.g., CT or MR) and the contour information the volume of interest as defined in the following.

The volume information of each training set $n = 1, \ldots, N$ is defined by a 3-dimensional matrix V_n where $V_n(i, h)$ contains the observed gray level in voxel (i, h), $i = (i_1, i_2) \in \{1, \ldots, I_1\} \times \{1, \ldots, I_2\}$ are the pixel indices in a slice, where typically $I_1 = I_2$, and $h \in \{1, \ldots, H\}$ is the number of slices per data set. The number of slices H is not necessarily the same for all patients in the training data sets. Therefore, we indicate H by H_n and in the same manner I_1 by I_{n1} and I_2 by I_{n2}, but for simplicity we use H, I_1 and I_2 if the meaning is clear.

In addition to the volume information, each training set $n = 1, \ldots, N$ consists of contour information of the prostate, manually drawn by a physician. The contour information can be modeled by a $(M \times K_n)$ configuration matrix $X_n := (X_{n1}, \ldots, X_{nK_n})$ with $X_{nk} = (x^n_{1k}, x^n_{2k}, x^n_{3k})^T \in \mathbb{R}^3, k = 1, \ldots, K_n$, where K_n

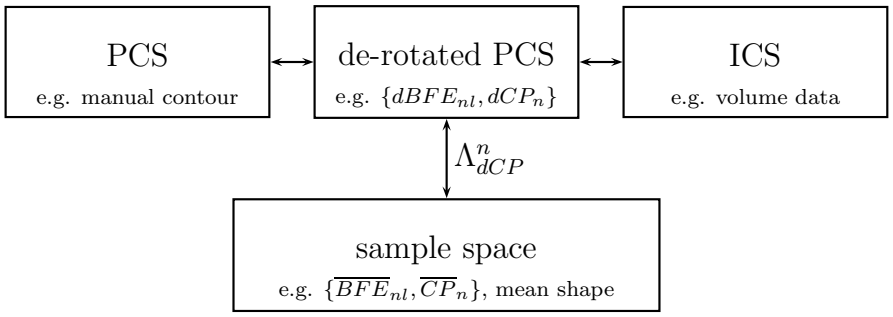


Figure 2: Different coordinate systems with example data. *PCS:* Patient based coordinate system (manual delineation line). *ICS:* Image coordinate system (volume data). *de-rotated PCS:* de-rotated patient based coordinate system with same scale and origin as PCS but same orientation as ICS (de-rotated best fitting ellipse $dBFE_{nl}$, de-rotated control points dCP_n). *sample space:* The transformation matrix Λ^n_{dCP} maps the de-rotated data $\{dBFE_n, dCP_n\}$ to $\{\overline{BFE}_n, \overline{CP}_n\}$ in the sample space.

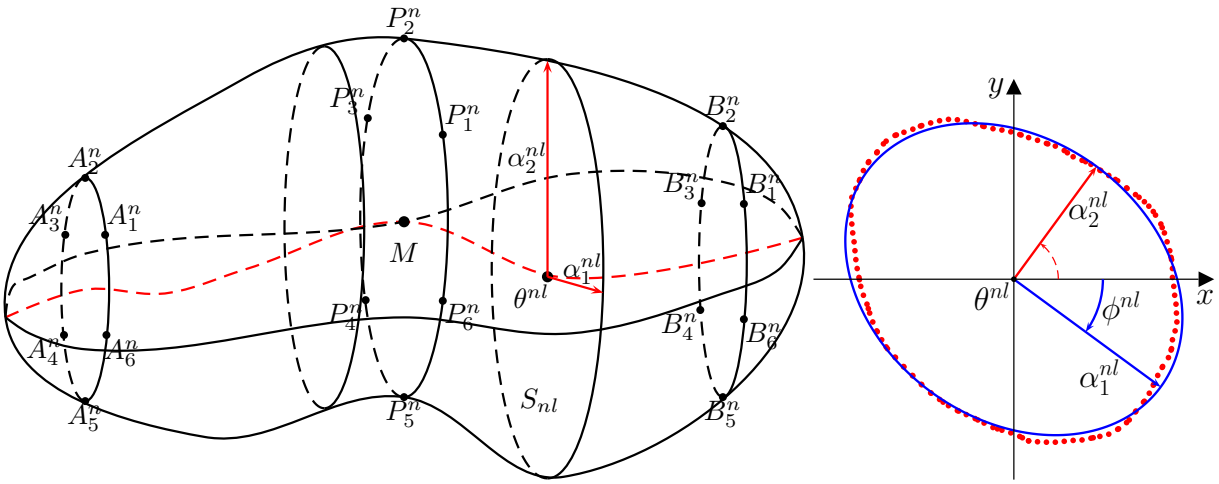
defines the total number of available contour information points in a data set and $M = 3$ defines the dimension. We assume the contour information for an object is defined in a sequentially sorted number L_n of equidistant slices whereas each contour slice contains $\tilde{K}_{nl}$ contour points, $l = 1, \ldots, L_n$. Hence it follows $K_n = \sum_l \tilde{K}_{nl}$ and $X_n = (\tilde{X}_{n1}, \ldots, \tilde{X}_{nL_n})$. The image information in slice l is denoted by S_{nl} and $S_n = \{S_{n1}, \ldots, S_{nL_n}\} \subseteq V_n$ and $\tilde{X}_{nl}$ defines the configuration matrix in slice S_{nl}.

In summary, the training population is given by the set $\{\mathbb{V}, \mathbb{X}\}$, with a set of volume information $\mathbb{V} = \{V_1, \ldots, V_N\}$ and configuration matrices $\mathbb{X} = \{X_1, \ldots, X_N\}$. We assume X_n defines the configuration matrix for the corresponding data set V_n and matches the volume information V_n exactly.

The contour information is often defined in a Patient based Coordinate System (PCS) whereas the volume information is defined in an Image based Coordinate System (ICS). The ICS can be transformed to PCS by a transformation matrix Λ_{DCM}, which transform an image coordinate $p^{im} = (i_1, i_2, h)^T$ to patient coordinate $p^p = (x, y, z)^T$. The definition of Λ_{DCM} and the relation between PCS and ICS (see Figure 2) is discussed in detail in the Supplementary Material. In addition, we introduce a de-rotated PCS where volume and contour information are aligned to each other.

Modeling

The prior information inferred from the training set is incorporated into a shape model. We assume a stacked ellipse model as a shape prior for the prostate. Specifically, the prostate outline in slice $S_{nl}, l = 1, \ldots, L_n$, $n = 1, \ldots, N$ is modeled by a slicewise best-fitting ellipse, as visualized in Figure 3. An ellipse in slice S_{nl} can be uniquely described by $\rho_{nl} = (\theta^{nl}, \alpha^{nl}, \phi^{nl})^T \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\frac{\pi}{2}, \frac{\pi}{2}]$ with



(a) Stacked ellipse model in 3D (b) Ellipse model in 2D

Figure 3: An illustration of the prostate model by slicewise best-fitting ellipses.

- **position** $\theta^{nl} = (\theta_1^{nl}, \theta_2^{nl})^T \in \mathbb{R}^2$ defines the center in slice S_{nl},
- **length of principle axes** $\alpha^{nl} = (\alpha_1^{nl}, \alpha_2^{nl})^T \in \mathbb{R}_+^2$ and
- **rotation angle** $\phi^{nl} \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ in slice S_{nl}.

The rotation parameter ϕ^{nl} is defined corresponding to the ICS with origin θ^{nl} in slice S_{nl}. The boundary of an ellipse ρ_{nl} centered at $\theta^{nl} \in \mathbb{R}^2$ in slice S_{nl} is defined by

$$C(\rho_{nl}) = \left\{ Rx + \theta^{nl} : \frac{x_1^2}{(\alpha_1^{nl})^2} + \frac{x_2^2}{(\alpha_2^{nl})^2} = 1, x \in \mathbb{R}^2 \right\} \tag{1}$$

$$\text{and } R = \begin{pmatrix} \cos\phi^{nl} & -\sin\phi^{nl} \\ \sin\phi^{nl} & \cos\phi^{nl} \end{pmatrix}$$

is a rotation matrix in $\mathbb{R}^2$ with rotation angle ϕ^{nl} and $x = (x_1, x_2)^T$.

The shape model described in this section requires the best fit of an ellipse $C(\rho_{nl})$ to the contour information $\tilde{X}_{nl}$ in each slice, i.e., we model $\tilde{X}_{nl} = C(\rho_{nl}) + \epsilon$ where ϵ is an error with mean zero. The best-fitting ellipses provide us with a slice-by-slice parametrization of the prostate for all slices in each training shape.

The problem of fitting an ellipse to geometric features like the contour is discussed widely in the literature (e.g., [18, 19]). This work follows Ahn et al. [18], who proposed a least-square minimizer for $\tilde{X}_{nl}$. The nonlinear estimate of parameters $\rho_{nl} = (\theta_1^{nl}, \theta_2^{nl}, \alpha_1^{nl}, \alpha_2^{nl}, \phi^{nl})^T$ given $\tilde{X}_{nl}$ must minimize the error

$$g(\hat{\rho}_{nl}) = \left(\tilde{X}_{nl} - \tilde{C}(\hat{\rho}_{nl})\right)^T \left(\tilde{X}_{nl} - \tilde{C}(\hat{\rho}_{nl})\right)$$

where $\tilde{C}(\hat{\rho}_{nl})$ is a set of nearest orthogonal points of $\tilde{X}_{nl}$ to $C(\hat{\rho}_{nl})$.

Definition 1 (Best Fitting Ellipse (BFE))**.** A best fitting ellipse for slice S_{nl} is defined by the set $BFE_{nl} := (\theta^{nl}, \alpha^{nl}, \phi^{nl})^T \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\frac{\pi}{2}, \frac{\pi}{2}]$, $l = 1, \ldots, L_n$, $n = 1, \ldots, N$ and minimizes the error function g, i.e., $BFE_{nl} = \hat{\rho}_{nl}$ with

$$g(\hat{\rho}_{nl}) = \min_{\rho_{nl} \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\frac{\pi}{2}, \frac{\pi}{2}]} g(\rho_{nl}). \tag{2}$$

The first and second principal axes must be reordered after calculation of $BFE_n = \{BFE_{n1}, \ldots, BFE_{nL_n}\}$ in order to establish correspondence between parameters of adjacent slices and across the population. Improved correspondence will support accurate statistics. The basic idea in our reordering procedure is to carry out the reordering corresponding to the lowest rotation angle of both principal axes to the first principal axis of the neighbor slice where the center slice is chosen as the basis. The rotation between the center slice M and an arbitrary slice is constrained by $\max(|\phi^i - \phi^M|) = \pi, i \in \{1, \ldots, L\}$ after reordering. Therefore, the set BFE_n of reordered best-fitting ellipses is an element of $(\mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\pi, \pi])^{L_n}$.

A further improvement of correspondence is achieved by the introduction of two additional constraints in the parameter model. First, we relax the rotation parameter ϕ^{nl} in case of circularity. If both principal axes have the same length, the orientation of an ellipse is undefined. Therefore we penalize ϕ^{nl} in the case of high circularity by taking $\phi^{nl'}$ from the neighboring slices into account. Second, smoothing is performed between neighboring slices to avoid large forward and backwards rotations between $\phi^{n(l-1)}, \phi^{nl}$ and $\phi^{n(l+1)}$. The reordering algorithm and implementation of constraints are described in detail in the Supplementary Material.

The current implementation assumes the definition of control points CP_n in the training data set $\{V_n, X_n, BFE_n\}$, where $BFE_n \in (\mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\pi, \pi])^{L_n}$ is a reordered set of best fitting ellipses, $n = 1, ..., N$. Furthermore, the control points have to be defined manually by a physician in a new patient data set. The control points are used to make the best fitting ellipses BFE_n comparable and to transform the parametrized ellipses model to a common position, scale and orientation by a transformation matrix Λ^n_{dCP}. The transformation matrix Λ^n_{dCP} maps the de-rotated prior data $\{dBFE_n, dCP_n\}$ to $\{\overline{BFE}_n, \overline{CP}_n\}$ in the sample space, as depicted in Figure 2. In this article, we assume 6 control points in the first, center and last slice at the boundary of the prostate, i.e.,

$$CP_n = \{A^n_1, \ldots, A^n_6, P^n_1, \ldots, P^n_6, B^n_1, \ldots, B^n_6\}$$

as visualized in Figure 3. In addition, we have tested alternative control point configurations. They are described together with the construction of Λ^n_{dCP} in the Supplementary Material.

After transformation we have obtained a reordered and comparable set of best fitting ellipses

$$\overline{BFE}_n = \{\overline{BFE}_{n1}, \ldots, \overline{BFE}_{nL_n}\}$$

with $\overline{BFE}_{nl} = (\overline{\theta}^{nl}, \overline{\alpha}^{nl}, \overline{\phi}^{nl})^T$, $n = 1, \ldots, N$, $l = 1, \ldots, L_n$. The statistical analysis of the training data requires an equal number $L_1 = \ldots = L_N$ to establish correspondence between the parameters of the best fitting ellipses. Therefore, we interpolate the set $\overline{BFE}_{nl}$ to a common number L. When L is chosen, interpolation is done by independent cubic interpolation in each dimension, i.e., we find points of a one-dimensional function that underlies the data $\overline{\theta}_1^{nl}, \overline{\theta}_2^{nl}, \overline{\theta}_3^{nl}, \overline{\alpha}_1^{nl}, \overline{\alpha}_1^{nl}$ and $\overline{\phi}^{nl}$. The final interpolated best fitting ellipses are denoted by

$$iBFE_n = \{iBFE_{n1}, \ldots, iBFE_{nL}\}. \tag{3}$$

These ellipses are used for the statistical analysis and computation of a mean shape model. To keep things simple, we denote such a reordered, transformed and interpolated set of best-fitting ellipses by $BFE_{nl} = (\theta^{nl}, \alpha^{nl}, \phi^{nl})^T$ for the number L of contour slices with $l = 1, \ldots, L$ and $n = 1, \ldots, N$. The comparable set of best fitting ellipses BFE_n is an element of the shape space $(\mathbb{R}^2 \times \mathbb{R}_+^2 \times (-\pi, \pi])^L$.

Statistical analyses

After reconstruction of our shape space we estimate the expectation and variance of the parameters of a mean shape model $\mu_{BFE} = \{\mu_{BFE}^1, \ldots, \mu_{BFE}^L\}$ with $\mu_{BFE}^l = (\mu_\theta^l, \mu_\alpha^l, \mu_\phi^l)$ from the training set BFE_{nl}, $l = 1, \ldots, L$. We denote the mean shape Mean Best Fitting Ellipses (MBFE). In addition to the described ellipse parameters we define the position $\theta^{nl} = (\theta_1^{nl}, \theta_2^{nl}, \theta_3^{nl})^T$ in terms of a distance vector η^{nl} of θ^{nl} to a center curve defined by the control points. We model $\theta^l = \xi^l + \eta^l$, where ξ^l is analytically defined by L intersection points of the curve within each slice. Thereby, we are describing the mean shape which is closest to the control points. This approach is reasonable under the assumption that the control points are well defined. In the Supplementary Material we explore various ways of describing the position parameter for different control point methods.

The mean curve of the expected location is given by

$$\mu_{\theta_j}^l = \frac{1}{N} \sum_{i=1}^{N} \theta_j^{il}, \quad j \in \{1, 2, 3\}, \tag{4}$$

where $\mu_\theta^l = (\mu_{\theta_1}^l, \mu_{\theta_2}^l, \mu_{\theta_3}^l)^T$, $l = 1, \ldots, L$. The variance and covariance are estimated by

$$(\sigma_{\theta_j}^l)^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\theta_j^{il} - \mu_{\theta_j}^l)^2, \quad j \in \{1,2,3\}, \text{ and} \tag{5}$$

$$\Sigma_\theta^l = \frac{1}{N-1}\sum_{i=1}^{N}(\theta^{il} - \mu_\theta^l)(\theta^{il} - \mu_\theta^l)^T. \tag{6}$$

The length parameter is modeled by a log-normal distribution because $\alpha \in \mathbb{R}_+^2$. Thus we estimate the mean and variance of $a = \log(\alpha) \in \mathbb{R}^2$. The estimation of means and variances of the remaining parameters a, ϕ, η is according to (4-5). Following Dryden and Mardia [9] we suggest a prior distribution for a new data set as

$$\begin{aligned}
&\theta_1^l \sim N\left(\mu_{\theta_1}^l, (\sigma_{\theta_1}^l)^2\right), \quad \theta_2^l \sim N\left(\mu_{\theta_2}^l, (\sigma_{\theta_2}^l)^2\right), \\
&a_1^l \sim N\left(\mu_{a_1}^l, (\sigma_{a_1}^l)^2\right) \Longleftrightarrow \alpha_1^l \sim \text{log-}N\left(\mu_{a_1}^l, (\sigma_{a_1}^l)^2\right) \text{ with } a_1^l = \log(\alpha_1^l), \\
&a_2^l \sim N\left(\mu_{a_2}^l, (\sigma_{a_2}^l)^2\right) \Longleftrightarrow \alpha_2^l \sim \text{log-}N\left(\mu_{a_2}^l, (\sigma_{a_2}^l)^2\right) \text{ with } a_2^l = \log(\alpha_2^l), \\
&\phi^l \sim N\left(\mu_\phi^l, (\sigma_\phi^l)^2\right),
\end{aligned}$$

$l = 1, \ldots, L$. If θ^l is defined according to the center curve given by the control points as described above, we model $\eta_i^l \sim N\left(\mu_{\eta_i}^l, (\sigma_{\eta_i}^l)^2\right), i = 1, 2$. Since the rotational parameter is expected to have small variance it is not necessary to apply a circular distribution, and we assume normality.

After constructing the shape model we estimate the best fitting ellipse BFE_l parametrized by $\rho_l = (\theta^l, \alpha^l, \phi^l)^T$, $l = l, \ldots, L$ in a new data set given the control points CP. This is obtained through the posterior $\pi(\rho \mid S)$ where $s_{il} \in S \subseteq V$ is the volume information and $i = (i_1, i_2) \in I(\rho)$ is a set of indices within the ellipses ρ. The control points CP are used to deform the prior model $\pi(\rho)$. Therefore we model the posterior by an empirical Bayes approach [20]. The posterior

$$\pi(\rho \mid S, CP) \propto L(S \mid \rho) * \pi(\rho \mid CP) \tag{7}$$

defines the posterior density of the deformed template $\pi(\rho \mid CP)$ given the the observed image. The Likelihood or image model $L(S \mid \rho)$ is the joint probability density function of the gray levels given the parametrized object $\rho|_{CP}$, while $\rho|_{CP}$ defines the ellipses ρ deformed by the control points CP. The prior $\pi(\rho \mid CP)$ models realistic variations from our mean shape $\mu_{BFE} \in (\mathbb{R}^2 \times \mathbb{R}_+^2 \times (-\pi, \pi])^L$ given the control points. We are estimating the posterior distribution using a Markov chain Monte Carlo (MCMC) approach. The method and results are discussed in detail in the Supplementary Material.

Evaluation

We have evaluated the proposed method using 33 patient case studies. The training data set consists of $N = 23$ T1-weigthed Fast Field Echo (FFE) 3D Magnetic Resonance (MR) data. The mean shape model and variance is calculated from the training data set and applied to a test data set of 10 MR FFE case studies. The splitting in test and training data is done according to the sequence of data acquisition. Each data set consists of H_n Digital Imaging and Communications in Medicine (DICOM) image files and one DICOM region structure file, while the contour information of the prostate is stored in the header of a DICOM file without any image information. The voxel size (l_x, l_y, l_z) is $(0.559\text{mm}, 0.559\text{mm}, 3\text{mm})$ with a slice distance of 3.3mm of the data sets V_n. Each slice consists of 288×288 voxels. The number of slices containing prostate information is 10.478 ± 2.626 (mean$\pm$standard deviation) in the training set and 10.5 ± 2.799 in the test set. Figure 1 illustrates test patient 3, whose image set consists of 24 MR FFE slices whereas 12 slices contain contour information. Slice 6 is the first slice where contour information of the prostate is available and the last slice is 17.

Three metrics are used to compare the manual and the semi-automatic contours. In the axial slices, where the expert manual delineations are present, we calculate the Hausdorff distance (HD) by

$$d_{hd}(X,Y) = \max\{\max_{x\in X}\min_{y\in Y} d(x,y), \max_{y\in Y}\min_{x\in X} d(x,y)\}. \tag{8}$$

The Hausdorff distance measures the maximum distance of a point in a set X to the nearest point in Y or vice versa. Generalization to 3D uses mean or median over all slices. The measure indicates how much manual corrections are required. An ideal value of HD equal to zero reflects complete agreement of the contours. A second criteria is the volume overlap (or Dice similarity coefficient) defined by

$$d_{vol}(X,Y) = 2\frac{|X \cap Y|}{|X| + |Y|}, \tag{9}$$

where $|\cdot|$ is the number of voxels contained in a region. Finally, accuracy is defined as

$$d_{acc}(X,Y) = 1 - \frac{|FP| + |FN|}{|TP| + |FN|} \tag{10}$$

with $TP = X \cap Y$ volume included in both X and Y (true positive), $FN = X \cap (\neg Y)$ volume of X not included in Y (false negative) and $FP = (\neg X) \cap Y$ volume of Y not included by X (false positive). Both values range from 0 to 1, with optimal value 1. Volume overlap indicates how much of the prostate has been detected by the approach while accuracy shows how incapable the method is to select the true prostate pixels.

Table 1: Evaluation metrics comparing Best Fitting Ellipses (BFE) to manual delineations, and comparing deformed Mean Best Fitting Ellipses (MBFE) to manual delineations (unit: HD mean in mm, dice 3D and accuracy in percentage).

Test set	1	2	3	4	5	6	7	8	9	10
	Best Fitting Ellipses									
HD mean	1.32	1.09	2.24	1.34	1.94	1.61	1.41	1.65	1.59	2.49
Dice 3D	0.96	0.97	0.95	0.96	0.94	0.94	0.96	0.96	0.97	0.93
Accuracy	0.93	0.94	0.91	0.91	0.88	0.88	0.92	0.93	0.94	0.85
	Mean Best Fitting Ellipses									
HD mean	2.38	2.58	6.12	2.76	3.79	4.32	2.73	5.59	4.88	4.63
Dice 3D	0.92	0.93	0.90	0.91	0.88	0.84	0.92	0.88	0.89	0.88
Accuracy	0.84	0.84	0.81	0.82	0.74	0.70	0.84	0.73	0.80	0.74

In addition to the quantitative metrics, we have performed a small pilot test on 8 new patients comparing time expenditure using the proposed method and manual delineation. The time expenditure for the proposed method includes the definition of control points and the correction of the contour obtained by the method for each patient. The used mean shape model and variance was calculated from the training data as decribed above. Time measurements were obtained by two independent physicians for each case. Manual delineations, definition of control points and corrections were performed using the treatment planning system Eclipse$^{\mathrm{TM}}$.

Results and Discussion

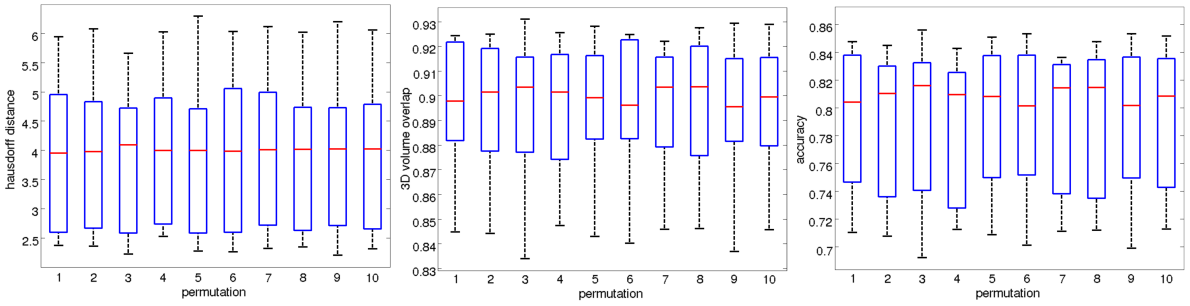
The evaluation is performed on the deformed mean best fitting ellipses, i.e., on $\pi(\rho \mid CP)$ in formula (7). Additional evaluations are done in the Supplementary Material for $\pi(\rho \mid S, CP)$.

Table 1 contains the distance metrics defined in (8) - (10) comparing the manual delineation and BFE for each test data set, and comparing the manual delineation and the deformed MBFE described by $\pi(\rho \mid CP)$. The high Dice similarity coefficient and accuracy values and small Hausdorff distances between manual delineations and BFE confirm the stacked ellipses model. The values show the best possible description of the test cases by the proposed model. The distance metrics between manual delineation and the deformed MBFE are fairly accurate. The values indicate that the estimates used as initial contours for final delineations will lower the time expenditure of the delineation procedure. Figures 1a to 1e illustrate 5 slices of the BFE evaluation of test patient 3 from Table 1. Figures 1f to 1j illustrate 5 slices of the deformed MBFE evaluation of test patient 3 from Table 1 with a volume overlap of the manual delineation line and the deformed mean shape of 0.90 and accuracy 0.81.

Table 2 shows the median and median absolute deviation (MAD) for the data groups “test data”, “training data” and “all data”. A BFE volume overlap for all data of 0.954 ± 0.010 (median$\pm$MAD) and accuracy of

Table 2: Evaluation metrics comparing BFE/MBFE and manual delineations for different data groups.

data group		test data		training data		all data	
		BFE	MBFE	BFE	MBFE	BFE	MBFE
HD mean [mm]	median	1.604	4.052	1.840	3.806	1.810	3.806
	MAD	0.274	1.305	0.286	0.427	0.289	0.538
Dice 3D [pct]	median	0.959	0.899	0.954	0.903	0.954	0.903
	MAD	0.008	0.021	0.007	0.013	0.010	0.019
Accuracy [pct]	median	0.918	0.807	0.908	0.800	0.908	0.800
	MAD	0.015	0.035	0.014	0.031	0.020	0.036



(a) Hausdorff (b) Dice 3D (c) Accuracy

Figure 4: Evaluation results between MBFE and manual delineations for 10 random permutations in training and test sets consisting of 23 and 10 patients, respectively.

0.908 ± 0.020 confirm the model further (ref. Table 2). Similar values in the subset of test data and training data are indicating model robustness. In addition to the BFE results, Table 2 summarizes the results by median and MAD of the distances between manual delineation and the deformed MBFE. A median volume overlap of 0.899 ± 0.021 and accuracy of 0.807 ± 0.035 of the test data show further the power of the prior. The deformation of the prior is done by the control points and can be computed directly since there is no sampling or estimation involved at this point.

To evaluate the robustness of the model, we randomly split 10-times the set of 33 patients into a training set with 23 cases and a test set with 10 cases. Figure 4 shows the evaluation distances between the manual delineation and the deformed MBFE. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the extreme data points. The figure shows only small variation between the different permutations, thereby demonstrating robustness of the stacked ellipses model.

Results from using MCMC to further optimize the delineation, as described in the Method section, are only presented in the Supplementary Material since a slight improvement comes at the cost of large computation time.

The time comparisons indicated an average of 30% time saving using the proposed method compared to manual delineation. The time measurement of the proposed method includes the definition of controls points as well as the correction of the estimated contour by the physician.

Conclusions

The presented results demonstrate the potential of the proposed method in modeling the prostate by slicewise best fitting ellipses. Deformation of the mean shape using control points gives very good results with little computational cost. Hence we believe that providing physicians with a good initial contour is beneficial in the clinical praxis of radiotherapy treatment.

The corrections of generated delineations based on few control points were not streamlined in the workflow of the physicians, and the task of correcting contours is not part of their everyday activity. Furthermore, corrections were not done directly after the definition of the control points and sometimes by different physicians, and physicians had to deal with a different orientation of the data set in the treatment planning system than in the diagnostic MRI. These issues must and can be solved for a well designed system. Therefore, a time saving of 30% likely represent a lower limit, and has to be validated in a well designed and properly powered study. Furthermore, we expect larger time savings in data sets where the prostate is imaged in a higher number of slices. In the extreme case, if the prostate is visible in only three slices, the BFE approach would not give any benefit using the current control point method.

In addition, the results show a precise description of the prostate by the BFE model with an average volume overlap of 95%. The high performance of the deformed mean shape model using the control points explains the small improvement by applying MCMC. Nevertheless, an improvement of the likelihood in the posterior distribution or by an active appearance model [21] is a field of further research as elaborated in the Supplementary Material. A clear disadvantage of an additional method like MCMC is the extra computation time.

Further improvements can be achieved in the constraint and regularization terms, e.g., by considering the surface curvature versus changes of the ellipses parameters. We do not expect abrupt changes between neighboring slices around the central slice, but larger changes between slices towards the ends can be permitted, particularly in the length of the first and second principal axis. Also, the reduction of manual interaction in the proposed method is left for future work.

Competing interests

The authors declare that they have no competing interests.

Authors contributions

JS conducted the design, development of the methodology and the implementation of the study resulting in a stacked ellipsoid model generation from a training data set and the registration of the model in a new patient data set. Furthermore, JS drafted and revised the manuscript. SOS and FG contributed to the conception and design of the study, and helped to draft the manuscript. VKT and KM coordinated and carried out the acquisition of patient data. All authors read and approved the final manuscript.

Acknowledgements

JS and FG are funded by the Norwegian Research Council through eVita program grant no 176872/V30. The research is performed as part of Tromsø Telemedicine Laboratory, funded by the Norwegian Research Council 2007-2014, grant no 174934. Special thanks goes to Jorunn Andrea Skjelvareid and Rune Sylvarnes at the Department of Oncology at University Hospital of North Norway.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics**. *CA- Cancer J Clin* 2011, **61**(2):69–90.
2. Smith WL, Lewis C, Bauman G, Rodrigues G, D'Souza D, Ash R, Ho D, Venkatesan V, Downey D, Fenster A: **Prostate volume contouring: A 3D analysis of segmentation using 3DTRUS, CT, and MR**. *Int J Radiat Oncol* 2007, **67**(4):1238–1247.
3. Ghose S, Oliver A, Martá R, Lladó X, Vilanova JC, Freixenet J, Mitra J, Sidibé D, Meriaudeau F: **A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images**. *Comp Meth Prog Bio* 2012, **108**:262–287.
4. Crouch J, Pizer SM, Chaney EL, Hu Y, Mageras GS, Zaider M: **Automated Finite Element Analysis for Deformable Registration of Prostate Images**. *IEEE T Med Imaging* 2007, **26**(10):1379–1390.
5. Dam EB, Fletcher P, Pizer SM: **Automatic Shape Model Building Based on Principal Geodesic Analysis Bootstrapping**. *Med Image Anal* 2008, **12**(2):136–151.
6. Merck D, Tracton G, Saboo R, Levy J, Chaney E, Pizer S, Joshi S: **Training models of anatomic shape variability**. *Med Phys* 2008, **35**(8):3584–3596.
7. Siddiqi K, Pizer SM: *Medial Representations*. Springer 2008.
8. Pizer SM, Jung S, Goswami D, Zhao X, Chaudhuri R, Damon JN, Huckemann S, Marron JS: **Nested Sphere Statistics of Skeletal Models**. In *Innovations for Shape Analysis: Models and Algorithms*. Edited by Breuß M, Bruckstein A, Maragos P, Springer Lecture Notes in Computer Science 2013:93–115.
9. Dryden IL, Mardia KV: *Statistical Shape Analysis*. John Wiley & Sons 1998.
10. Davies R, Twining C, Taylor C: *Statistical models of shape - optimisation and evaluation*. Springer 2008.
11. Heimann T, Meinzer HP: **Statistical shape models for 3D medical image segmentation: A review**. *Med Image Anal* 2009, **13**:543–563.

12. Saroul L, Bernard O, Vray D, Friboulet D: **Prostate segmentation in echographic images: A variational approach using deformable super-ellipse and Rayleigh distribution**. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008* 2008:129–132.

13. Mahdavi SS, Chng N, Spadinger I, Morris WJ, Salcudean SE: **Semi-automatic segmentation for prostate interventions**. *Med Image Anal* 2011, **15**(2):226–237.

14. Saboo R, Levy J, Chaney E, Pizer SM: **Medial Models of Populations of Nearly Tubular Objects**. In *Proceedings of the MICCAI Workshop on Probabilistic Models for Medical Image Analysis* 2009:232–243.

15. Betrouni N, Dewalle AS, Puech P, Vermandel M, Rousseau J: **Retraction notice to " 3D delineation of prostate, rectum and bladder on MR images**. *Comput Med Imag Grap* 2009, **32**(7):622–630.

16. Lu C, Chelikani S, Papademetris X, Knisely JP, Milosevic MF, Chen Z, Jaffray DA, Staib LH, Duncan JS: **An integrated approach to segmentation and nonrigid registration for application in image-guided pelvic radiotherapy**. *Med Image Anal* 2011, **15**(5):772–785.

17. Pasquier D, Lacornerie T, Vermandel M, Rousseau J, Lartigau E, Betrouni N: **Automatic Segmentation of Pelvic Structures From Magnetic Resonance Images for Prostate Cancer Radiotherapy**. *Int J Radiat Oncol* 2007, **68**(2):592–600.

18. Ahn SJ, Rauh W, Warnecke HJ: **Least-Squares Orthogonal Distances Fitting of Circle, Sphere, Ellipse, Hyperbola, and Parabola**. *Pattern Recogn* 2001, **34**(12):2283–2303.

19. Mulchrone KF, Choudhury KR: **Fitting an ellipse to an arbitrary shape: implications for strain analysis**. *J Struct Geol* 2004, **26**:143–153.

20. Casella G: **An Introduction to Empirical Bayes Data Analsysis**. *Am Stat* 1985, **39**(2):83–87.

21. Cootes T, Taylor C: **Statistical Models of Appearance for Computer Vision**. Tech. rep., Imaging Science and Biomedical Engineering 2004, [http://www.isbe.man.ac.uk].

Additional Files

Additional file 1 — Supplementary Materials

Article containing *i.)* a detailed description of the relative coordinates systems (e.g. ICS, PCS) on the basis of the DICOM file structure, *ii.)* post-processing procedures as for example reordering and introduction of constraints, *iii.)* a discussion of different control point method with construction of the transformation matrix Λ^n_{dCP} and the parameter η^{nl}, *iv.)* elaboration of the posterior distribution, and *v.)* a section with additional data analysis. (SupplementaryMaterial.pdf)

Supplementary material: A semiautomatic tool for prostate segmentation in radiotherapy treatment planning

Jörn Schulz*[1], Stein Olav Skrøvseth[2] and Fred Godtliebsen[1]

[1]Department of Mathematics and Statistics, University of Tromsø, 9037 Tromsø, Norway
[2]Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, 9038 Tromsø, Norway

Email: Jörn Schulz*- jorn.schulz@uit.no; Stein Olav Skrøvseth - stein.olav.skrovseth@telemed.no; Fred Godtliebsen - fred.godtliebsen@uit.no;

*Corresponding author

1 Data and relative coordinate systems

In this section, the volume data, particular on the basis of the DICOM data structure and the corresponding relative coordinate systems are discussed.

1.1 Basic principles

Rotations in the three dimensional space can be represented by the axis–angle representation [1] where an axis $c \in S^2$ is a direction that is left fixed by the rotation and angle θ is the amount of rotation. The pair (c, θ) represents a rotation in 3-space. As a convention, the orientation of the angle is determined by the right-hand rule, and a vector is a column vector. A vector $v \in \mathbb{R}^3$ rotated by (c, θ) can be obtained by pre-multiplication of the corresponding rotation matrix

$$R(c, \theta) = I_3 + \sin\theta[c]_\times + (1 - \cos\theta)(cc' - I_3), \tag{1}$$

where $[c]_\times$ is the cross product matrix, so that $[c]_\times x = c \times x$ for any $x \in \mathbb{R}^3$.

Definition 1 (Extended map)**.** The map ext maps a vector $v \in \mathbb{R}^n$ from a Cartesian coordinate system to a vector $(v, \tau)^T \in \mathbb{R}^n \times \{0, 1\}$ in an extended coordinate system. Furthermore we denote with ext^{-1} the related inverse map.

1.2 Volume information

We obtain from Magnetic Resonance (MR) tomography a discrete volume information of a continuous part of the lower male abdomen which contains the prostate. The discrete volume information of a patient data set is given by $I_1 \times I_2 \times H$ voxels on a regular grid in three dimensional space. We describe a voxel by a cube with middle point (x, y, z) and length l_x, width l_y and height l_z. Other approaches are possible too. An easy model is to assume the value of a voxel is given by the mean intensity about the volume with an overlap in x-y-direction and small gaps in z-direction, i.e., we assume

$$V(i, h) = \mu + \varepsilon_{ih} \text{ where } \varepsilon_{ih} \sim N(0, \sigma^2) \text{ iid}$$

and $i = (i_1, i_2) \in \{1, \ldots, I_1\} \times \{1, \ldots, I_2\}$, $h \in \{1, \ldots, H\}$. A more complex approach could up-rate the central areas of a voxel. The value of a voxel (i.e. the mean intensity) may describe different properties.

1.3 DICOM files and coordinate system

The Digital Imaging and Communications in Medicine (DICOM) file is a common standard container file used particularly in medical image processing. The file structure consists of a header part containing general data information and an image information part which contains mean intensities of the voxels. DICOM files uses a right handed Left-Posterior-Head (LPH) coordinate system. In a right handed coordinate system the x-axis is defined by the the thumb, the y-axis by the first finger and the z-axis by the second finger of the right hand. The coordinate system is related to the patient and not to the scanner. In a DICOM data set usually one file is defined per slice. This is the case in our data sets too. In terms of image processing, the most important variables in each file are

Rows (scalar) defines the number I_1 of voxels in row direction.

Columns (scalar) defines the number I_2 of voxels in column direction.

ImagePositionPatient (IPP) (3×1 vector) defines the coordinate of center of top left voxel in a slice.

ImageOrientationPatient (IOP) (6×1 vector) defines the unit vector along an image row and the unit vector along a column.

PixelSpacing (2×1 vector) which determine at first row spacing then column spacing, i.e. $PixelSpacing = (l_x, l_y)^T$.

These five variables define uniquely a slice in $\mathbb{R}^3$. Figure 1 visualizes some key parameters for a slice. The two variables PixelSpacing and IPP are defined in millimeter whereas IOP defines two unit vectors. Note, the variable SliceThickness defined in the DICOM file header may not be equal to the slice distance between two neighbor slices given by ImagePositionPatient because of small gaps in the z-direction as mentioned in the model description above.

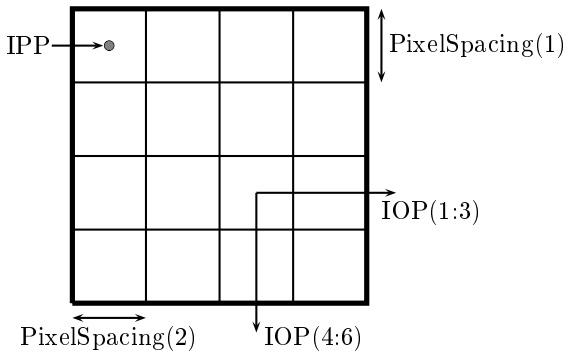


Figure 1: DICOM parameter definitions for a slice.

We have two different kinds of coordinate information, namely a patient based coordinate system (PCS) and an image based coordinate system (ICS). The ICS defines the discrete volume information $V(i_1, i_2, h)$ on a regular grid in the three dimensional space with grid points (x, y, z). Therefore, we can also write $V(x, y, z)$ instead of $V(i_1, i_2, h)$ and allocate the voxel indices (i_1, i_2, h) to volume coordinates $(x, y, z) \in \mathbb{R}^3$. $V(x, y, z)$ defines a voxel in PCS.
The PCS is defined by the above 5 variables, where IOP gives the directional cosines. We want to find a transformation matrix M between the two coordinate systems ICS and PCS. Without loss of generality we describe the transformation in an extended coordinate system (see Definition 1). The transformation matrix consists of rotation, translation and scaling whereas rotation is given by IOP, translation by IPP and scaling by PixelSpacing and IPP. We assume that IOP and PixelSpacing are equal between all slices H in a training data set. Furthermore we assume that the slices are

sorted and choose our global IPP candidate from the first slice.

The rotation matrix is defined by

$$R_{DCM} = \begin{pmatrix} dr & dc & ds & 0_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

where $0_3 = (0,0,0)^T$, $dr = (dr_x, dr_y, dr_z)^T = IOP(1:3)$ defines the directional cosines in row direction, $dc = (dc_x, dc_y, dc_z)^T = IOP(4:6)$ the directional cosines in column direction and $ds = (ds_x, ds_y, ds_z)^T = dr \times dc$ the directional cosines in slice direction whereas ds is defined by the cross product of dr and dc.

The scaling matrix is defined by

$$S_{DCM} = \begin{pmatrix} sc_x e_1 & sc_y e_2 & sc_z e_3 & 0_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

with $sc = (sc_x, sc_y, sc_z)^T = ext^{-1}(diag(S_{DCM}))$ and e_i is the 3×1 column vector with all elements being zeros except that the ith element is one. As mentioned above $sc_x = \frac{1}{l_x}$ and $sc_y = \frac{1}{l_y}$ are defined by PixelSpacing whereas sc_z is defined by the distance of IPP between two neighbor slices, e.g. the first and the second slice. Remember, we have assumed equidistant slices in a data set.

The translation matrix is defined by

$$T_t = \begin{pmatrix} e_1 & e_2 & e_3 & t \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4}$$

where $t = (t_x, t_y, t_z)^T$ defines the translation vector.

The rotation, scaling and translation matrix provide us to transform between ICS and PCS.

Definition 2 (DICOM transformation matrix)**.** The DICOM transformation matrix is defined by

$$\Lambda_{DCM} = T_{IPP} R_{DCM} S_{DCM} T_0 \tag{5}$$

and transform an image coordinate $p^{im} = (i_1, i_2, h)^T$ to patient coordinate $p^p = (x, y, z)^T$ by $ext(p^p) = \Lambda_{DCM} ext(p^{im})$. T_0 is the translation which shift the slice to make top left voxel centered at $(0,0,0)$, i.e. $T_0 := T_t$ with $t = (-1,-1,-1)^T$. R_{DCM} and S_{DCM} are defined by (2) and (3) and T_{IPP} translate the top left voxel from the first slice at IPP.

Remark 1 (Inverse DICOM transformation matrix). The inverse of Λ_{DCM} exist and transform a patient coordinate into an image coordinate by $ext(p^{im}) = \Lambda_{DCM}^{-1} ext(p^p)$.

Proof. The inverse of the rotation matrix is given by $R_{DCM}^{-1} = R_{DCM}^{T}$, the inverse of the translation matrix by $T_t^{-1} = T_{-t}$ and the inverse of scaling matrix by using $sc^{-1} = (\frac{1}{sc_x}, \frac{1}{sc_y}, \frac{1}{sc_z})$. □

Definition 3 (De-rotated patient coordinate system)**.** Corresponding to the PCS and the ICS we define a derotated patient coordinate system by $p^{dp} \in \mathbb{R}^3$ with $ext(p^{dp}) = R_{DCM}^{T} ext(p^{p}) = T_{IPP} S_{DCM} T_0 ext(p^{im})$. $IOP = (1, 0, 0, 0, 1, 0)^T$ in a derotated PCS.

The are no rotational differences between the de-rotated PCS and the ICS. Furthermore, the volume and contour information are aligned to each other in the de-rotated PCS. Therefore, it is used for the statistical analysis.

2 Post-processing of BFE

This section discusses the reordering algorithm and introduce constraint and regularization parameters into our shape model. Let $n \in \{1, \ldots, N\}$ be fixed in the following section and $l = 1, \ldots, L$ the number of contour slices. We assume equidistant slice distances, well defined and sorted slices. Calculation are carried out inside ICS.

2.1 Reordering of the first and second principal axis

After the computation of $BFE_l \in \mathbb{R}^2 \times \mathbb{R}_+^2 \times (-\frac{\pi}{2}, \frac{\pi}{2}]$ we have to reorder the first and second principal axis to establish correspondence between the parameter of adjacent slices. The reordering algorithm is visualized in Figure 2 and given by the the following steps:

1. Transformation of the orientation parameter ϕ^l to directional cosines $\psi^l = (\cos\phi^l, sin\phi^l)^T$ to avoid computational non-uniqueness by the direction of the rotation.

2. Computation of the position of the basis slice by $M = \text{round}(\theta_3^1 + \frac{1}{2}(\theta_3^L - \theta_3^1))$ where round means rounding to the closest integer.

3. Initialization of reordering for the basis slice S_M. The first principal axis is chosen to be closest to $[1, 0]$ by determination of the distance of the x-axis to the first and second principal axis. The coordinate origin is defined by θ^M and the direction of x and y-axis are given by $(1, 0)^T$ and $(0, 1)^T$ in the ICS. Note that each principal axis can be defined in the opposite

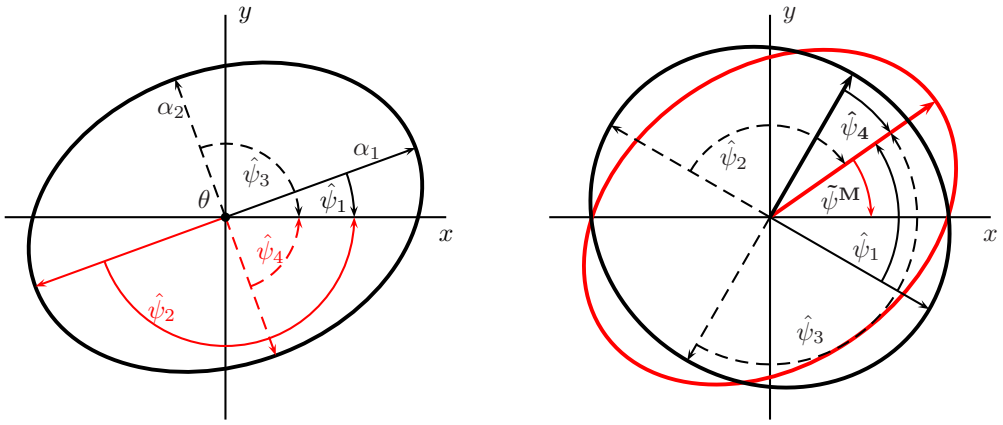


Figure 2: Visualization of the reordering procedure. Left image: Initialization of the reordering step for slice M. Right image: Reodering of slice $M-1$.

direction, i.e. the four possible directions are $\hat{\psi}_1^l = \psi^l$, $\hat{\psi}_2^l = -\psi^l$, $\hat{\psi}_3^l = (-sin\phi^l, \cos\phi^l)^T$ and $\hat{\psi}_4^l = -\hat{\psi}_3^l$. We obtain the reordered best fit $\underline{BFE_l} := (\theta^l, \tilde{\alpha}^l, \tilde{\psi}^l)^T$ by

$$\arccos((1,0)\tilde{\psi}^l) = \min_{i\in\{1,\dots,4\}} \arccos((1,0)\hat{\psi}_i^l) \tag{6}$$

where $\hat{\psi}_i^l$ is a normal vector by definition and $l = M$.

4. Reordering from the basis slice to first slice, i.e. $l = M-1,\dots,1$. The first principal axis $\tilde{\psi}^l$ is chosen to be closest to $\tilde{\psi}^{l+1}$ of the 4 axes defined by BFE_l, i.e., $\underline{BFE_l}$ is chosen to minimize $\arccos((\tilde{\psi}^{l+1})^T\hat{\psi}_i^l)$, $l = M-1,\dots,1$.

5. Reordering from the basis slice to last slice, i.e., $l = M+1,\dots,L$. The first principal axis $\tilde{\psi}^l$ is chosen to be closest to $\tilde{\psi}^{l-1}$ of the 4 axes defined by BFE_l, i.e., $\underline{BFE_l}$ is chosen to minimize $\arccos((\tilde{\psi}^{l-1})^T\hat{\psi}_i^l)$, $l = M+1,\dots,L$.

Note, that $\underline{BFE_l} \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\pi, \pi]$ compared to $BFE_l \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\frac{\pi}{2}, \frac{\pi}{2}]$. In fact, the rotation angle is not restricted by the above reordering procedure but we assume the maximal rotation between the basis and an arbitrary slice is $\max(|\phi^i - \phi^M|) = \pi, i \in \{1,\dots,L\}$ after reordering. This allows a full twist of the prostate which we do not expect. In the following, we do not distinguish between BFE_l and $\underline{BFE_l}$. For the sake of convenience, we assume a reordered set of best-fitting ellipses and denote it simply by $BFE_l := (\theta^l, \alpha^l, \phi^l)^T \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\pi, \pi]$.

2.2 Constraint and relaxation

We introduce two additional constraints in the parameter model BFE_l, $l = 1, \ldots, L$, which are relaxations of the rotation parameter ϕ in case of circularity and between large forward and backwards movings between neighbor slices.

Circularity: We define a cubic function $\lambda_l = c_l^p$ where $c_l = \frac{\min_l \alpha^l}{\max_l \alpha^l}$ is a measure of the circularity of the ellipse in slice $l = 1, \ldots, L$ and p is a parameter which defines the power of the cubic function. The circularity is 1 if both axes have the same length. Simulations have shown that a choice of $p = 12$ leads to reasonable results in our model. The orientation parameter ϕ^l is relaxed by

$$\tilde{\phi}^l = \begin{cases} (1-\lambda_l)\phi^l, & l = M, \\ \lambda_l \phi^{l+1} + (1-\lambda_l)\phi^l, & l < M, \\ \lambda_l \phi^{l-1} + (1-\lambda_l)\phi^l, & l > M. \end{cases}$$

That means, in case of high circularity c_l we do not trust the orientation in the current slice l.

Relaxation between neighbor slices: Penalizing of large changes of the rotation between neighbor slices, i.e., between the parameters $\phi^{(l-1)}, \phi^l$ and $\phi^{(l+1)}$. At first, we note that the largest possible change between two slices is maximal $\frac{\pi}{4}$ after reordering as mentioned above. A simple quadratic function is used for the relaxation by

$$\tilde{\phi}^l = \begin{cases} \lambda_l \phi^{l+1} + (1-\lambda_l)\phi^l \text{ with } \lambda_l = \left(\frac{4}{\pi}|\phi^{l+1} - \phi^l|\right)^{\frac{1}{2}}, & l < M \\ \lambda_l \phi^{l-1} + (1-\lambda_l)\phi^l \text{ with } \lambda_l = \left(\frac{4}{\pi}|\phi^{l-1} - \phi^l|\right)^{\frac{1}{2}}, & l > M. \end{cases}$$

3 Control point methods

In this section, we present three Control Point (CP) methods with construction of the transformation matrix Λ_{dCP}^n and the parameter η^{nl}. The control point method used in the main article is described by CP3 in Section 3.3 (page S13). In addition, two alternatives are presented by method CP1 in Section 3.1 (page S8) and CP2 in Section 3.2 (page S11). Both sections can be skipped if the reader is only interested in CP3.

Before we construct a transformation matrix Λ_{dCP}^n between the derotated PCS and the sample space from the control points (see Figure 2 in the main article), we have to transform the control

points CP_n and ellipses parametrization BFE_{nl} from the ICS to the derotated PCS by $ext(p^{dp}) = \Xi ext(p^{im})$ with

$$\Xi = T_{IPP} S_{DCM} T_0$$

using Definition 3 in Section 1.3. Therewith, all data sets have the same origin and rotation of the underlying coordinate system between different data sets and same distances between points. Note, the voxel size can be different between different data sets. Therefore, distances between coordinates in the ICS might be different for different data sets. For reasons of simplicity, the mapping between the normal and the extended coordinate system is not mentioned explicit in the following section. In general, we have to carry out this step if we convert with the transformation matrix Ξ. The derotated prior data are given by the set $\{dBFE_{nl}, dCP_n\}$ with

$$dCP_n = \Xi(CP_n) \text{ and} \tag{7}$$

$$dBFE_{nl}^{pos} = (\Xi(p_1^{nl}), \Xi(p_2^{nl}), \Xi(p_3^{nl}))^T \tag{8}$$

whereas $BFE_{nl}^{pos} = (p_1^{nl}, p_2^{nl}, p_3^{nl})^T \in \mathbb{R}^9$ is the positional description of a best-fitting ellipse as described in the following remark.

Remark 2 (Positional BFE). An alternative representation for a best fitting ellipse is given by the set $BFE_{nl}^{pos} := (p_1^{nl}, p_2^{nl}, p_3^{nl})^T \in \mathbb{R}^9$ with

$$\begin{aligned} p_1^{nl} &= \theta^{nl}, \\ p_2^{nl} &= \theta^{nl} + R(\alpha_1^{nl}, 0)^T, \\ p_3^{nl} &= \theta^{nl} + R(0, \alpha_2^{nl})^T, \end{aligned}$$

$$\text{and } R = \begin{pmatrix} \cos\phi^{nl} & -\sin\phi^{nl} \\ \sin\phi^{nl} & \cos\phi^{nl} \end{pmatrix}$$

is a rotation matrix. We call this representation the positional best fitting ellipse. Both representations are isomorph to each other.

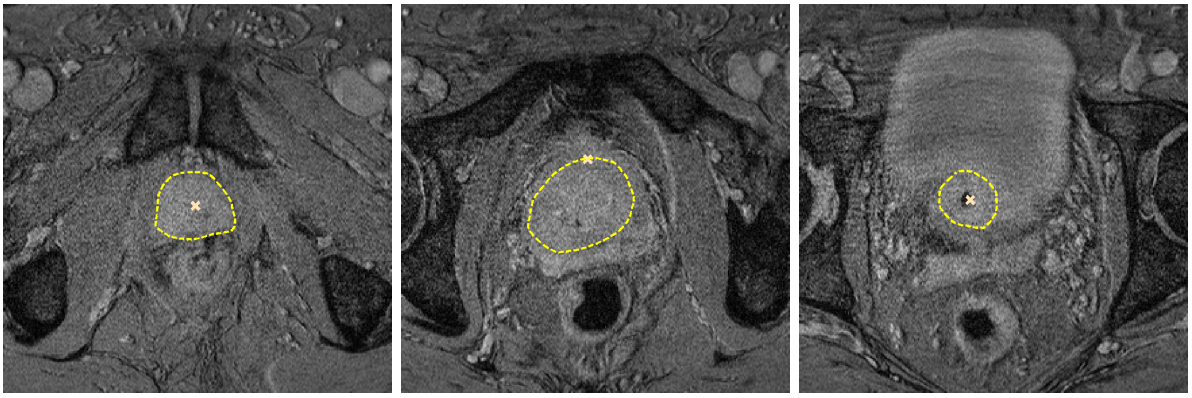
3.1 CP method 1

In the first approach, called CP1, we assume the definition of three control points $CP_n := (A^n, B^n, P_1^n) \in \mathbb{R}^{3\times 3}$. See Figure 3 for a visualization of the parameters which are defined as follow:

A^n centroid point of the prostate in the first contour slice of the prostate,

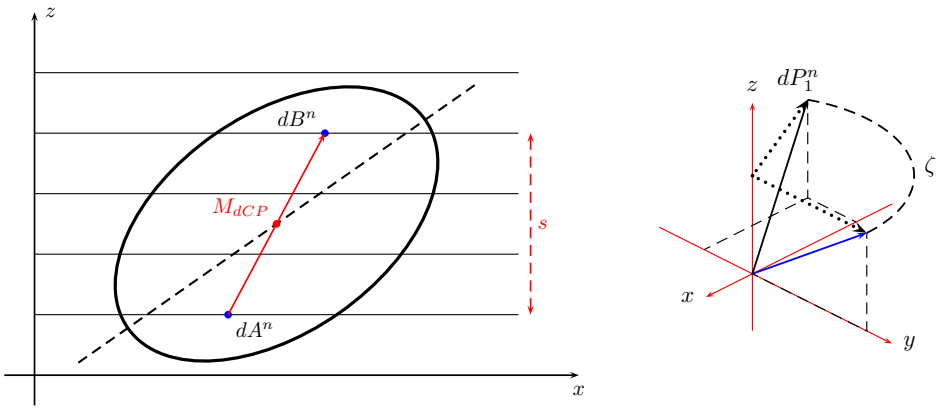
B^n centroid point of the prostate in the last contour slice of the prostate,

P_1^n point at the boundary of the prostate in posterior direction and in the base slice M (center slice, see Appendix 2.1).



(a) A^n (b) P_1^n (c) B^n

Figure 3: Definition of control points using method CP1. The dashed line defines the manual delineation of the prostate. (a) Definition of A^n (first visible contour slice). (b) Definition of P_1^n in basis slice M. (c) Definition of B^n (last visible contour slice).



(a) (b)

Figure 4: (a) Visualization of translation point M_{CP} and the scaling s from the control points dA and dB by a cross section in slice direction z. (b) Rotation of dP_1 around the z-axis to a point with x-coordinate zero.

Given the derotated prior data we construct the transformation matrix Λ_{dCP}^n to a set of comparable best fitting ellipses for fixed n by the following steps (see Figure 4):

1. Translation of $M_{dCP_n} = \frac{1}{2}(dB^n - dA^n)$ to $(0,0,0)$ where $dCP_n := (dA^n, dB^n, dP_1^n)$. This results in the translation matrix $T_{-M_{dCP_n}}$ using formula (4) on page S4.

2. Rotation of the point dP_1^n around the z-axis with angle ζ given by the angle between the two vectors dP_1^n and $(0,1,dP_1^n(z))^T$. The rotation matrix R_{dCP_n} is given by formula (1) on page S1 using $v = (0,0,1)^T$ and ζ as described before.

3. Scaling of the data by $s = |dB^n(z) - dA^n(z)|$. The scaling matrix is given by S_{dCP_n} with $sc = (\frac{\overline{L}-1}{s}, \frac{\overline{L}-1}{s}, \frac{\overline{L}-1}{s})^T$ using formula (3) on page S4 where $\overline{L}$ is defined as the common number of interpolated best fitting ellipses as described in section 2 of the main article. Note, we do not scale the length of the prostate by $\|dB^n - dA^n\|$ because we want to have the same length in z-direction. Figure 5 shows close correlation between both values for the training data set.

Finally, the transformation of the derotated prior data $\{dBFE_{nl}, dCP_n\}$ to $\{\overline{BFE}_{nl}, \overline{CP}_n\}$ is given by

$$\overline{BFE}_{nl}^{pos} = \left(\Lambda_{CP}^n(dp_1^{nl}), \Lambda_{CP}^n(dp_2^{nl}), \Lambda_{CP}^n(dp_3^{nl})\right)^T$$

and $\overline{CP}_n = \Lambda_{CP}^n(dCP^n)$ with

$$\Lambda_{dCP}^n = S_{dCP_n} R_{dCP_n} T_{-M_{dCP_n}}. \tag{9}$$

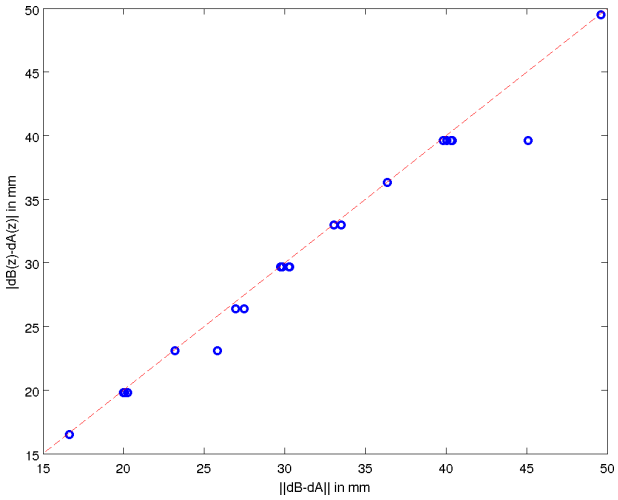


Figure 5: Comparison of the distance between the control points dA and dB in z-direction and $\|dB - dA\|$ (in mm). The dotted line shows equal values.

3.2 CP method 2

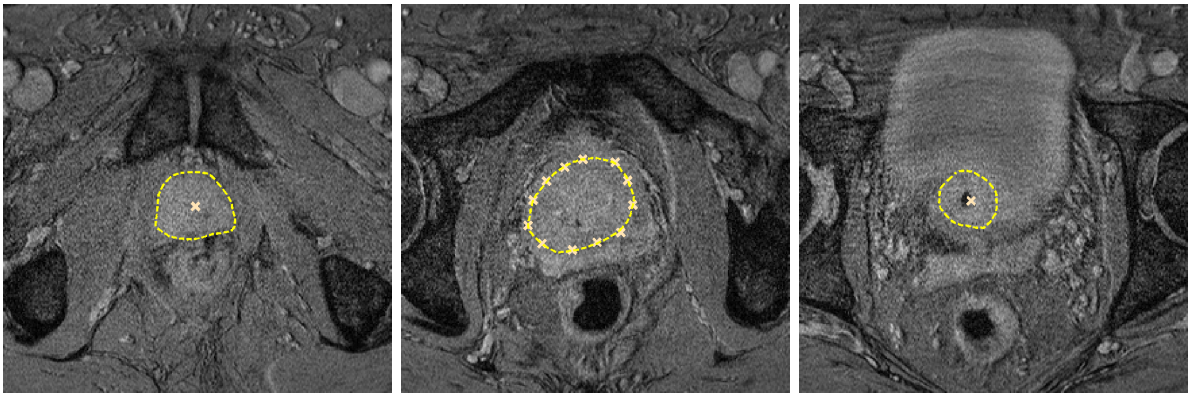
An alternative second approach, called CP2, assumes the definition of the control points $CP_n = (A^n, B^n, P_1^n, \ldots, P_{12}^n) \in \mathbb{R}^{3\times 14}, n = 1, \ldots, N$ with

A^n centroid point of the prostate in the first contour slice of the prostate,

B^n centroid point of the prostate in the last contour slice of the prostate,

$P_1^n, \ldots, P_{12}^n$ points at the boundary of the prostate in the base slice M whereas M is defined as described in Appendix 2.1.

Figure 6 visualizes the corresponding parameters plus the underlying manual delineation line which is not available in a new data set.



(a) A^n (b) P_m^n (c) B^n

Figure 6: Definition of control points using method CP2. The dashed line defines the manual delineation of the prostate. (a) Definition of A^n in the first visible contour slice. (b) Definition of P_m^n in basis slice M, $m = 1, \ldots, 12$. (c) Definition of B^n in the last visible contour slice.

Given the derotated prior data $\{dBFE_{nl}, dCP_n\}, l = 1, \ldots, L_n$ and $n = 1, \ldots, N$, we construct the transformation matrix Λ_{dCP}^n to a set of comparable best fitting ellipses for fixed n by the following steps:

1. Computation of the best fitting ellipse $CPBFE_n = (\theta^{CP_n}, \alpha^{CP_n}, \phi^{CP_n})^T \in \mathbb{R}^2 \times \mathbb{R}_+^2 \times (-\frac{\pi}{2}, \frac{\pi}{2}]$ for the set $\{P_1^n, \ldots, P_{12}^n\}$.

2. Transformation of $CPBFE_n$ to $dCPBFE_n = (\theta^{dCP_n}, \alpha^{dCP_n}, \phi^{dCP_n})^T$ in a derotated PCS.

3. The center slice is given by $s_n = \frac{L_n+1}{2}$ for each training data set n. If L_n is an odd number the center slice with the best-fitting ellipse is clearly defined. Otherwise, the best-fitting ellipses have to be interpolated to a number $L_n + 1$ before.

4. Estimation of the mean of the derotated best fitting ellipses in the center slice s_n from the training data, i.e.,

$$\mu^{dBFE} := (\mu_\theta^{dBFE}, \mu_\alpha^{dBFE}, \mu_\phi^{dBFE})^T = \frac{1}{N}\sum_n dBFE_{ns_n}. \tag{10}$$

   Furthermore, we estimate the mean length by

$$\mu_L^{dBFE} = \frac{1}{N}\sum_n \|\theta^{dBFE_{nL_n}} - \theta^{dBFE_{n1}}\|$$

   .

5. If L_n is an odd number then $M_{dCP_n} = \theta^{dCP_n}$. Otherwise M_{dCP_n} is determined from a cubic interpolation of the curve with support points in dA^n, θ^{dCP_n} and dB^n. The translation of $dCPBFE_n$ to the origin $(0, 0, 0)$ is given by the translation matrix $T_{-M_{dCP_n}}$ using formula (4) on page S4.

6. The rotation matrix R_{dCP_n} is given by formula (1) on page S1 using $v = (0, 0, 1)^T$ and $\zeta = -\phi^{dCP_n}$, i.e., we rotate the first principal axis of $dCPBFE_n$ to the direction of the x-axis.

7. Scaling of the data in x-direction by $sc_1 = \frac{\mu_{\alpha_1}^{dBFE}}{\alpha_1^{dCP_n}}$, y-direction by $sc_2 = \frac{\mu_{\alpha_2}^{dBFE}}{\alpha_2^{dCP_n}}$ and z-direction by $sc_3 = \frac{\mu_L^{dBFE}}{|dB^n(z) - dA^n(z)|}$. Thereby, the scaling matrix is given by S_{dCP_n} with $sc = (sc_1, sc_2, sc_3)^T$ using formula (3) on page S4. Note again, we do not scale the length of the prostate because we want to have the same length in z-direction.Figure 5 shows close correlation between both values for the training data set. Furthermore, Figure 7 illustrate the scaling parameters for our training data set and indicates correlations between them.

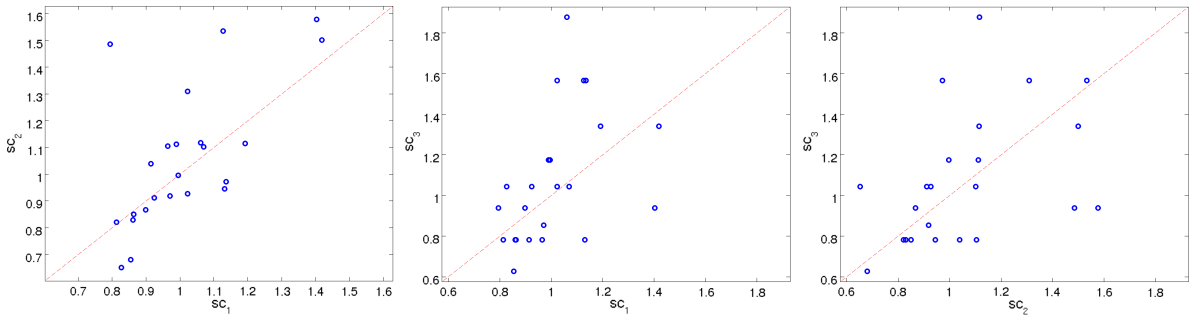
Finally, using CP method 2, the transformation of the derotated prior data $\{dBFE_{nl}, dCP_n\}$ to $\{\overline{BFE}_{nl}, \overline{CP}_n\}$ is given by

$$\overline{BFE}_{nl}^{pos} = \left(\Lambda_{dCP}^n(dp_1^{nl}), \Lambda_{dCP}^n(dp_2^{nl}), \Lambda_{dCP}^n(dp_3^{nl})\right)^T$$

and $\overline{CP}_n = \Lambda^n_{dCP}(dCP^n)$ with

$$\Lambda^n_{dCP} = S_{dCP_n} R_{dCP_n} T_{-M_{dCP_n}}. \tag{11}$$

An alternative to the above mentioned approach for computation of M_{dCP_n} is given by setting $M_{dCP_n} = \theta^{CP_n}$ for even and odd numbers L_n. Table 1 shows the average of the standard L1 and L2 norm of the distance vector between M_{dCP_n} and the interpolated center ellipse position from the corresponding training data. We argue for the implemented version using cubic interpolation due to a smaller error.



(a) sc_1 versus sc_2 (b) sc_1 versus sc_3 (c) sc_2 versus sc_3

Figure 7: Scaling parameters using transformation method 2 for the training data set.

Table 1: Average error of center ellipse position derived from control points versus center ellipse position derived from training data (unit: mm).

center ellipse position	alternative version	implemented version
$L1$ norm	0.5612	0.5470
$L2$ norm	0.4707	0.4568

3.3 CP method 3

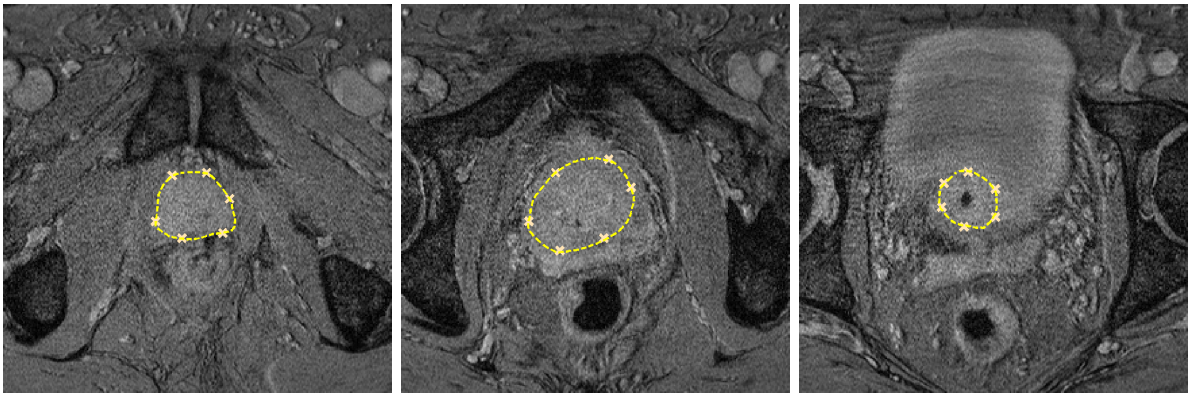
The third approach, called CP3, is the primary used approach in this paper and assumes the definition of control points $CP_n = (A^n_{m_1}, B^n_{m_2}, P^n_{m_3}) \in \mathbb{R}^{3\times(m_1+m_2+m_3)}, m_i \in \mathbb{N}$ with

$A^n_{m_1}$ points at the boundary of the prostate in the first contour slice of the prostate,

$B^n_{m_2}$ points at the boundary of the prostate in the last contour slice of the prostate,

$P^n_{m_3}$ points at the boundary of the prostate in the base slice M (see subsection 2.1).

We have to assume $m_i \geq 3$ to be able to fit ellipses to the first, base and last slice. In the current work $m_1 = m_2 = m_3 = 6$ is chosen to achieve robust fits of best-fitting ellipses. See Figure 8 for a visualization of the parameters.



(a) $A^n_{m_1}$ (b) $P^n_{m_3}$ (c) $B^n_{m_2}$

Figure 8: Definition of control points at the boundary of the prostate using method CP3 with $m_1 = m_2 = m_3 = 6$. (a) Definition of $\{A^n_1, \ldots, A^n_{m_1}\}$ (first visible contour slice). (b) Definition of $\{P^n_1, \ldots, P^n_{m_3}\}$ in basis slice M. (c) Definition of $\{B^n_1, \ldots, B^n_{m_2}\}$ (last visible contour slice).

Given the derotated prior data $\{dBFE_{nl}, dCP_n\}$, we construct the transformation matrix Λ^n_{dCP} similar to Section 3.2. The main difference is the additional deformation of the ellipsoid by the control point information in the first and last slice. Let $n \in \{1, \ldots, N\}$ be fixed. Λ^n_{dCP} is derived by the following steps:

1. Computation of the best fitting ellipses $CPBFE_{n,i} = (\theta^{i,CP_n}, \alpha^{i,CP_n}, \phi^{i,CP_n})^T \in \mathbb{R}^2 \times \mathbb{R}^2_+ \times (-\frac{\pi}{2}, \frac{\pi}{2}], i = 1, 2, 3$ from the set $\{A^n_1, \ldots, A^n_{m_1}\}$ if $i = 1$, $\{B^n_1, \ldots, B^n_{m_2}\}$ if $i = 2$ and $\{P^n_1, \ldots, P^n_{m_3}\}$ if $i = 3$.

2. Transformation of $CPBFE_{n,i}$ to $dCPBFE_{n,i}$ in a derotated PCS with $dCPBFE_{n,i} = (\theta^{i,dCP_n}, \alpha^{i,dCP_n}, \phi^{i,dCP_n})^T$.

3. The center slice is given by $s_n = \frac{L_n+1}{2}$ for each training data set n. We assume that L_n is an odd number where the center slice with the best-fitting ellipse is clearly defined. Otherwise,

the set of derotated best-fitting ellipses $dBFE_n$ has to be interpolated to an odd number $L_n + 1$ before.

4. Estimation of mean $\mu^{i,dBFE}, i = 1, 2, 3$, from the derotated best-fitting ellipses in the first, last and center slice s_n from the training data by

$$\mu^{1,dBFE} := (\mu_\theta^{1,dBFE}, \mu_\alpha^{1,dBFE}, \mu_\phi^{1,dBFE})^T = \frac{1}{N} \sum_n dBFE_{n1},$$
$$\mu^{2,dBFE} := (\mu_\theta^{2,dBFE}, \mu_\alpha^{2,dBFE}, \mu_\phi^{2,dBFE})^T = \frac{1}{N} \sum_n dBFE_{nL_n},$$
$$\mu^{3,dBFE} := (\mu_\theta^{3,dBFE}, \mu_\alpha^{3,dBFE}, \mu_\phi^{3,dBFE})^T = \frac{1}{N} \sum_n dBFE_{ns_n}.$$

   Furthermore, we estimate the mean length by

$$\mu_L^{dBFE} = \frac{1}{N} \sum_n \|\theta^{dBFE_{nL_n}} - \theta^{dBFE_{n1}}\|$$

   .

5. If L_n is an odd number then $M_{dCP_n} = \theta^{3,dCP_n}$. Otherwise M_{dCP_n} is determined from a cubic interpolation of the curve with support points in θ^{1,dCP_n}, θ^{3,dCP_n} and θ^{2,dCP_n}. The translation of $dCPBFE_n$ to the origin $(0, 0, 0)$ is given by the translation matrix $T_{-M_{dCP_n}}$ using formula (4) on page S4.

6. The rotation matrix R_{dCP_n} is given by formula (1) on page S1 using $v = (0, 0, 1)^T$ and $\zeta = -\phi^{3,dCP_n}$, i.e., we rotate the first principal axis of $dCPBFE_{n,3}$ to the direction of the x-axis.

7. Determination of scaling values sc_{1l} and sc_{2l} of the first and second principal axis for slices $l = 1, \ldots, L_n$. At first, the scaling values in the first, center and last slice are calculated, i.e., $sc_{11} = \frac{\mu_{\alpha_1}^{1,dBFE}}{\alpha_1^{1,dCP_n}}$, $sc_{21} = \frac{\mu_{\alpha_2}^{1,dBFE}}{\alpha_2^{1,dCP_n}}$, $sc_{1L_n} = \frac{\mu_{\alpha_1}^{2,dBFE}}{\alpha_1^{2,dCP_n}}$, $sc_{2L_n} = \frac{\mu_{\alpha_2}^{2,dBFE}}{\alpha_2^{2,dCP_n}}$, $sc_{1s_n} = \frac{\mu_{\alpha_1}^{3,dBFE}}{\alpha_1^{3,dCP_n}}$ and $sc_{2s_n} = \frac{\mu_{\alpha_2}^{3,dBFE}}{\alpha_2^{3,dCP_n}}$. Afterwards, the remaining scaling values sc_{1l} and sc_{2l} are interpolated from $(sc_{11}, sc_{1s_n}, sc_{1L_n})$ and $(sc_{21}, sc_{2s_n}, sc_{2L_n})$. In addition, the scaling value in z-direction is determine by $sc_3 = \frac{\mu_L^{dBFE}}{|\theta_3^{2,dCP_n} - \theta_3^{1,dCP_n}|}$. Thereby, the scaling matrices are given by $S_{l,dCP_n}, l = 1, \ldots, L_n$, with $sc = (sc_{1l}, sc_{2l}, sc_3)^T$ using formula (3) on page S4. We do not scale the length of the prostate because we want to have the same length in z-direction. Figure 5 shows close correlation between both values for the training data set.

The primary used CP method 3 in this paper transforms the derotated prior data $\{dBFE_{nl}, dCP_n\}$ to $\{\overline{BFE}_{nl}, \overline{CP}_n\}$ by

$$\overline{BFE}_{nl}^{pos} = \left(\Lambda_{dCP}^n(dp_1^{nl}), \Lambda_{dCP}^n(dp_2^{nl}), \Lambda_{dCP}^n(dp_3^{nl})\right)^T$$

and $\overline{CP}_n = \Lambda_{dCP}^n(dCP^n)$ with

$$\Lambda_{dCP}^n := S_{l,dCP_n} R_{dCP_n} T_{-M_{dCP_n}}. \tag{12}$$

3.4 Description of θ^{nl} in relation to the control points

In addition to the position $\theta^{nl} = (\theta_1^{nl}, \theta_2^{nl}, \theta_3^{nl})^T$ of each best-fitting ellipse, we describe the parameter in terms of a distance vector η^{nl} of θ^{nl} to a line or a curve, defined by the control points. Thereby, we are describing the mean shape which is closest to the control points. Corresponding to the different control point methods, see sections before, there are different ways of describing the positional parameter via η^{nl}. Let $n = 1, \ldots, N$ be fixed in this section.

3.4.1 Determination of η^{nl} using CP method 1

Given $dCP_n := (dA^n, dB^n, dP_1^n)$ the two control points dA^n and dB^n define a line $\overrightarrow{dA^n dB^n} \subset \mathbb{R}^3$ that cross each slice $l = 1, \ldots, L_n$ with prostate contour information. Now, κ^{nl} defines the intersection point of the line $\overrightarrow{dA^n dB^n}$ with each slice l. It is given by $\kappa^{nl} = dA^n + \frac{\tilde{l}}{L_n - 1} v^n$ with $\tilde{l} = l - 1$, $v^n = dB^n - dA^n$ and $l \in \{0, \ldots, L_n\}$. Thereby, the distance vector η^{nl} to θ^{nl} is defined by

$$\eta^{nl} = \theta^{nl} - \kappa^{nl},$$

with $l \in \{0, \ldots, L_n\}$. Simulations have shown that statistical inference of η^{nl} result in a much smaller variance than for θ^{nl}.

3.4.2 Determination of η^{nl} using CP method 2 and 3

Instead of construction a line $\overrightarrow{dA^n dB^n}$ as described before, we construct a cubic curve that connect three slices with contour points, i.e., the first, the basis and the last slice. Using CP method 2, dCP_n is defined by

$$dCP_n := (dA^n, dB^n, dP_1^n, \ldots, dP_{12}^n)$$

and $dCPBFE_n = (\theta^{dCP_n}, \alpha^{dCP_n}, \phi^{dCP_n})$. The three points dA^n, θ^{dCP_n} and dB^n define a curve $g_{dCP}(z)$. Similar, in case of CP method 3, $CP_n := (dA_1^n, \ldots, dA_{m_1}^n, dB_1^n, \ldots, dB_{m_2}^n, dP_1^n, \ldots, dP_{m_3}^n)$ results in the three best fitting ellipses $dCPBFE_{n,i} = (\theta^{i,dCP_n}, \alpha^{i,dCP_n}, \phi^{i,dCP_n}), i = 1, 2, 3$. Therewith, the three points θ^{i,dCP_n} define a cubic curve $g_{dCP}(z)$. The final curve is calculated by cubic interpolation with supporting points $\{A^n, \theta^{dCP_n}, B^n\}$ for CP2 or $\{\theta^{1,dCP_n}, \theta^{3,dCP_n}, \theta^{2,dCP_n}\}$ for CP3. Therewith, $g_{dCP}(z)$ cross each slice, whereas each slice is defined by an (x, y, z)-plane with $z = \theta_3^{nl}, l = 1, \ldots, L_n$. Let $g_{CP}(\theta^{nl})$ define the intersection points of the curve $g_{CP}(z)$ with each slice. The distance vector η^{nl} to θ^{nl} is calculated by

$$\eta^{nl} = \theta^{nl} - g_{CP}(\theta^{nl})$$

with $l \in \{0, \ldots, L_n\}$.

Figure 9 and 10 show μ_θ^l, μ_η^l, θ^{nl} and η^{nl} of the training data set in the sample space using CP3 with $l = 1, \ldots, 23$. The covariance matrices are visualized by 90% confidence ellipses. We can clearly decrease the variance using the alternative approach with identical variance for the base slice because of $\kappa^{nl} = 0 \quad \forall n = 1, \ldots, N$ for $l = M$.

4 Posterior distribution

In addition to construct a shape model and the corresponding shape space, we aim to estimate the best fitting ellipses BFE_l parametrized by $\rho_l = (\theta^l, \alpha^l, \phi^l)^T$, $l = l, \ldots, L$ in a new observed data set given a set of control points CP. In general, this can be obtained through the posterior $\pi(\rho \mid S)$ where $s_{il} \in S \subseteq V$ is the volume information and $i = (i_1, i_2) \in I(\rho)$ is a set of indices within the ellipses ρ. The control points CP are used to deform the prior model $\pi(\rho)$ as described in Section 3. Therefore, we model the posterior by an empirical Bayes approach [2]. The posterior

$$\pi(\rho \mid S, CP) \propto L(S \mid \rho) * \pi(\rho \mid CP)$$

defines the posterior density of the deformed template $\pi(\rho \mid CP)$ given the the observed image. The likelihood or image model $L(S \mid \rho)$ is the joint probability density function of the gray levels given the parametrized object $\rho|_{CP}$, whereas $\rho|_{CP}$ defines the ellipses ρ deformed by the control points CP. The prior $\pi(\rho \mid CP)$ models realistic variations from our mean shape $\mu_{BFE} \in (\mathbb{R}^2 \times \mathbb{R}_+^2 \times (-\pi, \pi])^L$ given the control points.

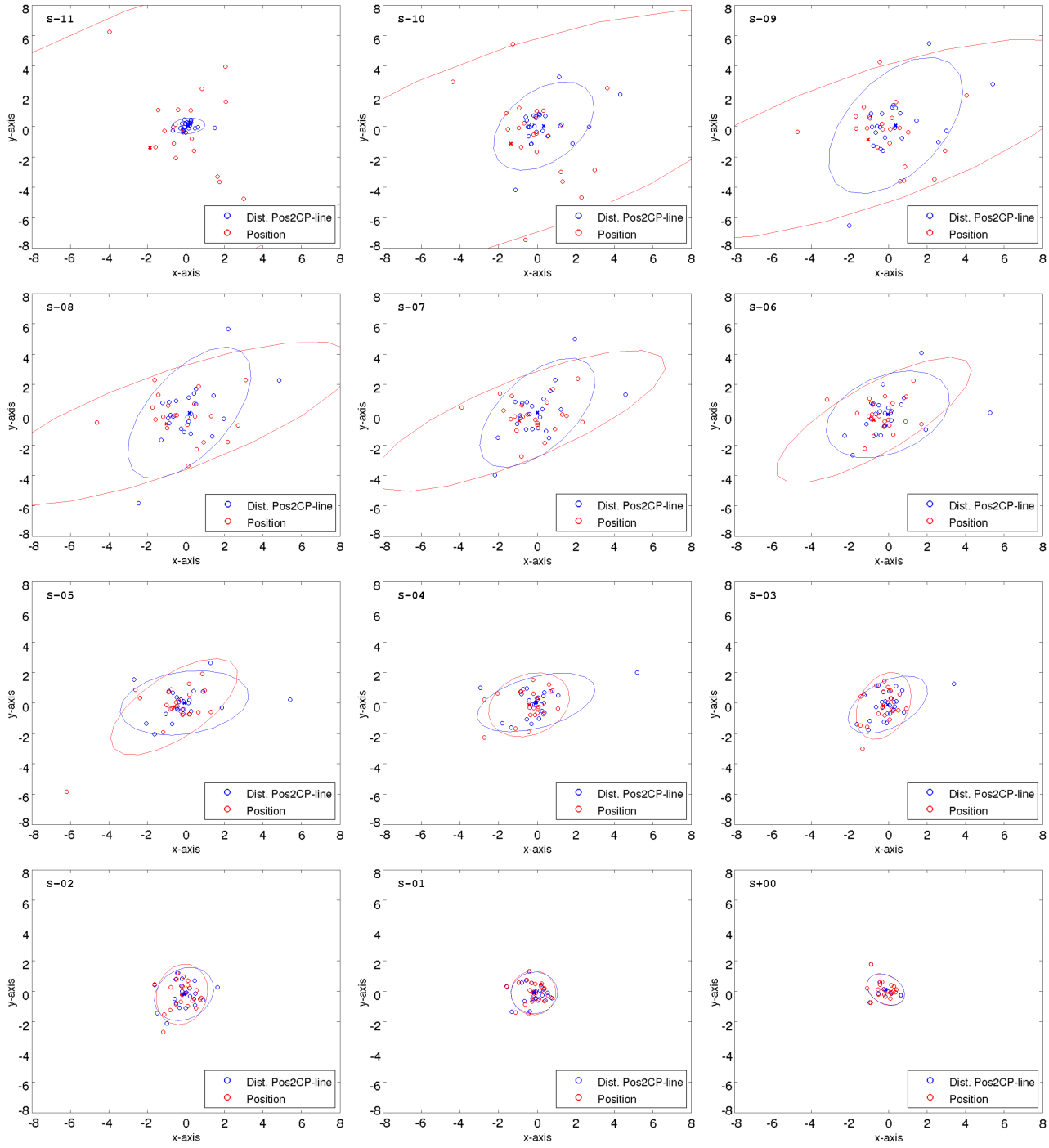


Figure 9: Slice S $= -11, \ldots, 0$ where 0 mark the base-slice (i.e. $-\frac{L-1}{2}, \ldots, 0$). Each plot contains 23 red and blue circles representing θ^{nl} (Position) and η^{nl} (Dist. Pos2CP-line) respectively, where $N = 23$ is the number of training cases and $n = 1, \ldots, N$. The means μ_θ^l and μ_η^l are depicted by small crosses. The corresponding covariance matrices are depicted by 90% confidence ellipses.

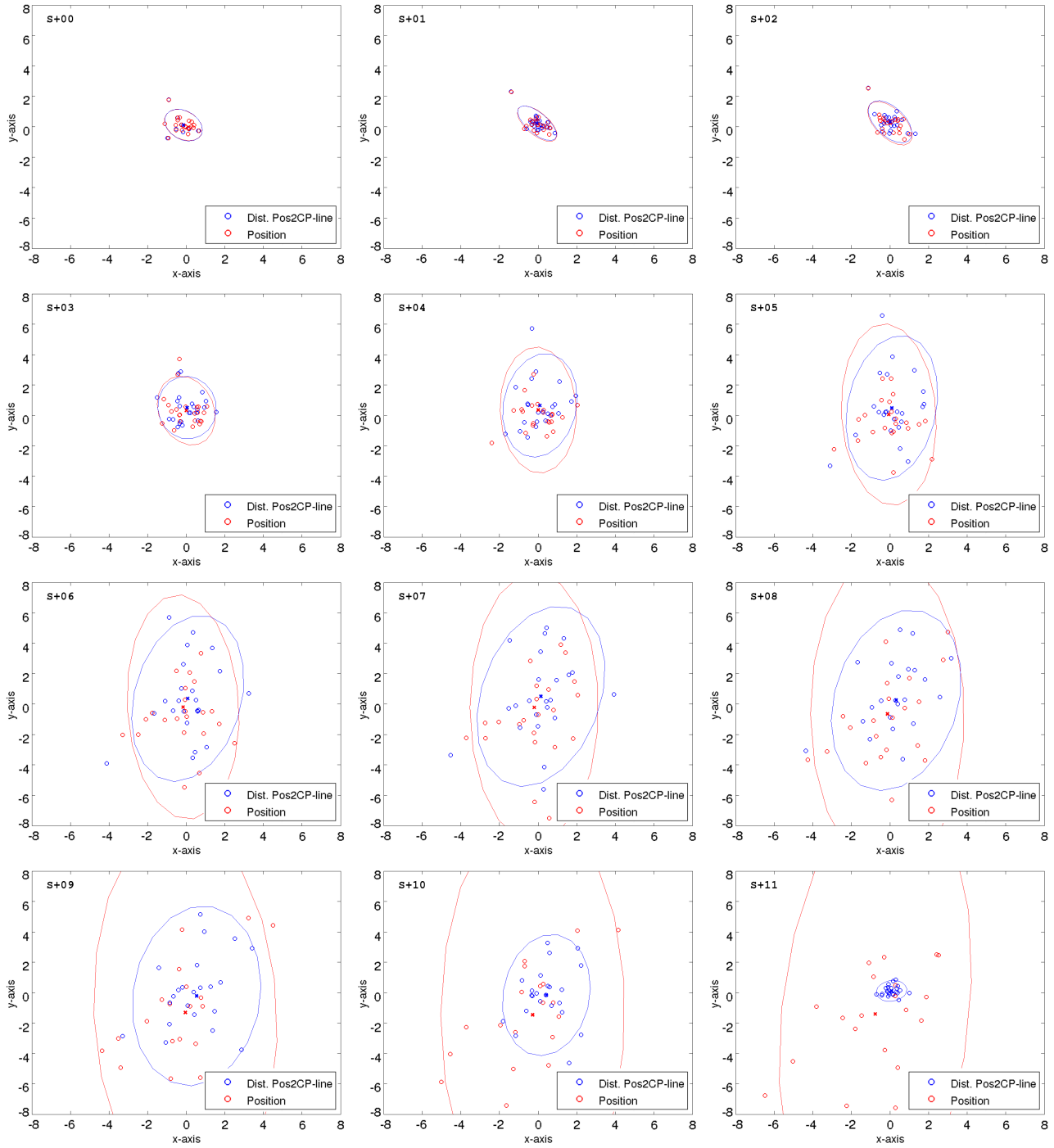


Figure 10: Slice S $= 0, \ldots, 11$ where 0 mark the base-slice (i.e. $0, \ldots, \frac{L-1}{2}$). Each plot contains 23 red and blue circles representing θ^{nl} (Position) and η^{nl} (Dist. Pos2CP-line) respectively, where $N = 23$ is the number of training cases and $n = 1, \ldots, N$. The means μ_θ^l and μ_η^l are depicted by small crosses. The corresponding covariance matrices are depicted by 90% confidence ellipses.

We assume different mean image gray values ν^0 inside and ν^1 outside of the object and constant variance τ^2 as proposed in Dryden and Mardia [3]. Therewith, the likelihood can be modeled by

$$L(S \mid \rho, CP) \propto \exp\left(-\frac{1}{2\tau^2} \sum_{l=1}^{L} \sum_{i \in I(\rho_l|_{CP})} (s_{il} - \nu^0)^2 \right.$$
$$\left. -\frac{1}{2\tau^2} \sum_{l=1}^{L} \sum_{i \notin I(\rho_l|_{CP})} (s_{il} - \nu^1)^2 \right)$$

with $s_i = \nu^0 + \epsilon_i$ if $i \in I(\rho|_{CP})$ and $s_i = \nu^1 + \epsilon_i$ if $i \notin I(\rho|_{CP})$ and an error distribution of $\epsilon_i \overset{\text{iid}}{\sim} N(0, \tau^2)$.

Using the suggest prior distribution from section 2 in the main article, the prior $\pi(\rho \mid CP) = \pi(\theta, \alpha, \phi \mid CP)$ is given by

$$\pi(\rho \mid CP) \propto \exp\left(-\beta \sum_l \frac{1}{2(\sigma^l_{\theta_1})^2} (\theta^l_1 - \mu^l_{\theta_1})^2 - \beta \sum_l \frac{1}{2(\sigma^l_{\theta_2})^2} (\theta^l_2 - \mu^l_{\theta_2})^2 \right)$$
$$* \exp\left(-\beta \sum_l \frac{\mu^l_{\alpha_1}}{2(\sigma^l_{\alpha_1})^2} (\alpha^l_1 - \ln(\mu^l_{\alpha_1}))^2 \right)$$
$$* \exp\left(-\beta \sum_l \frac{\mu^l_{\alpha_2}}{2(\sigma^l_{\alpha_2})^2} (\alpha^l_2 - \ln(\mu^l_{\alpha_2}))^2 \right)$$
$$* \exp\left(-\beta \sum_l \frac{1}{2(\sigma^l_{\phi})^2} (\phi^l - \mu^l_{\phi})^2 \right) \Bigg|_{CP}$$

with $\theta^l_i = \xi^l_i + \eta^l_i, i = 1, 2$ (see section 3.4). The weight parameter $\beta \in \mathbb{R}_+$ describes the importance of the prior compared to the likelihood. In addition, we extend the prior by taking into account the variation between neighbor slices by

$$\pi_1(\rho \mid CP) \propto \pi(\rho \mid CP)$$
$$* \exp\left(-\tilde{\beta}_1 \sum_{i=1,2} \sum_l (\theta^{l-1}_i - \theta^l_i)^2 + (\theta^{l+1}_i - \theta^l_i)^2 \right)$$
$$* \exp\left(-\tilde{\beta}_2 \sum_{i=1,2} \sum_l (\alpha^{l-1}_i - \alpha^l_i)^2 + (\alpha^{l+1}_i - \alpha^l_i)^2 \right)$$
$$* \exp\left(-\tilde{\beta}_3 \sum_l (\phi^{l-1} - \phi^l)^2 + (\phi^{l+1} - \phi^l)^2 \right) \Bigg|_{CP},$$

where $\tilde{\beta}_1 = \beta\beta_1, \tilde{\beta}_2 = \beta\beta_2, \tilde{\beta}_3 = \beta\beta_3$ and $\beta_1, \beta_2, \beta_3 \in \mathbb{R}_+$ are regularization parameters which have to be chosen. Therewith, we introduce a constraint that small changes between neighbor slices are

more likely than jumps. An alternative approach is to take into account neighbor relations by the second derivatives, i.e.,

$$\begin{aligned}\pi_2(\rho \mid CP) \propto \pi(\rho \mid CP) \\ * \exp\left(-\tilde{\beta}_1 \sum_l (\theta_1^{l+1} - 2\theta_1^l + \theta_1^{l-1})^2 + (\theta_2^{l+1} - 2\theta_2^l + \theta_2^{l-1})^2\right) \\ * \exp\left(-\tilde{\beta}_2 \sum_l (\alpha_1^{l+1} - 2\alpha_1^l + \alpha_1^{l-1})^2 + (\alpha_2^{l+1} - 2\alpha_2^l + \alpha_2^{l-1})^2\right) \\ * \exp\left(-\tilde{\beta}_3 \sum_l (\phi^{l+1} - 2\phi^l + \phi^{l-1})^2\right)\Big|_{CP}\end{aligned}$$

with $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3$ as before. Hence, the posterior becomes

$$\pi(\rho \mid S, CP) \propto L(S \mid \rho) * \pi_i(\rho \mid CP), \quad i \in \{1, 2\}. \tag{13}$$

Estimation of $\pi(\rho \mid S, CP)$ can be done by Markov chain Monte Carlo (MCMC) method using the Metropolis-Hastings algorithm [4]. The MCMC approach avoids the computation of the unknown normalization constant in the posterior $\pi(\rho \mid S, CP)$ by drawing independent samples from a proposal distribution and the generation of a Markov chain. The MCMC method is implemented as follows:

1. Choosing of an arbitrary initial estimate of $\rho|_{CP}$, e.g, by the prior model.

2. Generation of a random permutation $\mathbb{I}_p = \{i_1, \ldots, i_L\}$ from the index set $\{1, \ldots, L\}$ whereas L is the number of slices. Therewith, we update the slices in a random order.

3. Generation of a new random realization from the proposal distribution $g_(\rho_{new,i_k} \mid \rho_{old,i_k})$, e.g., by $\rho_{new,i_k} = (\theta^{new,i_k}, a^{new,i_k}, \phi^{new,i_k})^T \sim N(\rho_{old,i_k}, \chi^2_{i_k})$ where $\chi^2_{i_k}$ can be chosen depending on the prior distribution by $\chi_{i_k} = diag(\lambda_1 \sigma^{i_k}_{\theta_1}, \lambda_2 \sigma^{i_k}_{\theta_2}, \lambda_3 \sigma^{i_k}_{a_1}, \lambda_4 \sigma^{i_k}_{a_2}, \lambda_5 \sigma^{i_k}_{\phi})$, $k = 1, \ldots, L$ where

λ_i, $i = 1, \ldots, 5$ are weights that must be chosen. We assume a density

$$g(\rho_{new,i_k} \mid \rho_{old,i_k}) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{2} \frac{1}{(\sigma_{\theta_j}^{i_k})^2}(\theta_j^{new,i_k} - \theta_j^{old,i_k})^2 - \frac{1}{2}\sum_{j=1}^{2} \frac{1}{(\sigma_{a_j}^{i_k})^2}(a_j^{new,i_k} - a_j^{old,i_k})^2 - \frac{1}{2}\sum_{j=1}^{2} \frac{1}{(\sigma_{\phi_j}^{i_k})^2}(\phi_j^{new,i_k} - \phi_j^{old,i_k})^2 \right).$$

4. Calculation of the Hastings-ratio

$$p = \frac{\pi(\rho_{new} \mid S, CP) g(\rho_{old} \mid \rho_{new})}{\pi(\rho_{old} \mid S, CP) g(\rho_{new} \mid \rho_{old})}$$

using the derived formula 13 for the posterior density. Note, the used proposal density is symmetric and therefore

$$p = \frac{\pi(\rho_{new} \mid S, CP)}{\pi(\rho_{old,k} \mid S, CP)}.$$

5. Accept $\rho_{i_k} = \rho_{new,i_k}$ with probability $min(1, p)$ otherwise keep $\rho_{i_k} = \rho_{old,i_k}$.

6. Repeat steps 3 to 5 L-times until each slice is updated.

Typically the MCMC method consist of a burn-in and a sample period by an iteration of steps 2 to 6. The final $\rho|_{CP}$ is calculated by the average of the sample period. In our simulations we used a burn-in and sample period of 500 samples each. Furthermore, calculations are done inside a mask around the registered prior shape to decrease computation time and to avoid additional variance by other structures in the lower male pelvis.

5 Additional data analysis

This section reports additional results for the prior as well as for the MCMC method described in Section 4 before.

5.1 Prior results

In Section 3 different approaches of control points are discussed. We obtain a different mean shape as a prior corresponding to the used method. The results in Section 3 in the main article assume

the control point method CP3. Alternatively, Table 2 shows the the evaluation results between the manual delineation and the deformed MBFE to the manual contours for each test data set using CP2. All evaluation criterias in Table 3 (in the main article) show a better performance than in Table 2. We do not show the results of control point method CP1 here because the results are even less accurate than for CP2. Even if we observe a less accurate performance of CP2 compared to CP3, the method has the advantage of fewer control points in the outer slices where the definition of the delineation is more challenging. Therefore, this method is an object of further studies.

Table 2: CP2: Evaluation metrics between Mean Best Fitting Ellipse and manual delineations given by the physician. HD mean is given in mm, Dice 3D and Accuracy in percentage.

Test set	1	2	3	4	5	6	7	8	9	10	μ
HD mean	4.58	2.77	6.32	3.61	4.12	4.66	3.43	6.82	4.03	4.04	4.44
Dice 3D	0.83	0.90	0.85	0.87	0.87	0.83	0.89	0.84	0.92	0.89	0.87
Accuracy	0.69	0.81	0.74	0.76	0.73	0.63	0.80	0.64	0.84	0.78	0.74

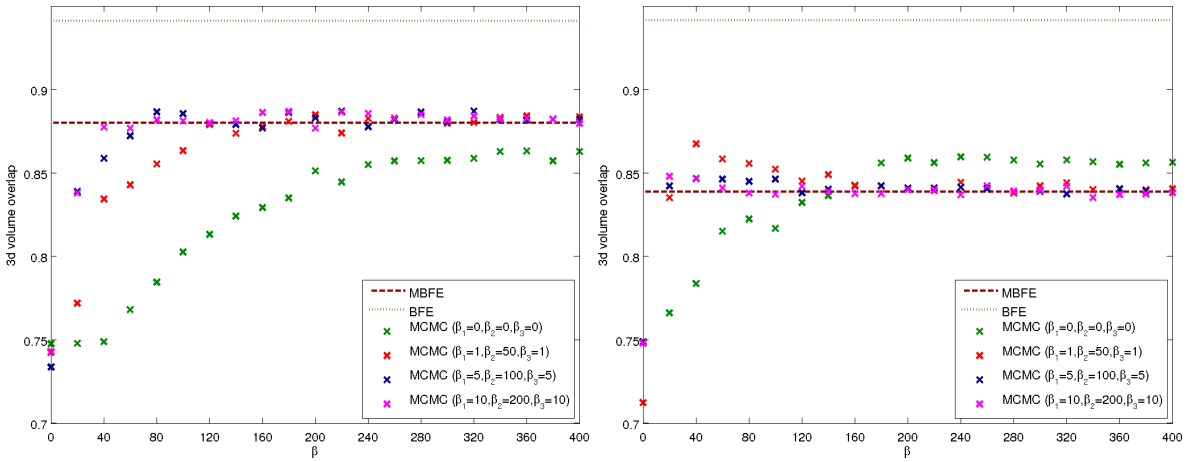
5.2 MCMC results

In this section we discuss results after MCMC as described in Section 4 with $\pi_2(\rho \mid CP)$ as the prior distribution whereas we assume control points method CP3 as described in Section 3.3. In the case of CP3, we do not update the first, center and last slice because the underlying contour is described reasonable well by the control point.

The weight β and regularization parameters β_1, β_2 and β_3 of the prior distribution have to be chosen in practice. A simulation study (reported below) motivates, e.g., a choice of $\beta = 40$ and $(\beta_1, \beta_2, \beta_3) = (2, 180, 10)$.

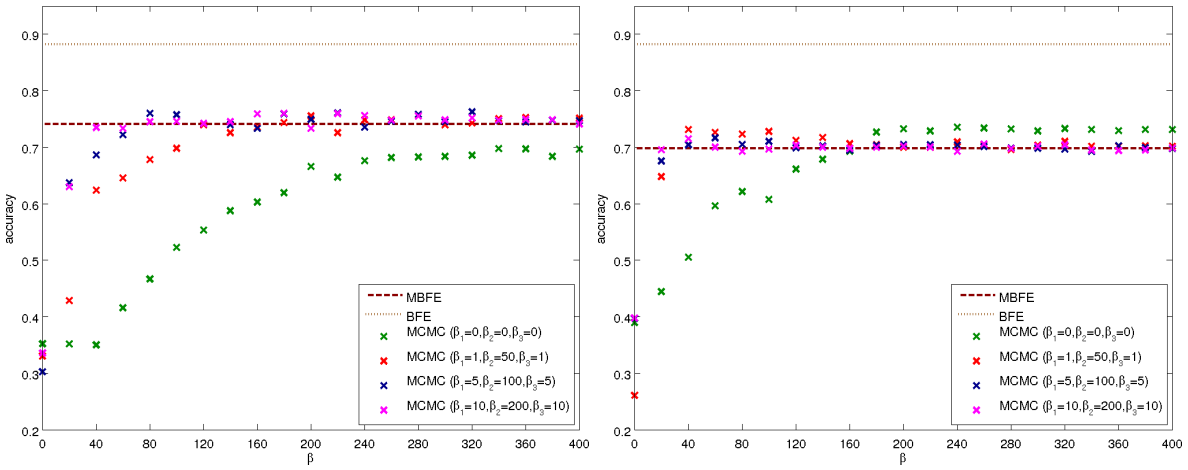
Figures 11 and 12 depict the influence of different β, β_1, β_2 and β_3 in formula (13) by a comparison of the distances (7)-(9) as described in the main article. $\beta = 0$ means that the posterior is only driven by the likelihood. On the other hand, $\beta \to \infty$ means we ignore the likelihood and the posterior is only determined by the prior. We can observe the volume overlap and accuracy converged against the prior for increasing weights β. The regularization parameters β_1, β_2 and β_3 control the speed of the convergence to the prior (MBFE) and stabilize the results.

Based on initial simulations with several parameter combinations, we replicated 200 times the MCMC step for 3 different patients from the test data set with $\beta \sim U(0, 100)$, $\beta_1 \sim U(0, 20)$,



(a) test data set 5 (b) test data set 6

Figure 11: Volume overlap after MCMC using different $(\beta, \beta_1, \beta_2, \beta_3)$ settings for two different test data sets compared to the volume overlap of BFE and MBFE with manual delineation line.



(a) test data set 5 (b) test data set 6

Figure 12: Accuracy after MCMC using different $(\beta, \beta_1, \beta_2, \beta_3)$ settings for two different test data sets compared to the accuracy of BFE and MBFE with manual delineation line.

$\beta_2 \sim U(0, 200)$ and $\beta_3 \sim U(0, 30)$ where $U(a, b)$ denote the standard uniform distribution on the open interval (a, b). The results with the two highest dice coefficients (volume overlap) are shown in Table 3.

On the basis of the simulation results we have chosen five different MCMC parameter settings $M_i = \{\beta, \beta_1, \beta_2, \beta_3\}$ by $M_1 = \{94, 6, 181, 28\}$, $M_2 = \{32, 1.6, 173, 1\}$, $M_3 = \{53, 1.1, 140, 12\}$, $M_4 = \{40, 2, 180, 10\}$ and $M_5 = \{80, 10, 180, 20\}$. Tables 4, 5 and 6 illustrate the distances between

Table 3: Prior distribution parameters ranked by the two highest dice 3D coefficients after 200 simulations for 3 different test data sets.

Test data set	Weights for the prior distribution β	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	Dice 3D	Accuracy
5	94.08	15.95	181.13	27.92	0.89	0.77
	92.17	13.22	118.95	10.91	0.89	0.76
6	53.64	1.14	138.84	12.48	0.87	0.74
	47.42	1.10	27.69	5.27	0.86	0.73
8	32.36	1.62	173.12	0.95	0.90	0.80
	28.31	1.75	181.46	19.35	0.90	0.80

the manual delineation and the MCMC result for each test data set and the five different parameter sets. For comparison M_0 shows the results from the registered mean shape without applying MCMC as we have seen in Table 2 and 3 in the main article. A comparison with the results indicate that we do not gain large improvement by MCMC compared to the registered mean shape using the control points. In fact, the 10 test data result in a median volume overlap and accuracy of 0.90, 0.81 for M_0 (MBFE) compared to 0.91, 0.82 for the parameter set M_4 (MCMC), with a median absolute deviation (MAD) of $0.021, 0.035$ compared to $0.021, 0.040$. The values show a slight overall improvement after MCMC. Particularly, there is an improvement of the accuracy in Table 6 for all patients in case of parameter set M_5. Furthermore, test data set 9 in Table 5 shows a significant improvement of the volume overlap after MCMC. The slices of this patient are depicted in Figure 13 together with the manual contour information, contour line from deformed mean shape and contour line after MCMC. Nevertheless, we believe the MCMC procedure is a too time consuming step in the praxis compared to the deformation of the mean shape given control points.

Table 4: HD mean between MCMC results and manual delineations given by the physician using different parameter sets $M_i = \{\beta, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3\}$ (in mm).

	Test data set											
	1	2	3	4	5	6	7	8	9	10	md	MAD
M_0	2.51	2.17	4.24	2.83	3.75	4.63	1.95	4.67	4.27	4.42	4.00	0.655
M_1	2.26	1.95	4.38	2.96	4.02	4.63	1.93	4.58	2.91	4.47	3.49	1.034
M_2	2.18	1.96	4.63	2.88	4.65	4.63	1.89	4.34	3.88	5.00	4.11	0.719
M_3	2.26	2.05	4.81	2.99	4.09	4.63	1.90	4.75	3.68	4.59	3.88	0.883
M_4	2.17	2.15	4.43	3.07	4.36	4.63	1.71	4.33	4.01	4.53	4.17	0.413
M_5	2.19	1.92	4.12	2.75	4.03	4.63	1.93	4.67	2.58	4.46	3.39	1.136

To verify further the proposed posterior distribution in formula (13), we compare the volume

Table 5: Dice 3D between MCMC results and manual delineations given by the physician using different parameter sets $M_i = \{\beta, \beta_1, \beta_2, \beta_3\}$ (in percentage).

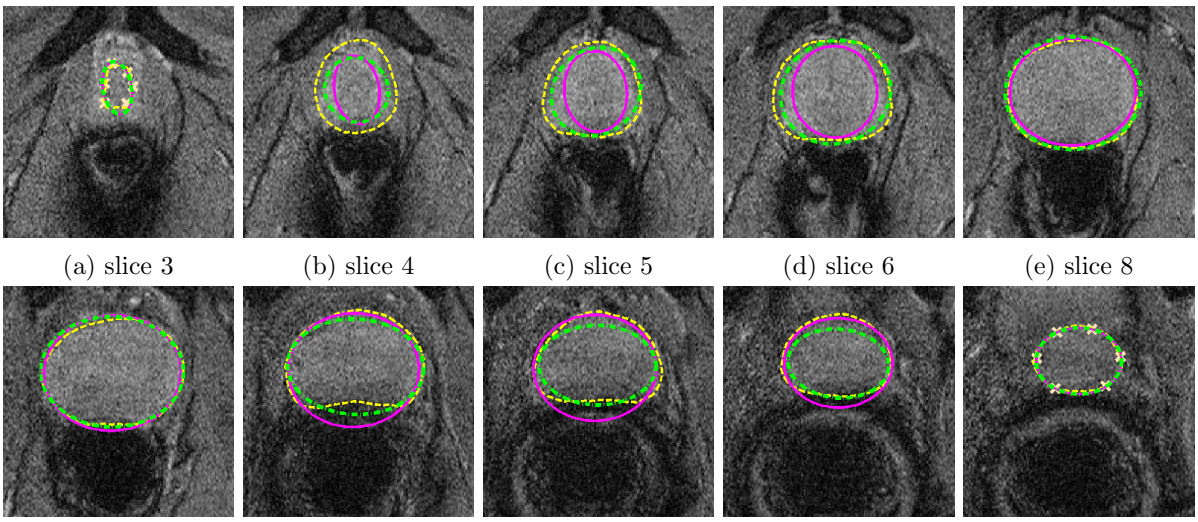
	Test data set											
	1	2	3	4	5	6	7	8	9	10	md	MAD
M_0	0.92	0.93	0.90	0.91	0.88	0.84	0.92	0.88	0.89	0.88	0.90	0.021
M_1	0.93	0.93	0.91	0.90	0.88	0.84	0.91	0.88	0.92	0.88	0.91	0.023
M_2	0.93	0.93	0.90	0.91	0.84	0.85	0.91	0.90	0.92	0.87	0.90	0.026
M_3	0.93	0.93	0.91	0.91	0.88	0.86	0.90	0.89	0.92	0.88	0.90	0.023
M_4	0.93	0.93	0.91	0.90	0.87	0.85	0.91	0.89	0.92	0.88	0.91	0.021
M_5	0.93	0.93	0.91	0.90	0.88	0.84	0.91	0.88	0.93	0.89	0.90	0.021

Table 6: Accuracy between MCMC results and manual delineations given by the physician using different parameter sets $M_i = \{\beta, \beta_1, \beta_2, \beta_3\}$ (in percentage).

	Test data set											
	1	2	3	4	5	6	7	8	9	10	md	MAD
M_0	0.84	0.84	0.81	0.82	0.74	0.70	0.84	0.73	0.80	0.74	0.81	0.035
M_1	0.86	0.86	0.82	0.81	0.75	0.70	0.83	0.74	0.85	0.75	0.82	0.047
M_2	0.87	0.86	0.80	0.83	0.64	0.72	0.82	0.80	0.85	0.71	0.81	0.046
M_3	0.86	0.85	0.81	0.83	0.73	0.74	0.81	0.79	0.85	0.74	0.81	0.042
M_4	0.87	0.86	0.82	0.82	0.71	0.73	0.84	0.78	0.85	0.75	0.82	0.040
M_5	0.86	0.86	0.82	0.82	0.75	0.71	0.84	0.76	0.86	0.76	0.82	0.047

overlap and accuracy in Figure 14 between MCMC results and MBFE result based on different volume information. The MCMC results are calculated for different β, fixed $(\beta_1, \beta_2, \beta_3) = (2, 180, 10)$ and on the basis of:

1. the original volume information,
2. a derived volume from the original by setting all gray values inside the manual delineation line to 100 and outside to 130 added by some error with the same variance inside and outside,
3. a derived volume from the original by setting all gray values inside the best fitting ellipse to 100 and outside to 130 added by some error with the same variance inside and outside,
4. a derived volume from the original by setting all gray values inside the manual delineation line to 80 and outside to 150 added by some error with the same variance inside and outside,
5. a derived volume from the original by setting all gray values inside the best fitting ellipse to 80 and outside to 150 added by some error with the same variance inside and outside.



(a) slice 3 (b) slice 4 (c) slice 5 (d) slice 6 (e) slice 8

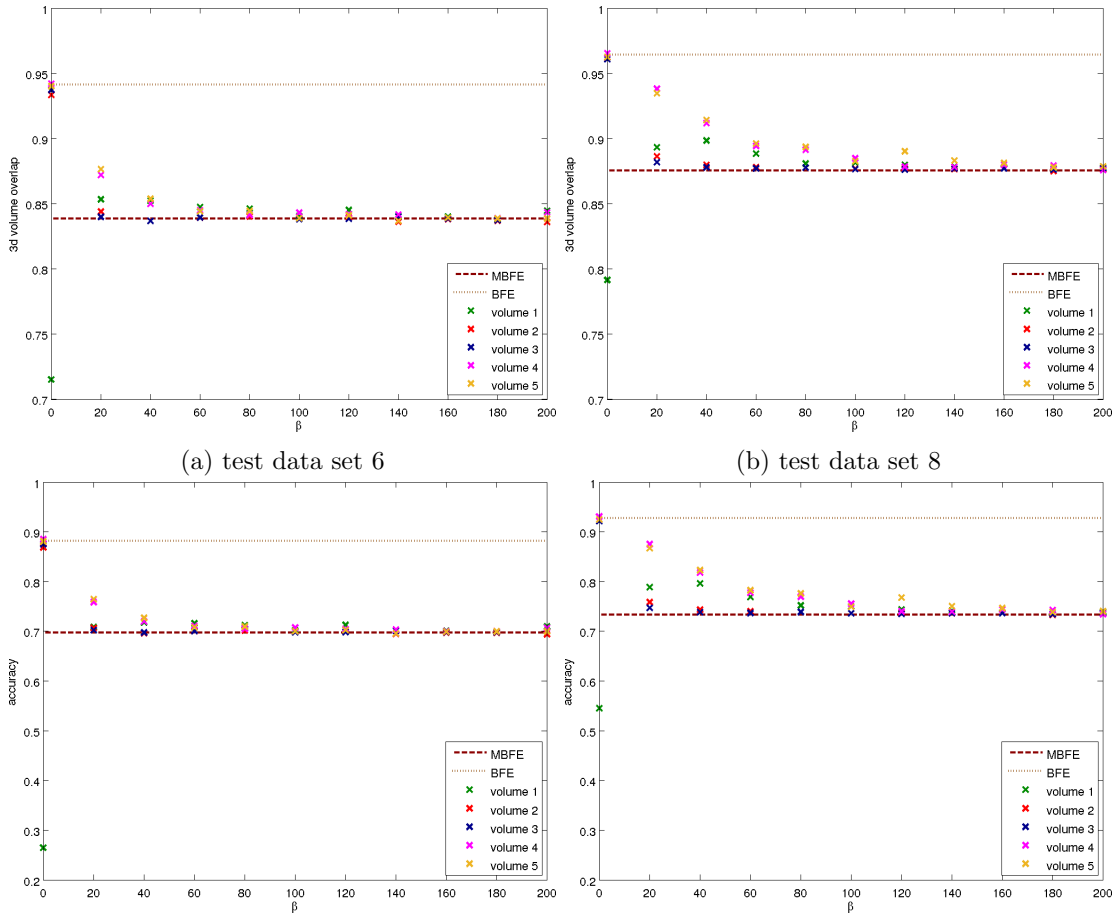
(f) slice 10 (g) slice 12 (h) slice 13 (i) slice 14 (j) slice 15

Figure 13: Selected MR slices of test data set 9 in Table 5 and 6. Slice 3 is the first slice and slice 15 the last slice with prostate information in the data set. The control point of the first and last slice are depicted in (a) and (j). Depicted are manual delineation line (yellow dashed line), deformed mean shape (magenta solid line) and result after MCMC (green dashed dotted line). Slices 7, 9 and 11 are not depitecd in the Figure.

The standard deviation of the error is chosen to be 50, i.e., $\epsilon \sim N(0, \sigma^2)$ with $\sigma = 50$.

In general, we can note the MCMC procedure works well in case of an ideal underlying image information. We do not need prior information if we use volume 2 to 5. The true underlying delineation can be found by the likelihood. Volume overlap and accuracy are close to the results obtained by the BFE for $\beta = 0$ which indicate the best possible values using the proposed model. Furthermore, we observe a decline of the volume overlap and accuracy for larger β using volume 2 to 5, and a convergence to the prior for $\beta \to \infty$. The convergence rate is different between both test data set which which can be explained by a less or more complex structure of the prostate. In addition, the convergence to the prior is faster using volume 2 and 3 compared to volume 4 and 5 due to the lower contrast in inside and outside of the object in volume 2 and 3. In case of the the original volume information (volume 1), the prior contains the most valuable information about the prostate contour. We can observe the largest improvement after MCMC for both data set in the range of $\beta = 20, 40$ and 60.

In Figure 14 is shown low information in the MR volumes is a limitation for the MCMC



(a) test data set 6 (b) test data set 8

(c) test data set 6 (d) test data set 8

Figure 14: Volume overlap ((a),(b)) and accuracy ((c),(d)) of the MCMC results with the manual delineation using different background images and for different β for test data sets 6 and 8. (volume 1): Using the original volume information. (volume 2): Derived volume by setting all gray values inside the best fitting ellipse to 100 and outside to 130 plus an error. (volume 3): Derived volume by setting all gray values inside the manual delineation line to 100 and outside to 130, plus an error. (volume 4): Derived volume by setting all gray values inside the best fitting ellipse to 80 and outside to 150, plus an error. (volume 5): Derived volume by setting all gray values inside the best fitting ellipse to 80 and outside to 150 plus an error.

approach. In addition, a careful choice of β, β_1, β_2 and β_3 is a difficult task. Several attempts to improve the MCMC results were carried out, e.g., different models for the likelihood, calculation of the Hastings ratio per slice instead of taking into account the entire volume and using the median instead of the mean to calculate the average of the gray values inside and outside the contour for the likelihood. Another idea was to select slices with a significant difference of the histogram inside

and outside of the deformed mean shape, e.g., by a Kolmogorv Smirnov test, a Andersson Darling test or by calculation of the earth mover's distance. All our attempts did not significantly improve the MCMC results for all test data sets so far.

Nevertheless, a potential improvement by a different likelihood model is a topic of further reaseach. Such a likelihood should model the underlying volume structure more carefully. Further possible directions are to use the MCMC method to sample from probability distributions on the space of curves as suggested by Fan et al. [5], to derive additional constraints from the training data set or to create an appearance model as suggested by Cootes and Taylor [6].

References

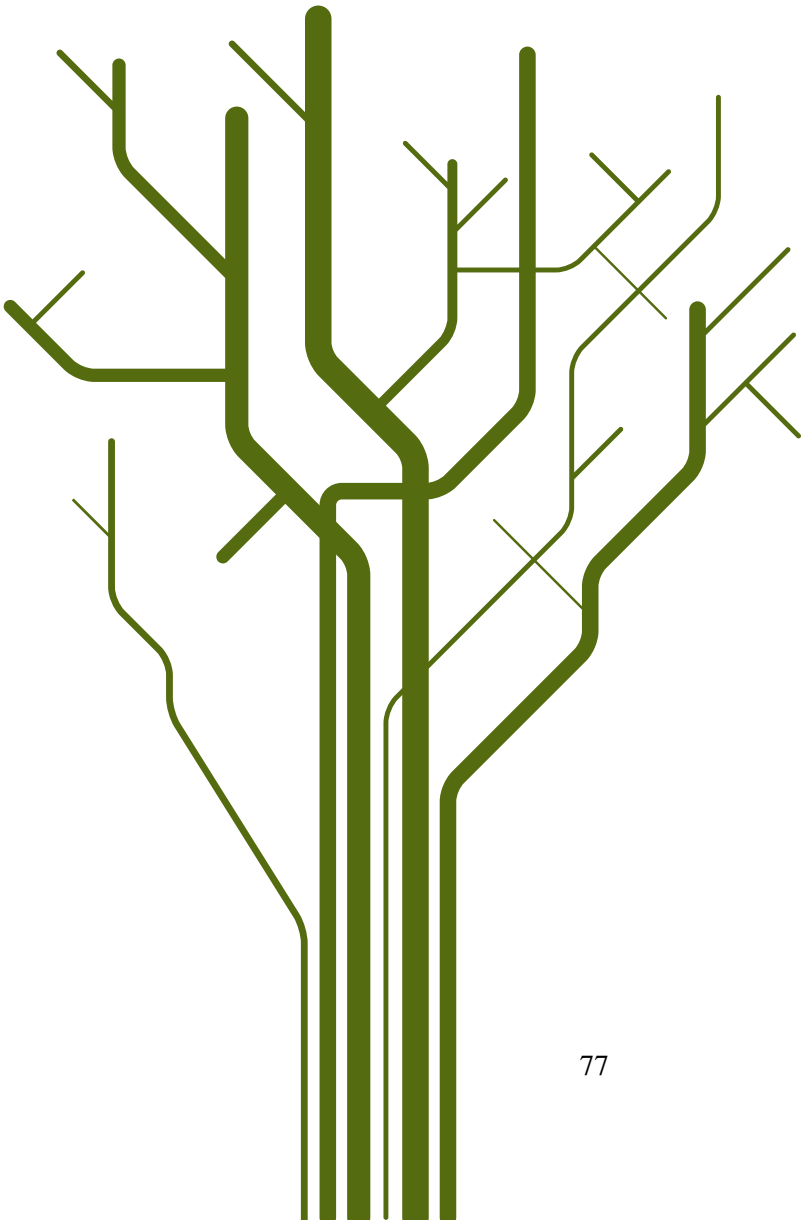
1. Altmann SL: *Rotations, quaternions, and double groups*. Dover books on mathematics, Dover Publications 2005.
2. Casella G: **An Introduction to Empirical Bayes Data Analsysis**. *The American Statistician* 1985, **39**(2):83–87.
3. Dryden IL, Mardia KV: *Statistical Shape Analysis*. John Wiley & Sons 1998.
4. Robert CP, Casella G: *Monte Carlo Statistical Method*. Springer, 2 edition 2004.
5. Fan AC, Fisher JWI, Kane J, Willsky AS: **MCMC curve sampling and geometric conditional simulationSPIE**. In *Computational Imaging VI, Volume 6814* 2008:681407–681407–12.
6. Cootes T, Taylor C: **Statistical Models of Appearance for Computer Vision**. Tech. rep., Imaging Science and Biomedical Engineering 2004, [http://www.isbe.man.ac.uk].

Paper II

Analysis of rotational deformations from directional data

Schulz, Jörn, Jung, Sungkyu, Huckemann, Stephan, Pierrynowski, Michael, Marron, J.S. and Pizer, Stephen M.

revised and re-submitted to *Journal of Computational and Graphical Statistics*



Analysis of Rotational Deformations from Directional Data

Jörn Schulz[1] *, Sungkyu Jung[2], Stephan Huckemann[3], Michael Pierrynowski[4], J. S. Marron[5] and Stephen M. Pizer[5]

[1]University of Tromsø, [2]University of Pittsburgh, [3]University of Göttingen, [4]McMaster University, Hamilton [5]University of North Carolina at Chapel Hill

Abstract

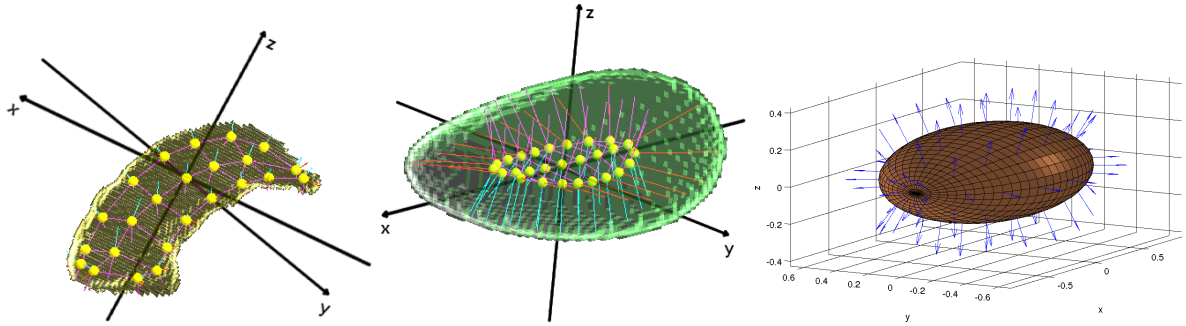
This paper discusses a novel framework to analyse rotational deformations of real 3D objects. The rotational deformations such as twisting or bending have been observed as the major variation in some medical applications, where the features of the deformed 3D objects are directional data. We propose modelling and estimation of the global deformations in terms of generalized rotations of directions. The proposed method can be cast as a generalized small circle fitting on the unit sphere. We also discuss the estimation of descriptors for more complex deformations composed of two simple deformations. The proposed method can be used for a number of different 3D object models. Two analyses of 3D object data are presented in detail: one using skeletal representations in medical image analysis as well as one from biomechanical gait analysis of the knee joint. Supplementary Materials are available online.

Keywords: 3D object, axis of rotation, directional statistics, skeletal model, small circle.

1 Introduction

Modeling deformations of a real object is a central issue in computer vision, biomechanics and medical imaging. In a number of applications, generalized rotations appear to be the

*This research was funded by the Norwegian Research Council through grant 176872/V30 in the eVita program and performed as part of Tromsø Telemedicine Laboratory.



(a) (b) (c)

Figure 1: 3D objects and their models. (a) S-rep of a hippocampus (b) S-rep of a rotationally deformed ellipsoid. (c) Attached boundary normals on the meshed surface of an ellipsoid.

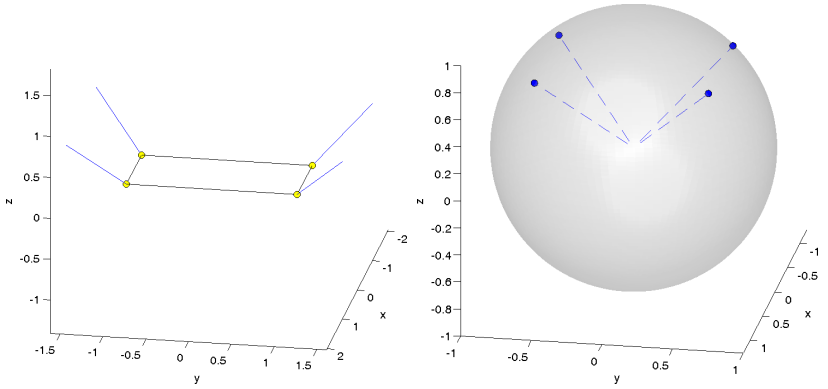
major forms of deformation. For instance, the major variation of shapes of hippocampi in the human brain has been shown to be bending of the object (Joshi et al., 2002; Pizer et al., 2013); Human joint movements, such as the motion of the knee or the elbow, consist of bending and twisting about the joint (Rivest, 2001; Rivest et al., 2008; Oualkacha and Rivest, 2012). A direct modeling of such rotational deformations will promote a precise description of object variation and will be important for surgery or treatment planning.

In this paper, we propose an estimation procedure for descriptors of underlying rotational deformations from a random sample of objects. Specifically, the descriptors are parameters of the model we introduce in Section 2; they include rotational axes of a rotational model. Our model embraces a number of different types of deformations including rigid rotation, bending, twisting and a mixture of the last two. Although we aim to analyze variations in sophisticated human organs such as the hippocampus (Fig. 1a), we work with a simpler object resembling ellipsoids (Fig. 1b) to show the validity of the proposed method.

A major challenge in modeling rotational deformation is that such variations are typically mixed with translational and scaling effects. We address this issue by only considering direction vectors, which are invariant to translation and size changes. It will be shown that the rotational deformation can be sufficiently modeled using directional data. Another advantage of our approach is that well-studied directional data techniques can be applied (Fisher et al., 1993; Mardia and Jupp, 2000; Chang and Rivest, 2001; Jung et al., 2011).

Before we introduce our method, we point out several modeling approaches of 3D objects that are relevant to our framework of directional data, as follows:

Point distribution model A solid object is modeled by the positions of the sampled surface points on which directions normal to the surface can be attached (Cootes et al., 1992; Dryden and Mardia, 1998; Kurtek et al., 2011). See Fig. 1c.



(a) A simple 3D object (b) Direction vectors

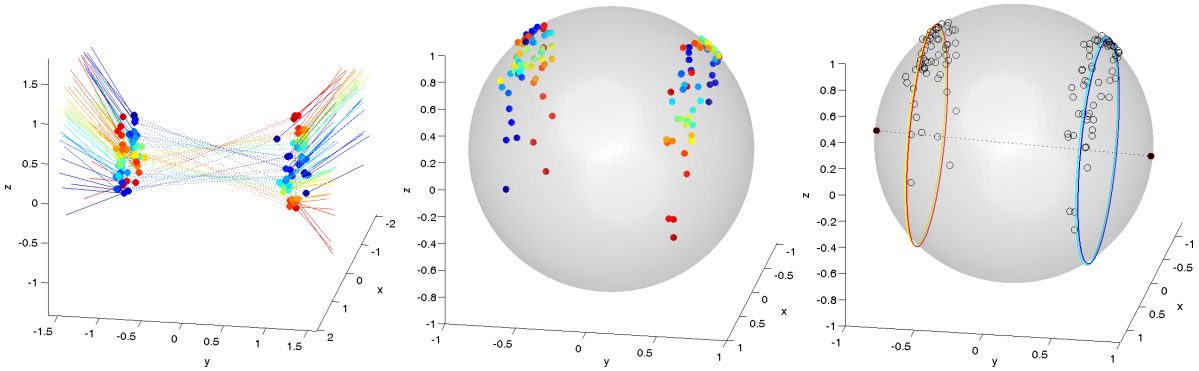
Figure 2: Toy example. (a) A toy object, to be deformed. (b) Each direction vector is a point on the surface of the unit sphere.

Large deformations The shape changes of an object in images are modeled by the deformations of a template image (Pennec, 2008; Rohde et al., 2008). The deformation can be understood as a vector field, where each vector contains the direction.

S-rep In skeletal representations (s-rep), a 3D object is modeled by skeletal positions lying inside of the object and spoke vectors pointing to the boundary of the object (Siddiqi and Pizer, 2008; Pizer et al., 2013). See Fig. 1a and Fig. 1b. We describe s-rep data analysis in more detail in Section 5.

The framework of our analysis can be understood by considering a simple example of a 3D object (Fig. 2). The object is modeled by four surface points (or skeletal positions) with attached direction vectors μ_j for $1 \leq j \leq 4$ (Fig. 2a). Consider random twists of the object, where the left and right sides are rotated about a common axis by a common angle but in opposite directions (Fig. 3a). After collecting the directional data (Fig. 3b), our method finds an estimate of the axis (overlaid in Fig. 3c) as well as the mean directions μ_j and the rotation angles. The proposed method provides an simple interpretation of the underlying rotational deformation and accurate estimation of the parameters. The estimated axis in the toy example is close to the true axis with relatively small sample size $n = 30$. See Section 2.3 for detailed discussion of this example.

To the best of our knowledge, this paper is the first attempt to model rotational deformations and to estimate the axis of rotations from directional data. There are, however, several methods available for the estimation of the axis of rotation based on 3D landmarks, especially in the area of biomechanics (Halvorsen et al., 1999; Rancourt et al., 2000; Rivest, 2001; Gamage and Lasenby, 2002; Teu and Kim, 2006; Rivest et al., 2008; Ball and Greiner,



(a) Random twists (b) Random directions (c) Estimated axis

Figure 3: Toy example. (a) $n = 30$ realizations of random twists. Different colors represent different observations. (b) Directional features. (c) The estimated axis of the twists (dotted line) is close to truth. The four concentric circles, discussed in more detail in Section 2.3, are also overlaid.

2012). In the statistical literature, estimation of rotation matrices has been studied in terms of spherical regression and its generalizations (Chang, 1986, 1988, 1989; Rivest, 1989, 2006; Chang and Rivest, 2001). In spherical regression both the axis c and the angle θ are estimated. In contrast, our model treats θ as a random variable. Hanna and Chang (2000) used the quaternion representation to model a smooth path of rotations. However, in our problem of estimation of axes of rotational deformations, it is straightforward to use the axis–angle representation. As will be pointed out later and may be guessed from Fig. 3c, the estimation of the axis of rotation based on directional vectors has a close relationship with the estimation of small circles on the unit sphere, which was studied in various contexts (Mardia and Gadsden, 1977; Rivest, 1999; Jung et al., 2011).

This paper is a part of a bigger project in modeling and estimation of deformations. We leave relevant asymptotic theory as future work.

The rest of the paper is organized as follows. We begin with introducing some necessary facts on rotations and their connections to circles on the unit sphere. In Section 2, a simple rotation model is introduced, and our estimation procedure is discussed. Non-rigid deformations such as twisting and bending of the object are also discussed in that section. A hierarchical rotation model is introduced in Section 3, where the object is assumed to be deformed sequentially by different rotations. In Section 4, simulation results are reported to show the effectiveness of the estimator. In Section 5 and 6, the merits of the proposed methods are further illustrated using applications from models that represent organs and knee motion during gait.

1.1 Rotations, circles and spheres

In the axis–angle representation of rotations, an axis c is the unit vector that is left fixed by the rotation and an angle θ gives the amount of rotation. A unit vector lies on the unit sphere $S^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$. The axis–angle pair $(c, \theta) \in S^2 \times [0, 2\pi)$ represents a rotation in 3-space, where a vector $x \in \mathbb{R}^3$ is rotated by (c, θ) by applying $x \mapsto R(c, \theta)x$ with

$$R(c, \theta) = I_3 + \sin\theta[c]_\times + (1 - \cos\theta)(cc' - I_3), \tag{1}$$

where c' denotes the transpose of c, and $[c]_\times$ is the cross product matrix satisfying $[c]_\times v = c \times v$ for any $v \in \mathbb{R}^3$.

A useful observation in our analysis is that the direction vectors follow circles when they are rotated. In particular, when $x \in S^2$ is rotated about an axis $c \in S^2$, the trajectory of such rotation is precisely a circle, centered at c with radius $r = \arccos(x'c)$, and is denoted by $\delta(c, r) = \{x \in S^2 : x'c = \cos(r)\} \subset S^2$. Since $\delta(c, r) = \delta(-c, \pi - r)$ we may assume that $r \leq \pi/2$. We call $\delta(c, r)$ a great circle if $r = \pi/2$ and a small circle if $r < \pi/2$.

If a K-tuple of $K \geq 2$ direction vectors $\mathbf{x} = (x_1, \ldots, x_K) \in (S^2)^K$ are rotated together about a common axis c, then each of the rotated direction vectors is on a circle with common center c but with different radii $r_j = \arccos(c'x_j)$, $j = 1, \ldots, K$. Denote the collection of concentric circles with a common center c and radii tuple $\mathbf{r} = (r_1, \ldots, r_K) \in [0, \pi/2] \times [0, \pi]^{K-1}$ by

$$\delta(c, \mathbf{r}) = \{(x_1, \ldots, x_K) \in (S^2)^K : x_j'c = \cos(r_j),\ j = 1, \ldots, K\}.$$

To work with observations on S^2, the geodesic distance function $d_g : S^2 \times S^2 \to [0, \pi]$ is defined by the arc length of the shortest great circle segment joining $x, y \in S^2$, and is $d_g(x, y) = \arccos(x'y)$. We further define $d_g(x, A) = \inf_{y \in A} d_g(x, y)$ for $x \in S^2, A \subset S^2$. For a random element X whose domain is S^2, a sensible notion of mean $\mu(X)$ is defined by a minimizer of mean squared distance,

$$\mu(X) = \operatorname*{argmin}_{x \in S^2} \mathbb{E}\{d_g^2(x, X)\},$$

often called the geodesic or Fréchet mean (Fréchet, 1948; Karcher, 1977). A useful measure of dispersion is geodesic variance which is defined as $\mathrm{Var}(X) = \mathbb{E}\{d_g^2(\mu(X), X)\} = \min_{x \in S^2} \mathbb{E}\{d_g^2(x, X)\}$ provided that $\mu(X)$ exists.

2 Single rotational deformations

In this section, an estimation procedure for rotational deformation models is proposed. We begin with a discussion on the simpler rigid rotation model.

2.1 Rigid rotation model

Suppose we have a K-tuple of random direction vectors $\mathbf{X} = (X_1, \ldots, X_K)$. For some unknown constants $c, \mu_j \in S^2$ and a latent random variable $\theta \in [-\pi/2, \pi/2)$, we model $X_j \in S^2$ $(j = 1, \ldots, K)$ as noisy observations of rotations of μ_j by $R(c, \theta)$, that is,

$$X_j = R(c, \theta)\mu_j \oplus \epsilon_j \quad (j = 1, \ldots, K). \tag{2}$$

Here, the ϵ_j are independently distributed random error terms, and the $\oplus$ sign defines a specific action of the error distribution as defined in the following.

There are several ways to define random spherical points $X \sim \mu \oplus \epsilon \in S^2$. A natural way is to introduce an S^2-valued distribution, e.g., the von Mises–Fisher distribution (Mardia and Jupp, 2000, p. 36) with the density $f_{\text{vMF}}(x; \mu, \kappa) \propto \exp(\kappa \mu' x)$ with respect to the uniform measure on S^2 for $\mu \in S^2$, $\kappa > 0$. Alternatively, one can utilize the tangent space at $\mu \in S^2$, allowing a distribution on the tangent space to be mapped to S^2. Another approach is to use the embedding of S^2 into $\mathbb{R}^3$, by scaling a three-dimensional random vector to unit length. This approach is often called a perturbation model (Goodall, 1991). It is well-known that a perturbation model introduces a bias in the estimation of the geodesic mean unless the distribution is isotropic (Kent and Mardia, 1997; Le, 1998; Huckemann, 2011a). In this paper in Section 4, we use the von Mises–Fisher distribution and in Section 5 the perturbation model. The latter is justified because we consider only isotropic distributions. In the following discussion, we do not specify a particular distribution for ϵ, but require that the geodesic mean of $X \sim \mu \oplus \epsilon$ is uniquely found at μ, i.e., $\mu = \operatorname{argmin}_x \mathbb{E}\{d_g^2(x, X)\}$. The geodesic variance $\operatorname{Var}(\epsilon) := \operatorname{Var}(\mu \oplus \epsilon)$ is then well defined.

In model (2), several different combinations of θ and μ_j lead to the same model. Specifically, replacing θ and μ_j by $\theta^*(a) = \theta - a$ and $\mu_j^*(a) = R(c, a)\mu_j$ for any $a \in \mathbb{R}$ gives the same model as (2). Therefore, we assume

$$\mathbb{E}\theta = 0. \tag{3}$$

The trajectory of rotated direction vectors forms a small circle (*cf.* Section 1.1), which

is approximately true in the presence of the noise. In other words, the collection of X_j in (2) are distributed along concentric circles with common center at c, as the following lemma states.

Lemma 1. *Let the η-neighborhood of concentric circles $\delta(c, \mathbf{r})$ be*

$$\delta_\eta(c, \mathbf{r}) = \{(X_1, \ldots, X_K) \in (S^2)^K : d_g(\delta(c, r_j), x_j) < \eta \text{ for all } j = 1, \ldots, K\},$$

for $\eta > 0$. If $\mu \oplus \epsilon_j$ are independent and identically distributed and spherically symmetric about μ, then

$$P\{\mathbf{X} \in \delta_\eta(c, \mathbf{r})\} \geq \left\{1 - \frac{Var(\epsilon)}{\eta^2}\right\}^K.$$

The auxiliary parameters $r_j = \arccos(c'\mu_j)$ represent the radii of the concentric circles, and are obtained from c and μ_j, the parameters of (2). A proof of Lemma 1 is given in the Appendix.

Lemma 1 suggests that $\mathbf{X}$ and $\delta(c, \mathbf{r})$ are close with high probability, which motivates to define the population concentric circles $\delta(c_0, \mathbf{r}_0)$ as a minimizer of squared loss. In the view of this estimation strategy, the capability of identifying parameters as minimizers leads to a natural estimation strategy, namely the M–estimation or the sample Fréchet mean (Karcher, 1977; Huckemann, 2011b). The rest of this section is devoted to the identification of the population parameters c, r_j and μ_j as population Fréchet means.

First, the distance function ρ between $\delta(c, \mathbf{r})$ and $\mathbf{x}$ is defined as the Cartesian product metric based on d_g by

$$\rho^2(\delta(c, \mathbf{r}), \mathbf{x}) = \sum_{j=1}^K d_g^2\left(\delta(c, r_j), x_j\right) = \sum_{j=1}^K \left(\arccos(x_j'c) - r_j\right)^2.$$

The collection of population concentric circles $\delta(c_0, \mathbf{r}_0)$ is defined as the Fréchet ρ-mean set

$$\operatorname*{argmin}_{c \in S^2, \mathbf{r} \in [0, \pi/2] \times [0, \pi]^{K-1}} \mathbb{E}\rho^2(\delta(c, \mathbf{r}), \mathbf{X}), \tag{4}$$

where the expectation $\mathbb{E}$ is with respect to the random directions $\mathbf{X}$. We assume in the following that there is a unique minimizer $\delta(c_0, \mathbf{r}_0)$. The center c_0 of the circles also represents the axis of rotation.

It should be noted that there is no guarantee for the true axis of rotation $\check{c}$ of (2) to be the same as c_0 from (4). Simulation studies, reported in the Supplementary Material, have suggested that the case $c_0 = \check{c}$ occurs when $\mathrm{Var}(r_j\theta_j)$ is large enough compared to the error

variance $\mathrm{Var}(\epsilon_j)$ for all j. In our simulation studies in Section 4, the effect of this bias is shown to be small.

While the axis of rotation c is the center of the concentric circles $\delta(c, \mathbf{r})$, each base point μ_j is also a point on $\delta(c, r_j)$, $j = 1, \ldots, K$. The assumption of isotropy of ϵ_j implies that

$$\mu(Y_j^{\theta_0}) = R(c, \theta_0)\mu_j$$

for $Y_j^{\theta_0} = R(c, \theta_0)\mu_j + \epsilon_j$ with deterministic angle $\theta_0 \in [-\pi/2, \pi/2)$. In particular, $\mu(Y_j^0) = \mu_j$. For random θ define

$$\mu(X_j|\theta) := R(c, \theta)\mu_j \,.$$

With the distance function $\rho_{\delta(c,r)}(x, y)$ which measures the shortest arc-length between $x, y \in \delta(c, r)$ along the (small) circle via $\rho_{\delta(c,r)}(x, y) = \sin(r) \arccos[(x'y - \cos^2(r))/\sin^2(r)]$ (Jung et al., 2012) we have by definition

$$\psi_0 = \underset{\psi \in [-\pi/2, \pi/2)}{\operatorname{argmin}} \mathbb{E}\rho^2_{\delta(c,r_j)}(\mu(X_j|\theta), R(c, \psi)\mu_j) = \underset{\psi \in [-\pi/2, \pi/2)}{\operatorname{argmin}} \mathbb{E}(\theta - \psi)^2,$$

which leads to the minimizer $\psi_0 = 0$ due to the assumption (3). Thus,

$$\mu_j = \underset{\mu \in \delta(c,r_j)}{\operatorname{argmin}} \mathbb{E}\rho^2_{\delta(c,r_j)}(\mu(X_j|\theta), \mu). \tag{5}$$

Finally, we view $\mu(X_j|\theta)$ as the expectation of X_j conditioned on the unobserved random variable θ which represents the amount of rotation. Then, by solving the equation $\mu(X_j|\theta) = R(c, \theta)\mu_j$ for θ, using the Rodrigues' rotation formula (Gray, 1980; Altmann, 2005) $R(c, \theta)\mu_j = \mu_j \cos\theta + (c \times \mu_j) \sin\theta + \langle c, \mu_j \rangle c(1 - \cos\theta)$, we get

$$\theta = \mathrm{atan2}[\langle \mu(X_j|\theta), c \times \mu_j \rangle, \langle \mu(X_j|\theta), \mu_j - c\cos(r_j) \rangle], \quad (j = 1, \ldots, K), \tag{6}$$

where the two argument function $\mathrm{atan2}(x_2, x_1) \in (-\pi, \pi]$ is the signed angle between two vectors $e_1 = (1, 0)$ and $(x_1, x_2) \in \mathbb{R}^2$.

2.2 Estimation

Suppose we have n independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from model (2). Each $\mathbf{X}_i$ is a collection of K directions $\mathbf{X}_i = (X_{ij})_{j=1,\ldots,K}$. The estimates of parameters c, r_j, μ_j are obtained

as the sample Fréchet means as follows:

$$(\hat{c}, \hat{\mathbf{r}}) = \underset{c, r_1, \ldots, r_K}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{K} d_g^2\{\delta(c, r_j), X_{ij}\}, \tag{7}$$

$$\hat{\mu}_j = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{n} \rho^2_{\delta(\hat{c}, \hat{r}_j)}(P_{(\hat{c}, \hat{r}_j)} X_{ij}, \mu) \quad (j = 1, \ldots, K). \tag{8}$$

Note that in (8), we have used $P_{(\hat{c},\hat{r}_j)}X_{ij}$, the projection of X_{ij} onto $\delta(\hat{c}, \hat{r}_j)$, instead of $\mathbb{E}(X_{ij}|\theta_i)$ used in (5). The projection $P_{\delta(c,r)}x$ is a point on $\delta(c, r)$ with the minimal geodesic distance to x, given by (Mardia and Gadsden, 1977, Eq. (3.3))

$$P_{\delta(c,r)}x = \underset{v \in \delta(c,r)}{\operatorname{argmin}}\, d_g(v, x) = \frac{x \sin(r) + c \sin\{d_g(x, c) - r\}}{\sin\{d_g(x, c)\}}.$$

The predicted values of the latent variable θ_i are obtained using (6) by substituting the estimates for the parameters. The predictor for θ_i is $\hat{\theta}_i = K^{-1} \sum_{j=1}^{K} \theta_{ij}$ for each $i = 1, \ldots, n$ with

$$\theta_{ij} = \operatorname{atan2}\{\langle P_{(\hat{c},\hat{r}_j)}X_{ij}, \hat{c} \times \hat{\mu}_j \rangle, \langle P_{(\hat{c},\hat{r}_j)}X_{ij}, \hat{\mu}_j - \hat{c}\cos(\hat{r}_j) \rangle\}. \tag{9}$$

The least squares problems (7-8) do not have closed form solutions. The problem (8) is simpler and the same as finding the geodesic mean of angles, since both $P_{(\hat{c},\hat{r}_j)}X_{ij}$ and μ are on the one-dimensional circle $\delta(\hat{c}, \hat{r}_j)$. Solutions to this type of problem are combinatorial (Moakher, 2002) but also found efficiently by numerical methods (Le, 2001; Fletcher et al., 2003). The problem (7) is precisely the fitting of concentric (small) circles. Therefore, numerical algorithms for (7) are generalized algorithms of the well-studied fitting of small circles (Mardia and Gadsden, 1977; Rivest, 1999; Jung et al., 2011, 2012) and are discussed in the Appendix.

2.3 Rotational deformations

The single rotation model (2) describes rigid rotations of objects. We extend the model to more general cases so that the generalized rotational model can explain, for example, non-rigid twisting or bending.

Suppose two direction vectors x_1 and x_2 are rotated about the same axis c but by different angles θ_1 and θ_2. This allows the underlying object to deform. In general, the assumption of a single rotation angle θ in (2) is relaxed to possibly different angles $\theta_1, \ldots, \theta_K$, which may be either independent or dependent of each other. To incorporate such general situations,

the single rotation model is generalized to

$$X_j = R(c, \theta_j)\mu_j \oplus \epsilon_j \quad (j = 1, \ldots, K). \tag{10}$$

The relationships among the θ_j can be specified using prior knowledge about the specific rotational deformation. As a special case, when a rigid rotation is assumed, it is reasonable to set $\theta_1 = \cdots = \theta_K$, which goes back to the model (2). The general model (10) includes other important physical deformations. The twisting or bending of the object can be modeled by different rotations with a common axis of rotation. As an example, when an object (and its attached direction vectors) is twisted, one group of direction vectors is rotated clockwise, while the other group is rotated counter–clockwise. Let I_1 and I_2 be a partition of the indices $\{1, \ldots, K\}$ representing groups of the direction vectors that rotate together. A simple twisting or bending motion can be obtained by assuming $\theta_i = -\theta_j$ for all $i \in I_1, j \in I_2$. Another example is the scenario of independent rotations where all directions in the same group rotate together ($\theta_i = \theta_j, i, j \in I_l$) but two angles in different groups are independent (θ_i and θ_j are independent for $i \in I_l, j \in I_k, 1 \leq l \neq k \leq 2$).

In all cases above, we assume that some functions f_j are known in advance, so that the relationships between θ_j are modeled through known functions, i.e., $\theta_j = f_j(\theta)$.

In the estimation of the parameters in (10), we use the fact that the estimation procedure (7) does not depend on specific assumptions of the latent variable θ_j. Therefore, the same least squares estimators $\{\hat{c}, \hat{r}_j, \hat{\mu}_j\}$ can be used to estimate the parameters of (10). When $f_j(\theta)$ is known and invertible, the prediction of the ith sample of θ, $\theta_{(i)}$, can be obtained. Since each θ_{ij} of (9) is a perturbed version of $f_j(\theta_i)$, the prediction of $\theta_{(i)}$ is then

$$\hat{\theta}_i = \frac{1}{K} \sum_{j=1}^{K} f_j^{-1}(\theta_{ij}).$$

Remark 1. A misspecification of the function f_j does not affect the estimation procedure (7), i.e., the estimation of the rotation axis. Nevertheless, the specification of f_j models the relationships between the rotation angles θ_j and is therefore crucial for their prediction as elaborated in Section 4 of the Supplementary Material. The partition I_1 and I_2 models f_j and is not a parameter of (7).

Example 1. The toy example presented in Fig. 2 is now discussed in more detail. The dataset consists of $n = 30$ observations of random twisting. The axis of twist is $c = (0, 1, 0)'$. The random angle θ follows $N(0, \sigma^2)$ with $\sigma \approx 22.5^\circ$ with $\theta_1 = \theta_2 = \theta$ and $\theta_3 = \theta_4 = -\theta$. The noise is independently added by a perturbation of $N_3(0, 0.1^2)$ on both the head and tail of

the direction vectors and then projected onto S^2.

The estimate $(\hat{c}, \hat{\mathbf{r}})$ was obtained by (7). The corresponding four concentric circles and the axis estimates $\hat{c}$ are overlaid in Fig. 3c. The estimate $\hat{c} = (0.007, 0.999, -0.031)'$ is only 1.8 degrees away from the truth. The base point estimates $\hat{\mu}_j$, predictions of θ_i, and the estimate of σ are also obtained, which are close to the truth. For example, $\hat{\sigma} = 21.5°$.

Despite a relatively small sample size ($n = 30$), the proposed estimator successfully estimated the axis of rotation, and leads to a clear visualization of the underlying rotational deformation, as depicted in Fig. 3.

3 Hierarchical rotations

We now discuss an estimation procedure for rotational deformations that consist of two independent generalized rotations. Such deformations include twisting and bending of the objects about different axes.

Suppose a set of base points μ_j is rotated by $R(c_1, \theta_j)$ and then by $R(c_2, \psi_j)$. The rotated random direction vector X_j is represented by

$$X_j = R(c_2, \psi_j)R(c_1, \theta_j)\mu_j \oplus \epsilon_j \quad (j = 1, \ldots, K), \tag{11}$$

with some error ϵ_j as seen in (2). The axes c_1, c_2 and the base points μ_j are unknown parameters and θ_j, ψ_j are independent latent variables representing rotation angles. The random direction vectors X_j have the same distribution as in

$$R'(c_2, \psi_j)X_j = R(c_1, \theta_j)\mu_j \oplus \epsilon_j \quad (j = 1, \ldots, K), \tag{12}$$

provided that the distribution of $\mu_j \oplus \epsilon_j$ is spherically symmetric about μ_j. The order of these rotations is not interchangeable because $R(c_2, \psi_j)R(c_1, \theta_j) \neq R(c_1, \theta_j)R(c_2, \psi_j)$ in general. Therefore, call the first rotation operation $R(c_1, \theta_j)$ the primary rotation, and $R(c_2, \psi_j)$ the secondary rotation.

With n observations, we have $X_{ij} = R(c_2, \psi_{ij})R(c_1, \theta_{ij})\mu_j \oplus \epsilon_{ij}$ $(i = 1, \ldots, n, j = 1, \ldots, K)$ and we wish to estimate the axes of rotations c_1, c_2 and predict the unobserved random variables ψ_{ij} and θ_{ij}. It is required to constrain the relationship among the ψ_j as a function of ψ (also for θ_j). Otherwise, θ_{ij} and ψ_{ij} will catch all sample fluctuations, leading to an overfit of c_1 and c_2. Let $\theta_j = f_{1j}(\theta)$ and $\psi_j = f_{2j}(\psi)$ for some known functions f_{1j} and f_{2j} $(j = 1, \ldots, K)$. For example, when the deformation is composed of rigid rotation and

twisting, the two functions can be modeled by $f_{1j}(\theta) = \theta$ and $f_{2j}(\psi) = 1_{j\in I_1}\psi - 1_{j\in I_2}\psi$, where 1 denotes the indicator function for a partition I_1, I_2 of $\{1, \ldots, K\}$. See also Section 4.

The estimates of the axes of rotation c_1, c_2 might be obtained as a solution of the least-squares problem

$$\min_{c_1,c_2} \sum_{i=1}^{n} \sum_{j=1}^{K} \rho_g^2\{X_{ij}, R(c_2, \psi_{ij})R(c_1, \theta_{ij})\mu_j\}.$$

Since the above problem is challenging to solve directly, we divide it into two subproblems, which can be solved iteratively. In a search for two subproblems, suppose first that (c_2, ψ_{ij}) of the secondary rotation are known. Then we would de-rotate X_{ij} by the action of $R(c_2, -\psi_{ij}) = R(c_2, \psi_{ij})'$ so that the unrotated direction vectors $X_{ij}^* = R(c_2, \psi_{ij})'X_{ij}$ are solely expressed by the primary rotation of μ_j. In other words,

$$X_{ij}^* = R(c_1, \theta_{ij})\mu_j \oplus \epsilon_{ij}. \tag{13}$$

Noticing the structure of the model is the same as (2), the least squares estimators of c_1, μ_j and the auxiliary parameters $r_{1j} = \arccos(c_1'\mu_j)$, for $j = 1, \ldots, K$, are given by

$$\begin{aligned} (\hat{c}_1, \hat{r}_{11}, \ldots, \hat{r}_{1K}) &= \operatorname*{argmin}_{c_1,(r_{11},\ldots,r_{1K})} \sum_{i=1}^{n} \sum_{j=1}^{K} d_g^2\{\delta(c_1, r_{1j}), X_{ij}^*\}, \\ \hat{\mu}_j &= \operatorname*{argmin}_{\mu} \sum_{i=1}^{n} d^2_{\delta(\hat{c}_1,\hat{r}_{1j})}(P_{(\hat{c}_1,\hat{r}_{1j})}X_{ij}^*, \mu). \end{aligned} \tag{14}$$

On the other hand, suppose we know in advance (c_1, θ_{ij}) of the primary rotation as well as the base points μ_j. Then we would rotate μ_j so that the secondary rotation is only needed to reach for the observations X_{ij} from the rotated base points $\mu_{ij}^* = R(c_1, \theta_{ij})\mu_j$. That is,

$$X_{ij} = R(c_2, \psi_{ij})\mu_{ij}^* \oplus \epsilon_{ij}. \tag{15}$$

The model (15) is different from models (2) and (13) as the base points μ_{ij}^* are different for different observations and assumed to be known. To estimate c_2, we modify (14) to incorporate the fact that μ_{ij}^* are varying but known, which leads to

$$\hat{c}_2 = \operatorname*{argmin}_{c_2} \sum_{i=1}^{n} \sum_{j=1}^{K} d_g^2\{\delta(c_2, r_{ij}^*(c_2)), X_{ij}\}, \tag{16}$$

where $r_{ij}^*(c_2) = \arccos(c_2'\mu_{ij}^*)$ is a function of c_2. For the estimation of c_1 and c_2 in (11),

we iteratively update $\hat{c}_1$ given $\hat{c}_2$ and then update $\hat{c}_2$ given $\hat{c}_1$. With the kth estimates $\hat{c}_1^{(k)}, \hat{c}_2^{(k)}, \tilde{\theta}_{ij}^{(k)}, \tilde{\psi}_{ij}^{(k)}$ of $c_1, c_2, \theta_{ij}, \psi_{ij}$ and the pre-specified functions $f_{1j}(\theta)$, $f_{2j}(\psi)$, the $(k+1)$th estimate is obtained as follows.

Step 1 Using $\hat{c}_2^{(k)}$ and $\hat{\psi}_{ij}^{(k)}$, obtain the de-rotated observations $X_{ij}^{(k)} = R(\hat{c}_2, -\hat{\psi}_{ij}^{(k)})X_{ij}$.

Step 2 Obtain $\hat{c}_1^{(k+1)}, \hat{r}_{1j}^{(k+1)}, \hat{\mu}_j^{(k+1)}$ using (14) with $X_{ij}^{(k)}$ in place of X_{ij}^*, and also the predictions $\tilde{\theta}_{ij}^{(k+1)}$ of θ_{ij} using (9) with $X_{ij}^{(k)}, \hat{c}_1^{(k+1)}, \hat{r}_{1j}^{(k+1)}, \hat{\mu}_j^{(k+1)}$ in place of $X_{ij}, \hat{c}, \hat{r}_j, \hat{\mu}_j$. Afterwards, use the function f_{1j} to predict θ_i as $\hat{\theta}_i^{(k+1)} = K^{-1}\sum_{j=1}^K f_{1j}^{-1}(\tilde{\theta}_{ij}^{(k+1)})$, so the predictions for θ_{ij} are updated as $\hat{\theta}_{ij}^{(k+1)} = f_{1j}(\hat{\theta}_i^{(k+1)})$.

Step 3 Rotate the base points to $\tilde{\mu}_{ij}^{(k+1)} = R(\hat{c}_1^{(k+1)}, \hat{\theta}_{ij}^{(k+1)})\hat{\mu}_j^{(k+1)}$ using the partial estimates $\hat{c}_1^{(k+1)}$ and predictions $\hat{\theta}_{ij}^{(k+1)}$.

Step 4 Obtain $\hat{c}_2^{(k+1)}$ as the minimizer of (16) with $\tilde{\mu}_{ij}^{(k+1)}$ in place of μ_{ij}^*, and also the predictions of rotation angles $\tilde{\psi}_{ij}^{(k+1)}$ using

$$\tilde{\psi}_{ij}^{(k+1)} = \text{atan2}(\langle X_{ij}, \hat{c}_2^{(k+1)} \times \tilde{\mu}_{ij}^{(k+1)} \rangle, \langle X_{ij}, \tilde{\mu}_{ij}^{(k+1)} - \langle \hat{c}_2^{(k+1)}, \tilde{\mu}_{ij}^{(k+1)} \rangle \hat{c}_2^{(k+1)} \rangle), \qquad (17)$$

which is similar to (9). Given $\tilde{\psi}_{ij}^{(k+1)}$, use the function f_{2j} to predict ψ_i by $\hat{\psi}_i^{(k+1)} = K^{-1}\sum_{j=1}^K f_{2j}^{-1}(\tilde{\psi}_{ij}^{(k+1)})$, leading to updated predictions for ψ_{ij} as $\hat{\psi}_{ij}^{(k+1)} = f_{2j}(\hat{\psi}_i^{(k+1)})$.

Step 5 If both $d_g(\hat{c}_1^{(k)}, \hat{c}_1^{(k+1)})$ and $d_g(\hat{c}_2^{(k)}, \hat{c}_2^{(k+1)})$ are negligible, stop. Otherwise, return to Step 1.

Numerical algorithms to solve the least-squares optimizations (14, 16) are similar to the problem (7), and are presented in the Appendix. In updating $\hat{c}_1^{(k+1)}$ and $\hat{c}_2^{(k+1)}$, one can use $\hat{c}_1^{(k)}$ and $\hat{c}_2^{(k)}$ as initial values for the optimization.

Remark 2. The initial values $\hat{c}_1^{(0)}$ and $\hat{c}_2^{(0)}$ may be suggested by a practitioner or obtained from an exploratory analysis. A careful choice is important to avoid convergence into local minima and is a topic of further studies. If the two rotational deformations are uncorrelated, a principal component analysis (PCA) may be used to find two major components as initial values. We propose to use the principal arc analysis (PAA) of Jung et al. (2011), which is a generalized PCA for data lying on $(S^2)^K$. Jung et al. (2011) argued that non-linear variation along small circles is better captured by PAA than by other extensions of PCA including Fletcher et al. (2004) and Huckemann et al. (2010). PAA is well suited to our problem, since the X_{ij} are distributed along small circles.

We now discuss how to use PAA to obtain initial values. For $\mathbf{X}_i = (X_{i1}, \ldots, X_{iK})' \in (S^2)^K$ $(i = 1, \ldots, n)$, PAA gives the mean $\mu^{PAA} = (\mu_1^{PAA}, \ldots, \mu_K^{PAA})' \in (S^2)^K$ and the

projections $\widetilde{\mathbf{X}}_{i(m)} = (\widetilde{X}_{i1(m)}, \ldots, \widetilde{X}_{iK(m)})' \in (S^2)^K$ onto the mth component, $m \in \{1, \ldots, M\}$ and M is the minimum of $2K$ and $n-1$. The first two components will be used to provide initial values. Which component corresponds to which rotational motion depends on the variance of θ and ψ. If $\mathrm{Var}(\theta)$ of the primary rotation is assumed to be larger than $\mathrm{Var}(\psi)$ of the secondary rotation, then the first component will provide an initial value for the primary rotation. In such a case, the solutions of (7) and (9) with $\widetilde{X}_{ij(1)}$ in place of X_{ij} are used as the initial values of $\hat{c}_1^{(0)}, \tilde{\theta}_{ij}^{(0)}$. Likewise, $\widetilde{X}_{ij(2)}$ are used to evaluate $\hat{c}_2^{(0)}, \tilde{\psi}_{ij}^{(0)}$. On the other hand, if $\mathrm{Var}(\theta) < \mathrm{Var}(\psi)$, then $\widetilde{X}_{ij(1)}$ is used for $\hat{c}_2^{(0)}$, and $\widetilde{X}_{ij(2)}$ for $\hat{c}_1^{(0)}$.

Remark 3. In contrast to single rotational deformations the function f_j effects the estimation of the rotation axes by the iterative back-and-forward rotation between two deformations which depend on the angle predictions. The order of the hierarchical deformation is specified by the primary and secondary information as well as by the functions f_{1j}, f_{2j}. Simulation studies, reported in Section 4 of the Supplementary Material, discuss the misspecification of f_j and a misspecified deformation order.

4 Numerical studies

In this section, we turn to the numerical performance of the proposed estimators. As our modeling approach is novel, there is no competing method to compare with. We study performance over several different rotational deformation situations.

Two different objects are studied. The first object (body 1), illustrated in Fig. 2, consists of $K = 4$ directions, while the second object (body 2) contains $K = 8$ direction vectors. The von Mises–Fisher distribution (Mardia and Jupp, 2000, p. 36) with concentration parameter κ, denoted as $\mathrm{vMF}(\kappa)$, is used for the distribution of errors. Three models (indexed by equation number) are considered for each object:

- Model (2)–*Rigid rotation*: $c = (1, 0, 0)'$, $\theta \sim N(0, \sigma_\theta^2)$ and $\sigma_\theta = \pi/12 \approx 15°$.
- Model (10)–*Twisting*: $c = (0, 1, 0)'$, $\theta_j = f_j(\theta) = 1_{j \in I_1}\theta - 1_{j \in I_2}\theta$, where $\theta \sim N(0, \sigma_\theta)^2$, $\sigma_\theta = \pi/8 \approx 22.5°$. Here, I_1 and I_2 are the partitions of $\{1, \ldots, K\}$ according to the right and left sides.
- Model (11)–*Hierarchical deformations*: $c_1 = (1, 0, 0)'$, $c_2 = (1/\sqrt{2}, -1/\sqrt{2}, 0)'$, $\theta_j = \theta$ and $\psi_j = f_j(\psi) = 1_{j \in I_1}\psi - 1_{j \in I_2}\psi$, where $\theta \sim N(0, \sigma_\theta)^2$, $\sigma_\theta \approx 22.5°$ and $\psi \sim N(0, \sigma_\psi)^2$, $\sigma_\psi \approx 15°$. The I_1 and I_2 are the same partition used in the twisting model above.

Table 1: Numerical performance over 1000 replications, for single deformation models.

Rotation (unit: degrees)		Body 1 $d_g(\hat{c}, c)$	Body 1 $\hat{\sigma}_\theta$ ($\sigma_\theta = 15$)	Body 2 $d_g(\hat{c}, c)$	Body 2 $\hat{\sigma}_\theta$ ($\sigma_\theta = 15$)
$\kappa = 100$	$n = 30$	4.133 (2.269)	15.248 (1.909)	2.905 (1.602)	15.018 (1.891)
	$n = 100$	2.235 (1.182)	15.365 (1.138)	1.560 (0.836)	15.114 (1.138)
$\kappa = 1000$	$n = 30$	1.166 (0.641)	14.896 (1.974)	0.841 (0.466)	14.881 (1.966)
	$n = 100$	0.655 (0.344)	15.012 (1.040)	0.448 (0.227)	14.982 (1.041)
Twisting (unit: degrees)		Body 1 $d_g(\hat{c}, c)$	Body 1 $\hat{\sigma}_\theta$ ($\sigma_\theta = 22.5$)	Body 2 $d_g(\hat{c}, c)$	Body 2 $\hat{\sigma}_\theta$ ($\sigma_\theta = 22.5$)
$\kappa = 100$	$n = 30$	2.761 (1.510)	22.647 (2.820)	4.055 (3.624)	22.785 (2.702)
	$n = 100$	1.482 (0.781)	22.753 (1.687)	1.883 (1.336)	22.705 (1.672)
$\kappa = 1000$	$n = 30$	0.803 (0.439)	22.344 (2.939)	1.017 (0.743)	22.336 (2.935)
	$n = 100$	0.446 (0.234)	22.484 (1.564)	0.536 (0.362)	22.474 (1.561)

In the last model, c_1 and c_2 are not orthogonal. Recall that we do not require any orthogonality of the two axes. The model thus represents hierarchical deformations by a rigid rotation and oblique twist.

For each model, we generate $n = 30, 100$ rotationally deformed objects with different error concentration parameters $\kappa = 100, 1000$. The proposed method is then applied to obtain the estimates $\hat{c}, \hat{\mu}_j$ and the predictions $\hat{\theta}_i$. These are replicated 1000 times, and the estimation quality is measured by $d_g(\hat{c}, c)$ and $\hat{\sigma}_\theta = \sum_{i=1}^n \hat{\theta}_i^2/n$ because the empirical mean is zero by equation (3).

Table 1 reports the mean and standard deviations of the measures, for the single deformation models. The quality of estimation is improved upon larger sample size and smaller error variance (larger κ). In the rigid rotation model, the performance is better for body 2. A main difference between the two bodies is the number of directions: $K = 4$ for body 1 and $K = 8$ for body 2. The higher accuracy observed for body 2, can be explained as the additional directions yielding more information. On the other hand, the performance in fitting the twisting model is inferior for body 2. An explanation is the close proximity of the axis c and the base points μ_j (to be rotated) in body 2, which is further illustrated in Fig. 4.

Figure 4 compares the estimates $\hat{c}$ from the single deformation models (2) and (10). Since $\hat{c} \in S^2$, these points are approximated by their images under the inverse exponential map (see Appendix) on the tangent plane $T_c S^2$ centered at c. For comparison, the scatter of vMF(100) is plotted at the top left panel. The rest of the top row shows the scatter of $\hat{c}$ from the rigid rotation model. In the bottom row, where the scatter of $\hat{c}$ from the twist model is plotted, the estimates of body 2 show a different pattern of scatter compared to body 1. This is due to a smaller radius $r_j = d_g(c, \mu_j)$ for some j, where the dispersion of the error is large compared to small r_j. When fitting the concentric circles, the smaller radius

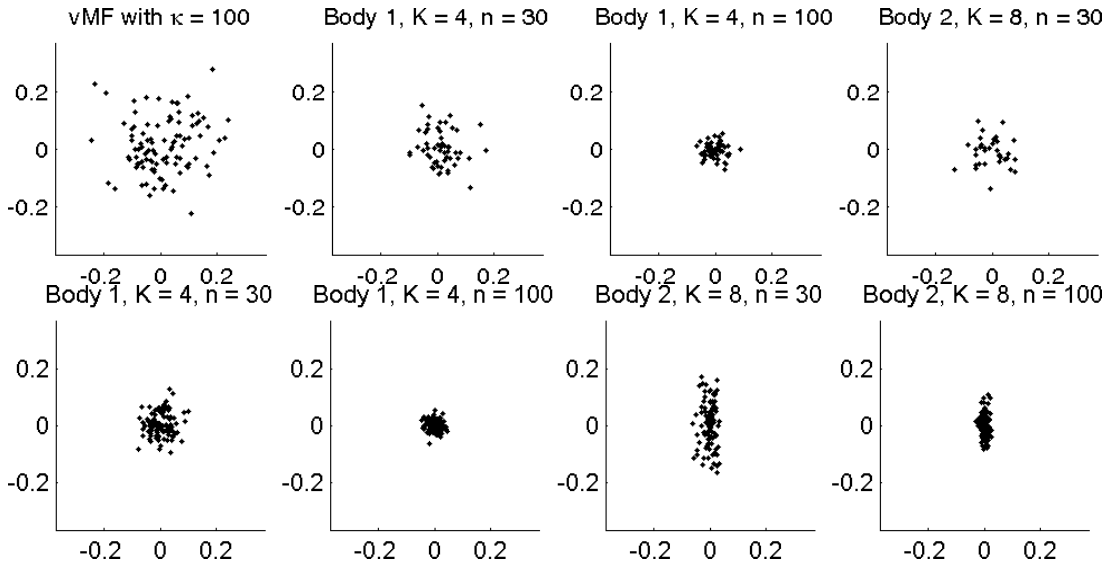


Figure 4: Scatters under their inverse exponential map of vMF(100) (top left panel) and of estimates $\hat{c}$ for the rigid rotation model (top row) and the twisting model (bottom row). The shape of the empirical covariance of $\hat{c}$ is different for the body 2–twist pair due to the proximity of the axis and base points.

$d_g(c, \mu_j)$ introduces large variance of the estimate $\hat{c}$.

The numerical performance in estimation of the hierarchical deformation model (11) is comparable to the simpler models. The results can be found in the Supplementary Material.

The asymptotic time complexity of the proposed algorithm is $O(nK)$ if a finite number of iterations is assumed as elaborated in Section 5 of the Supplementary Material. Furthermore, it is exemplified that the computation time increases, approximately linear in the number of samples n and the number of direction vectors K.

5 Analysis of s-rep deformations

In this section, an application of the proposed method to s-rep data is discussed.

5.1 S-reps of deformed ellipsoids

The skeletal representation (s-rep) gives a rich and efficient description of 3D objects (Siddiqi and Pizer, 2008; Pizer et al., 2013). The s-rep of human organs has been used to study structural and statistical properties and to promote precise segmentation of the organ from images. Accurate understanding of the deformations of prostates and hippocampi is crucial

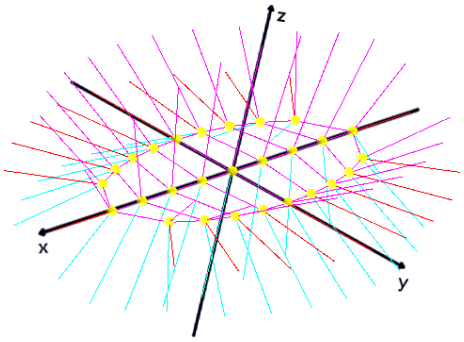


Figure 5: Fitted s-rep of a plain ellipsoid.

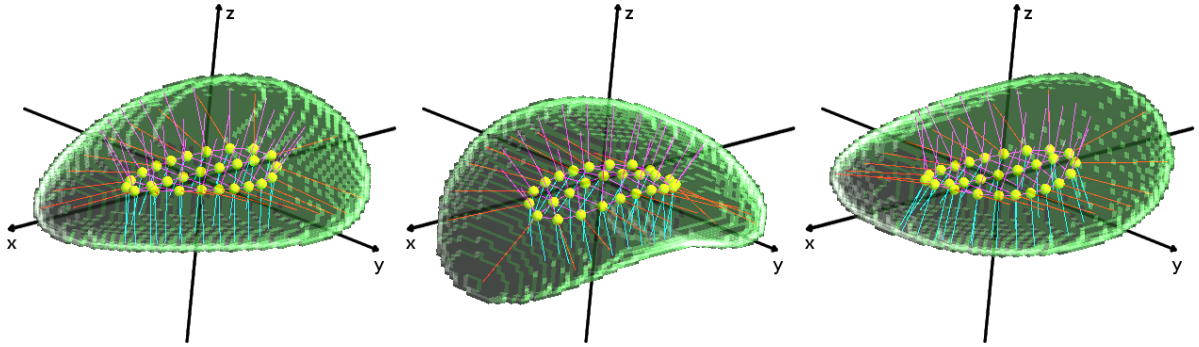
in medical operations. It has been observed that the major variation in the shape of these organs appears to be rotational deformations (Joshi et al., 2002; Jeong et al., 2008; Pizer et al., 2013), which motivated our analysis of s-rep data.

While our final goal is to analyze s-reps of real human organs, we work here with s-reps of deformed ellipsoids to validate the proposed method. A number of human organs, including hippocampi, prostates and bladders, are similar in shape to bent and twisted ellipsoids. Therefore, in our analysis, the rotationally deformed ellipsoids were used as a surrogate of real human organs. This enables us to the compare the estimate with the underlying truth.

An s-rep of a 3D object consists of a two-sided sheet of skeletal positions with spokes connecting the skeletal sheets to the boundary of the object. The skeletal sheet is nearly medial and the spokes are roughly normal to the boundaries, as defined in Pizer et al. (2013). We work with discrete s-reps, where the skeletal positions and the corresponding spokes are evaluated over a finite grid (9 by 3 in our analysis), as shown in Fig. 5.

The raw data are binary images of rotationally deformed ellipsoids, to which s-reps are fitted using the s-rep fitting procedure described in Pizer et al. (2013). The binary images are pre-processed by the anti-aliasing method of Niethammer et al. (2013). Figure 5 shows the fitted s-rep of a plain ellipsoid. The plain ellipsoid is centered at the origin with axes of lengths 3/4, 1/2 and 1/4, in x, y, z coordinates, and is a template for further deformation. The s-rep fitting for deformed ellipsoids is an iterative procedure using the s-rep of the plain ellipsoid as the initial value. The Supplementary Material contains a detailed description of the plain and deformed ellipsoids.

Three different rotational deformations of ellipsoids are examined: twisting, bending and a mix of those. In each experiment, $n = 30$ randomly deformed ellipsoids were obtained, followed by s-rep fitting. See Fig. 6 for the result of s-rep fitting for randomly deformed ellipsoids. The $K = 74$ spoke directions of each s-rep were recorded.



(a) Twisted ellipsoid (b) Bent ellipsoid. (c) Bent and twisted ellipsoid.

Figure 6: Examples of fitted s-reps for rotationally deformed ellipsoids. The surface of the raw ellipsoid, to which the s-rep is fitted, is overlaid.

5.2 Results

The first data set consists of fitted s-rep directions of twisted ellipsoids. The twisting was implemented as the rotation of the plain ellipsoid about the x-axis with angles proportional to $f_x(\theta) = \theta x$, where $x \in [-\frac{3}{4}, \frac{3}{4}]$ is the x-coordinate of the skeletal position of the spokes. The random angle θ is a zero-mean normal random variable with standard deviation $\sigma_\theta = 0.3$. From the $n = 30$ samples, we obtained an estimate of the rotation axis $\hat{c} = (0.99, 0.05, 0.12)'$, with $d_g(\hat{c}, c) = 7°$. The standard deviation estimate, $\hat{\sigma}_\theta = 0.29$, is close to the truth.

The bending deformation in the second experiment was realized as the rotation about the y-axis with angles proportional to $f_x(\alpha) = \alpha x$. Here, α follows $N(0, \sigma_\alpha^2)$ with $\sigma_\alpha = 0.4$. Similar to the estimation of twisting, an accurate estimate of the axis of rotation $\hat{c} = (0.01, 1.00, 0.01)'$ with distance $0.7°$ to the true axis was obtained from the sample of size $n = 30$. However, the estimate of σ_α was not consistent with the truth. This is so because the rotation angle of each direction is not consistent with $f_x(\alpha)$ due to the additional swing of the direction. The additional swing is a consequence of the change in surface curvature. Dealing with such an issue is beyond the scope of the current paper; it is discussed further in the Supplementary Material.

Finally, we report the results for bent and twisted ellipsoids. The raw ellipsoids were sequentially deformed by bending about the y-axis, then twisting about the x-axis. The initial values chosen by the data-driven method (see Remark 2 in Section 3) are $\hat{c}_1^0 = (-0.13, -0.99, -0.00)'$ and $\hat{c}_2^0 = (-0.07, -0.99, 0.02)'$, which are almost the same. A uniformly randomly chosen initial value for c_2 was used instead. In particular, a uniform random direction $\tilde{c}_2^0$ was used, provided that $\tilde{c}_2^0$ is at least 11 degrees away from $\hat{c}_1^0$. With this alternative initial value, the iterative estimation leads to estimates $\hat{c}_1 = (0.01, -1, -0.00)'$

and $\hat{c}_2 = (-0.99 - 0.05, 0.00)'$, both of which are close to their corresponding population counterparts. A simulation study, the report of which is omitted, confirms that the estimates are stable with respect to different choices of initial value $\tilde{c}_2^0$.

As we have pointed out in the introduction, the ellipsoid considered here can be understood as a template for many real human organs. The accurate estimation of the parameters of rotational deformations of ellipsoids indicates the potential of this type of analysis of deformed objects in real 3D images obtained from, e.g., magnetic resonance imaging. Further experiments cover surface point distribution models and a more general deformation; they are discussed in the Supplementary Material.

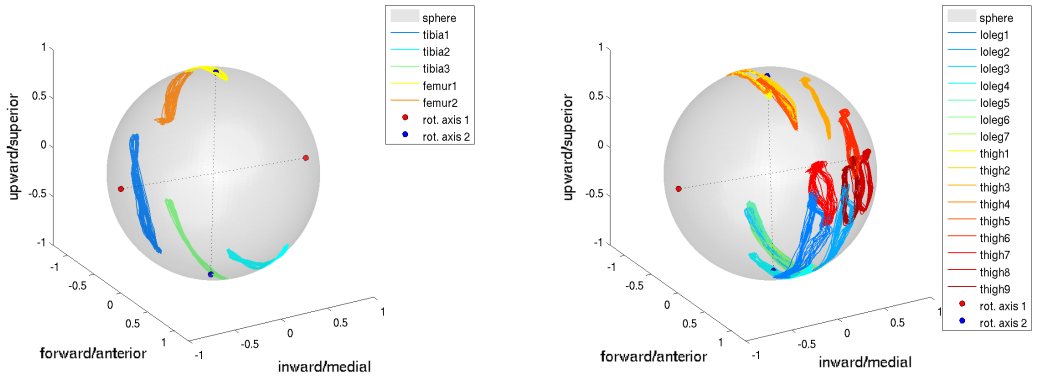
6 Application to knee motion during gait

In order to further support the validity of the proposed estimation procedure, this section presents findings from experimentally collected biomechanical data as a part of a larger project.

The estimation of two rotation axes of the knee joint is a well-studied problem in biomechanics (e.g., Ball and Greiner (2012)). The two estimated rotation axes model the primary and secondary rotation axes of the upper and lower leg relative to each other. The dominant rotation axis defines the flexion-extension motion at the knee. This axis is approximately directed right-to-left (lateral-to-medial). The secondary rotation axis defines the internal-external motion of the lower leg relative to the upper leg. This axis is approximately directed down-to-up (distal-to-proximal) along the long axis of the tibia (ankle-to-knee joint centers).

The motion of 25 markers placed on the right lower extremity of one healthy male volunteer was collected following informed consent. The volunteer consented to have two 6 mm pins surgically inserted into his femur and tibia. Insertion sites were selected to minimize neuro-muscular effects that could influence natural knee motion. Three and four markers were then attached to these rigid pins which allowed us to measure the true motion of the hidden femur and tibia bones. Additional markers were also placed on the surface of the thigh (10 skin markers) and lower leg (8 skin markers). In each of the four segments (femur, tibia, thigh, lower leg) one marker was chosen as a basis point and directions were derived between the basis point and the remaining markers of that segment. The coordinate system for this experiment was defined when the volunteer stood at attention and faced forwards. The XYZ axes were in the directions Forward, Inward, Upward (FIU).

The volunteer walked at 2.5 mph on a motor driven treadmill. After a familiarization



(a) bone (b) skin

Figure 7: Estimation of first (flexion-extension) and second (internal-external) rotation axes of the knee from bone and skin marker directions on the upper (femur; thigh) and lower (tibia; lower leg) extremity. In addition, the path of each marker direction is depicted. (a) Estimated rotation axes of directions derived from bone markers. (b) Estimated rotation axes of directions derived from skin markers.

time period, the motion of the markers were collected for approximately 20 seconds at 50 Hz. Within this data collection period, 16 complete gait cycles were identified. A gait cycle is defined from right foot contact with the floor to the next right foot contact. In total 976 time points were used within the following analyses.

Figure 7 shows the path of $K = 5$ bone directions and $K = 16$ skin directions of all 16 walking cycles. In addition to the least square estimations of (c_1, c_2), the rotation axes c_{1j} and c_{2j} were estimated for each marker $j = 1 \ldots, K$. Therewith, we can quantify the variation in the estimations by the dispersion measure $\sigma_{d_g(\hat{c}_1, \hat{c}_{1j})}$ of the geodesic distance $d_g(\hat{c}_1, \hat{c}_{1j})$ and $\sigma_{d_g(\hat{c}_2, \hat{c}_{2j})}$ respectively. The estimates are obtained by the least squares estimator (7) and the procedure is described in detail in Section 6 of the Supplementary Material.

The estimated rotation axes from the bone directions are $\hat{c}_1 = (-0.00, -1.00, 0.06)$ with dispersion $\sigma_{d_g(\hat{c}_1, \hat{c}_{1j})} = 9.41$ degrees and $\hat{c}_2 = (0.06, -0.06, -1.00)$ with $\sigma_{d_g(\hat{c}_2, \hat{c}_{2j})} = 12.30$ degrees. The unsigned directions of these two axes correspond to the anatomically expected flexion-extension and internal-external knee rotation axes (Grood and Suntay, 1983, Figure 6, page 142).

The two rotation axes estimated from the skin marker directions are very similar to the bone marker direction estimates. Quantitatively these were, $\hat{c}_1 = (0.01, -1.00, 0.07)$ with $\sigma_{d_g(\hat{c}_1, \hat{c}_{1j})} = 12.80$ degrees and $\hat{c}_2 = (0.01, 0.03, -1.00)$ with $\sigma_{d_g(\hat{c}_2, \hat{c}_{2j})} = 25.82$ degrees. In both cases, the higher dispersion of the second rotation axis can be explained by the smaller range of rotation angles. As expected, we observe more dispersion of the rotation axes using

the skin data because of the complex deformation of the skin surface. Estimation results of the rotation angles can be found in the Supplementary Material. Future work should improve the estimations and reduce the dispersion by a more careful time modelling of knee motions such as that proposed by Rivest (2001) who examined elbow motion.

7 Discussions

The paper proposes a novel method to estimate rotational deformations from directional data by concentric small circles. The estimation procedure is independent of the latent variable θ_j for single rotational deformations. In addition, the paper proposed an estimation procedure for hierarchical deformations depending on functions f_{1j}, f_{2j} and the ordering of the initial estimation. An interesting topic for future research is improvement of the prediction of θ_j, which includes the automatic classification of directions into a partition I_1 and I_2 as well as the development of methods to predict f_j from the data. The geometry of composing deformations has to be studied in further detail in order to extend the estimation method to more than two rotational deformations. A first step in decreasing the relevance of the order of deformations would be the implementation of an expectation maximization (EM) based optimization procedure.

Appendix

Proof of Lemma 1

Let $A_j = \{v_j \in \delta_\eta(c, r_j)\}$, where $\delta_\eta(c, r) = \{x \in S^2 : d_g(\delta(c, r), x) < \eta\}$. For $R = R(c, \theta)$,

$$\begin{aligned} P(A_j) &= P[d_g\{R\mu_j \oplus \epsilon_j, \delta(c, r_j)\} < \eta] = P[d_g\{R^T(R\mu_j \oplus \epsilon_j), R^T\delta(c, r_j)\} < \eta] \\ &= P[d_g\{\mu_j \oplus \epsilon, \delta(c, r_j)\} < \eta] \geq P[d_g(\mu_j \oplus \epsilon, \mu_j) < \eta] \geq 1 - \mathrm{Var}(\epsilon)/\eta^2, \end{aligned}$$

where $\mu_j \oplus \epsilon$ has the same distribution as $R^T(R\mu_j \oplus \epsilon_j)$ because of the spherical symmetry. A Markov inequality is used. Since the A_js are independent, $P(\bigcap_{j=1}^K A_j) = \prod_{j=1}^K P(A_j) \geq \{1 - \mathrm{Var}(\epsilon)/\eta^2\}^K$.

Numerical Algorithms for (7), (14), and (16)

The optimization problems (7) and (14) are identical and can be understood as fitting concentric circles on the unit sphere. The problem (16) is a more general nonlinear least squares problem, which however can be solved in a similar manner to the former two problems. We propose a variant of the doubly iterative algorithm used in fitting small circles in S^m (Jung et al., 2011, 2012).

We first introduce some notation. For $m \geq 2$, the tangent space at $c \in S^m$ is denoted by $T_c S^m$, which can be parametrized by $\mathbb{R}^m$. Let $c = e_{m+1}$ without loss of generality. The exponential map $\mathrm{Exp}_c : \mathbb{R}^m \to S^m$ is defined for $v_1 \in \mathbb{R}^m$ by

$$\mathrm{Exp}_c(v_1) = \left(\frac{v_1}{\|v_1\|} \sin \|v_1\|, \cos \|v_1\| \right),$$

with a convention of $\mathrm{Exp}_c(0) = c$. The exponential map has an inverse, called the log map, and is denoted by $\mathrm{Log}_c : S^m \to T_c S^m$.

For problems (7) and (14), the following iterative algorithm can be used. The algorithm finds a suitable point of tangency $\hat{c}_0$, which is also the center of the fitted concentric circles. Given the candidate $\hat{c}_0$, the data x_{ij} are mapped to the tangent space $T_{\hat{c}_0} S^2$ by the Log map. Let $x_{ij}^\dagger = \mathrm{Log}_{\hat{c}_0}(x_{ij})$. Since the Log map preserves distance, we have $\arccos(\hat{c}_0' x_{ij}) = \|x_{ij}^\dagger\|$. Then we solve a non-linear least-squares problem

$$\min_{c^\dagger, r_j} \sum_{i=1}^{n} \sum_{j=1}^{K} (\|x_{ij}^\dagger - c^\dagger\| - r_j)^2. \tag{18}$$

Since the optimization problem (18) does not have any constraint, it can be numerically solved by, e.g., the Levenberg–Marquardt algorithm (Scales, 1985). The solution $c^\dagger$ is then mapped to S^2 by the exponential map at c and becomes $\hat{c}_1$. This procedure is repeated until $\hat{c}$ converges.

The optimization problem (16) can be solved in a similar way. We use the fact that $\rho_g^2(\delta(c, r^*(c)), x) = (\arccos(c'x) - \arccos(c'\mu^*))^2 = (\|\mathrm{Log}_c x\| - \|\mathrm{Log}_c \mu^*\|)^2$. Thus for fixed c, $\rho_g^2(\delta(c, r^*(c)), x) \geq \min_y (\|\mathrm{Log}_c x - y\| - \|\mathrm{Log}_c \mu^* - y\|)^2$. The minimizer y leads to a better candidate for c through the exponential map. The algorithm to solve (16) follows the same lines as the algorithm to solve (7), except instead of (18) we minimize

$$\min_{c^\dagger} \sum_{i=1}^{n} \sum_{j=1}^{K} (\|\mathrm{Log}_{\hat{c}} x - c^\dagger\| - \|\mathrm{Log}_{\hat{c}} \mu - c^\dagger\|)^2.$$

Supplementary Materials

Additional discussions and data analyses: Article containing *i.)* additional data analysis results, *ii.)* simulation results for the hierarchical deformation model described in Section 4, *iii.)* further discussion of the model bias, brought up in Section 2.1, *iv.)* study of the estimator behaviour using misspecified parameters, *v.)* a computational complexity study of the algorithm and *vi.)* the estimation procedure for knee motion analysis during gait as discussed in Section 6. (SupplementaryMaterialSJ.pdf)

Matlab code: A set of Matlab code for application of the proposed method. The code also contains all datasets used as examples in the article. (estRotDeformation.zip)

References

Altmann, S. L. (2005), *Rotations, Quaternions, and Double Groups*, Dover books on mathematics, Dover Publications.

Ball, K. A. and Greiner, T. M. (2012), "A Procedure to Refine Joint Kinematic Assessments: Functional Alignment," *Computer Methods in Biomechanics and Biomedical Engineering*, 15, 487–500.

Chang, T. (1986), "Spherical Regression," *The Annals of Statistics*, 14, 907–924.

— (1988), "Estimating the Relative Rotation of Two Tectonic Plates from Boundary Crossings," *Journal of the American Statistical Association*, 83, 1178–1183.

— (1989), "Spherical Regression with Errors in Variables," *The Annals of Statistics*, 17, 293–306.

Chang, T. and Rivest, L.-P. (2001), "M-Estimation for Location and Regression Parameters in Group Models: A Case Study using Stiefel Manifolds," *The Annals of Statistics*, 29, 784–814.

Cootes, T. F., Taylor, C., Cooper, D., and Graham, J. (1992), "Training Models of Shape from Sets of Examples," in *Proc. British Machine Vision Conference*, eds. Hogg, D. and Boyle, R., Berlin. Springer-Verlag, pp. 9–18.

Dryden, I. L. and Mardia, K. V. (1998), *Statistical Shape Analysis*, Chichester: Wiley.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1993), *Statistical Analysis of Spherical Data*, Cambridge: Cambridge University Press.

Fletcher, P., Lu, C., and Joshi, S. (2003), "Statistics of Shape via Principal Geodesic Analysis on Lie Groups," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1, pp. I–95–I–101 vol.1.

Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. (2004), "Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape," *IEEE Transactions on Medical Imaging*, 23, 995–1005.

Fréchet, M. (1948), "Les Éléments Aléatoires de Nature Quelconque dans un Espace Distancié," *Annales de l'Institut Henri Poincaré*, 10, 215–310.

Gamage, S. S. and Lasenby, J. (2002), "New Least Squares Solutions for Estimating the Average Centre of Rotation and the Axis of Rotation," *Journal of Biomechanics*, 35, 87–93.

Goodall, C. R. (1991), "Procrustes Methods in the Statistical Analysis of Shape (with discussion)," *Journal of the Royal Statistical Society: Series B*, 53, 285–339.

Gray, J. J. (1980), "Olinde Rodrigues' Paper of 1840 on Transformation Groups," *Archive for History of Exact Sciences*, 21, 375–385.

Grood, E. and Suntay, W. (1983), "A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee," *Journal of Biomechanical Engineering*, 105, 136–144.

Halvorsen, K., Lesser, M., and Lundberg, A. (1999), "A new Method for Estimating the Axis of Rotation and the Center of Rotation," *Journal of Biomechanics*, 32, 1221–1227.

Hanna, M. S. and Chang, T. (2000), "Fitting Smooth Histories to Rotation Data," *Journal of Multivariate Analysis*, 75, 47–61.

Huckemann, S. (2011a), "Inference on 3D Procrusted Means: Tree Bole Growth, Rank Deficient Diffusion Tensors and Perturbation Models," *Scandinavian Journal of Statistics*, 38, 424–446.

— (2011b), "Intrinsic Inference on the Mean Geodesic of Planar Shapes and Tree Discrimination by Leaf Growth," *Annals of Statistics*, 39, 1098–1124.

Huckemann, S., Hotz, T., and Munk, A. (2010), "Intrinsic Shape Analysis: Geodesic PCA for Riemannian Manifolds modulo Isometric Lie Group Actions," *Statistica Sinica*, 20, 1–58.

Jeong, J.-Y., Stough, J. V., Marron, J. S., and Pizer, S. M. (2008), "Conditional-Mean Initialization Using Neighboring Objects in Deformable Model Segmentation," in *SPIE Medical Imaging*.

Joshi, S., Pizer, S. M., Fletcher, P., Yushkevich, P., Thall, A., and Marron, J. S. (2002), "Multiscale Deformable Model Segmentation and Statistical Shape Analysis Using Medial Descriptions," *IEEE Transactions on Medical Imaging*, 21, 538–550.

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568.

Jung, S., Foskey, M., and Marron, J. S. (2011), "Principal Arc Analysis on Direct Product Manifolds," *Annals of Applied Statistics*, 5, 578–603.

Karcher, H. (1977), "Riemannian Center of Mass and Mollifier Smoothing," *Communications on Pure and Applied Mathematics*, 30, 509–541.

Kent, J. T. and Mardia, K. V. (1997), "Consistency of Procrustes Estimators," *Journal of the Royal Statistical Society: Series B*, 59, 281–290.

Kurtek, S., Ding, Z., Klassen, E., and Srivastava, A. (2011), "Parameterization-Invariant Shape Statistics and Probabilistic Classification of Anatomical Surfaces," in *Information Processing in Medical Imaging*, vol. 22, pp. 147–158.

Le, H. (1998), "On the Consistency of Procrustean Mean Shapes," *Advances in Applied Probability*, 30, 53–63.

— (2001), "Locating Fréchet Means with Application to Shape Spaces," *Advances in Applied Probability*, 33, 324–338.

Mardia, K. V. and Gadsden, R. J. (1977), "A Circle of Best Fit for Spherical Data and Areas of Vulcanism," *Journal of the Royal Statistical Society: Series C*, 26, 238–245.

Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics*, Chichester: Wiley.

Moakher, M. (2002), "Means and Averaging in the Group of Rotations," *SIAM Journal on Matrix Analysis and Applications*, 24, 1–16 (electronic).

Niethammer, M., Juttukonda, M. R., Pizer, S. M., and Saboo, R. R. (2013), "Anti-Aliasing Slice-Segmented Medical Images via Laplacian of Curvature Flow," *In preparation*.

Oualkacha, K. and Rivest, L.-P. (2012), "On the Estimation of an Average Rigid Body Motion," *Biometrika*, 99, 585–598.

Pennec, X. (2008), "Statistical Computing on Manifolds: from Riemannian Geometry to Computational Anatomy," *Emerging Trends in Visual Computing*, 5416, 347–386.

Pizer, S. M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013), "Nested Sphere Statistics of Skeletal Models," in *Innovations for Shape Analysis: Models and Algorithms*, eds. Breuß, M., Bruckstein, A., and Maragos, P., Springer Lecture Notes in Computer Science, pp. 93–115.

Rancourt, D., Rivest, L.-P., and Asselin, J. (2000), "Using Orientation Statistics to Investigate Variations in Human Kinematics," *Journal of the Royal Statistical Society: Series C*, 49, 81–94.

Rivest, L.-P. (1989), "Spherical Regression for Concentrated Fisher-Von Mises Distributions," *The Annals of Statistics*, 17, 307–317.

— (1999), "Some Linear Model Techniques for Analyzing Small-Circle Spherical Data," *Canadian Journal of Statistics*, 27, 623–638.

— (2001), "A Directional Model for the Statistical Analysis of Movement in Three Dimensions," *Biometrika*, 88, 779–791.

— (2006), "Regression and Correlation for 3×3 Rotation Matrices," *Canadian Journal of Statistics*, 34, 1–17.

Rivest, L.-P., Baillargeon, S., and Pierrynowski, M. (2008), "A Directional Model for the Estimation of the Rotation Axes of the Ankle Joint," *Journal of the American Statistical Association*, 103, 1060–1069.

Rohde, G. K., Ribeiro, A. J. S., Dahl, K. N., and Murphy, R. F. (2008), "Deformation-Based Nuclear Morphometry: Capturing Nuclear Shape Variation in HeLa Cells," *Cytometry A*, 73, 341–350.

Scales, L. E. (1985), *Introduction to Nonlinear Optimization*, New York: Springer-Verlag.

Siddiqi, K. and Pizer, S. (2008), *Medial Representations: Mathematics, Algorithms and Applications*, Computational Imaging and Vision, Vol. 37, Dordrecht, Netherlands: Springer, 1st ed.

Teu, K. K. and Kim, W. (2006), "Estimation of the Axis of a Screw Motion from Noisy Data–A New Method Based on Plücker Lines," *Journal of Biomechanics*, 39, 2857–2862.

Supplementary Material: Analysis of Rotational Deformations from Directional Data

Jörn Schulz[1] *, Sungkyu Jung[2], Stephan Huckemann[3], Michael Pierrynowski[4], J. S. Marron[5] and Stephen M. Pizer[5]

[1]University of Tromsø, [2]University of Pittsburgh, [3]University of Göttingen, [4]McMaster University, Hamilton [5]University of North Carolina at Chapel Hill

1 Additional data analysis

In this section, some additional data analyses are discussed, in particular using datasets from point distribution models.

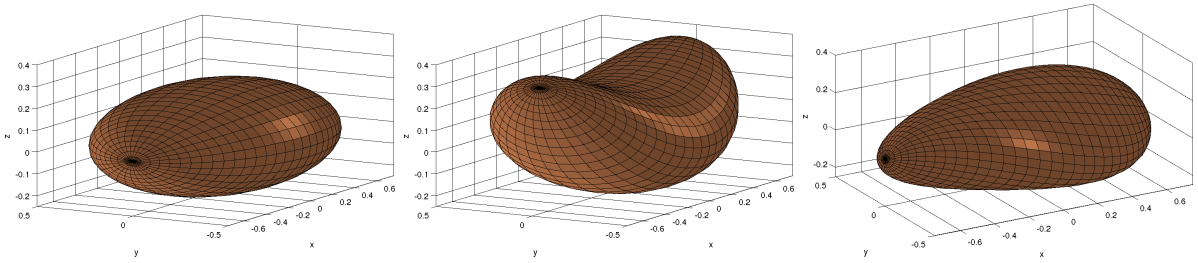
1.1 Rotationally deformed ellipsoids

We first discuss the procedures of obtaining the raw ellipsoid data. To avoid notational confusion, we denote a random vector by X_{ij} and their observed values by χ_{ij}, $i = 1, \ldots, n, j = 1, \ldots, K$. A point in $\mathbb{R}^3$ is described by $(x, y, z) \in \mathbb{R}^3$ in a fixed Cartesian coordinate system.

The surface of an ellipsoid can be parameterized by

$$\mathbf{s}(u,v) = \begin{pmatrix} x(u,v) \\ y(u,v) \\ z(u,v) \end{pmatrix} = \begin{pmatrix} r_a \sin(v) \\ r_b \sin(u)\cos(v) \\ r_c \cos(u)\cos(v) \end{pmatrix}, \qquad u \in [-\pi, \pi), v \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \tag{1.1}$$

*This research was funded by the Norwegian Research Council through grant 176872/V30 in the eVita program and performed as part of Tromsø Telemedicine Laboratory.



(a) Original (b) Bent (c) Twisted

Figure 1.1: Tube views of meshed surfaces of (a) an original ellipsoid with $r_a = 0.75$, $r_b = 0.5$ and $r_c = 0.25$, (b) bent ellipsoid with $\alpha = 0.6$, (c) twisted ellipsoid with $\theta = 0.6$.

where $r_a \geq r_b \geq r_c > 0$ are the length of the axes. We assume a default ellipsoid of $r_a = 0.75$, $r_b = 0.5$ and $r_c = 0.25$. For a parameter space $\Omega = [-\pi, \pi) \times [-\frac{\pi}{2}, \frac{\pi}{2}] \subset \mathbb{R}^2$, $\mathbf{s} : \Omega \to \mathbb{R}^3$ is a surface map in $\mathbb{R}^3$ with $(u, v) \xrightarrow{\mathbf{s}} (x(u, v), y(u, v), z(u, v))$. Without loss of generality, we assume that the first to third principal axes are x, y, z-axis in $\mathbb{R}^3$, respectively. The centroid of the ellipsoid is at origin $(0, 0, 0)'$.

Rotational bending around the y-axis ($c_b = (0, 1, 0)'$) is given by

$$B(\mathbf{s}) = R(c_b, f_b(x))\mathbf{s}(u, v), \tag{1.2}$$

and twisting around the x-axis ($c_t = (1, 0, 0)'$) is given by

$$T(\mathbf{s}) = R(c_t, f_t(x))\mathbf{s}(u, v), \tag{1.3}$$

where

$$\begin{aligned} R(c_b, f_b(x)) &= \begin{pmatrix} \cos(f_b(x)) & 0 & -\sin(f_b(x)) \\ 0 & 1 & 0 \\ \sin(f_b(x)) & 0 & \cos(f_b(x)) \end{pmatrix}, \\ R(c_t, f_t(x)) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(f_t(x)) & -\sin(f_t(x)) \\ 0 & \sin(f_t(x)) & \cos(f_t(x)) \end{pmatrix}. \end{aligned}$$

Here, $f_b(x) = \alpha x$ and $f_t(x) = \theta x$ for some $\alpha, \theta \in [-\pi/2, \pi/2]$ representing the overall size of the deformation. The amount of bending and twisting depends on the location on the x-axis. In addition to the rotational bending, we also consider quadratic bending around the

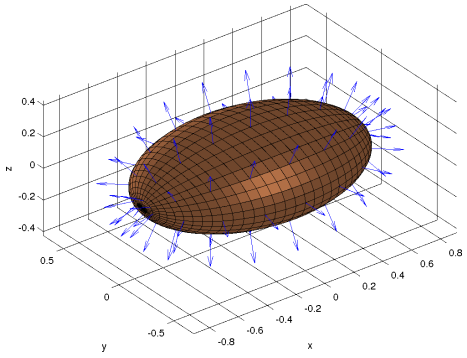


Figure 1.2: Visualization of a standard ellipsoid with attached boundary normals.

y-axis which is defined by

$$B_q(\mathbf{s}) = (x, y, z + \alpha x^2)', x, y, z, \in \mathbb{R} \tag{1.4}$$

using the ellipsoid parametrization (1.1). In the following, the term *bending* is used for rotational bending, and *quadratic bending* will be specified explicitly.

Figure 1.1 shows an example of an original ellipsoid, bent ellipsoid with $\alpha = 0.6$ and twisted ellipsoid with $\theta = 0.6$.

1.2 Point distribution and boundary normals

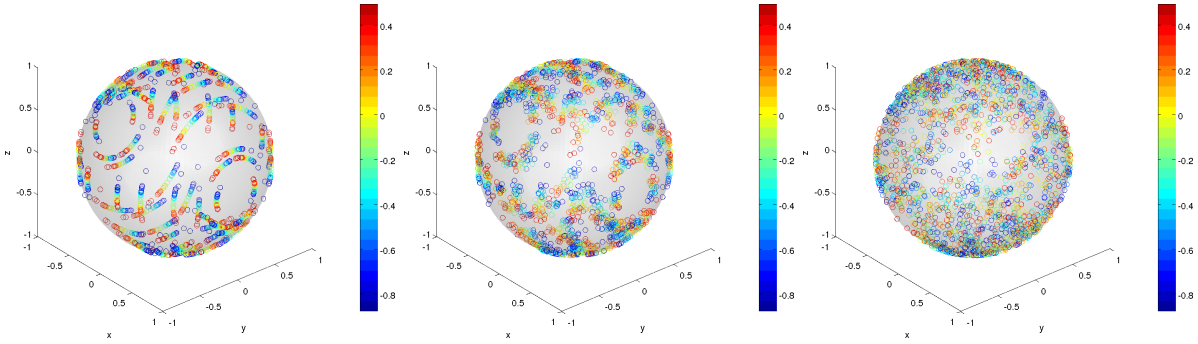
We now discuss a parameterization of ellipsoids by a point distribution model. In particular, a mesh grid and attached boundary normals of the surface $\mathbf{s}(u, v)$ will be used. See Fig. 1.2.

The size of surface mesh we used is 37×33. We chose $K = 9 \times 8 = 72$ vertices to attach normal direction vectors $\chi_{k_1 k_2} \in S^2, k_1 = 1, \ldots, 9, k_2 = 1, \ldots, 8$. For each k_1, the vertices where $\chi_{k_1 k_2}$ are attached have common x-coordinates

$$\{-0.738, -0.649, -0.482, -0.256, 0, 0.256, 0.482, 0.649, 0.738\},$$

the values of which influence the amount of deformation. Note that there are some directions that will not vary when the object is deformed. For example, the normals χ_{5k_2} $(1 \leq k_2 \leq 8)$, which are attached to vertices with zero x-coordinate, will stay still when the object is twisted or bent.

In the following we report four sets of experiments on the boundary normal ellipsoid data. As opposed to the s-rep data analysis, we are working with the raw data directly. The noise in the data is from the von Mises–Fisher distribution (Mardia and Jupp, 2000)



(a) without noise (b) $\kappa = 1000$ (c) $\kappa = 100$

Figure 1.3: (Twisting) Boundary normals using different concentration parameter κ. The rotation around the x-axis is clearly visible in (a). Different colors represent different amounts of twisting parameter θ.

with concentration parameter $\kappa > 0$. For each experiment, two levels of noise are used: $\kappa = 100, 1000$. The four sets of models are

- *Twisting* by (1.3), with $c_t = (1, 0, 0)'$, $\theta \sim N(0, \sigma_\theta^2)$, $\sigma_\theta = 0.3 \approx 17.2°$. See Fig. 1.3.
- *Bending* by (1.2) with $c_b = (0, 1, 0)'$, $\alpha \sim N(0, \sigma_\alpha^2)$, $\sigma_\alpha = 0.4 \approx 22.9°$. See Fig. 1.4.
- *Quadratic bending* by (1.4) with above c_b and σ_α. See Fig. 1.5.
- *Hierarchical deformation* by bending (primary) and twisting (secondary) with the same set of parameters above. See Fig. 1.7.

From each model a random sample of size $n = 30$ or 100 is obtained, from which the estimate $\hat{c}$ of the axis and $\hat{\sigma}$ are obtained. This is repeated for 1000 times.

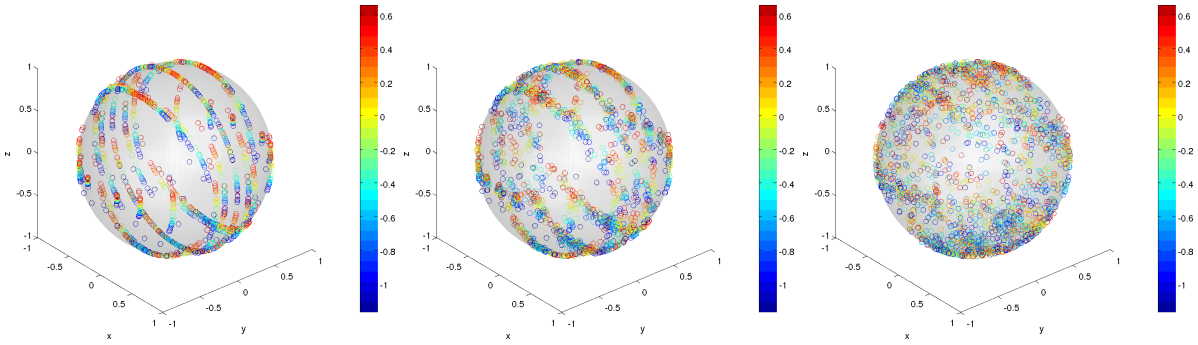
Twisting

Figure 1.3 shows 30 samples of 72 different normal directions from sets of twisted ellipsoids with different noise levels. Different colors represent different amounts of twisting parameter θ. Therefore, the number of colors correspond to the sample size.

Table 1.1 shows the performance of our estimator based on 1000 simulations. The performance is measured by the mean and standard deviation of the absolute error $d_g(\hat{c}, c)$ and those of the estimated twisting parameter $\hat{\sigma}_\theta$. The accuracy is increased for larger n or κ. In general, we observe quite accurate estimates even for a larger noise level ($\kappa = 100$).

Table 1.1: Twisting: Mean absolute error for $\hat{c}$, and the estimates $\hat{\sigma}_\theta$

Twisting (unit: degrees)			
κ	n	$d_g(\hat{c}, c)$	$\hat{\sigma}_\theta(\sigma_\theta = 17.189)$
100	30	3.174 (2.294)	17.209 (2.152)
100	100	1.563 (0.890)	17.324 (1.250)
1000	30	0.561 (0.317)	17.045 (2.133)
1000	100	0.289 (0.164)	17.173 (1.235)



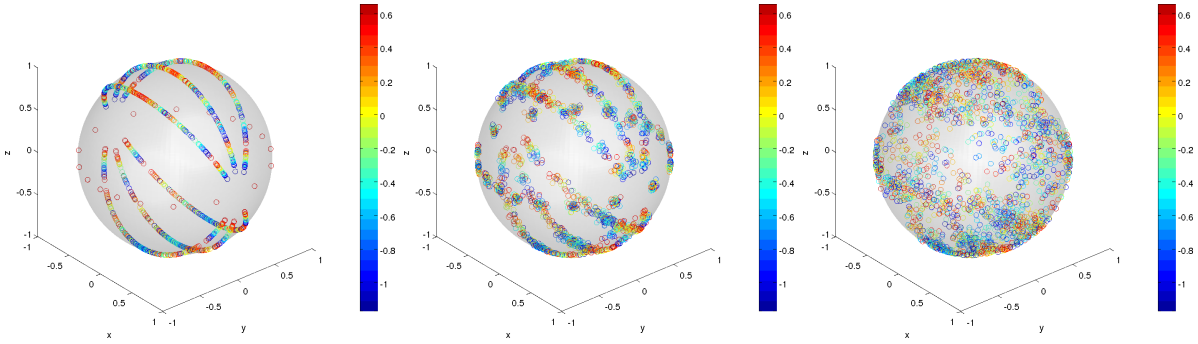
(a) without noise (b) $\kappa = 1000$ (c) $\kappa = 100$

Figure 1.4: (Rotational bending) Boundary normals using different concentration parameter κ. Different colors represent different amounts of bending parameter α.

Bending

Figure 1.4 shows 30 samples of 72 different normal directions after rotational bending using different noise levels. The case of quadratic bending is illustrated in Fig. 1.5. Different colors represent different amounts of bending parameter α. The different effects of rotational and quadratic bending to the boundary normals can be compared by examination of Fig. 1.4a and Fig. 1.5a. Rotationally bent directions are at concentric small circles (Fig. 1.4a) while quadratically bent directions are at circles with different centers (Fig. 1.5a). Despite the major violation of our assumption of concentric circles, the proposed method surprisingly works well for quadratic bending models, as Table 1.2 summarizes.

Table 1.2 shows the performance of our estimator for the rotational and quadratic bending models. The absolute errors between the true axis c and the estimated rotation axis $\hat{c}$ are virtually small for both models. The performance of the estimator is enhanced for larger κ and n. Moreover, as expected, the empirical errors are smaller for rotational bending than quadratic bending. Note that the estimates $\hat{\sigma}_\alpha$ of rotational bending model are biased upwards, which we discuss in section 1.2.1. The parameter σ_α affects the quality of estimators.



(a) without noise (b) $\kappa = 1000$ (c) $\kappa = 100$

Figure 1.5: (Quadratic bending) Boundary normals using different concentration parameter κ. Different colors represent different amounts of bending parameter α.

Table 1.2: Rotational bending: Mean absolute error for $\hat{c}$, and the estimates $\hat{\sigma}_\alpha$

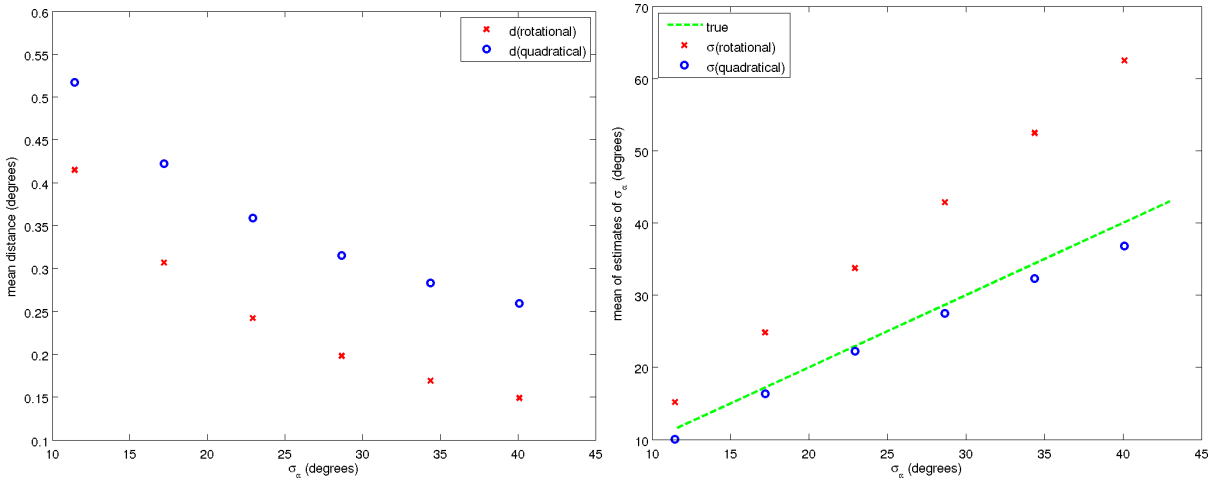
(unit: degrees)		Rotational bending		Quadratic bending	
κ	n	$d_g(\hat{c}, c)$	$\hat{\sigma}_\alpha(\sigma_\alpha = 22.918)$	$d_g(\hat{c}, c)$	$\hat{\sigma}_\alpha(\sigma_\alpha = 22.918)$
100	30	0.898 (0.492)	34.133 (4.429)	1.494 (0.871)	23.277 (3.389)
100	100	0.467 (0.261)	34.179 (2.592)	0.789 (0.470)	22.880 (2.454)
1000	30	0.242 (0.127)	33.739 (4.448)	0.359 (0.213)	22.203 (3.184)
1000	100	0.127 (0.069)	33.973 (2.581)	0.193 (0.112)	22.276 (1.891)

In particular, larger σ_α leads to a greater bias of $\hat{\sigma}_\alpha$, meanwhile it yields a better estimate of $\hat{c}$ (Fig. 1.6).

Hierarchical motion

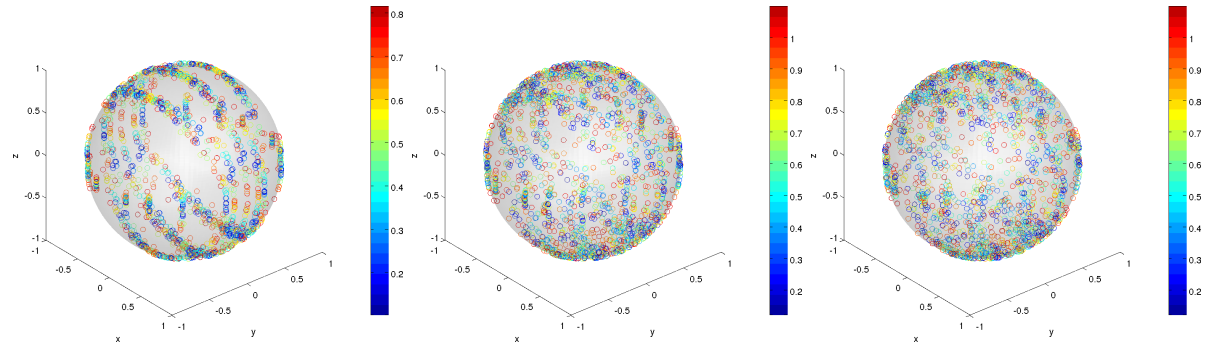
Figure 1.7 shows 30 samples of 64 different normal directions using different concentration parameters κ, twistings σ_θ and a fixed bending parameter $\sigma_\alpha = 0.4$. We have excluded 8 normal directions attached at $x = 0$. Different colors represent different amounts of absolute rotation, which are $|\alpha_j| + |\theta_j|$. In Fig. 1.7a we see rotations of normals along small circles around the y-axis. The clear motion pattern disappeared after an increase of σ_θ and κ in Figure 1.7b and Figure 1.7c.

Table 1.3 shows the performance of our estimator based on 1000 simulations under hierarchical rotational bending and twisting. As expected, the rotation axis estimates are less accurate than for single motions. Moreover, the estimate of the secondary rotation axis is less accurate than the estimate of the primary axis, particularly for $\kappa = 100$. The bias in the estimates of the rotation angle will be further discussed shortly.



(a) (b)

Figure 1.6: (a) Mean absolute error $d_g(\hat{c}, c)$ after quadratic and rotational bending for different σ_α values. (b) Mean of estimated bending parameter $\hat{\sigma}_\alpha$ after quadratic and rotational bending for different σ_α values. The green dashed line marks the first diagonal.



(a) $\epsilon_j = 0, \sigma_\theta = 0.1$ (b) $\epsilon_j = 0, \sigma_\theta = 0.3$ (c) $\kappa = 1000, \sigma_\theta = 0.3$

Figure 1.7: (Hierarchical deformation) Boundary normals using $\sigma_\alpha = 0.4$, different concentration parameter κ and twisting σ_θ.

Two initial value choices (from Principal Arc Analysis and random directions) are applied. In contrast to the s-rep analysis in the main article, the results in Table 1.3 are very similar for both choices.

In the four sets of analyses, we have shown accurate estimation results of the proposed method for the boundary normal data. In particular, the estimation from the quadratic bending model is surprisingly accurate despite the misspecified model. We believe that the proposed method will lead to robust estimates in other deformation models, which are similar to a rotational deformation.

Table 1.3: (Rotational bending and twisting) Mean absolute error for $\hat{c}_1$, $\hat{c}_2$, and the estimates $\hat{\sigma}_\alpha$, $\hat{\sigma}_\theta$.

(unit: degrees)		1st rotation axis		2nd rotation axis	
κ	n	$d_g(\hat{c}_1, c_1)$	$\hat{\sigma}_\alpha(\sigma_\alpha = 22.918)$	$d_g(\hat{c}_2, c_2)$	$\hat{\sigma}_\theta(\sigma_\theta = 17.189)$
initialization by 1st and 2nd principal component					
100	30	3.526 (2.775)	18.125 (2.370)	20.047 (16.717)	9.232 (1.598)
100	100	1.902 (1.444)	18.268 (1.337)	11.081 (13.477)	9.239 (1.143)
1000	30	2.683 (2.272)	17.785 (2.377)	8.570 (9.031)	9.119 (1.323)
1000	100	1.637 (1.126)	18.101 (1.349)	3.901 (2.459)	9.367 (0.691)
initialization by 1st principal component and a random direction					
100	30	3.496 (2.762)	18.125 (2.367)	19.133 (15.445)	9.295 (1.498)
100	100	1.866 (1.390)	18.260 (1.342)	8.944 (6.753)	9.390 (0.798)
1000	30	2.678 (2.272)	17.785 (2.377)	8.479 (8.702)	9.125 (1.299)
1000	100	1.635 (1.127)	18.102 (1.349)	3.892 (2.451)	9.367 (0.691)

1.2.1 Estimation of σ_α

A precise estimation of σ_α under the bending model is an interesting open problem. We have observed that the amount of swing is nonlinear, and conjecture that the change of the surface curvature in the object is responsible for the additional swing of the directions. Figure 1.8 exemplifies such a non-linear relationship.

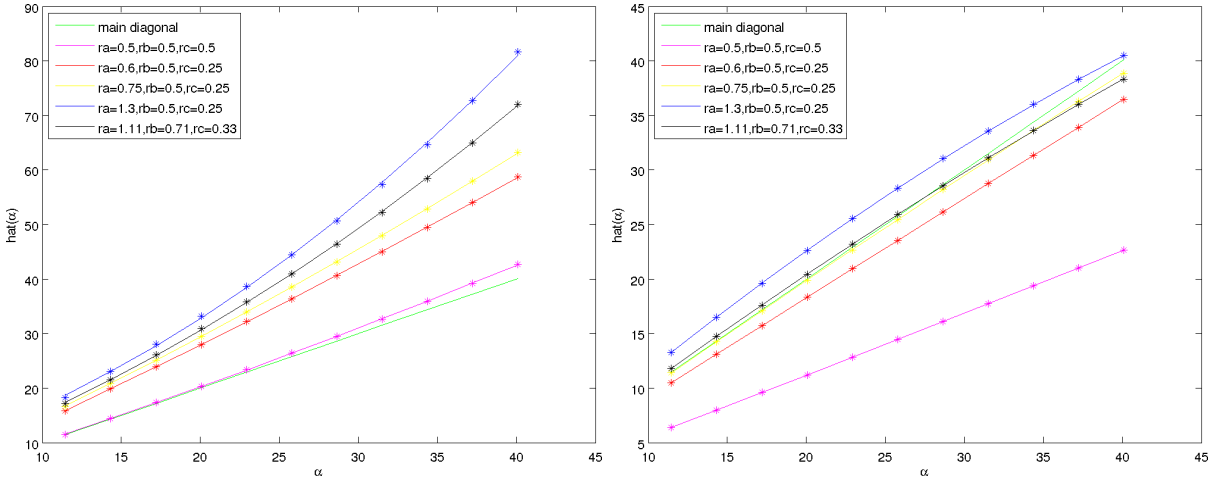
In the case of rotational bending, we may assume that our estimate $\hat{\sigma}_\alpha$ and the parameter σ_α are related through a quadratic function as Fig.1.8a suggests. If such a quadratic function $f(\sigma_\alpha) = p_0 + p_1\sigma_\alpha + p_2\sigma_\alpha^2 = \hat{\sigma}_\alpha$ is known or can be estimated efficiently by a least square quadratic polynomial, one can correct the estimate of σ_α for fixed r_a, r_b and r_c by

$$\hat{\sigma}_\alpha^{\text{new}} = -\frac{1}{2p_2} \pm \sqrt{\frac{1}{4p_2^2}(p_1^2 - 4p_0p_2 + 4p_2\hat{\sigma}_\alpha)},$$

with $f''(\hat{\sigma}_\alpha^{\text{new}}) \geq 0$. A similar modification can be made for quadratic bending (Fig.1.8b).

In general, we believe that modeling based on the curvatures will improve our current estimator. Such a modeling is beyond the scope of this paper, and we list a few important notions of curvature that have potential in modeling.

Most common curvature measures are the principal curvatures (κ_1, κ_2) with $\kappa_1 \geq \kappa_2$, Gaussian curvature, and mean curvature. These measures are calculated from the first and



(a) Rotational bending (b) Quadratic bending

Figure 1.8: Relationship between $\hat{\alpha}$ and α for various (r_a, r_b, r_c).

second fundamental form (see Gray (1998) and Kühnel (2006)). Koenderink (1990) indicated that Gaussian curvature and mean curvature are not representatives of local shapes because Gaussian curvature is identical for all local approximations for which the ratio of the principal curvatures (κ_1, κ_2) is equal. Therefore, he introduced the two alternative quantities: *shape index* S and *curvedness* C,

$$S = \frac{2}{\pi} \tan^{-1}\left(\frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1}\right), \qquad (\kappa_1 \geq \kappa_2) \tag{1.5}$$

$$C = \frac{2}{\pi} \ln\left(\sqrt{\frac{1}{2}(\kappa_1^2 + \kappa_2^2)}\right). \tag{1.6}$$

1.3 Quadratic bending of s-reps

The proposed method is also applied to the fitted s-reps of quadratically bent ellipsoids. After fitting s-reps to the raw images discussed in Section 1.2, we obtained the estimated axis of rotation of $\hat{c}_b = (-0.026, 0.999, 0.002)$ with a distance of 1.517° to the true y-axis $c_b = (0, 1, 0)$. Similar to mesh ellipsoid surfaces, the distance of the rotation axis estimate to the true axis is lower in case of rotational bending compared to quadratic bending.

Table 2.1: Numerical performance over 1000 replications, for hierarchical deformations.

Hierarchical rotations (unit: degrees)			$d_g(\hat{c}_1, c_1)$	$d_g(\hat{c}_2, c_2)$	$\hat{\sigma}_\theta(\sigma_\theta = 22.5)$	$\hat{\sigma}_\psi(\sigma_\psi = 15)$
Body 1	$\kappa = 100$	$n = 30$	4.527 (3.591)	4.929 (3.007)	22.693 (3.031)	15.890 (2.096)
		$n = 100$	2.201 (1.206)	2.944 (1.550)	22.698 (1.589)	15.844 (1.126)
	$\kappa = 1000$	$n = 30$	2.084 (1.364)	1.275 (0.701)	22.385 (2.984)	14.940 (1.941)
		$n = 100$	1.123 (0.741)	0.652 (0.338)	22.492 (1.542)	15.030 (1.110)
Body 2	$\kappa = 100$	$n = 30$	2.617 (1.762)	3.066 (3.099)	22.440 (2.959)	15.094 (2.011)
		$n = 100$	1.366 (0.746)	1.682 (0.870)	22.542 (1.562)	15.219 (1.073)
	$\kappa = 1000$	$n = 30$	1.099 (1.171)	0.921 (2.349)	22.339 (2.983)	14.872 (1.945)
		$n = 100$	0.568 (0.354)	0.438 (0.236)	22.470 (1.543)	14.981 (1.099)

2 Simulation results for the hierarchical deformation model

Table 2.1 summarizes the numerical performance of estimation of the hierarchical deformation model (11) as discussed in Section 4 of the main article. We have used the data-driven method to choose the initial values as described in Section 3 of the main article. The results are comparable to those from the simpler models in Section 4 and are fairly successful.

3 Discussion of model bias

A possibly important issue that is left unanswered is whether the parameters c_0 and $r = d_g(c_0, \mu)$ of the model

$$X = R(c_0, \theta)\mu \oplus \epsilon \tag{3.1}$$

are the minimizer $(\tilde{c}, \tilde{r})$ of the least squares problem

$$\min_{c,r} \mathbb{E}\rho^2\{\delta(c, r), X\} = \min_{c,r} \mathbb{E}\{d_g(c, X) - r\}^2. \tag{3.2}$$

Rivest (1999) has shown that when c_0 is known, the minimizer $\tilde{r} = \text{argmin}_r\, \mathbb{E}\{d_g(c_0, X) - r\}^2$ is biased towards $\pi/2$, i.e. $\tilde{r} > r = d_g(c_0, \mu)$ if $r < \pi/2$. In fact, given any c for the axis of rotation, $\tilde{r}_c = \mathbb{E}\{d_g(c, X)\}$ minimizes $\mathbb{E}\{d_g(c, X) - r\}^2$. Now to focus on c, $\min_{c,r} \mathbb{E}\{d_g(c, X) - r\}^2 = \min_c \mathbb{E}\{d_g(c, X) - \tilde{r}_c\}^2 = \min_c \text{Var}\{d_g(c, X)\}$. Therefore c_0 of (3.1) is the minimizer of (3.2) if for all $c \in S^2$,

$$\text{Var}\{d_g(c_0, X)\} \leq \text{Var}\{d_g(c, X)\}. \tag{3.3}$$

Table 3.1: Distance to true axis (measured in degrees)

	σ_θ				
r	0.01	0.1	0.2	0.5	1
$\pi/16 = 11.25$	11.25	11.19	10.99	9.10	2.29
$\pi/8 = 22.50$	22.50	22.36	21.89	2.55	0.20
$\pi/4 = 45.00$	44.98	44.81	42.73	0.22	0.00
$\pi/3 = 60.00$	59.97	59.79	3.72	0.16	0.00
$\pi/2 = 90.00$	90.00	90.00	0.49	0.02	0.02

This inequality may be satisfied when $r\sigma_\theta$ is large compared to the standard deviation of the error ϵ. If σ_θ or r is 0, there is no variation due to the rotation of $R(c_0, \theta)$, which makes the model unidentifiable. Heuristically, small σ_θ makes the estimation difficult. Likewise, the variation due to rotation is small if the rotation radius $r = d_g(c, \mu)$ is small. The standard deviation of the length $r\theta$ of the arc on $\delta(c, r)$ is $r\sigma_\theta$. Hence, the hypothesis:

(H) If (3.3) is not satisfied, the least-squares estimator may be biased.

While we have not succeeded in finding analytic forms of conditions that satisfy (3.3), a simulation study has been carried out to support our hypothesis (H). For $c_0 = e_3 = (0, 0, 1)'$, $\mu_r = \mu(r) = \cos(r)c_0 + \sin(r)e_1$, and $\theta \overset{\text{iid}}{\sim} N(0, \sigma_\theta^2)$, X is distributed as the von Mises–Fisher distribution with center $R(c, \theta)\mu_r$ and the concentration parameter $\kappa = 100$. We then evaluated the minimizer $\tilde{c}$ of $\text{Var}\{d_g(c, X)\}$ based on 5×10^5 Monte-Carlo random observations of X, for different combinations of $r \in \{\pi/16, \pi/8, \pi/4, \pi/3, \pi/2\}$ and $\sigma_\theta \in \{0.1, 0.2, 0.5, 1\}$ in radian. The result of the experiment is summarized as the distance between $\tilde{c}$ and c_0 in Table 3.1.

It can be checked from Table 3.1 that the distance between $\tilde{c}$ and c_0 is smaller for larger values of σ_θ and r. Moreover, for sufficiently large σ_θ and r, $d_g(\tilde{c}, c_0) = 0$, which leads to c_0 from the model (3.1) satisfying (3.3). On the other hand, when σ_θ and r are small with respect to the variance of the error, the rotation (3.1) becomes unidentifiable. This is further illustrated at Figure 3.1, which shows the scatter of 100 random observations from model (3.1), with $(r, \sigma_\theta) = (\pi/16, 0.2)$ on the left and $(\pi/4, 0.5)$ on the right panel. The left panel suggests that when (r, σ_θ) are small, the rotation about the axis c_0 is not distinguished and the optimal $\tilde{c}$ is near μ and $d_g(\tilde{c}, c_0)$ is large (cf. Table 3.1). The right panel illustrates that the rotation is visually identified for large values of (r, σ_θ), and for such a case, the parameter c_0 is the solution of the least squared problem (3.2).

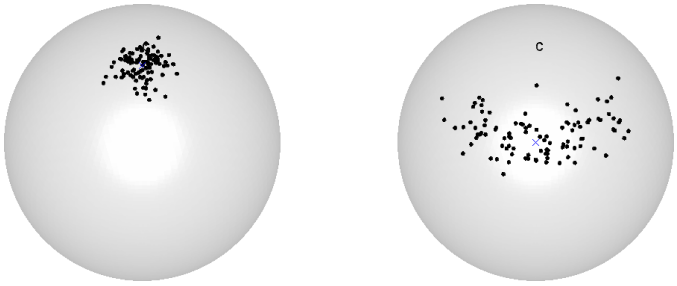


Figure 3.1: The scatter of 100 random observations from the model (3.1), with $(r, \sigma_\theta) = (\pi/16, 0.2)$ on the left and $(\pi/4, 0.5)$ on the right panel. The north pole, depicted as c in the figure, is the true axis of rotation c_0. The blue `x` visualizes μ.

4 Numerical studies with misspecified parameters

In this section, we study the impact of parameter misspecification of the estimator, particularly of the function f_j that model the relationships between the rotation angles θ_j. We study parameter misspecification over different rotational deformation situations as described in Section 4 in the main article.

Recall, Section 4 in the main article reports studies for two different objects. The first object (Body 1) consists of $K = 4$ directions, while the second object (Body 2) contains $K = 8$ directions. The von Mises–Fisher distribution is used for the distribution of errors. Three rotation models (indexed by equation number in the main article) are considered for each object:

- Model (2)–*Rigid rotation*: $c = (1, 0, 0)'$, $\theta_j = f_j(\theta) = \theta$, where $\theta \sim N(0, \sigma_\theta^2)$ and $\sigma_\theta = \pi/12 \approx 15°$.
- Model (10)–*Twisting*: $c = (0, 1, 0)'$, $\theta_j = f_j(\theta) = 1_{j\in I_1}\theta - 1_{j\in I_2}\theta$, where $\theta \sim N(0, \sigma_\theta)^2$, $\sigma_\theta = \pi/8 \approx 22.5°$. Here, I_1 and I_2 are the partitions of $\{1, \ldots, K\}$ according to the right and left sides whereas $I_1 = \{1, 2\}$ and $I_2 = \{3, 4\}$ for Body 1 and $I_1 = \{1, \ldots, 4\}$ and $I_2 = \{5, \ldots, 8\}$ for Body 2.
- Model (11)–*Hierarchical deformations*: $c_1 = (1, 0, 0)'$, $c_2 = (1/\sqrt{2}, -1/\sqrt{2}, 0)'$, $\theta_j = \theta$ and $\psi_j = f_j(\psi) = 1_{j\in I_1}\psi - 1_{j\in I_2}\psi$, where $\theta \sim N(0, \sigma_\theta)^2$, $\sigma_\theta \approx 22.5°$ and $\psi \sim N(0, \sigma_\psi)^2$, $\sigma_\psi \approx 15°$. The I_1 and I_2 are the same partition used in the twisting model above.

The hierarchical model represents deformations by a rigid rotation and oblique twist. For each model, we generate $n = 30, 100$ rotationally deformed objects with different error concentration parameters $\kappa = 100, 1000$. These are replicated 1000 times, and the estimation quality is measured by $d_g(\hat{c}, c)$ and $\hat{\sigma}_\theta = \sum_{i=1}^n \hat{\theta}_i^2/n$.

Table 4.1: Numerical performance over 1000 replications, for single deformation models.

Rotation (unit: degrees)		Body 1 $d_g(\hat{c}, c)$	Body 1 $\hat{\sigma}_\theta$ ($\sigma_\theta = 15$)	Body 2 $d_g(\hat{c}, c)$	Body 2 $\hat{\sigma}_\theta$ ($\sigma_\theta = 15$)
$\kappa = 100$	$n = 30$	4.133 (2.26)	3.314 (0.44)	2.905 (1.60)	6.771 (0.85)
	$n = 100$	2.235 (1.18)	3.308 (0.24)	1.560 (0.83)	6.816 (0.51)
$\kappa = 1000$	$n = 30$	1.166 (0.64)	1.037 (0.13)	0.841 (0.46)	6.698 (0.88)
	$n = 100$	0.656 (0.34)	1.039 (0.07)	0.448 (0.22)	6.744 (0.46)
Twisting (unit: degrees)		Body 1 $d_g(\hat{c}, c)$	Body 1 $\hat{\sigma}_\theta$ ($\sigma_\theta = 22.5$)	Body 2 $d_g(\hat{c}, c)$	Body 2 $\hat{\sigma}_\theta$ ($\sigma_\theta = 22.5$)
$\kappa = 100$	$n = 30$	2.761 (1.51)	3.669 (0.48)	4.062 (3.67)	17.150 (2.12)
	$n = 100$	1.482 (0.78)	3.658 (0.25)	1.889 (1.35)	17.055 (1.27)
$\kappa = 1000$	$n = 30$	0.803 (0.43)	1.139 (0.15)	1.017 (0.74)	16.760 (2.19)
	$n = 100$	0.446 (0.23)	1.147 (0.08)	0.536 (0.36)	16.857 (1.17)

The estimation results using correct parameters are reported for the single deformation models in Table 1 in the main article and for the hierarchical deformation model in Table 2.1 above.

4.1 Parameter f_j

The modelling of the function f_j can be challenging as discussed in Section 1.2.1 or for the s-rep model in case of bent, and bent and twisted ellipsoids in chapter 5 in the main article. Therefore, it is crucial to study the impact of misspecification of f_j.

Table 4.1 reports mean and standard deviations of the measures for the single deformation models. The true underlying rigid rotation deformation is indicated by $f_j(\theta) = \theta, j = 1, \ldots, 4$ for Body 1 and is misspecified by $f_j(\theta) = \theta, j = 1, 2$ and $f_j(\theta) = -\theta, j = 3, 4$. Body 2 is misspecified by $f_j(\theta) = (j/10)\theta$ instead of $f_j(\theta) = \theta$ for $j = 1, \ldots, 8$. In both cases the misspecification does not effect the estimated rotation axis $\hat{c}$ but leads to a poor prediction of $\hat{\sigma}_\theta$. The true underlying twisting motion of model (10) is given by $f_j(\theta) = \theta, j = 1, 2,\ f_j(\theta) = -\theta, j = 3, 4$ for Body 1 and is misspecified by $f_j(\theta) = \theta, j = 1, 4,\ f_j(\theta) = -\theta, j = 2, 3$. Body 2 is misspecified by $f_j(\theta) = \theta,\ f_l(\theta) = -0.5\theta$ whereas the true parameter is given by $f_j(\theta) = \theta,\ f_l(\theta) = -\theta$ for $j = 1, \ldots, 4$ and $k = 5, \ldots, 8$. The comparison of Table 4.1 to Table 1 in the main article shows that a misspecification of the parameter f_j does not effect the rotation axis estimation of a single deformation by fitting concentric small circles whereas the predicted rotation angle is biased by misspecification of f_j. The specification of f_j models the relationships between the rotation angles θ_j and is therefore crucial for their prediction.

On the other hand, the partition I_1 and I_2 is not explicitly used in the estimation proce-

Table 4.2: True and misspecified parameter f_j for Body 1 in the hierarchical deformation model (11).

	$f_{11}(\theta)$	$f_{12}(\theta)$	$f_{13}(\theta)$	$f_{14}(\theta)$	$f_{21}(\psi)$	$f_{22}(\psi)$	$f_{23}(\psi)$	$f_{24}(\psi)$
True	θ	θ	θ	θ	ψ	ψ	$-\psi$	$-\psi$
Table 4.4	θ	θ	$-\theta$	$-\theta$	ψ	ψ	ψ	ψ
Table 4.5	θ	$-\theta$	$-\theta$	θ	$-\psi$	ψ	ψ	$-\psi$

Table 4.3: True and misspecified parameter f_j for Body 2 in the hierarchical deformation model (11).

Hierarchical rotations		j							
		1	2	3	4	5	6	7	8
True	$f_{1j}(\theta)$	θ	θ	θ	θ	θ	θ	θ	θ
	$f_{2j}(\psi)$	ψ	ψ	ψ	ψ	$-\psi$	$-\psi$	$-\psi$	$-\psi$
Table 4.4	$f_{1j}(\theta)$	0.5θ	0.5θ	0.5θ	0.5θ	0.5θ	0.5θ	0.5θ	0.5θ
	$f_{2j}(\psi)$	ψ	ψ	ψ	ψ	$-\psi$	$-\psi$	$-\psi$	$-\psi$
Table 4.5	$f_{1j}\theta)$	0.1θ	0.2θ	0.3θ	0.4θ	θ	θ	θ	θ
	$f_{2j}(\psi)$	ψ	ψ	0.7ψ	0.7ψ	-0.2ψ	-0.2ψ	-0.3ψ	-0.3ψ

dure of the rotation axis. The partition I_1 and I_2 is implied by the function f_j. A partition I_1 and I_2 of indices $\{1, \ldots, K\}$ represents sets of direction vectors that rotate together, i.e., the partition models the deformation type. In the simulated examples, the partitions are chosen to model the bending and twisting deformation accordingly. The specification of f_j gives the possibility to incorporate additional prior knowledge about the statistical model of the rotation angle θ_j, e.g., linear or quadratic relation by the distance of the direction to the rotation axis. Nevertheless, the modelling of the function f_j can be challenging as discussed before. There are real data applications where the definition of a partition is naturally motivated, e.g., by the physical structure of the body. An example is to estimate the rotational deformation between the upper and lower leg as studied in Section 6 in the main article. This example can be extended to all joints inside the human body and to all objects which are connected by a joint. On the other hand, there is a group of data sets where such a partition is not obvious, e.g., in the s-rep model of the hippocampus. A first approach could be to define the partition on the basis of an observation of a medical expert. An automatic detection of partitions and clusters is an interesting future research direction.

In contrast to the single deformation model, a misspecification of f_j has an impact in the hierarchical deformation model by the iterative back-and-forward deformations of the random direction vectors. Therefore, a misspecification of the parameter might guide the iterative estimation procedure to fall in a local minimum, and leads to a poor estimation of the rotation axes and angles. Table 4.2 and Table 4.3 report the true and misspecified f_j for the simulation studies whereas the estimation results are summarized in Table 4.4

Table 4.4: Numerical performance over 1000 replications, for hierarchical deformations using moderate misspecified parameters.

Hierarchical rotations (unit: degrees)			$d_g(\hat{c}_1, c_1)$	$d_g(\hat{c}_2, c_2)$	$\hat{\sigma}_\theta(\sigma_\theta = 22.5)$	$\hat{\sigma}_\psi(\sigma_\psi = 15)$
Body 1	$\kappa = 100$	$n = 30$	45.223 (3.61)	44.792 (3.69)	15.013 (2.00)	22.725 (3.01)
		$n = 100$	45.188 (1.73)	44.423 (1.79)	14.992 (1.06)	22.731 (1.58)
	$\kappa = 1000$	$n = 30$	43.648 (1.29)	44.257 (2.67)	14.454 (1.89)	22.442 (2.99)
		$n = 100$	43.705 (0.72)	43.981 (1.42)	14.482 (1.06)	22.510 (1.54)
Body 2	$\kappa = 100$	$n = 30$	2.617 (1.76)	3.066 (3.09)	22.440 (2.96)	15.094 (2.01)
		$n = 100$	1.100 (1.17)	0.921 (2.35)	22.339 (2.98)	14.872 (1.94)
	$\kappa = 1000$	$n = 30$	1.366 (0.74)	1.683 (0.87)	22.542 (1.56)	15.219 (1.07)
		$n = 100$	0.569 (0.35)	0.438 (0.23)	22.470 (1.54)	14.981 (1.09)

Table 4.5: Numerical performance over 1000 replications, for hierarchical deformations using drastically misspecified parameters..

Hierarchical rotations (unit: degrees)			$d_g(\hat{c}_1, c_1)$	$d_g(\hat{c}_2, c_2)$	$\hat{\sigma}_\theta(\sigma_\theta = 22.5)$	$\hat{\sigma}_\psi(\sigma_\psi = 15)$
Body 1	$\kappa = 100$	$n = 30$	11.576 (4.66)	33.737 (4.94)	5.423 (2.58)	2.619 (3.45)
		$n = 100$	11.272 (2.38)	33.372 (2.48)	4.749 (1.02)	0.966 (1.44)
	$\kappa = 1000$	$n = 30$	11.228 (3.91)	33.586 (4.06)	3.597 (1.98)	1.186 (2.34)
		$n = 100$	11.183 (2.19)	33.445 (2.25)	3.295 (0.32)	0.413 (0.26)
Body 2	$\kappa = 100$	$n = 30$	12.204 (5.01)	33.456 (5.11)	15.917 (1.88)	214.219 (20.16)
		$n = 100$	11.337 (2.46)	33.959 (2.47)	16.180 (1.15)	214.553 (10.75)
	$\kappa = 1000$	$n = 30$	11.838 (4.40)	33.787 (4.46)	15.727 (2.01)	219.414 (20.99)
		$n = 100$	11.289 (2.30)	33.978 (2.34)	16.316 (1.20)	219.587 (10.45)

and Table 4.5. At first we have changed the order of the original parameters f_1, f_2 for Body 1 in Table 4.4. The first estimated rotation axis $\hat{c}_1$ is around $(1/\sqrt{2}, -1/\sqrt{2}, 0)'$ and the second estimated rotation axis $\hat{c}_2$ is around $(1, 0, 0)'$, i.e., the estimator has interchanged the true underlying deformations which results in a distance of approximately 45 degree of $\hat{c}_1$ to c_1 and $\hat{c}_2$ to c_2. Nevertheless, the order of generalized rotational deformations are not interchangeable in general, and a misspecification might bias the results. In a second set, we have misspecified f_{1j} globally by a factor of 0.5 and kept the original f_{2j} for Body 2. The means and standard deviations in Table 4.4 show only small changes compared to Table 2.1 and are very accurate. A global scale change does not change the relationships between the rotation angles. Both cases demonstrate the performance of the hierarchical estimation procedure in case of a moderate misspecification by reasonable estimates.

In addition, two cases with drastically misspecified parameters are reported. In a third scenario, f_1 and f_2 are misspecified so as to generate different deformations for Body 1 in Table 4.5. Both the means and the standard deviations show poor estimation results. In a last set we modified f_1 and f_2 by keeping the underlying deformation direction of each

Table 4.6: Numerical performance over 1000 replications, for hierarchical deformations with misspecified order of primary and secondary rotation axis.

Hierarchical rotations (unit: degrees)			$d_g(\hat{c}_1, c_1)$	$d_g(\hat{c}_2, c_2)$	$\hat{\sigma}_\theta(\sigma_\theta = 22.5)$	$\hat{\sigma}_\psi(\sigma_\psi = 15)$
Body 1	$\kappa = 100$	$n = 30$	4.624 (2.66)	4.876 (2.85)	22.720 (3.00)	15.896 (2.10)
		$n = 100$	2.258 (1.30)	2.949 (1.54)	22.701 (1.58)	15.847 (1.12)
	$\kappa = 1000$	$n = 30$	2.044 (1.27)	1.279 (0.70)	22.382 (2.98)	14.947 (1.94)
		$n = 100$	1.124 (0.72)	0.656 (0.33)	22.492 (1.54)	15.033 (1.11)
Body 2	$\kappa = 100$	$n = 30$	2.590 (1.47)	2.992 (1.75)	22.439 (2.96)	15.094 (2.01)
		$n = 100$	1.323 (0.72)	1.688 (0.87)	22.541 (1.56)	15.220 (1.07)
	$\kappa = 1000$	$n = 30$	1.063 (0.67)	0.849 (0.47)	22.336 (2.98)	14.873 (1.94)
		$n = 100$	0.567 (0.35)	0.438 (0.23)	22.470 (1.54)	14.982 (1.09)

direction vector for Body 2 but changing the amount of deformation locally. Also in this scenario the hierarchical estimator shows poor estimation results in Table 4.5.

4.2 Primary and secondary rotation

The hierarchical model is a first attempt to model and estimate more then one rotational deformation. The order of two rotations is not interchangeable and is considered as a part of the statistical model which attempts to describe the nature as well as possible. Therefore, we assume a primary rotation $R(c_1, \theta_j)$ and a secondary rotation $R(c_2, \theta_j)$ in our hierarchical model. The order of rotation can be misspecified in two different ways in the proposed approach. At first, we might interchange f_1 and f_2 as discussed in Section 4.1 above for Body 1 in Table 4.4. In addition, the primary and secondary rotation has to be specified for the initialization.

Table 4.6 summarizes the results in estimation of the hierarchical deformation model (11) with interchanged primary and secondary rotation for the initialization. The results are similar to Table 2.1. The estimator converges to the same results in this example. Nevertheless, a misspecification might increase the risk that the estimation procedure converges in a different local minimum.

4.3 Discussion

The introduction of f_j in our model has advantages and disadvantages. We study generalized rotational deformations on the basis of directional data, and the rotation of a direction from one to another place on the sphere is not uniquely defined in the hierarchical case. Therefore, different functions f_j can describe different rotational deformations. Prior knowledge is nec-

essary in order to restrict the rotational directions to avoid the convergence of the optimizer into a different local minimum and to avoid an overfit of c_1 and c_2 as mentioned Section 4 in the main article.

The development of a method to predict the function f_j from a given data set is an interesting open research question. Furthermore, an automatic classification of spoke directions into a partition I_1 and I_2 and a final classification of the deformation type into rigid rotation, bending or twisting are of future interest.

5 Computational complexity of the algorithm

The computational complexity of the proposed estimation procedure is now reported in two forms: the asymptotic time complexity and finite sample time measurements.

The asymptotic time complexity of the proposed algorithm depends on the number of samples n and the number of direction vectors K. In particular the optimization problem,

$$(\hat{c}, \hat{\mathbf{r}}) = \underset{c, r_1, \ldots, r_K}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{K} d_g^2\{\delta(c, r_j), X_{ij}\},$$

is the dominant part of the algorithm. Our algorithm to solve this nonlinear least squares problem is doubly iterative. The outer loop consists of applications of the exponential and inverse exponential maps whose time complexity is $O(nK)$. The inner loop iteratively updates $c^\dagger \in \mathbb{R}^3$ and $r_j^\dagger \in (0, \pi/2)$ by the Levenberg–Marquardt algorithm, where each iteration requires the asymptotic time complexity of $O(nK)$ mainly due to the computation of Jacobian matrix. Notice, that the computation time for inverting a 3×3 matrix is dominated by other terms and is ignored. Overall, if a finite number of iterations is assumed, then the asymptotic time complexity of the proposed algorithm is $O(nK)$.

We now turn to our attention to the complexity of the algorithm in real time, which we believe is more useful for practitioners. Computation times were measured by a personal computer with a Intel(R) Xeon(TM) 3.73GHz processor.

We have tested the estimations of three different rotational deformations

- *Model (2) - Rigid rotation*,
- *Model (10) - Twisting* and
- *Model (11) - Hierarchical deformations*

Table 5.1: Median time measurements in seconds over 100 replications.

n	Model	K					
		4	8	16	32	64	128
30	(2)	0.04	0.07	0.13	0.26	0.61	1.81
	(10)	0.11	0.16	0.26	0.40	0.79	2.00
	(11)	0.54	0.84	1.61	3.17	6.91	19.23
60	(2)	0.04	0.07	0.13	0.28	0.68	2.03
	(10)	0.10	0.18	0.29	0.42	0.85	2.18
	(11)	0.75	1.39	2.72	6.55	13.84	30.80
120	(2)	0.04	0.07	0.13	0.29	0.76	2.34
	(10)	0.12	0.18	0.27	0.44	0.92	2.50
	(11)	1.29	2.50	4.90	10.03	20.37	53.91

Table 5.2: Minimal time measurements in seconds over 100 replications..

n	Model	K					
		4	8	16	32	64	128
30	(2)	0.02	0.05	0.08	0.19	0.48	1.59
	(10)	0.03	0.07	0.12	0.25	0.61	1.75
	(11)	0.23	0.40	0.74	1.50	3.21	7.71
60	(2)	0.02	0.04	0.08	0.18	0.54	1.78
	(10)	0.03	0.05	0.14	0.27	0.68	1.89
	(11)	0.37	0.68	1.32	2.61	5.52	12.37
120	(2)	0.02	0.04	0.09	0.22	0.63	2.12
	(10)	0.03	0.09	0.15	0.33	0.79	2.38
	(11)	0.64	1.21	2.39	4.85	9.99	21.61

as described in Section 4 (the numbers correspond to the equation number in the main article), with two different types of objects to be deformed.

We first examined the computation times for a set of well-controlled objects. Using the Body 1 (as plotted in Fig. 2 of the main article) consisting of the original $K = 4$ directions, we duplicated the same direction vectors to increase the number of directions ($K = 8, 16, 32, 64, 128$) so that the effect on computation time of the different locations of direction vectors is minimized. For each sample size $n = 30, 60, 120$, we have repeated the estimation $R = 100$ times to measure the computation time required to obtain the estimates $\hat{c}$. Note that we have used von Mises-Fisher distribution with $\kappa = 1000$ for the i.i.d. errors.

Tables 5.1 and 5.2 report the median computation time and the minimal computation time among the R repetitions, respectively. With large numbers of sample and directions, the computation requires less than a minute on average, while it takes less than a second for moderate sizes of sample and directions.

From an inspection of Table 5.1, there is a trend for the computation time increasing

Table 5.3: Quality of the repeated rotation axis estimations for the time measurements by $d_g(\hat{c}, c)$ in degree.

n	Model	rotation axis	K					
			4	8	16	32	64	128
30	(2)	1	1.18	0.92	0.63	0.40	0.33	0.22
	(10)	1	0.88	0.61	0.40	0.29	0.20	0.14
	(11)	1	2.18	2.10	1.99	1.89	1.86	1.81
	(11)	2	1.43	1.09	0.77	0.60	0.48	0.43
60	(2)	1	0.74	0.63	0.42	0.30	0.21	0.16
	(10)	1	0.59	0.42	0.29	0.18	0.14	0.10
	(11)	1	1.42	1.28	1.24	1.20	1.16	1.17
	(11)	2	0.88	0.64	0.49	0.35	0.28	0.21
120	(2)	1	0.60	0.39	0.30	0.20	0.15	0.10
	(10)	1	0.37	0.30	0.20	0.14	0.10	0.07
	(11)	1	1.09	0.97	0.99	0.90	0.89	0.89
	(11)	2	0.61	0.44	0.29	0.22	0.18	0.12

approximately linear with K and also with n for all models.

By comparing the minimal time (Table 5.2) and the median time (Table 5.1), we have noticed that the computation time varies by a large amount. See for example model (11) with $n = 60$, $K = 128$; The median time is over 30 seconds while the shortest time is only 12 seconds. This is due to the slow convergence of the iterative algorithm for a particular choice of observations.

Finally, Table 5.3 reports the quality of the repeated rotation axis estimations by the mean distance of $d_g(\hat{c}, c)$, which shows higher accuracy for larger K and n as discussed in Section 4 of the article.

We also have examined the computation times with another set of objects whose base direction vectors are determined randomly. As shown in the following, the additional randomness leads to more variation in the computation times. Nevertheless, the computation time exhibits again the approximate linear increase for K and n.

The second set of objects to be deformed has $K = 8, 16, 32, 64, 128$ directions vectors, each of which is obtained from uniform distribution on a hemisphere. With $n = 30, 60, 120$ samples, we also report the time measurements from $R = 100$ replications.

Table 5.4 and 5.5 report the median and the minimum computation time in seconds, respectively. As expected, the time increases with larger K and larger n. Due to the uncontrolled model complexity (originated from the random base directions) the time difference between the median and the minimum time is larger than the previous controlled case. We conjecture that the computation times are heavily dependent on the convergence of the

Table 5.4: Median time measurements in seconds over 100 replications.

n	Model	K				
		8	16	32	64	128
30	(2)	0.08	0.14	0.26	0.63	1.83
	(10)	0.18	0.29	0.49	0.94	2.10
	(11)	1.99	6.39	17.18	51.44	130.85
60	(2)	0.08	0.14	0.28	0.70	2.04
	(10)	0.22	0.32	0.51	0.96	2.27
	(11)	2.74	6.75	24.25	63.20	265.95
120	(2)	0.08	0.14	0.30	0.77	2.40
	(10)	0.22	0.32	0.48	0.98	2.54
	(11)	9.79	18.86	42.81	129.75	305.54

Table 5.5: Minimal time measurements in seconds over 100 replications..

n	Model	K				
		8	16	32	64	128
30	(2)	0.05	0.08	0.20	0.58	1.78
	(10)	0.07	0.12	0.25	0.59	1.84
	(11)	0.42	0.84	2.30	3.45	8.03
60	(2)	0.05	0.10	0.23	0.59	1.80
	(10)	0.07	0.16	0.31	0.72	1.99
	(11)	0.72	1.36	2.75	5.99	18.78
120	(2)	0.04	0.10	0.24	0.63	2.23
	(10)	0.06	0.17	0.32	0.71	2.31
	(11)	1.25	2.43	4.91	15.24	33.16

Table 5.6: Quality of the repeated rotation axis estimations for the time measurements by $d_g(\hat{c}, c)$ in degree.

n	Model	rotation axis	K				
			8	16	32	64	128
30	(2)	1	0.92	0.64	0.45	0.33	0.21
	(10)	1	0.67	0.44	0.30	0.22	0.15
	(11)	1	1.20	0.59	0.40	0.23	0.16
	(11)	2	2.57	0.73	0.55	0.37	0.25
60	(2)	1	0.64	0.53	0.31	0.23	0.15
	(10)	1	0.43	0.26	0.22	0.14	0.10
	(11)	1	1.05	0.37	0.25	0.18	0.11
	(11)	2	2.67	0.60	0.35	0.23	0.17
60	(2)	1	0.42	0.33	0.23	0.14	0.11
	(10)	1	0.31	0.21	0.15	0.10	0.07
	(11)	1	0.86	0.27	0.17	0.11	0.07
	(11)	2	2.39	0.37	0.25	0.17	0.12

Levenberg–Marquardt algorithm.

In addition to the computation time, Table 5.6 reports the quality of the repeated rotation axis estimations by the mean distance of $d_g(\hat{c}, c)$, which again shows higher accuracy for larger K and n as discussed in Section 4 of the article.

6 Estimation procedure for knee motion during gait

In section 6 in the main article we have studied a real data example: the deformation of the upper and lower leg by two potentially non-orthogonal rotations at the knee joint during gait. These two rotations are flexion-extension about a right-to-left (medial-lateral) axis and the internal-external rotation of the lower-leg relative to the upper leg about an axis directed along the long axis of the lower leg. The data set consists of T time dependent observations $\mathbf{M}_1, \ldots, \mathbf{M}_T$ whereas each $\mathbf{M}_i$ is a collection of markers $\mathbf{M}_i = (M_{i1}, \ldots, M_{i\widetilde{K}})$ on the upper and lower leg with $M_{ij} \in \mathbb{R}^3, j = 1, \ldots, \widetilde{K}$. Let $\widetilde{I}_1$ and $\widetilde{I}_2$ be a partition of indices $\{1, \ldots, \widetilde{K}\}$ representing groups of $\widetilde{K}_1$ markers $\widetilde{I}_1$ on the upper leg and $\widetilde{K}_2$ markers $\widetilde{I}_2$ on the lower leg. Two markers $\mathbf{M}_{\iota_1}, \iota_1 \in \widetilde{I}_1$ and $\mathbf{M}_{\iota_2}, \iota_2 \in \widetilde{I}_2$ are chosen as basis points at the upper and lower leg. Set $I_1 = \widetilde{I}_1 \setminus \{\iota_1\}$, $I_2 = \widetilde{I}_2 \setminus \{\iota_2\}$, $K = \widetilde{K} - 2$, $K_1 = \widetilde{K}_1 - 1$ and $K_2 = \widetilde{K}_2 - 1$ then directions $\mathbf{X}_i = (X_{i1}, \ldots, X_{iK})$ are derived by

$$X_{ij} = \frac{M_{ij} - M_{i\iota_1}}{\|M_{ij} - M_{i\iota_1}\|}, j \in I_1, \quad X_{ij} = \frac{M_{ij} - M_{i\iota_2}}{\|M_{ij} - M_{i\iota_2}\|}, j \in I_2$$

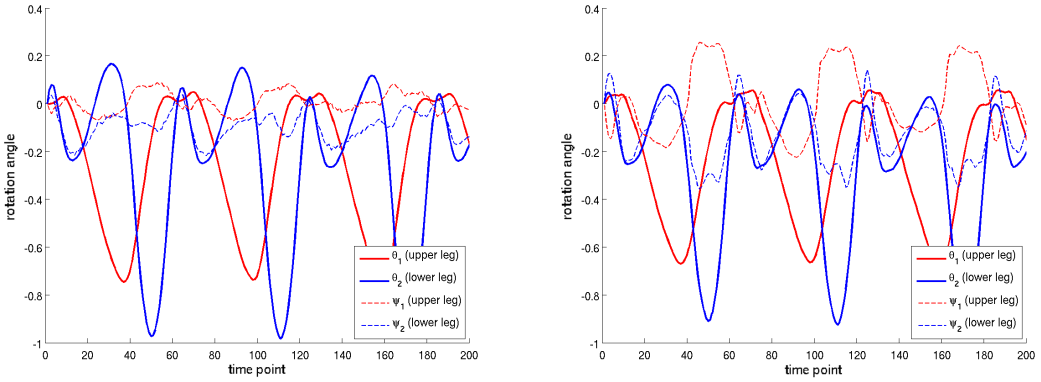
for $i = 1, \ldots, T$, which are invariant to translation and size changes.

For the sake of convenience, we assume the observations $\mathbf{X}_1, \ldots, \mathbf{X}_T$ are independent and modify the single rotation model (10) in the main article to

$$X_{ij} = R(c, \theta_{ij}) X_{1j} \oplus \epsilon_{ij} \quad (j = 1, \ldots, K). \tag{6.1}$$

A more careful modelling of the time dependencies is left for future work, e.g., by an autoregressive model as suggested by (Rivest, 2001, Section 4.1).

The first (flexion-extension) rotation axis c_1 is estimated by the estimation procedure (7) in the main article and describes a bending deformation of the upper and lower leg around the knee. The lower leg rotates relative to the upper leg whereas the upper leg rotates relative to the pelvis. At first, we estimate the rotation angles θ_i^u of the upper leg in order to estimate the rotation angles θ_i^l of the lower leg relative to the upper leg. The least squares



(a) bone (b) skin

Figure 6.1: Predicted rotation angles $(\theta_1, \psi_1) = (\hat{\theta}^u, \hat{\psi}^u)$ of directions on the upper leg and angles $(\theta_2, \psi_2) = (\hat{\theta}^l, \hat{\psi}^l)$ of directions on the lower leg for the first 200 time points. (a) Estimated rotation angles of directions derived from bone markers. (b) Estimated rotation angles of directions derived from skin markers. (angle units in radian)

estimators $(\hat{c}_1, \hat{\mathbf{r}}_1)$ can be used to estimate the parameters of (6.1) with

$$\theta_{ij} = \text{atan2}\{\langle P_{(\hat{c}_1,\hat{r}_{1j})} X_{ij}, \hat{c}_1 \times X_{1j} \rangle, \langle P_{(\hat{c}_1,\hat{r}_{1j})} X_{ij}, X_{1j} - \hat{c}_1 \cos(\hat{r}_j) \rangle\}. \tag{6.2}$$

The predictor of $\theta_i^u, i = 1, \ldots, T$ is

$$\hat{\theta}_i^u = \frac{1}{K_1} \sum_{j=1}^{K_1} \theta_{ij}, \quad j \in I_1.$$

Next, the directions are de-rotated by $\widetilde{X}_{ij} = R(\hat{c}_1, -\hat{\theta}_i^u) X_{ij}$ for $j = 1 \ldots, K$ and $i = 1, \ldots, T$. The directions $\widetilde{X}_{ij}, j \in I_1$ are directions on the upper leg and do not rotate about $\hat{c}_1$ after the inverse deformation. The remaining rotation of the lower leg relative to the upper leg is then

$$\hat{\theta}_i^l = \frac{1}{K_2} \sum_{j=1}^{K_2} \tilde{\theta}_{ij}, \quad j \in I_2 \text{ with}$$

$$\tilde{\theta}_{ij} = \text{atan2}\{\langle P_{(\hat{c}_1,\hat{r}_{1j})} \widetilde{X}_{ij}, \hat{c}_1 \times \widetilde{X}_{1j} \rangle, \langle P_{(\hat{c}_1,\hat{r}_{1j})} \widetilde{X}_{ij}, \widetilde{X}_{1j} - \hat{c}_1 \cos(\hat{r}_j) \rangle\}.$$

Finally, we obtain a set of de-rotated directions $\mathbf{Z}$ by $Z_{ij} = \widetilde{X}_{ij}, \ j \in I_1$ and $Z_{ij} = R(\hat{c}_1, -\hat{\theta}_i^l) \widetilde{X}_{ij}, \ j \in I_2$ for $i = 1, \ldots, T$.

The estimation of the second (internal-external) rotation axis $(\hat{c}_2, \hat{\mathbf{r}}_2)$ and $\hat{\psi}^u, \hat{\psi}^l$ follows the same steps of the above paragraph using using $\mathbf{Z}_i$ instead of $\mathbf{X}_i$ for $i = 1, \ldots, T$.

In addition to the estimates $\hat{c}_1$ and $\hat{c}_2$, we estimate the rotation axes $\hat{c}_{1j}$ and $\hat{c}_{2j}$ for each

marker $j = 1 \ldots, K$. Therewith, we can quantify the estimations by the dispersion $\sigma_{d_g(\hat{c}_1, \hat{c}_{1j})}$ of the geodesic distance $d_g(\hat{c}_1, \hat{c}_{1j})$ and $\sigma_{d_g(\hat{c}_2, \hat{c}_{2j})}$ respectively.

Figure 6.1 reports the predicted rotation angles $(\hat{\theta}_i^u, \hat{\psi}_i^u)$ of the upper leg relative to the laboratory reference frame and $(\hat{\theta}_i^l, \hat{\psi}_i^l)$ of the lower leg relative to the upper leg for the first 200 time points. The larger observed rotation angles around the second rotation axis for the skin data is due to the well-known deformation of the skin surface.

References

Gray, A. (1998), *Modern Differential Geometry of Curves and Surfaces with Mathematica*, CRC Press, 2nd ed.

Koenderink, J. J. (1990), *Solid Shape*, The MIT Press.

Kühnel, W. (2006), *Differential Geometry*, vol. 16, Student Mathematical Library, 2nd ed.

Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics*, Chichester: Wiley.

Rivest, L.-P. (1999), "Some Linear Model Techniques for Analyzing Small-Circle Spherical Data," *Canadian Journal of Statistics*, 27, 623–638.

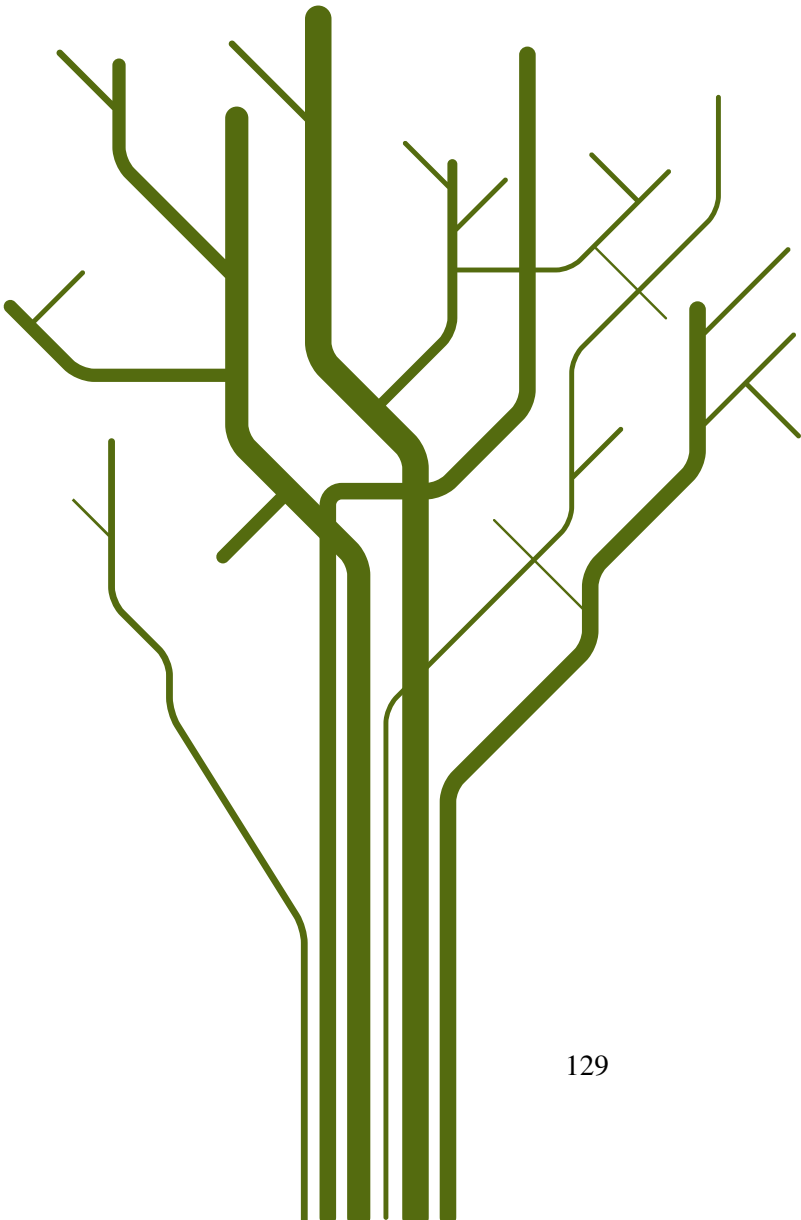
— (2001), "A Directional Model for the Statistical Analysis of Movement in Three Dimensions," *Biometrika*, 88, 779–791.

Paper III

Nonlinear hypothesis testing of geometrical object properties of shapes applied to hippocampi

Schulz, Jörn, Pizer, Stephen M., Marron, J.S. and Godtliebsen, F.

preprint



Nonlinear Hypothesis Testing of Geometrical Object Properties of Shapes Applied to Hippocampi

Jörn Schulz · Stephen M. Pizer · J.S. Marron · Fred Godtliebsen

Abstract This paper presents a novel method to test mean differences of geometrical object properties (GOPs). The method is designed for data whose representations include both Euclidean and non-Euclidean elements. It is based on advanced statistical analysis methods such as backward means on spheres. We develop a suitable permutation test to find global and local morphological differences between two populations based on the GOPs. To demonstrate the sensitivity of the method, an analysis exploring differences between hippocampi of first episode schizophrenics and controls is presented. Each hippocampus is represented by a discrete skeletal representation (s-rep). We investigate important model properties using the statistics of populations. These properties are highlighted by the s-rep model that allows accurate capture of the object interior and boundary while, by design, being suitable for statistical analysis of populations of objects. By supporting non-Euclidean GOPs such as direction vectors, the proposed hypothesis test is novel in the study of morphological shape differences. Suitable difference measures are proposed for each GOP. Both global and local analyses showed statistically significant differences between the first episode schizophrenics and controls.

J. Schulz
Department of Mathematics and Statistics, University of Tromsø, Norway Tel.: +47 45696867
E-mail: jorn.schulz@uit.no

F. Godtliebsen
Department of Mathematics and Statistics, University of Tromsø, Norway, E-mail: fred.godtliebsen@uit.no

Stephen M. Pizer
Department of Computer Science, University of North Carolina at Chapel Hill (UNC), USA, E-mail: smp@cs.unc.edu

J.S. Marron
Department of Statistics & Operations Research, UNC, USA, E-mail: marron@unc.edu

Keywords Hippocampus · Hypothesis test · Permutation test · Principal nested sphere · Schizophrenia · Skeletal representation

1 Introduction

Statistical analysis of anatomical shape differences has been broadly reported in the literature (e.g., [2,5,10, 15]). In medical settings, the study of morphological changes of human organs and body structures is of great interest. An important subfield in medical imaging is to understand neuroanatomical structures of the human brain (e.g., [12,14,42]). Morphological changes of brain structures can provide the physician with information about neuropsychiatric diseases such as Alzheimer's and schizophrenia. A common interest of medical shape analysis is to test for morphological differences between healthy and diseased populations. In addition, the study of drug effects is of high interest in epidemiology. Volumetric measurements often can not distinguish between brain structure differences of two populations [45]. Therefore, sophisticated mathematical shape models with properties that support an accurate statistical analysis are required.

Shape differences can be quantified by hypothesis tests. A statistical hypothesis test requires a null hypothesis H_0 and an alternative hypothesis H_1; a standard null hypothesis assumes no differences between the populations. In this paper, we propose a novel approach for a hypothesis test on geometrical object properties (GOPs) of shapes with application to the hippocampus of the human brain.

Such a hypothesis test of populations of medical shapes depends on 1) the type of medical data, 2) extraction of the object and the following shape representation by a model, 3) selecting object properties for the shape model, 4) statistics necessary to perform population comparison of the models and, 5) a

method for constructing a hypothesis test based on given difference measures.

We use a discrete *skeletal representation*, abbreviated as s-rep [36] as a shape model. The amenities of s-reps relative to other shape representations are described in Section 3. The s-reps are fit to a set of binary images of the hippocampus extracted from a magnetic resonance imaging (MRI) data set. All skeletal shape models have Euclidean as well as non-Euclidean components. Thus, a hypothesis test based on skeletal models must support the two different types of features. The approach presented in this paper allows a sensitive hypothesis test between the components of s-reps. By using this approach, local and global shape differences of the hippocampi between schizophrenia and control populations are investigated.

The hypothesis test requires *i)* fair correspondence between all skeletal models within a population and across populations, *ii)* a method to compute means of populations of skeletal models, *iii)* a test statistic with appropriate distance measures for the Euclidean and non-Euclidean components of the means, *iv)* a method to calculate a test statistic and the empirical distribution of the test statistic, and *v)* a procedure to correct for multiple comparison of local and global testing of GOPs.

The paper is presented as follows. The data set for the schizophrenia study describing two shape populations is presented in Section 2. The skeletal model is discussed in Section 3 in addition to required statistical properties for shape analysis of populations. Section 4 introduces the method composite principal nested great spheres (CPNG), which allows statistical analysis of the Euclidean and non-Euclidean components of skeletal models such as the calculation of means. Section 5 describes the model fitting procedure for the two shape populations, which produces the statistical properties required for each model. A permutation test is introduced in Section 6 and specified for skeletal models together with required statistics. Finally in Section 7, hypothesis test results of the hippocampus study are reported.

2 Schizophrenia study data set

The data consist of MRI assessments of hippocampi from patients with schizophrenia and a similar set from a healthy control group as described in [28,40]. In the original study, 238 first-episode schizophrenics and 56 controls were enrolled. First-episode schizophrenia patients have not received medical treatment prior to the MRI assessment. The hippocampi were segmented from the aligned MRI scans with an automated atlas based segmentation tool developed at the University of North Carolina [16,28].

Statistical analysis must be performed on either the left or right hippocampus as a combined analysis could bias the result. Accordingly, the left hippocampus is evaluated in this paper. Records of the the left hippocampus were not available for 17 patients from the schizophrenia group. Therewith, the data set consists of 221 first-episode schizophrenia cases (**SG**) and 56 control cases (**CG**) and is represented by binary files which reflect the segmented hippocampi. In the data provided, the hippocampi have been normalized in volume but the original volumes were reported as separate scaling features.

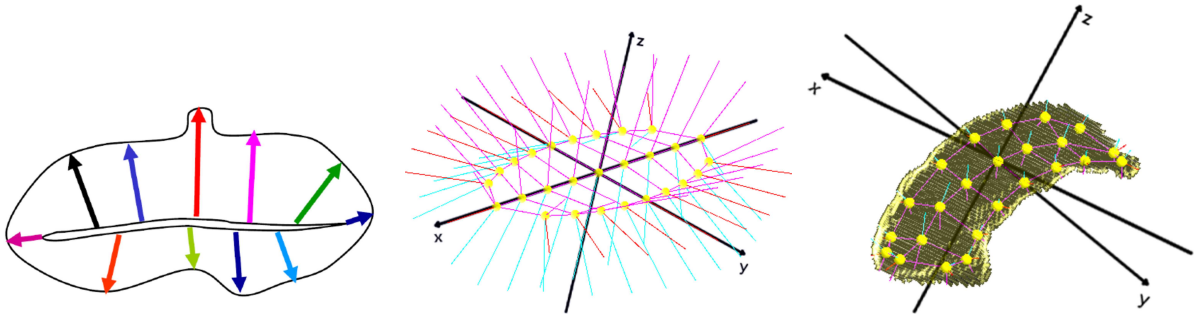
3 Object representation

The representation by a shape model allows calculation of shape statistics of the hippocampus. The type of model, chosen to compare two shape populations, should capture a rich collection of GOPs presented in the data. In addition, small deformations in objects should be reflected by small deformations in the models. Finally, the model should not introduce artificial variation across a population which is not present in the objects themselves.

As discussed by [36], a model that fulfills these three properties is an interior-filling s-rep as depicted in Figure 1. Starting with the continuous case, this model and its properties will be discussed in the following.

The desired GOPs of the model can be categorized into three groups. The first group (G1) should capture locational information of the object boundary. The second group (G2) should reflect the local surface curvature by incorporating directional information into the shape model. The shape model should accurately depict the local orientation of the object. The third group (G3) should describe how the object boundary is connected by the interior in order to reflect the relationship across the interior of the object. The thickness of an object is one property of the interior among others. *Skeletal models* are designed to obtain these geometric properties.

The family of skeletal models has been widely studied in computer vision and medical image analysis. In Section 3 of [41] it is shown that the medial locus [4] of an object $\Omega \subseteq \mathbb{R}^n$ can be described by an inward "grassfire" that starts at the boundary and dies out at a folded version of the medial locus called M_Ω. Given a folded medial locus M_Ω, the medial representation of an object Ω is determined by a set of *spoke* directions from points of M_Ω to the corresponding points of tangency on the boundary $\partial\Omega$. The collection of spoke end points capture locational information of the object boundary as postulated in (G1). The second group (G2) is captured by the directional information



(a) skeletal representation (b) ellipsoid (c) hippocampus

Fig. 1: Continuous skeletal representation and fitted s-reps. (a) Sectional view of a two-sided skeletal 3D object representation. Colored spokes emanate from the skeletal sheet (which is not medial) to the surface. In the continuous form there is a spoke on each point on the skeletal sheet. (b) Discrete s-rep of a non-deformed 3D ellipsoid. (c) Discrete s-rep of a hippocampus.

of the spokes. The points of M_Ω describe the inherent symmetry of an object and therewith (G3) above.

Strictly medial representations are limited by the fact that every protruding boundary kink results in additional medial branches. Thus, two versions of the same object with small noise can have drastically different medial representations. Skeletal models achieve additional stability by relaxing the medial constraint.

Figure 1a visualizes a sectional view of a two-sided skeletal object representation in $\mathbb{R}^3$ composed of a *skeletal sheet* and spokes which emanate from a *skeletal position* on the skeletal sheet to the surface. The skeletal sheet is close to midway but is not medial. An exactly medial representation of this object would require the set M_Ω to include an additional long branch. Elimination of such branches in M_Ω is the goal of the skeletal representation. Figures 1b and 1c will be discussed later.

Stability in the branching structure and stability in the skeletal sheet ensure structural case-by-case stability of the model and thus good correspondence across the samples in the full data set. The branching constraint can be tightened for specific classes of objects where the shape is known. For an ellipsoid-like object shape, such as the hippocampus, the constraint of no branching is reasonable and is adopted. Yet, we want to retain as much as possible the medial properties, such as orthogonal spokes to the boundary, as equal as possible skeletal positions and approximately equal spoke length on both sides of the skeletal sheet. Therefore, the family of skeletal models is restricted by the class of interior-filling s-reps that are modeled as medial as possible [36].

In addition to the case-by-case stability, we require population stability to avoid artificial variance across a population that is solely an artifact of the individual s-rep fittings; such variance is not connected to the objects themselves. Population stability can be achieved by a re-fitting step of the s-rep to the object using an estimated shape probability distribution of the population. The re-fitting step reduces the variance of the s-rep population as described in Section 5. Both case-by-case stability and population stability ensure that the shape models have improved correspondence of both the spokes and the skeletal locations between objects, which will support accurate statistics across a population.

A discrete s-rep, as required for the numerical analysis of slab-shaped objects, consists of a two-sided (folded) sheet of skeletal positions sampled as a grid of atoms, whose skeletal positions are depicted as small spheres in Figures 1b and 1c. On each side of the sheet, there is a spoke, a vector with direction and length on the top and on the bottom connecting the skeletal sheet to the boundary. Also, for each edge grid point there is an additional spoke vector connecting the skeletal sheet folded to the crest of the slab. The sheet is close to midway consistent to the fixed branching constraint between the two sides of the slab, and the spokes are approximately orthogonal to the object boundary. Each discrete s-rep is described by a feature vector

$$\mathbf{s} = (p_1, \ldots, p_{n_a}, r_1, \ldots, r_{n_s}, u_1, \ldots, u_{n_s}) \tag{1}$$

with $n_a = n_a^{ext} + n_a^{int}$ the number of atoms and $n_s = 3n_a^{ext} + 2n_a^{int}$ is the number of radii and spoke directions. A slabular s-rep consists of n_a^{ext} exterior (edge grid points) and n_a^{int} interior atoms. An interior atom consists of a skeletal position $p \in \mathbb{R}^3$, two spoke directions $u \in S^2$ and two spoke lengths $r \in \mathbb{R}_+$ (top, bottom) where $S^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$ is the unit sphere. An exterior atom consists of a skeletal position $p \in \mathbb{R}^3$, three spoke directions $u \in S^2$ and three spoke lengths $r \in \mathbb{R}_+$ (top, crest, bottom). As a result, the shape space of $\mathbf{s} \in \mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s} \times (S^2)^{n_s}$ is a product of Euclidean and non-Euclidean spaces. Each s-rep can

also be described in the space $\mathbb{R}^{n_s+1} \times S^{3n_a-4} \times (S^2)^{n_s}$ together with a scaling factor $\gamma \in \mathbb{R}_+$. This representation is derived from a pre-shape space [22] as discussed in Section 4.

Another popular class of modeling 3D objects is a boundary point distribution model (PDM) where a solid object is defined by the positions of the sampled surface points [8,10,24]. In general, normal directions can be attached on the surface points of a PDM but to the best of our knowledge, it has not been used in practice. In addition, deformation-of-atlas models are well known, wherein the shape changes of an object in images are modeled by the deformations of a template image [34,38]. Such models can capture the local orientation of an object. Nevertheless, both approaches are less suitable for shape statistics of populations by the lack of the interior description of an object.

We restrict our analysis to discrete slabular s-reps which are organized into a (3×8) grid of skeletal positions, i.e., each s-rep consists of 24 atoms. The choice of the grid size defines how exactly the binary images can be described by the s-rep model. We have chosen a grid of 3×8 atoms as a trade-off between capturing important object features, avoiding an overfitting and keeping the dimension of the shape space low. A hippocampus example with bumps that are not tightly described by a (3×8) grid is visualized in Section 1.1 of the Supplementary Material. However, we do not look at individual s-reps that may not be perfectly correct but rather at differences between groups which are not biased versus the other.

4 CPNG analysis

A hypothesis test on mean differences requires a method to calculate means from populations of shape models. The method should incorporate all geometrical components of such models. We have presented in Section 3 an s-rep as a suitable model with Euclidean components and components which live on spheres. This section will discuss an approach to produce means, in addition to shape distributions of populations of s-reps.

First of all, we need to understand the shape space of a discrete s-rep to apply a proper statistical analysis. Each discrete s-rep is described by a feature vector $\mathbf{s} \in \mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s} \times (S^2)^{n_s}$ as defined in (1), and lives in a product of Euclidean and non-Euclidean spaces. Each element of $\mathbf{s}$ corresponds across the population. The points $X_p = (p_1, \ldots, p_{n_a})' \in \mathbb{R}^{3n_a}$ form an $(n_a \times 3)$ matrix and a PDM that can be centered and normalized at the origin by $Z_H = HX_p/\|HX_p\|$ with H a Helmert sub-matrix which removes the origin [10,22]. H is an $((n_a - 1) \times n_a)$ matrix with row $i-1$ defined by the vector

$$(H)_{i-1} = (d_i, \ldots, d_i, -id_i, 0, \ldots, 0)$$

with $d_i = (i(i+1))^{-\frac{1}{2}}, i = 2, \ldots, n_a$ where d_i is repeated i-times. Z_H is called a pre-shape with information of location and scale removed. Therewith, the Cartesian product of $p_i \in \mathbb{R}^3, i = 1, \ldots, n_a$ can be described by the pre-shape Z_H and by a scaling term $\gamma = \|HX_p\|$. The pre-shape Z_H lives on the $(3n_a - 4)$ dimensional unit sphere $S^{3n_a-4} \subset \mathbb{R}^{3n_a-3}$. Each spoke direction $u_i, i = 1, \ldots, n_s$ of $\mathbf{s}$ lives on the unit sphere S^2. The radii $r_i \in \mathbb{R}_+, i = 1, \ldots, n_s$ and scale factor $\gamma \in \mathbb{R}_+$ are log-transformed to the Euclidean space $\mathbb{R}$. Thus, a discrete s-rep $\mathbf{s}$ can be described in the shape space $\mathbb{R}^{n_s+1} \times S^{3n_a-4} \times (S^2)^{n_s}$ composed of several spheres and a real space. Jung et al. [20] and Pizer et al. [36] have proposed a method to analyze a population of s-reps that are living in such an abstract space. This method is called *composite principal nested spheres* (CPNS) and will be discussed in the following.

Suppose we have a population of N s-reps. In order to analyze the covariance structure of such a population, we have to find a common coordinate system. CPNS consists of two main parts. First, the spherical parts are analyzed by *principal nested spheres* (PNS) [18,19], which analyzes data on spheres in decreasing dimension, i.e., using a *backward view*. Therewith, the pre-shape Z_{Hj} and each u_{ij} can be mapped to a Euclidean space with $j = 1, \ldots, N$. Second, the Euclideanized variables are concatenated with the $\log r_i$ and $\log \gamma$ to give a matrix Z_{comp} and an array of scale factors to make all variables commensurate as discussed in detail in [36]. Finally, the structure of the covariance is investigated from the scaled matrix Z_{comp}. PNS is a novel method to estimate the joint probability distribution of data on a d-dimensional sphere S^d by a backward view along the dimensions. The backward view allows dealing with one dimension at a time and thus produces better probability distributions.

In Euclidean space, the forward and backward approaches to *principal component analysis* (PCA) are equivalent, which is not true in general non-Euclidean spaces, such as the d-dimensional unit sphere S^d. Damon and Marron [9] have studied generalizations (e.g., PNS) of PCA across a variety of contexts, and have shown that backwards is generally more amenable to analysis, because it is equivalent to a simple adding of constraints.

PNS is a fully backward approach that fits the best lower dimensional subsphere in each dimension starting with S^d. The subsphere can be great (a sphere with radius 1) or small (less than 1). Figure 2 visualizes the method which takes into account variation along small circles (non-geodesic variations) as well as variation along geodesics. Thus, the decomposition

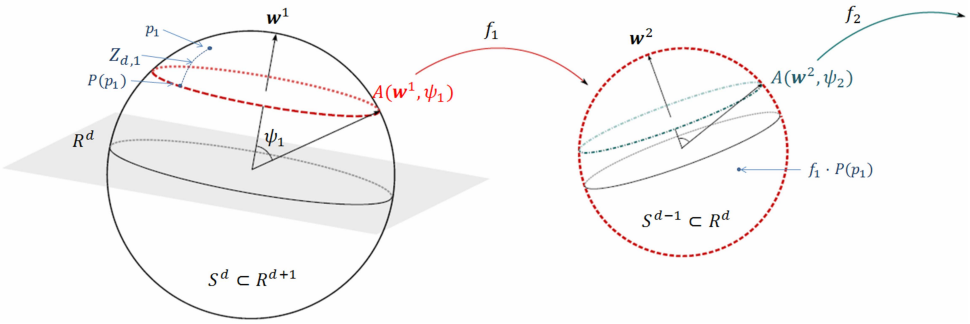


Fig. 2: Backwards PNS by computation of the nearest lower dimensional subsphere. In addition, the projection $P(p_1)$ of a data point p_1 with score $Z_{d,1}$ on the subsphere $A(w^1, psi_1)$ is depicted.

is non-geodesic. Each subsphere $A(w^l, \psi_l)$ of S^{d-l+1} is defined by an axis (polar position) w^l with latitude angle $\psi_l, l = 1, \ldots, d$. The lowest principal nested sphere $A(w^d, \psi_d)$ is a circle. The Fréchet mean [13,21] on the remaining 1-dimensional subsphere can be seen as the best-fitting 0-dimensional subsphere (a point) of the data. Finally, this mean is projected back to S^d resulting in a backward Fréchet mean.

CPNS has been shown to be powerful in the analysis of a single s-rep population. In this study, our hypothesis test of local or global shape differences involves the comparison of two or more CPNS statistics, which is facilitated by stable statistics and correspondence between populations, using a common coordinate system. In fact, PNS is a non-geodesic method which fits small and great subspheres. The fitting of small spheres has advantages in describing the amount of data variation inside a population with fewer principal components but two or more populations can have different decompositions into small and great spheres, which introduce additional variation across the populations, e.g., reflected by a larger variation between CPNS means of several populations. Therefore, all CPNS analyses in this paper are constrained by fitting principal nested great spheres, called PNG and CPNG respectively. Therewith, we avoid additional variation, ensure correspondence and a common coordinate system across several populations. A preliminary simulation study on permuted populations of s-reps confirmed the improvements of CPNG compared to CPNS. Notice, CPNG is identical to [17] in the two dimensional case. We are leaving a commensurate CPNS analysis for populations using small spheres for future work and discuss a possible approach in Section 1.2 of the Supplementary Material.

5 Model fitting

The application of the proposed hypothesis test to the hippocampi study, introduced in Section 2, requires a procedure to generate s-rep fittings with statistical object and population properties as discussed in Section 3. This section will introduce such a procedure.

Assume we have a set of binary images and a set of corresponding signed distance images. The distance images are used during the fitting process as the target data to which reference models are fitted. Following [36], the fitting procedure can be described with 5 consecutive steps: initial alignment, atom stage, spoke stage, CPNG stage and the final spoke stage.

Initial alignment: A reference s-rep is translated, rotated and scaled into the space of distance images by matching of moments to the boundary.

Atom stage: The atom stage defines the geometry of the object and accordingly, the case-by-case stability. Each atom, i.e., each skeletal grid point and its set of spokes are fit one by one with multiple iterations through these atoms. For each atom, an objective function is optimized [36]. The objective function reflects the goodness of the fit and is calculated by a weighted sum of different optimization criteria. The function penalizes factors which are making the s-rep structurally improper, such as irregularity in the grid and crossing of adjacent spokes. In addition, it penalizes the spoke ends deviating from the object boundary and their directions deviating from the boundary normal (both implied by the input distance image). The spokes are further penalized from failing to match the geometry of the crest implied by the distance image. The penalties are summed over spokes which are interpolated from the original s-rep.

Spoke stage: The spoke stage optimizes the spoke lengths to match the object boundary more closely. The skeletal grid points and the resulting geometry of the s-rep will not be changed during this stage. The atom and spoke stage provide appropriate s-rep fittings to the data with case-by-case stability.

CPNG stage: The CPNG stage is designed to provide improved correspondence across a population of s-reps. The fits of the spoke stage are used to calculate CPNG statistics as described in Section 4. Improved correspondence is achieved by restricting the fits to a

shape space which results from the CPNG analysis. CPNG estimates a mean s-rep from the population but also yields eigenmodes and modes of variation [18]. Consequently, any s-rep can be expressed by the score of the eigenmodes in the CPNG space. Hence, correspondence across a population is achieved by initialization of each fitting with the CPNG mean s-rep and by restricting the shape space to the CPNG space. In addition, s-rep candidates from the CPNG space are penalized by the Mahalanobis distance between the candidate and the CPNG mean. As a result improved correspondence is achieved.

Final spoke stage: The final spoke stage adjusts the spokes of the CPNG stage fits to match closely the object boundary. Consequently, s-reps can be generated which are not an element of the CPNG space.

The first three stages form a *preliminary stage* in the fitting procedure. The fitting stages are implemented in a software called Pablo, developed at the University of North Carolina. The program is available at [32].

6 Multiple hypothesis testing

A sensitive hypothesis test is useful for the quantification of shape differences, both to compare populations globally and locally. The introduction of a suitable shape model in Section 3, a method to calculate means from populations in Section 4 and a procedure to generate s-rep fittings in Section 5 provide us with tools to generate models and means that contain the desired properties for a sensitive hypothesis test. An important challenge is that the geometric object elements of each model are spatially correlated. Furthermore, a suitable hypothesis test should correct for multiple comparisons.

6.1 An overview of multiple comparison corrections

The problem of false positives with multiple statistical tests is well recognized. Statistical shape analysis must deal with a large number of hypotheses, each derived from a GOP element, for example of the s-rep. Two common categories of multiple comparison correction are familywise error rate (FWER) and false discovery rate (FDR) [3]. Let V be the number of rejected hypotheses when the null is true (type 1 error), and S the number rejected hypotheses when the null is false. The FWER is defined as the probability of at least one type 1 error by $\mathrm{P}(V \geq 1)$. The FDR is defined as the expected proportion of type 1 errors among the total number of rejected hypotheses by $\mathrm{E}(V/(S+V)$ with $V/(S+V) = 0$ if $(S+V) = 0$. There are several approaches to control FWER and FDR. A commonly used one is the Bonferroni correction. Another approach is using typical wavelet coefficient selection methods [1,6,44]. In addition, variable selection based on threshold random field theory (RFT) have been used [7,23,33]. Permutation tests allow multiple comparison correction by estimating the empirical null-distribution and the covariance structure of the test statistics [30,35,43]. This paper uses multiple comparison correction by FWER.

The Bonferroni correction has several major drawbacks; the Bonferroni threshold can be conservative if the GOPs are dependent of each other. In particular, spatial autocorrelations result in fewer effective variables. Spatial correlation can be expected between neighbor spokes and skeletal positions of an s-rep. In addition, the Bonferroni correction reduces the power of a test as the probability of false negatives increases, because it controls only the probability of false positives. RFT requires strong assumptions such as the same parametric distribution at each spatial location (e.g., multivariate Gaussian), sufficient smoothness as well as stationarity. The assumption of a parametric distribution can not be fulfilled in case of an s-rep model and the assumption of stationarity can also be doubtful.

Permutation tests have advantages over the approaches above that make them particularly suitable for s-reps. S-reps are defined on a product of Euclidean and non-Euclidean spaces with unknown probability distributions of the geometric object elements. In contrast to standard parametric methods such as Bonferroni and RFT, a permutation test is a non-parametric approach using the data to estimate the sampling distribution of the test statistic under the null-hypothesis H_0. Permutation tests are also adaptive to underlying correlation patterns in the data.

A minimal assumption of permutation testing is the exchangeability under H_0 such as identical distributions of populations 1 and 2. The underlying idea of a permutation test is that any permutation of the observations has the same probability to occur under the assumption H_0. Given the permuted populations, a common test statistic measures differences between population means. The test statistic may calculate feature by feature differences or combine features to measure differences between GOPs. The permuted populations can be used to estimate the distribution of the test statistic as well as to estimate the correlation structure.

6.2 A permutation test for s-reps

Suppose we have two populations of s-reps described by a set $\tilde{A}_1 = \{\tilde{\mathbf{s}}_{11}, \ldots, \tilde{\mathbf{s}}_{1N_1}\}$ of N_1 s-reps and a set $\tilde{A}_2 = \{\tilde{\mathbf{s}}_{21}, \ldots, \tilde{\mathbf{s}}_{2N_2}\}$ of N_2 s-reps with $\tilde{\mathbf{s}}_{il}$ as defined in (1). We assume without loss of generality $N_1 \geq N_2$.

The permutation test for populations of s-reps can be divided into four steps.

First, observed and permuted population CPNG means are generated as described in Section 6.2.2. Second, appropriate Euclidean or non-Euclidean GOP differences are calculated between the means of the observed populations $\tilde{A}_1$ and $\tilde{A}_2$, and between the means of corresponding permuted populations as described in Sections 6.2.3 and 6.2.4. Third, p-values are calculated for each GOP difference as described in Section 6.2.5. Each of these p-value is uniformly distributed and mapped by probability integral transformations to standard normal distributed variables. Hence, the GOPs can be mapped from a non-linear to a linear space with the same coordinate system for each GOP. Finally, the covariance matrix of these standard normal distributed variables is estimated, in order to incorporate the true multivariate nature of the data and the correlation between the GOPs as described in Section 6.2.6. As a result, the partial tests of the GOPs can be combined into a single summary statistic by the Mahalanobis distance. In addition, a feature-by-feature test can be constructed as described in Section 6.2.7.

6.2.1 Pre-processing

In a first pre-processing step, global translational and rotational variations should be removed from all s-reps in order to analyze only shape variations. To make the alignment unbiased with respect to the population, the overall backwards CPNG mean $\tilde{\mu}$ is estimated from the set union

$$\tilde{A} = \tilde{A}_1 \bigcup \tilde{A}_2 = \{\tilde{\mathbf{s}}_{11}, \ldots, \tilde{\mathbf{s}}_{1N_1}, \tilde{\mathbf{s}}_{21}, \ldots, \tilde{\mathbf{s}}_{2N_2}\}.$$

The CPNG mean $\tilde{\mu}$ is translationally aligned by the subtraction of the mean of the locational components. In addition, the eigenvectors of the second moments about the center of the skeletal positions yields a rotational alignment to the x, y and z-axis. The translationally and rotationally aligned CPNG mean $\tilde{\mu}$ is called μ. Afterwards, each s-rep $\tilde{\mathbf{s}} \in \tilde{A}$ is translated, rotated and scaled to μ by standard Procrustes alignment (see [10]) based on the hub-positions of each s-rep. For each aligned s-rep $\mathbf{s}$, the scaling factor $\tau \in \mathbb{R}_+$ is kept as a variable. The global translation and rotation information is not considered of interest in the shape analysis of hippocampi. Moreover, we have chosen to use features which can be understood by the user (e.g., physicians). Therefore, the skeletal positions are considered in $\mathbb{R}^{3n_a}$ instead on S^{3n_a-4} as in Section 4. Thus, each aligned s-rep is described by a feature vector $\mathbf{t} = (\tau, \mathbf{s})$, where $\mathbf{t}$ contains $n = 1 + n_a + 2n_s$ features and is an element of the shape space $\mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s+1} \times (S^2)^{n_s}$. Set $A_1 = \{\mathbf{t}_{11}, \ldots, \mathbf{t}_{1N_1}\}$, $A_2 = \{\mathbf{t}_{21}, \ldots, \mathbf{t}_{2N_2}\}$ and $A = A_1 \bigcup A_2$.

6.2.2 Generation of observed and permuted sample means

First, a method to calculate means for the observed and permuted samples of the two populations is required in order to create a hypothesis test of mean differences.

Observed sample means. For each set A_i, $i = 1, 2$ the observed sample mean is $\hat{\mu}_i = (\bar{\tau}_i, \bar{\mu}_i) \in \mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s+1} \times (S^2)^{n_s}$. The component $\bar{\mu}_i$ is a CPNG backwards mean as described in Section 4. The mean scaling factor $\bar{\tau}_i \in \mathbb{R}_+$ is computed as a geometric mean (which is natural for scaling factors) by

$$\bar{\tau}_i = \exp\left(\frac{1}{N_i}\sum_{j=1}^{N_i} \log(\tau_{ij})\right), \qquad i = 1, 2. \tag{2}$$

In fact, the CPNG backwards mean $\bar{\mu}_i$ consists of n_s+1 PNG backwards means, one for the skeletal position and n_s for the spoke directions, and n_s means for the spoke lengths respective to (2).

Permuted sample means. The number of all possible permutations of the index set $I = \{1, \ldots, N_1+N_2\}$ is

$$\binom{N_1 + N_2}{N_1} = \frac{(N_1 + N_2)!}{N_1!N_2!}.$$

Random sample sets $I_l, l = 1, \ldots, P$ of $P = 30,000$ permutations of the index set I were generated, a number comparable to the suggested number in [11] and [26]. Larger numbers of permutations increase the accuracy of the p-values but require more computation time. The permutation group $A_{1l} \subset A$ contains all s-reps indexed by the first N_1 indices of I_l. The group $A_{2l} = A \setminus A_{1l}$ contains the remaining N_2 s-reps. For each permutation I_l the means $\hat{\nu}_{il}$, $i = 1, 2$ are estimated by $\hat{\nu}_{1l} = (\bar{\kappa}_{1l}, \bar{\nu}_{1l})$ for the group A_1, $\hat{\nu}_{2l} = (\bar{\kappa}_{2l}, \bar{\nu}_{2l})$ for the group A_2. $\bar{\nu}_{il}$ is estimated by the CPNG backwards mean and $\bar{\kappa}_{il}$ is the mean scaling factor of the corresponding permutation respective to (2).

6.2.3 Test statistics

Equality of distributions between populations A_1 and A_2 can be tested by a nonparametric combination of a finite number of dependent *partial tests* as proposed in Pesarin [35]. The global null hypothesis is given by $H_0 : \{A_1 \stackrel{d}{=} A_2\}$, where $\stackrel{d}{=}$ denotes the equality in distribution. Let H_1 be the global alternative hypothesis. In general, the test requires the definition of a *statistic* T in testing a null hypothesis. A natural test statistic is

$$T(A_1, A_2) = d(\hat{\mu}_1, \hat{\mu}_2), \tag{3}$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the observed sample means as defined in Section 6.2.2 and d is a difference measure on the nonlinear manifold describing the GOPs. The test statistic T consists of K different partial tests depending on the difference measure. Thus, the global null hypothesis can be written in terms of K subhypotheses $H_0 : \{\bigcap_{k=1}^{K} H_{0k}\}$ and the alternative as $H_1 : \{\bigcup_{k=1}^{K} H_{1k}\}$. Usually, the dependence relation among partial tests are unknown even though they are functions of the same data. Pesarin [35] has shown that a suitable combining function (described in Section 6.2.6) will produce an unbiased test for the global hypothesis H_0 against H_1 if all partial tests are assumed to be marginally unbiased, consistent and significant for large values. The partial tests $T_k, k = 1, \ldots, K$ are defined by the *partial difference measures*. Therewith, a hypothesis test for identical statistical distribution of two s-rep populations is given by mean differences,

$$H_0 : \{\mu_1 = \mu_2\} \text{ versus } H_1 \ : \ \{\mu_1 > \mu_2\} \tag{4}$$

for a one-sided test in case the difference measures are unsigned and

$$H_0 : \{\mu_1 = \mu_2\} \text{ versus } H_1 \ : \ \{\mu_1 \neq \mu_2\} \tag{5}$$

for a two-sided test in case of signed differences.

The hypothesis H_0 will be rejected if the probability of observing $T(A_1, A_2)$ under H_0 from the empirical distribution is smaller than a chosen significance level α; otherwise we do not reject. The significance level describes the probability of type 1 error, i.e., H_0 is wrongly rejected. Alternatively, the type 2 error occurs when H_0 is not rejected but it is in fact false.

6.2.4 Difference measures

This section defines a signed difference measure d^2 for the test statistic (3). An alternative unsigned difference measure d^1 is defined in Section 1.3 of the Supplementary Material. Suppose we have two s-reps

$$\mathbf{t}_i = (\tau_i, p_{i1}, \ldots, p_{in_a}, r_{i1}, \ldots, r_{in_s}, u_{i1}, \ldots, u_{in_s})',$$

$i = 1, 2$ with the skeletal positions $p_{ij}, \in \mathbb{R}^3$ and the scale factors $\log(\tau_i), \log(r_{ij}) \in \mathbb{R}$ as Euclidean GOPs and the spoke directions $u_{ij} \in S^2$ as non-Euclidean GOPs. Thus, a suitable difference measure is required as defined in the following.

The measure d^2 is a vector of differences

$$\begin{aligned} d^2(\mathbf{t}_1, \mathbf{t}_2) := (&d_1(\tau_1, \tau_2), \\ &d_2(p_{11}, p_{21}), \ldots, d_2(p_{1n_a}, p_{2n_a}), \\ &d_3(r_{11}, r_{21}), \ldots, d_3(r_{1n_s}, r_{2n_s}), \\ &d_4(u_{11}, u_{21}), \ldots, d_4(u_{1n_s}, u_{2n_s}))' \end{aligned} \tag{6}$$

with appropriate partial difference measures: d_1 for the scaling factors τ_i, d_2 for the positions p_{ik}, d_3 for the spoke lengths r_{ij} and d_4 for the spoke directions u_{ij} with $i = 1, 2$, $k = 1, \ldots, n_a$ and $j = 1, \ldots, n_s$ by

$$\begin{aligned} d_1(\tau_1, \tau_2) &= \log(\tau_2) - \log(\tau_1), \\ d_2(p_{1k}, p_{2k}) &= p_{2k} - p_{1k}, \\ d_3(r_{1j}, r_{2j}) &= \log(r_{2j}) - \log(r_{1j}), \\ d_4(u_{1j}, u_{2j}) &= d_{gs}(u_{1j}, u_{2j}). \end{aligned}$$

The partial difference measure d_{gs} is defined by longitude and latitude differences of the spoke directions (u_{1j}, u_{2j}) using a normalization by the shift of the geodesic mean as explained in the next paragraph. The components of

$$\begin{aligned} d^2 : &(\mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s+1} \times (S^2)^{n_s}) \times \\ &(\mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s+1} \times (S^2)^{n_s}) \longrightarrow \mathbb{R}^{3n_a+3n_s+1} \end{aligned}$$

are not metrics because they can take on negative values.

Shift by the geodesic mean. The spoke directions $(u_{1j}, u_{2j}) \in S^2 \times S^2$ can be mapped by spherical parametrization to latitudes ϕ_{1j}, ϕ_{2j} and longitudes θ_{1j}, θ_{2j} in the base coordinate system of all aligned hippocampi. The spherical mapping can be defined by

$$\begin{aligned} \phi_{ij}(u_{ij}) &= \text{atan2}(u_{ij3}, \sqrt{u_{ij1}^2 + u_{ij2}^2}), \\ \theta_{ij}(u_{ij}) &= \text{atan2}(u_{ij2}^2, u_{ij1}^2), \end{aligned}$$

with $\phi_{ij} \in [-\pi/2, \pi/2]$ and $\theta_{ij} \in (-\pi, \pi]$; the two-argument function $\text{atan2}(x_2, x_1) \in (-\pi, \pi]$ is the signed angle between two vectors $e_1 = (1, 0)'$ and $(x_1, x_2)' \in \mathbb{R}^2$. The longitude ϕ is measured from the x-y plane.

The spherical mapping is not uniquely defined in general. Furthermore, it does not establish an appropriate correspondence. Two points close to the equator with identical geodesic distance as two points close to the north pole have different latitude and longitude differences, and are therefore not commensurate. For that reason, longitude and latitude pair differences will be normalized by shifting the geodesic mean of (u_{1j}, u_{2j}) along its meridian to the equator by a rotation about an axis $c \in S^2$ with rotation angle $\psi \in [0, \pi/2)$. Then, the directions (u_{1j}, u_{2j}) are rotated along small circles on the sphere about the same axis c with the same rotation angle ψ towards the equator.

In more detail, consider a pair (u_{1j}, u_{2j}) of spoke directions on S^2 with northpole $N_p = (0, 0, 1)'$. At first, find its geodesic mean by

$$\mu_g(u_{1j}, u_{2j}) = \frac{u_{1j} + u_{2j}}{\|u_{1j} + u_{2j}\|}.$$

We assume $\text{acos}(|\mu_g' N_p|)) > 1e - 3$; otherwise choose a different northpole. Given a rotation matrix $R_1 :=$

$R_1(c,\psi)$, the rotation of μ_g along its meridian to the equator is $\tilde{\mu}_g = R_1\mu_g$ with

$$R_1(c,\psi) = I_3 + \sin\psi[c]_\times + (1-\cos\psi)(cc' - I_3), \quad (7)$$

where I_3 is the three-dimensional unit matrix and $[c]_\times$ is the cross product matrix satisfying $[c]_\times v = c \times v$ for any $v \in \mathbb{R}^3$. To avoid discontinuity problems between $-\pi$ and π for θ_{ij}, let R_2 be the rotation matrix as defined by (7) that rotates $\tilde{\mu}_g$ towards $(1,0,0)'$, i.e., $R_2\tilde{\mu}_g = (1,0,0)'$. Now, shift each pair (u_{1j}, u_{2j}) by applying $\tilde{u}_{1j} = R_2R_1u_{1j}$ and $\tilde{u}_{2j} = R_2R_1u_{2j}$. Finally, we calculate the latitudes $\phi_{1j}(\tilde{u}_{1j}), \phi_{2j}(\tilde{u}_{2j})$ and longitudes $\theta_{1j}(\tilde{u}_{1j}), \theta_{2j}(\tilde{u}_{2j})$ and define the differences of the transported spoke directions by the delta latitude $\Delta\phi_j = \phi_{2j} - \phi_{1j}$ and delta longitude $\Delta\theta_j = \theta_{2j} - \theta_{1j}$. Therewith, the difference measure d_{gs} is defined by

$$d_{gs}(u_{1j}, u_{2j}) := (\Delta\phi_j, \Delta\theta_j).$$

6.2.5 Mapping of GOP differences to standard normally distributed variables

Suppose we have the test statistic $T_0 := T(A_1, A_2)$ of the underlying observed sample. The idea is to estimate the sampling distribution of the statistic T_0 from test statistics of the permuted samples

$$T_l := T(A_{1l}, A_{2l}), \quad l = 1, \ldots, P.$$

The test statistic measures the GOP differences in different units. The vector $T_l = (T_{l1}, \ldots, T_{lK})$ contains K partial tests, where K is the number of components of the difference measure d^2. The elements of the vector T_l are not commensurate as required for the estimation of the covariance structure. Thus, the GOP differences must be normalized and mapped to a common coordinate system in a way that preserves the multivariate dependence structure between the GOPs. The procedure is explained in the following and depicted in Figure 3 on the basis of a selected GOP using distance measure d^2. The figure is discussed further in the text.

Calculating p-values for GOP differences. After the calculation of T_l, we estimate for each GOP difference $k = 1, \ldots, K$ the empirical cumulative distribution function (CDF) by

$$C_k(T_{lk}) = \frac{1}{P}\sum_{l'=1}^{P} I(T_{l'k}, T_{lk})$$

$$\text{with } I(T_{l'k}, T_{lk}) = \begin{cases} 1 & : \quad T_{l'k} \le T_{lk}, \\ 0 & : \quad \text{otherwise.} \end{cases}$$

Respectively, we can calculate $C_k(T_{0k})$.

Mapping of p-values to $\mathcal{N}(0,1)$. By construction the p-values have a uniform distribution. Thus, the GOP differences can be represented as

$$U_{lk} = \Phi^{-1}\left(\tilde{C}_k(T_{lk})\right), \quad (8)$$

where Φ^{-1} is the inverse standard Gaussian CDF,

$$\tilde{C}_k(T_{lk}) = \frac{sc-2}{sc}C_k(T_{lk}) + \frac{1}{sc}$$

and $k = 1, \ldots, K$, $l = 1, \ldots, P$. The inverse standard Gaussian CDF requires values greater than 0 or less than 1; otherwise $U_{lk} = \pm\infty$. Therefore, all p-values are scaled by $\tilde{C}_k(T_{lk})$ with $sc = 10000$. Simulations have shown numerical instabilities for larger values of sc. The marginal distribution of U_{lk} is standard Gaussian for every k, i.e, $U_{lk} \sim \mathcal{N}(0,1)$.

Using the estimated inverse empirical CDF C_k, the observed GOP differences T_{0k} are mapped to U_{0k}, respectively.

6.2.6 Global test with multivariate comparisons correction

Given $U_{lk} \sim \mathcal{N}(0,1)$, the $K \times K$ covariance matrix Σ_U of the $P \times K$ matrix $U = (U_1, \ldots, U_P)'$ with $U_l = (U_{l1}, \ldots, U_{lK})$, $l = 1, \ldots, P$ is estimated by

$$\hat{\Sigma}_U = \frac{1}{K-1}U'U.$$

A corrected test statistic is then given by the Mahalanobis distances

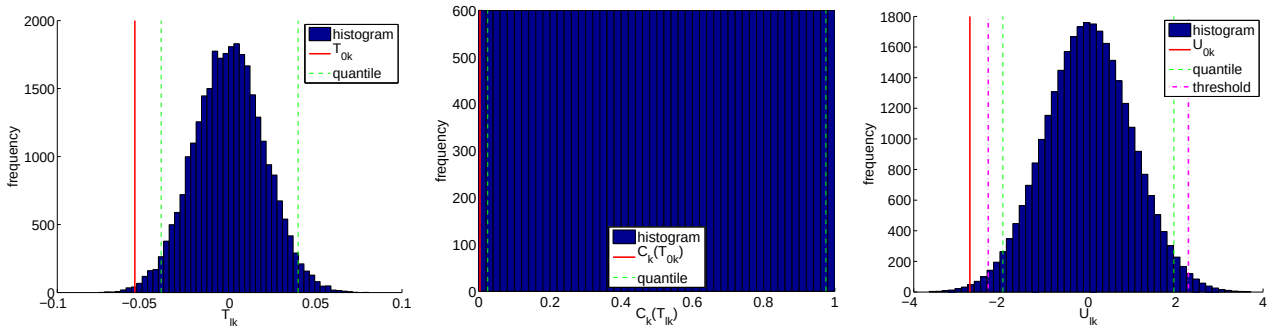
$$M_0 = U_0'\hat{\Sigma}_U^{-1}U_0, \quad M_l = U_l'\hat{\Sigma}_U^{-1}U_l, \quad l = 1, \ldots, P,$$

which defines a suitable combining function [35, Section 6.2.4] that includes the GOP correlation structure. The sampling distribution of the final test statistic under the null-hypothesis H_0 can be estimated from M_l by an empirical CDF. The probability of observing M_0 under H_0 from the empirical null-distribution is given by

$$p(M_0) = \frac{1}{P}\sum_{l=1}^{P} H(M_l, M_0), \quad (9)$$

$$\text{with } H(M_l, M_0) = \begin{cases} 1 & : \quad M_l \ge M_0, \\ 0 & : \quad M_l < M_0. \end{cases}$$

Equation (9) defines the p-value of the final global test by rejecting H_0 if $p(M_0) < \alpha$.



(a) GOP difference (b) p-value (c) std with corrected threshold

Fig. 3: Mapping of GOP differences to standard normally distributed variables using the example of a skeletal z-position ($k = 3$) with distance measure d^2. The 2.5% and 97.5% quantiles are visualized in each plot. (a) GOP differences T_{lk} of the permuted samples with $l = 1, \ldots, 30,000$ and T_{0k} of the underlying observed sample. (b) Calculated p-values $C_k(T_{lk})$ and $C_k(T_{0k})$ using the empirical cumulative distribution function. (c) Standard normal distributed variables U_{lk} and U_{0k}. The dotted-dashed line depict the corrected threshold λ for the GOP as described in Section 6.2.7.

6.2.7 Feature-by-feature test with multivariate comparisons correction

The global shape analysis in the previous section can not indicate local shape differences which motivates the introduction of an FWER threshold correction for a feature-by-feature test. The permutation test approach on each variable T_{lk} yields an empirical distribution C_k, dependent standard Gaussian variables U_{lk} and the empirical covariance matrix $\hat{\Sigma}_U$. As a result, $U_l = (U_{l1}, \ldots, U_{lK})$ is approximately distributed as $\mathcal{N}_K(0, \hat{\Sigma}_U)$, where $\mathcal{N}_K$ is a multivariate Gaussian distribution with mean 0, covariance $\hat{\Sigma}_U$ and density function ψ such that each marginal is $U_{lk} \sim \mathcal{N}(0, 1)$.

Because each random variable U_{lk} is standard Gaussian, the threshold for each standard Gaussian variable should be the same. Thus, given a significance level α, we wish to find the threshold λ such that

$$P(U_{l1} < \lambda, \ldots, U_{lK} < \lambda) = 1 - \frac{\alpha}{2}.$$

The function P is a multiple integral from $-\infty$ to λ in each variable of $U_l \sim \mathcal{N}_K(0, \hat{\Sigma}_U)$ and can be understood as a function $g(\lambda)$ of the single variable λ. The function $g(\lambda)$ is monotonic increasing with asymptotes at 0 and 1. The numerical calculation of the p-values is based on the approximation over an appropriate interval of λ. Recall that $\lambda \geq \lambda_{corr}$ with

$$\lambda_{corr} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

is the threshold for a single standard Gaussian variable. Let $l \in \{1, \ldots, P\}$ be fixed, the threshold λ_{corr} is applicable if all $U_{\cdot k}$ are perfectly correlated. Furthermore, we know that $\lambda \leq \lambda_{indep}$ with

$$\lambda_{indep} = \Phi^{-1}\left(\left(1 - \frac{\alpha}{2}\right)^{1/K}\right)$$

because the threshold λ_{indep} is applicable if all $U_{\cdot k}$ are independent. The desired level $1 - \alpha/2$ will be rather near 1. Thus, the function $g(\lambda)$ will be concave downward in the interval $[\lambda_{corr}, \lambda_{indep}]$.

The values $g(\lambda_{corr})$ and $g(\lambda_{indep})$ can be estimated from a large number N_{Samp} of random samples $Y_n \sim \mathcal{N}_K(0, \hat{\Sigma}_U)$ with $n = 1, \ldots, N_{Samp}$ by

$$\hat{g}(\lambda) = \frac{\sum_{n=1}^{N_{Samp}} I_\lambda(\psi(y_{n1}, \ldots, y_{nK}))}{\sum_{n=1}^{N_{Samp}} \psi(y_{n1}, \ldots, y_{nK})}$$

$$\text{with } I_\lambda(\psi(y_n)) = \begin{cases} \psi(y_n) & : \quad \psi(y_n) < \lambda, \\ 0 & : \quad \text{otherwise}, \end{cases}$$

and $y_n = (y_{n1}, \ldots, y_{nK})$. We have chosen a number of $N_{Samp} = 200,000$ samples.

The computation of $g(\lambda_{indep})$ requires the comparison of y_n values only for those identified as not in the accepted subset for the smaller value λ_{corr} and adding into the accumulated sum for the newly accepted samples. Finally, the standard regula falsi method can be used to iteratively solve the equation $g(\lambda) = 1 - \alpha/2$ with initial evaluations $g(\lambda_{corr})$ and $g(\lambda_{indep})$.

The dashed-dotted line in Figure 3c shows the corrected threshold λ for a selected GOP.

7 Results

7.1 Fitting of s-reps to hippocampi

The hippocampus data set consists of binary images of 221 first-episode schizophrenia cases and 56 control cases as described in Section 2. Antialiased distance images were generated from the binary images according to [31]. Based on the distance images, appropriate preliminary fits by an initial alignment and

an atom and spoke stage are produced as described in Section 5. This preliminary stage is described in detail in Section 1.4 of the Supplementary Material. In order to control the manual work during the preliminary stage, we considered only the first 96 of 221 cases of SG as discussed in the Supplementary Material. Let $\tilde{A}_1$ be the set of 96 preliminary fits for SG and $\tilde{A}_2$ be the set of 56 preliminary fits for CG. All preliminary fits were translated and rotated to the CPNG mean of the set union $\tilde{A}_1 \cup \tilde{A}_2$ by standard Procrustes alignment [10] in order to remove global variation from the preliminary fits. Let $\bar{A}_1$ be the set of 96 aligned SG preliminary fits and $\bar{A}_2$ the set of 56 aligned CG preliminary fits. Finally, CPNG statistics were calculated for the s-rep populations as described in Sections 4 and 5.

A challenging question is the appropriate estimation of the shape distributions of both populations (SG and CG) during the CPNG stage. An option, is to calculate the CPNG statistic of each population ($\bar{A}_1$ and $\bar{A}_2$) resulting in two means and shape distributions. Another option, is to calculate the CPNG statistic of the pooled population ($\bar{A}_1 \cup \bar{A}_2$) resulting in a single mean and shape distribution. The use of two individual shape distributions result in independent fittings between the two populations. On the other hand, the fittings should not be biased and have good correspondence between the populations, which is provided by a pooled shape distribution. A pooled CPNG statistic also removes possible bias from the manual adjustments during the preliminary stage.

The final fitting results obtained from two separate shape distributions showed extraordinary high separation properties and indicated a large bias. Thus, the main focus was the analysis of fittings using a pooled CPNG statistic from $\bar{A}_1 \cup \bar{A}_2$. In addition, we have generated a second group of final fittings derived from CPNG stages using a pooled shape distribution, two individual shape distributions and two individual interchanged shape distributions. The second group is a compromise between independence and a small bias, and is discussed in Section 3 of the Supplementary Material.

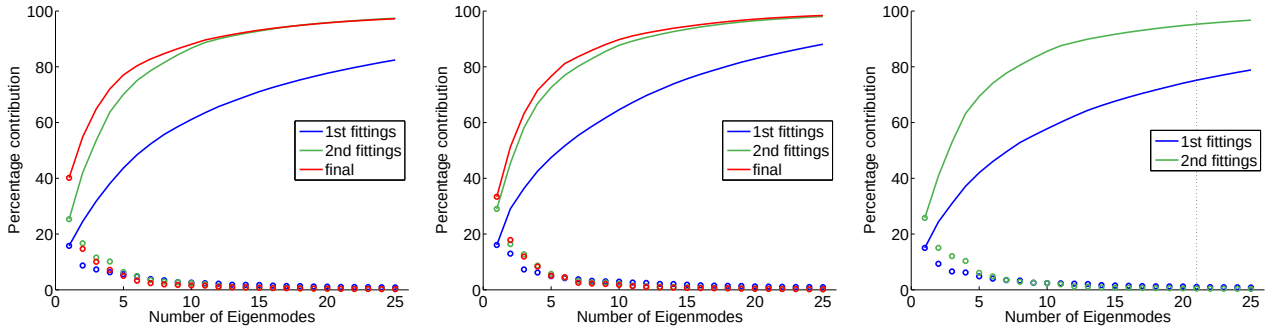
Each CPNG statistic contains a backward mean, the eigenmodes and the corresponding CPNG scores. Figure 4 shows the explained amount of variation by the first 25 eigenmodes for the aligned preliminary fittings after atom and spokes stages (1st fittings), i.e., for $\bar{A}_1$ (subset of SG), $\bar{A}_2$ (CG) and $\bar{A}_1 \cup \bar{A}_2$. The number of eigenmodes was selected to describe more than 75% of the total cumulative variance. This number compromises on capturing enough shape variation while limiting the shape space in order to avoid overfitting. Accordingly, the first 21 eigenmodes of the pooled shape distribution were selected for the CPNG stage describing 75.2% total variance. 18 eigenmodes are required to describe 75.3% of the total cumulative variance of $\bar{A}_1$, and 15 eigenmodes to describe 75.7% for $\bar{A}_2$.

In the CPNG stage, the obtained backward mean of $\bar{A}_1 \cup \bar{A}_2$ was translationally and rotationally aligned to the 221 SG cases and the 56 CG cases. An additional scaling of the means would bias the CPNG statistic because the principal components already contain size information. Afterwards, the aligned means were optimized inside the CPNG shape space and under the penalty of a Mahalanobis distance match term. A high penalty term leads to better correspondence between cases but to less accurate fits. An appropriate penalty term was chosen by a simulation study, the report of which is omitted. At the end, the final spoke stage was performed to ensure that the spoke directions match the boundary.

Figure 4 also shows CPNG analyses for the obtained 2nd fittings of the corresponding cases to $\bar{A}_1$, $\bar{A}_2$ and $\bar{A}_1 \cup \bar{A}_2$ using a pooled shape space during the CPNG stage. The respective numbers of eigenmodes explain an increased amount of variation compared to the first fittings as a result of improved correspondence across the populations. Now, 18 eigenmodes describe 94.9% of the total variance for the subset of SG, 15 eigenmodes describe 93.5% for CG and 21 eigenmodes describe 95.3% for the pooled group.

The final fittings were re-scaled into a world coordinate system (units of mm) with the stored scaling factor from the normalization step described in Section 2. We denote the re-scaled sets of final fittings by the set A_1 of 221 s-reps for SG and the set A_2 of 56 s-reps for CG. The total cumulative variance of the CPNG analysis of A_1 (SG) and A_2 (CG) is depicted in Figure 4 (final). Now, 18 eigenmodes describe 94.8% of the total variance of SG and 15 eigenmodes describe 94.5% for CG. More than 75% of the total cumulative variance of CPNG shape space is now described by using only 5 eigenmodes compared to 18 and 15 as shown previously.

The average volume in mm^3 (and standard deviation) of the final fittings was $3,036$ (343) for SG and $3,137$ (295) for CG. The observed hippocampal volume reduction for schizophrenia patients is consistent with previous studies (e.g., [25]). The average volume overlap between fittings and binary images was 94% for SG and CG (depicted in Section 3 of the Supplementary Material) which is fairly accurate. The percent-volume overlap was measured by the Dice coefficient as defined in the Supplementary Material. The variance of the Dice coefficient is small for both groups. Nevertheless, a larger variance inside SG is observed. Schizophrenia is a heterogeneous disease and also contains hippocampi variations between healthy patients.



(a) SG (b) CG (c) joint group

Fig. 4: CPNG analysis of s-reps before the CPNG stage (1st fittings), after CPNG and final spoke stage (2nd fittings) and after scaling into the world coordinate system (final). The variance contribution of the first 25 eigenmodes are depicted together with the cumulative variance for the CPNG analysis of the (a) SG group, (b) CG group and (c) joint group. The set of 1st and 2nd fittings consist of 96 s-reps for SG, 56 s-reps for CG and 152 s-reps for the pooled group. The set of final fittings consist of 221 s-reps for SG and 56 s-reps for CG. The dotted vertical line in (c) depicts the number 21 of used eigenmodes for the description of the shape space during the CPNG stage.

In Figure 5, the distributions of the SG and CG fittings are visualized by the projections of the scaled CPNG scores matrix Z_{Comp} (see Section 4) onto the distance weighted discrimination (DWD) direction. DWD is a discrimination method which avoids the data piling problems of support vector machine [27, 37]. The projected distributions of SG and CG fittings for the pooled class are estimated by kernel density estimates (KDEs). The different areas under the CG and SG curves are due to unbalanced population sizes (56 for CG compared to 221 for SG). A difference between the populations is visible but not very strong. Thus, it is an interesting question whether the proposed hypothesis test in Section 6.2 will be able to find significant differences between SG and CG for both fittings classes.

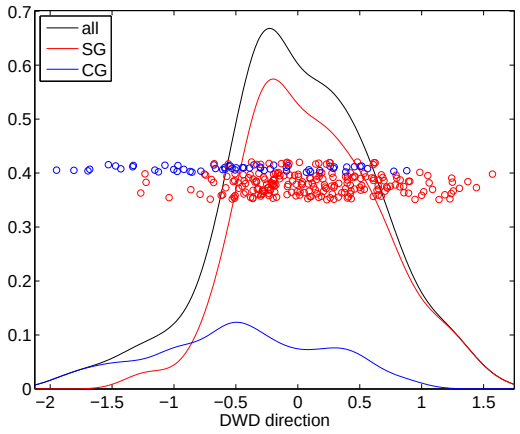


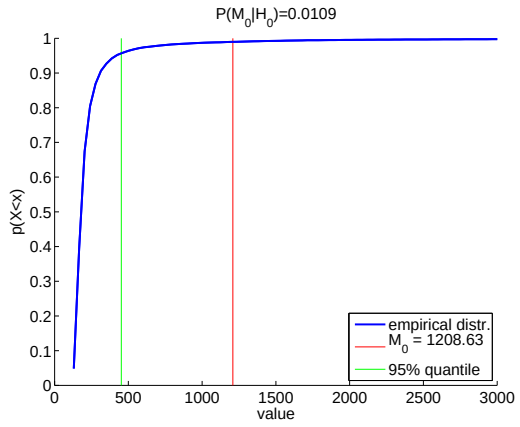
Fig. 5: Jitterplot and KDEs showing the distributions of final SG and CG fittings projected onto the DWD direction. Additionally, the KDE of the pooled distribution of SG and CG is shown (all). A difference between the populations is visible but not very strong.

7.2 Global test results

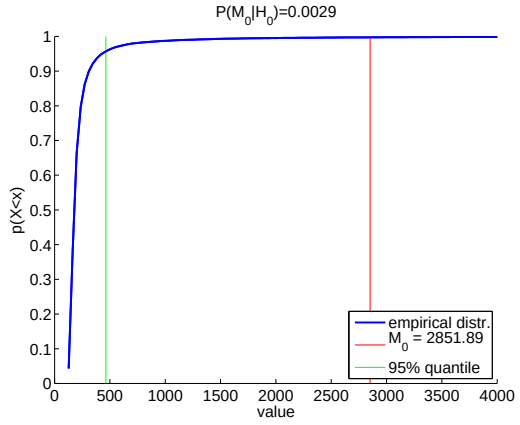
The obtained final fittings were used to test the hypothesis (5) by the proposed procedure in Section 6.2 with a significance level of $\alpha = 0.05$. An alternative pre-processing step (called PP2) is applied in addition to the pre-processing described in Section 6.2.1 (called PP1 in the following). PP2 translates and rotates each s-rep $\tilde{\mathbf{s}} \in \tilde{A}$ to an overall CPNG backward mean μ without scaling. Thus, each aligned s-rep is described by a feature vector $\mathbf{t} = \mathbf{s}$. The global scaling information was previously described by the feature τ in PP1. In contrast, this is captured by the skeletal positions and spoke length using PP2.

Figure 6 shows the global test results for the difference measure d^2 using PP1 and PP2. The global hypothesis of equal sample means is rejected and a statistically significant difference between the shape distribution of SG and CG is established ($p = 0.0109$ for PP1 and $p = 0.0029$ for PP2 with $p = P(M_0|H_0)$). The smaller p-value for PP2 seems to be due to the volume information being spread into the 24 skeletal positions instead of into a single feature. The feature-by-feature test will highlight this fact in the next section. Intermediate results of the proposed hypothesis test procedure are shown in Figure 3 on the basis of a selected GOP. Further visualizations of the procedure can be found in the Supplementary Material.

A detailed power and simulation study is beyond the scope of this paper and left for future work. However, the power of the proposed hypothesis test is



(a) PP1



(b) PP2

Fig. 6: Global test results using PP1 in (a) and PP2 in (b). The empirical distribution of $M_l, l = 1, \ldots, 30,000$ is shown together with M_0 and the 95% quantile of the empirical distribution.

demonstrated on the basis of a real data example. Furthermore, the results are compared with a direction projection permutation (DiProPerm) based mean hypothesis test [46]. The DiProPerm test is based on the evaluation of the scaled CPNG scores matrix Z_{Comp} (see Section 4). The CPNG scores matrix is calculated for SG and CG using both pre-processing methods. Thus, the DiProPerm test is calculated in Euclidean space using standard Euclidean statistics in contrast to the proposed hypothesis test, which is performed in the non-Euclidean s-rep space using the CPNG backward means. An interesting open problem is to extend a method such as DiProPerm in an intrinsic way: in other words to perform DiProPerm using Manifold geodesic distances.

Table 1 summarizes all global test results. We used 30,000 permutations in all settings to be consistent with Section 6.2.2. The DiProPerm test does not require such a high number of permutations in contrast to the proposed global test. Simulations, reported in the Supplementary Material, reveals that a large permutation size is needed to obtain stable results because of the Mahalanobis distance. DiProPerm was carried out using a mean difference (MD) test statistic as recommended in [46]. The DWD-DiProPerm performance was comparable to the Mahalanobis distance results. The support vector machine (SVM) results of DiProPerm were less powerful, probably due to data pilling effects. All results are statistically significant at the level of $\alpha = 0.05$.

Table 1: Empirical p-value results using difference measure d^2 for the proposed global hypothesis test in comparison with results obtained by DiProPerm. Two different pre-processing steps were applied: (PP1) Full Procrustes alignment with scaling. (PP2) Full Procrustes alignment without scaling. Three different projection directions were used for DiProPerm.

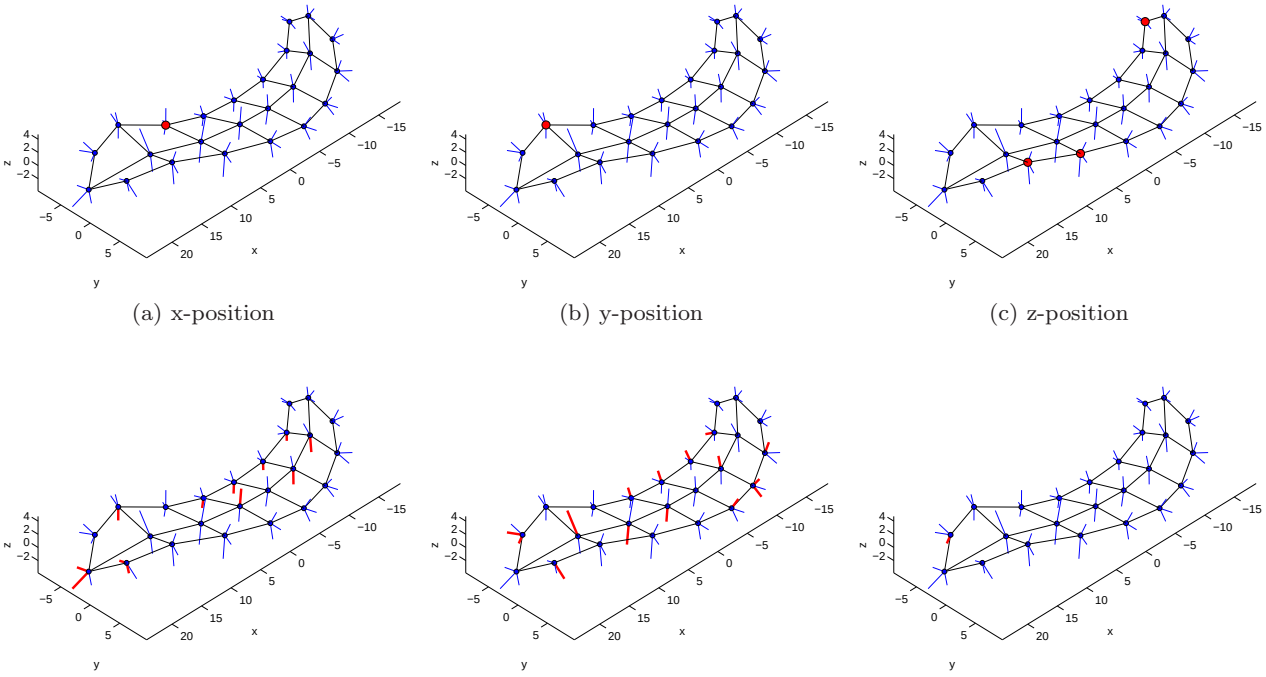
method	empirical p-value PP1	PP2
Mahalanobis distance		
difference measure d^2	0.0109	0.0029
DiProPerm using MD-statistic		
DWD direction vector	0.0074	0.0038
SVM direction vector	0.0119	0.0136

7.3 Single GOP test results

The global shape analysis of hippocampi in the previous section can not indicate local shape differences. Interesting structural changes of the surface are often reflected by a few GOPs, e.g., the local bending of an area. Therefore, the proposed threshold correction for a feature-by-feature test in Section 6.2.7 is useful. Such a feature-by-feature test is not available from DiProPerm.

As our feature-by-feature test approach is novel for nonlinear hypotheses, there is no competing method to compare with. However, a method to evaluate the test is needed. The performance of the feature-by-feature test was evaluated using Receiver Operating Characteristic (ROC) curves. Selected examples of this analysis are reported in Section 2.4 of the Supplementary Material. For each permutation, an ROC curve was generated from the cumulative histograms of the two permuted samples which results in an envelope under the null distribution. In addition, an ROC curve between the two true observed samples was obtained. A significant feature is indicated if the ROC curve of the observed data is close to the boundary or outside the envelope, otherwise not. A comparison of the hypothesis test results to this reveals the high quality of the proposed method.

Figures 7 and 8 visualize the feature-by-feature test results for PP1. Test results are shown on the basis of the skeletal grid given by the CPNG back-



(a) x-position (b) y-position (c) z-position

(d) latitude direction (e) longitude direction (f) spoke length

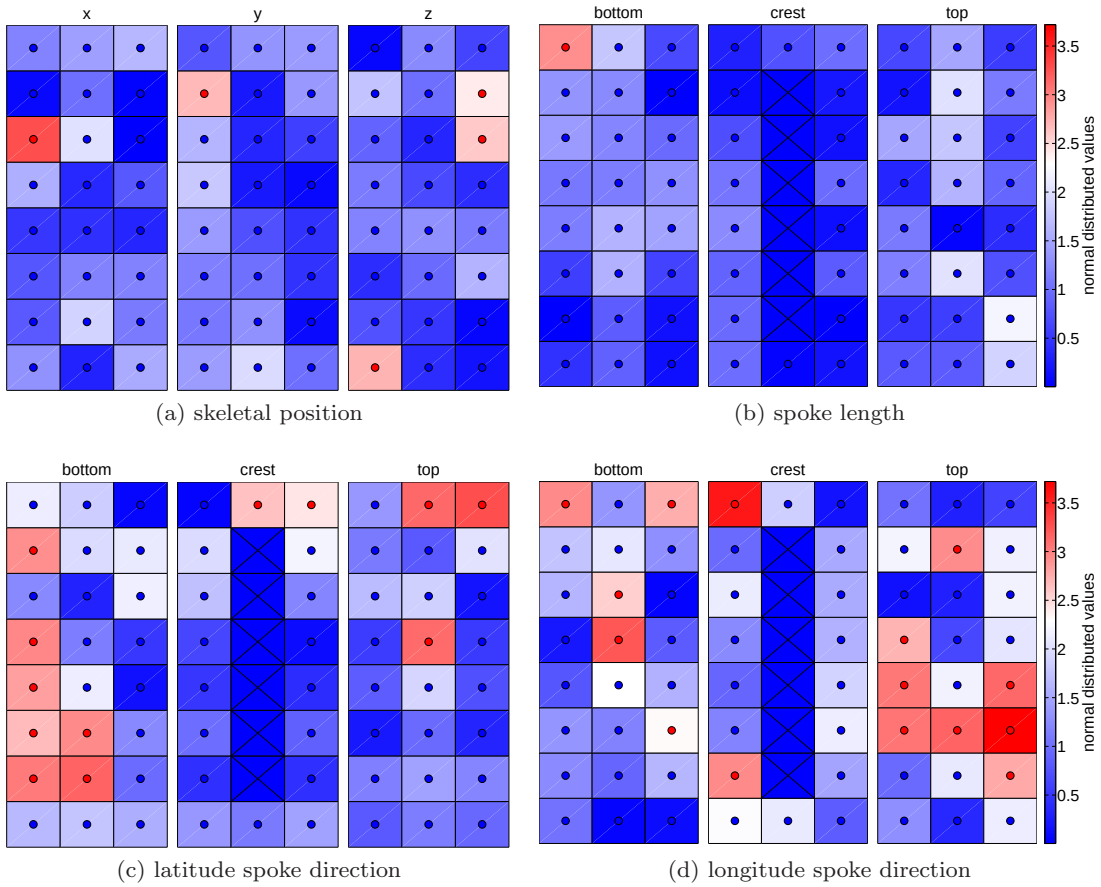
Fig. 7: Significant GOPs using PP1 based on the 3×8 skeletal sheet of the SG CPNG mean. Test results are shown in (a)-(c) for the skeletal x, y and z-positions, in (d) for the latitude spoke directions, in (e) for the longitude spoke directions and in (f) for the spoke lengths. Non-significant skeletal positions are marked by small blue points and significant skeletal positions are marked by large red points. Similar, non-significant spoke directions and lengths are marked by small blue lines whereas significant spoke directions and lengths are marked by wide red lines.

ward mean of SG. Recall that each discrete slabular s-rep is organized into 24 atoms in a 3×8 grid (see Section 3). This results in 271 GOPs with 72 GOPs corresponding to the skeletal positions of the s-rep (x, y and z-positions), 66 GOPs for the latitude spoke directions (bottom, crest and top), 66 GOPs for the longitude spoke directions (bottom, crest and top), 66 GOPs for the spoke lengths (bottom, crest and top) and 1 GOP for the global scaling factor. The corrected threshold is $\lambda = 2.2917$ as defined in Section 6.2.7.

Figure 7 indicates local shape changes by highlighting local parts of the s-rep. Red points mark significant skeletal x, y and z-positions in the Figures (a)-(c). Non-significant skeletal positions are marked by smaller blue points in these figures. Five significant skeletal positions can be observed at the crest of the sheet, one in the x and y-direction and three in the z-direction. Moreover, significant spoke directions and lengths are marked by wide red lines and non-significant by thinner blue lines in the Figures (d)-(f). Several latitude and longitude spoke directions indicate locally significant deformations between the two groups in the Figures (d)-(e). The most latitude differences are statistically significant on the bottom side of the skeletal sheet whereas more longitude differences are significant on the top side. Furthermore, we observe no spoke direction with simultaneously significant latitude and longitude. This behavior should be investigated in future studies. The latitude and longitude differences could indicate local bending around the y and z-axis, respectively. Figure (f) highlights one significant spoke length on the front bottom side of the skeletal sheet.

In addition to the results presented in Figures 7, the global scaling factor τ between SG and CG was found statistically significant. The GOP $|U_{0K}|$ was 2.7627 where the index K corresponds to the global scale factor.

These observations and results are also emphasized by Figure 8 which shows the magnitude of significance of all GOPs except the scaling factor. In order to simplify the visualization all standard normal values $U_{0k}, k = 1, \ldots, K-1$ are presented in absolute values. The color map is non-linear defined from blue to white to red. The corrected threshold λ defines the color white. Blue and red visualize non-significant and significant values, respectively. The blue small circle inside a block marks whether a U_{0k} is less than or equal to the threshold λ. Red small circles mark if a



(a) skeletal position (b) spoke length

(c) latitude spoke direction (d) longitude spoke direction

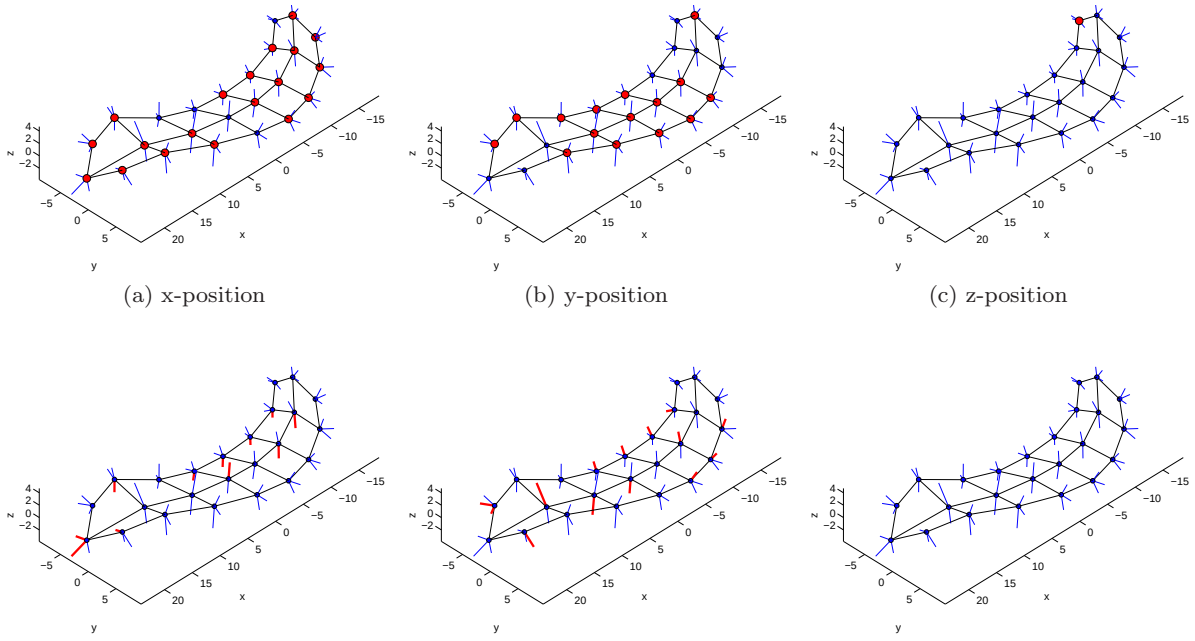
Fig. 8: Colored significant map of U_{0k} with a corrected threshold $\lambda = 2.2917$ using PP1. Each box represents a GOP which correspond to a skeletal atom. The color map on the left side is non-linear and has a range from blue (not significant) to white (λ) to red (significant). The circle inside each box marks whether an U_{0k} is less or equal than the threshold λ (symbolized by blue) or if an U_{0k} is greater than the threshold λ (symbolized by red).

U_{0k} is greater than the threshold λ and thus statistically significant, showing up as red in Figure 7. Particularly, several latitude and longitude spoke directions show a highly significant magnitude in Figure 8.

Figures 9 and 10 are identical to the two previous figures except for the use of PP2 instead of PP1. Several skeletal x and y-positions are statistically significant in contrast to Figures 7 and 8 with only one significant skeletal x and y-position. The volume difference between the two populations is reflected by the skeletal x and y-positions using PP2. Thus, the significant skeletal x and y-positions show rather significant differences from a global deformation than from local deformations. However, we observe only one statistically significant skeletal z-position because the skeletal sheet of the hippocampus is rather flat, as medial as possible and therefore located close to the x-y plane, where $z = 0$. As a result, several skeletal z-coordinates are scaling invariant.

Nevertheless, the observation of only one significant skeletal z-position in addition to no observed statistically significant spoke length in Figure 9f implies that we only observe statistically significant volume differences in the x-y direction but not in the z-direction. Skeletal x and y-positions equal to $x = 0$ and $y = 0$ are scaling invariant in the x and y-directions respectively. As a result, no statistically significant x-positions can be observed close to $x = 0$ in Figure 9a. Moreover, Figures 10c and 10d show only small differences compared to Figures 8c and 8d. Similar results between spoke directions are expected because of the scaling invariance of $u_{ij} \in S^2$. The slightly different color scheme is also due to a different threshold.

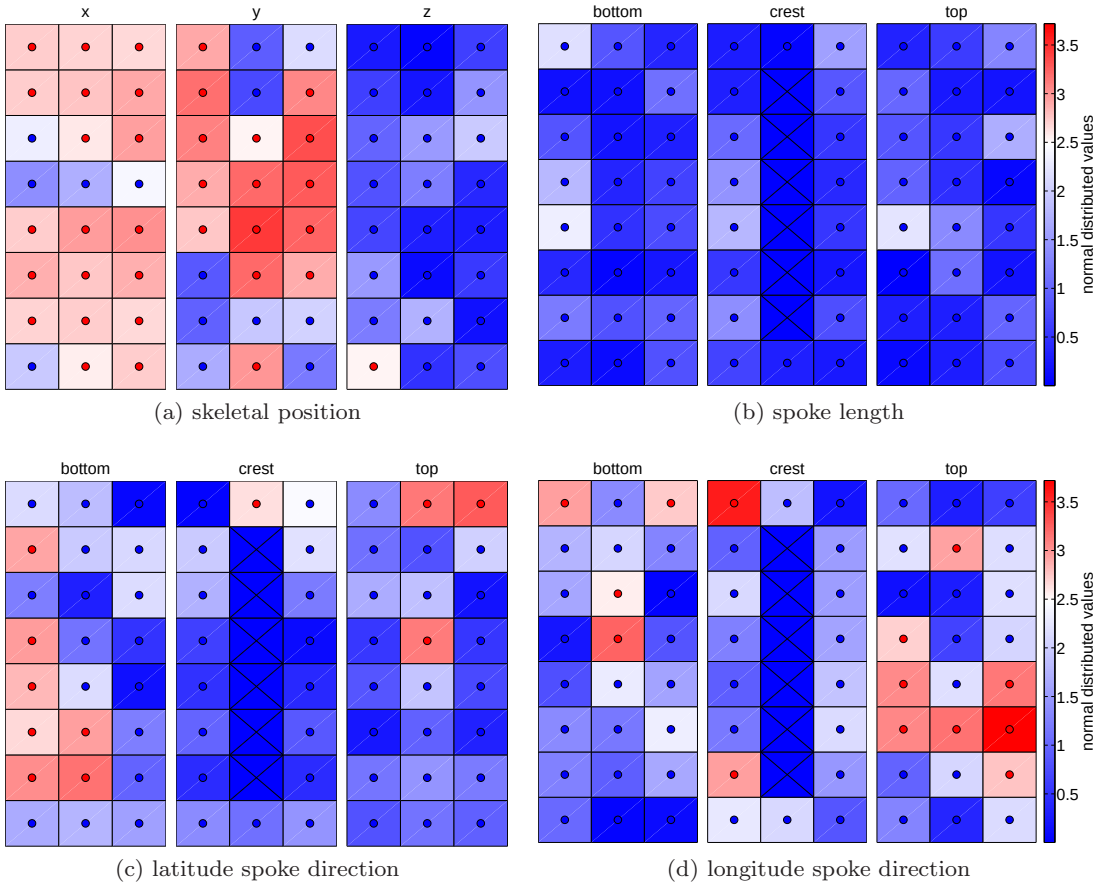
Additional computations and results are shown in the Supplementary Material. Section 2.5 of the Supplementary Material presents results using an alternative measure d^1 defined by a vector of unsigned partial differences such as the Euclidean distance between two skeletal positions. That difference measure changes the GOPs, i.e., how the single s-rep features are combined to GOPs. The difference measure d^2 closely reflects each s-rep feature. The choice of an appropriate difference measure depends on the nature of the medical



(a) x-position (b) y-position (c) z-position

(d) latitude direction (e) longitude direction (f) spoke length

Fig. 9: As Figure 7, now based on PP2.



(a) skeletal position (b) spoke length

(c) latitude spoke direction (d) longitude spoke direction

Fig. 10: As Figure 8, now based on PP2 with a corrected threshold $\lambda = 2.4837$.

research question. In addition, hypothesis test results using a second group of final fittings are presented, derived from 5 independent CPNG stages in Section 5 by using a pooled shape distribution, two individual shape distributions and two individual interchanged shape distributions. The second group of final fittings is described in detail in Section 3 of the Supplementary Material.

8 Discussion

This paper proposes a novel method to test global and local hypotheses on Euclidean and non-Euclidean data. Important requirements of shape models are pointed out in order to test for population differences. Furthermore, suitable statistical methods are proposed to analyze the Euclidean and non-Euclidean elements of the models. In addition, the estimation of appropriate shape distributions of populations is worked out. Finally, the analysis of first episode schizophrenia patients compared to controls demonstrated the power of the hypothesis test given a proper pre-processing. The effect of different pre-processings of the data are highlighted. The developed feature-by-feature test is novel and important for physicians in order to understand local shape changes. The method can easily be adapted for desired GOPs depending on underlying research questions. A difference measure for the analysis of s-reps is proposed. The visualization of local shape changes is of great interest for the study of local rotational deformations [39] which is a subject of future studies.

The s-rep model, statistics and the fitting procedure resulted in accurate fittings with a high concentration of variance in relatively few eigenmodes. This reflects the high correspondence between the s-reps. The introduced test found significant differences between the two populations. First, a statistically significant loss of hippocampal volume was observed by the global scaling factor which is in agreement with [25,28,29]. Second, a significant volume difference was observed in the x and y-directions but not in the z-direction for the aligned hippocampi. Third, several spoke directions were found as statistically significant.

This study is the first study that examines directional information using s-reps. The significant differences of several spoke directions confirm the importance of our contribution in the research of morphological shape changes and encourages further research. Later studies should more deeply investigate if spoke direction differences are due to independent local deformation of GOPs or due to local rotational deformation. Styner et al. [42] indicated a potential local bending of the hippocampi between the two groups. Furthermore, this study is the first study that could identify directions driving the volume change.

In general, results are challenging to compare between studies of brain morphology because of different models, features and metrics. Narr et al. [29] calculated a radial distance measure in addition to a measure that examined the signal intensity on the basis of a surface based mesh modelling method. Also, Mamah et al. [25] used a triangulated graph representation of the hippocampi. Such models are limited compared to s-reps because the interior of an object is not described by the model itself. The model representation in McClure et al. [28] is a skeleton type which leads to less correspondence between populations and contains further disadvantages, e.g., all spoke length are identical on each atom. Furthermore, McClure et al. [28] applied an FDR based test approach in contrast to the FWER based approach proposed in this paper. The FDR is a less strict multiple testing criteria than the FWER. However, the discussed results are consistent between the studies. The s-rep model provides a relatively rich description of an object. Moreover, the proposed test procedure offers global and local nonlinear hypothesis tests based on Euclidean and non-Euclidean GOPs. Thus, the test supports more consistent and sensitive interpretations of morphological changes.

This paper motivates several areas of further research. 1) A simplification of the s-rep fitting procedure is desirable that depends on correct choice of several fitting parameters. The choice of a large number of parameters might be an avoidable difficulty in the use of s-reps in clinical practice. 2) The definition of an adaptive s-rep model that finds an optimal skeletal grid could be of relevance for the future. The grid need not to be rectangular but must correspond across cases of a population. 3) The hypothesis test might be extended by including image intensities in addition to morphological features. An interesting research question is the study of correlation between morphological changes and intensities. 4) An alternative combining function might decrease the required large number of permutations for a global test. 5) A power study on the basis of simulated data to elaborate further the proposed method. 6) A comparison of hippocampi between two treatment groups of first episode schizophrenia. 7) Extension of the method to hippocampi from longitudinal data. In addition, a similar hypothesis test based on the sample variance instead of the sample mean could be of future interest.

Acknowledgements The following researchers have also contributed to this work: Jared Vicory (UNC) gave advice on running Pablo and provided earlier fits of 62 hippocampi, Juan Carlos Prieto (CREATIS-INSA, France) provided the implementation of a crest interpolation term in Pablo and removed bugs from the program, Sungkyu Jung

(University of Pittsburgh, USA) provided Figure 2, program code and additional discussions about CPNS, Martin Styner (UNC) provided the hippocampus data set and answered questions. The first author acknowledges support from the Norwegian Research Council through grant 176872/V30 in the eVita program and additional support from the Tromsø Telemedicine Laboratory and the Department of Electrical Engineering and Computer Science at the University of Stavanger, Norway.

References

1. Abramovich, F., Benjamini, Y.: Adaptive thresholding of wavelet coefficients. Comput. Statist. Data Anal. **22**(4), 351–361 (1996)
2. Albertson, R.C., Streelman, J.T., Kocher, T.D.: Genetic basis of adaptive shape differences in the cichlid head. J. Hered. **94**(4), 291–301 (2003)
3. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. **57**(1), 289–300 (1995)
4. Blum, H., Nagel, R.: Shape description using weighted symmetric axis features. Pattern Recognit. **10**(3), 167–180 (1978)
5. Bookstein, F.L.: Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med. Image Anal. **10**(3), 225–243 (1996)
6. Bullmore, E., Fadili, J., Breakspear, M., Salvador, R., Suckling, J., Brammer, M.: Wavelets and statistical analysis of functional magnetic resonance images of the human brain. Stat. Methods Med. Res. **12**(5), 375–399 (2003)
7. Chumbley, J.R., Friston, K.J.: False discovery rate revisited: FDR and topological inference using Gaussian random fields. NeuroImage **44**(1), 62–70 (2009)
8. Cootes, T.F., Taylor, C., Cooper, D., Graham, J.: Training models of shape from sets of examples. In: D. Hogg, R. Boyle (eds.) Proc. British Machine Vision Conference, pp. 9–18. Berlin. Springer-Verlag (1992)
9. Damon, J., Marron, J.S.: Backwards principal component analysis and principal nested relations. J. Math. Imaging Vision (2013). DOI: 10.1007/s10851-013-0463-2
10. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. John Wiley & Sons, Chichester (1998)
11. Edgington, E.: Randomization Tests, 3rd edn. Dekker, New York (1995)
12. Ferrarini, L., Palm, W.M., Olofsen, H., van Buchem, M.A., Reiber, J.H., Admiraal-Behloul, F., et al.: Shape differences of the brain ventricles in Alzheimer's disease. NeuroImage **32**(3), 1060–1069 (2006)
13. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. Ann. Inst. Henri Poincaré **10**, 215–310 (1948)
14. Gerig, G., Styner, M., Shenton, M.E., Lieberman, J.A.: Shape versus size: improved understanding of the morphology of brain structures. MICCAI pp. 24–32 (2001)
15. Goodall, C.: Procrustes methods in the statistical analysis of shape. J. R. Stat. Soc. Ser. B Stat. Methodol. **53**(2B), 285–339 (1991)
16. Gouttard, S., Styner, M., Joshi, S., Gerig, G.: Subcortical structure segmentation using probabilistic atlas prior. In: Proceedings of the SPIE Medical Imaging, vol. 65122, pp. J1–J11 (2007)
17. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. Statist. Sinica **20**(1), 1–58 (2010)
18. Jung, S., Dryden, I.L., Marron, J.S.: Analysis of principal nested spheres. Biometrika **99**(3), 551–568 (2012)
19. Jung, S., Foskey, M., Marron, J.S.: Principal arc analysis on direct product manifolds. Ann. App. Statist. **5**(1), 578–603 (2011)
20. Jung, S., Liu, X., Marron, J.S., Pizer, S.M.: Generalized PCA via the backward stepwise approach in image analysis. In: J.A. et al. (ed.) Brain, Body and Machine: Proceedings of an International Symposium on the 25th Anniversary of McGill University Centre for Intelligent Machines, *Advances in Intelligent and Soft Computing*, vol. 83, pp. 111–123 (2010)
21. Karcher, H.: Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math. **30**(5), 509–541 (1977)
22. Kendall, D.G., Barden, D., Carne, T.K., Le, H.: Shape and Shape Theory. Wiley, Chichester (1999)
23. Kilner, J.M., Kiebel, S.J., Friston, K.J.: Applications of random field theory to electrophysiology. Neurosci. Lett. **374**, 174–178 (2005)
24. Kurtek, S., Ding, Z., Klassen, E., Srivastava, A.: Parameterization-invariant shape statistics and probabilistic classification of anatomical surfaces. Inf. Process. Med. Imaging **22**, 147–158 (2011)
25. Mamah, D., Harms, M.P., Barch, D.M., Styner, M.A., Lieberman, J., Wang, L.: Hippocampal shape and volume changes with antipsychotics in early stage psychotic illness. Front Psychiatry **3**(96), 1–10 (2012)
26. Marozzi, M.: Some remarks about the number of permutations one should consider to perform a permutation test. Statistica **64**(1), 193–202 (2004)
27. Marron, J.S., Todd, M.J., Ahn, J.: Distance weighted discrimination. J. Amer. Statist. Assoc. **102**(480), 1267–1271 (2007)
28. McClure, R.K., Styner, M., Maltbie, E., Liebermann, J.A., Gouttard, S., Gerig, G., Shi, X., Zhu, H., et al.: Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type. Psychiatry Res. Neuroimaging **211**(1), 1–10 (2013)
29. Narr, K.L., Thompson, P.M., Szeszko, P., Robinson, D., Jang, S., Woods, R.P., Kim, S., Hayashi, K.M., Asunction, D., Toga, A.W., Bilder, R.M.: Regional specificity of hippocampal volume reductions in first-episode schizophrenia. NeuroImage **21**(4), 1563–1575 (2004)
30. Nichols, T.E., Hayasaka, S.: Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat. Methods Med. Res. **12**(5), 419–446 (2003)
31. Niethammer, M., Juttukonda, M.R., Pizer, S.M., Saboo, R.R.: Anti-aliasing slice-segmented medical images via Laplacian of curvature flow. In preparation (2013)
32. Nitrc: S-rep fitting, statistics, and segmentation. http://www.nitrc.org/projects/sreps (2013)
33. Pantazis, D., Nichols, T.E., Baillet, S., Leahy, R.M.: A comparison of random field theory and permutation methods for the statistical analysis of MEG data. NeuroImage **25**(2B), 383–394 (2005)
34. Pennec, X.: Statistical computing on manifolds: from Riemannian geometry to computational anatomy. Emerging Trends in Visual Computing **5416**, 347–386 (2008)
35. Pesarin, F.: Multivariate Permutation Tests with Applications to Biostatistics. John Wiley & Sons, Chichester (2001)
36. Pizer, S.M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J.N., Huckemann, S., Marron, J.S.: Nested sphere statistics of skeletal models. In: Innovations for Shape Analysis: Models and Algorithms, Lecture Notes in Comput. Sci., pp. 93–115. Springer (2013)

37. Qiao, X., Zhang, H.H., Liu, Y., Todd, M.J., Marron, J.S.: Weighted distance weighted discrimination and its asymptotic properties. J. Amer. Statist. Assoc. **105**(489), 401–414 (2010)
38. Rohde, G.K., Ribeiro, A.J.S., Dahl, K.N., Murphy, R.F.: Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells. Cytometry A **73**(4), 341–350 (2008)
39. Schulz, J., Jung, S., Huckemann, S., Pierrynowski, M., Marron, J.S., Pizer, S.M.: Analysis of rotational motion from directional data. Submitted (2013)
40. Shi, X., Ibrahim, J.G., Lieberman, J., Styner, M., Li, Y., Zhu, H.: Two-stage empirical likelihood for longitudinal neuroimaging data. Ann. Appl. Stat. **5**(2B), 1132–1158 (2011)
41. Siddiqi, K., Pizer, S.: Medial Representations: Mathematics, Algorithms and Applications, 1 edn. Computational Imaging and Vision, Vol. 37. Springer, Dordrecht, Netherlands (2008)
42. Styner, M., Lieberman, J., Pantazis, D., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophrenia. Med. Image Anal. **8**(3), 197–203 (2004)
43. Terriberry, T., Joshi, S., Gerig, G.: Hypothesis Testing with Nonlinear Shape Models. In: G. Christensen, M. Sonka (eds.) Information Processing in Medical Imaging, *Lecture Notes in Computer Science*, vol. 3565, pp. 15–26. Springer Berlin Heidelberg (2005)
44. Van De Ville, D., Blu, T., Unser, M.: Integrated wavelet processing and spatial statistical testing of fMRI data. NeuroImage **23**(4), 1472–1485 (2004)
45. Wang, L., Joshi, S.C., Miller, M.I., Csernansky, J.G.: Statistical analysis of hippocampal asymmetry in schizophrenia. NeuroImage **14**(3), 531–545 (2001)
46. Wei, S., Lee, C., Wichers, L., Li, G., Marron, J.S.: Direction-projection-permutation for high dimensional hypothesis tests (2013). ArXiv:1304.0796

Supplementary Material

Nonlinear Hypothesis Testing of Geometrical Object Properties of Shapes Applied to Hippocampi

Jörn Schulz · Stephen M. Pizer · J.S. Marron · Fred Godtliebsen

1 Model fitting and statistics

1.1 Limitation of a 3×8 grid of skeletal positions

A hippocampus example with bumps which are not tightly described by a (3×8) grid is visualized in Figure 1. An s-rep model with a larger number of skeletal positions, i.e., with a finer grid could solve such problems. The example depicts a limitation only in specific cases since the shape of the hippocampus differs from person to person. Furthermore, we do not look at individual s-reps that may not be perfectly correct but rather at differences between groups which are not biased versus the other.

1.2 Discussion on CPNS analysis across populations

In Section 4 in the main article, we have pointed out the difference between CPNS and CPNG. CPNG uses only great subsphere fittings whereas the best fitting subspheres can be small or great in CPNS. We have observed an increased variance of the CPNS means across several populations, e.g., for a large number of permutation sets as used in the proposed hypothesis test. Jung et al. [2] pointed out a potential overfitting of the data because PNS tends to find smaller spheres than great spheres. Therefore, a sequential test was proposed in [2, Section 3]. This section will propose a modification of the test in [2] and refers to the paper for detailed descriptions. The sequential test procedure consists of a likelihood ratio test and a parametric bootstrap test in order to test the significance of a "small" subsphere fitting as explained in the following.

1. Test $H_{0a} : r = \pi/2$ versus $H_{1a} : r < \pi/2$ by the likelihood ratio test where $r = \pi/2$ indicates a great sphere and $r < \pi/2$ a small sphere. If H_{0a} is accepted, then fit a great sphere with $r = \pi/2$ and proceed to the next layer.
2. If H_{0a} is rejected, then test the isotropy of the distribution by the parametric bootstrap test with $H_{0b} : F_X$ is an isotropic distribution with a single mode, versus $H_{0b} :$ not H_{0b} (i.e., anisotropic) given a distribution function $F_X, X \in S^d$. If H_{0b} is accepted, then use great spheres for all further subsphere fittings.

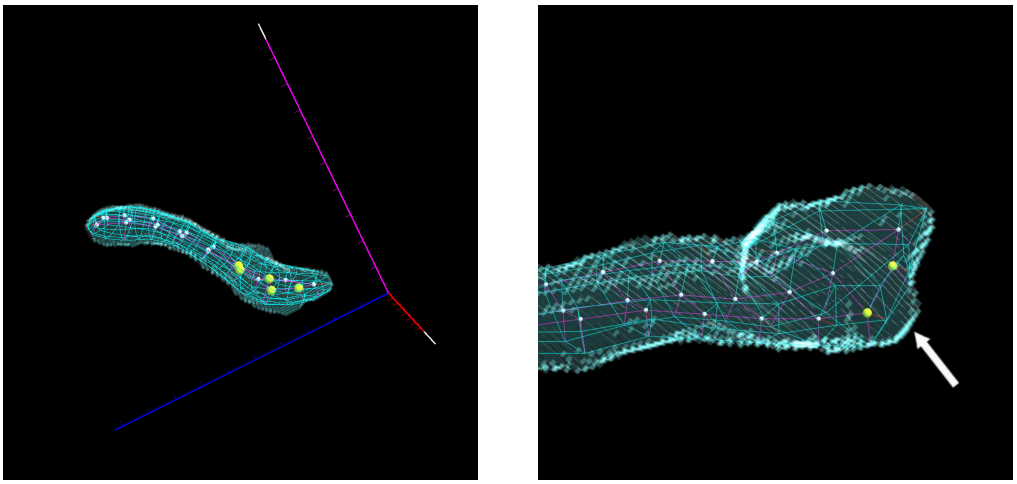
In calculation of CPNS statistics for several populations, the sequential test will be carried out independently for each population leading to potential different decompositions. Thus, the test must be modified, because the analysis of CPNS means across populations requires commensurate coordinate systems. Suppose we have two populations G_1 and G_2 with samples on S^d and P permutations of the set union $G_1 \bigcup G_2$. Each permuted set union can be split into two subgroups G_{1l} and G_{2l} with the same number of elements as G_1

J. Schulz
Department of Mathematics and Statistics, University of Tromsø, Norway Tel.: +47 45696867
E-mail: jorn.schulz@uit.no

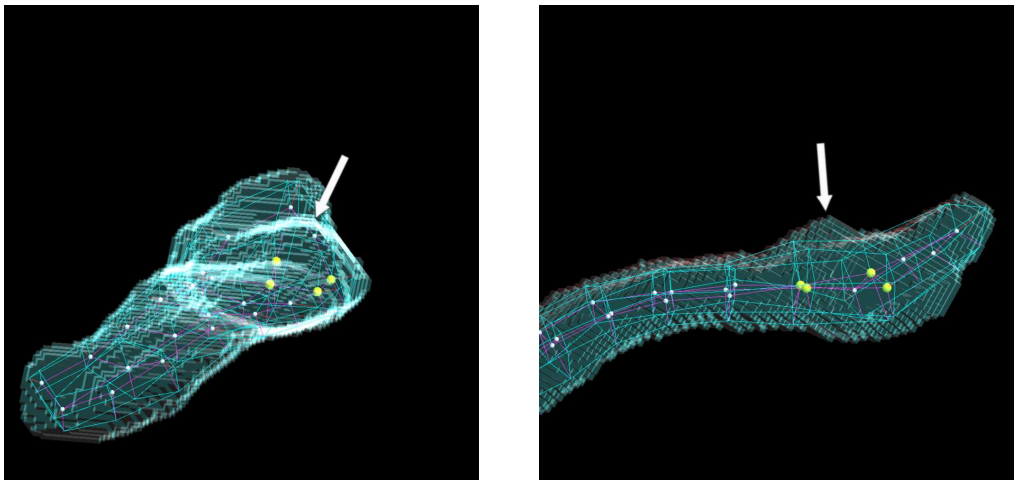
F. Godtliebsen
Department of Mathematics and Statistics, University of Tromsø, Norway, E-mail: fred.godtliebsen@uit.no

Stephen M. Pizer
Department of Computer Science, University of North Carolina at Chapel Hill (UNC), USA, E-mail: smp@cs.unc.edu

J.S. Marron
Department of Statistics & Operations Research, UNC, USA, E-mail: marron@unc.edu



(a) (b)



(c) (d)

Fig. 1: Final fit of a hippocampus with bumps that are not well described by an s-rep based on a (3×8) grid. (a) Entire 3D view to the s-rep with corresponding coordinate system. (b) Bump on the side located between two hub positions. (c-d) Bump on the top located between four hub positions.

and G_2, $l = 1, \ldots, P$. In order to analyze mean difference, the CPNS mean must be calculated for each permutation group $G_{il}, i = 1, 2$. We propose a modified sequential test by the following procedure.

1. Test $H_{0a} : \bigcap_i \bigcap_l H_{0a}^{i,l}$ versus $H_{1a} : \bigcup_i \bigcup_l H_{1a}^{i,l}$ by the likelihood ratio test with $i = 1, 2$ and $l = 1, \ldots, P$, whereas $H_{0a}^{i,l}$ is the sub-hypothesis for the lth permutation of group i. If H_{0a} is accepted, then fit a great sphere with $r = \pi/2$ and proceed to the next layer.
2. If H_{0a} is rejected, then test the isotropy of the distribution by the parametric bootstrap test. If $H_{0b} : \bigcap_i \bigcap_l H_{0b}^{i,l}$ is accepted, then use great spheres for all further subsphere fittings.

The implementation of such a test is left for future work. In this article we have used CPNG to analyze populations of s-reps.

1.3 An alternative unsigned difference measure d^1

This section introduce an alternative difference measure d^1 in addition to d^2 as described in Section 6.2.4 in the main article. The measure d^2 is defined by signed differences whereas the measure d^1 is defined by unsigned differences which turning each GOP into a single non-negative value. Suppose we have two s-reps

$$\mathbf{t}_i = (\tau_i, p_{i1}, \ldots, p_{in_a}, r_{i1}, \ldots, r_{in_s}, u_{i1}, \ldots, u_{in_s})',$$

$i = 1, 2$ with the skeletal positions $p_{ij}, \in \mathbb{R}^3$ and the scale factors $\log(\tau_i), \log(r_{ij}) \in \mathbb{R}$ as Euclidean GOPs and the spoke directions $u_{ij} \in S^2$ as non-Euclidean GOPs. The vector d^1 of differences is defined by

$$d^1(\mathbf{t}_1, \mathbf{t}_2) := (d_1(\tau_1, \tau_2),\ d_2(p_{11}, p_{21}), \ldots, d_2(p_{1n_a}, p_{2n_a}),\ d_3(r_{11}, r_{21}), \ldots, d_3(r_{1n_s}, r_{2n_s}), \\ d_4(u_{11}, u_{21}), \ldots, d_4(u_{1n_s}, u_{2n_s}))' \tag{1}$$

with appropriate partial difference measures: d_1 for the scaling factors τ_i, d_2 for the positions p_{ik}, d_3 for the spoke lengths r_{ij} and d_4 for the spoke directions u_{ij} with $i = 1, 2$, $k = 1, \ldots, n_a$ and $j = 1, \ldots, n_s$ by

$$\begin{aligned} d_1(\tau_1, \tau_2) &= |\log(\tau_2) - \log(\tau_1)|, \\ d_2(p_{1k}, p_{2k}) &= \left(\sum_{m=1}^{3} (p_{2km} - p_{1km})^2 \right)^{1/2}, \\ d_3(r_{1j}, r_{2j}) &= |\log(r_{2j}) - \log(r_{1j})|, \\ d_4(u_{1j}, u_{2j}) &= d_g(u_{1j}, u_{2j}) = \arccos(u'_{1j} u_{2j}). \end{aligned}$$

The geodesic distance function $d_g : S^2 \times S^2 \to [0, \pi]$ is defined by the arc length of the shortest great circle segment joining $u_{1j}, u_{2j} \in S^2$ and is invariant to rotation. The Euclidean metric $d_2 : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}_+$ is invariant to translation and $d_1, d_3 : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ are invariant to scale. All GOP differences of

$$d^1 : (\mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s+1} \times S^{2n_s}) \times (\mathbb{R}^{3n_a} \times \mathbb{R}_+^{n_s+1} \times S^{2n_s}) \longrightarrow \mathbb{R}_+^{n_a+n_s+1} \times [0, \pi]^{n_s}$$

are single non-negative values. Therewith, the hypothesis test of identical statistical distributions of two s-rep populations is given by an one-sided test,

$$H_0 : \{\mu_1 = \mu_2\} \text{ versus } H_1 \ : \ \{\mu_1 > \mu_2\}. \tag{2}$$

Given d^1, we can calculate the p-values $C_k(T_{lk})$ as described in Section 6.2.5 in the main article. In the case of a one-sided test by using difference measure d^1, we map the p-values $C_k(T_{lk})$ to the positive half of a standard Gaussian CDF by

$$\widetilde{U}_{lk} = \Phi^{-1} \left(0.5 + 0.5 \tilde{C}_k(T_{lk}) \right), \tag{3}$$

where Φ^{-1} is the inverse standard Gaussian CDF,

$$\tilde{C}_k(T_{lk}) = \frac{sc - 2}{sc} C_k(T_{lk}) + \frac{1}{sc}$$

and $sc = 10000$, $k = 1, \ldots, K$, $l = 1, \ldots, P$ similar to Section 6.2.5 in the main article.

An open problem is a sensitive mapping of $\widetilde{U}_{lk}$ to a full multivariate distribution that preserve the correlation structure of the variables. Given an appropriate mapping, the global and feature-by-feature test can be applied as described in Section 6.2.6 and 6.2.7 of the main article.

The results presented in Section 2.5 below use random signs $\tau_{lk} \in \{-1, 1\}$ that are generated for each permutation and GOP in order to map $\tilde{C}_k(T_{lk})$ to a full multivariate distribution by $U_{lk} = \tau_{lk} \widetilde{U}_{lk}$ with standard normal marginals. Thereby, we dot not preserve the correlation structure between the GOPs which results in a conservative test.

1.4 Preliminary fitting stage of s-reps to hippocampi

The hippocampus data set consists of binary images of 221 first-episode schizophrenia cases and 56 control cases as described in Section 2 in the main article. Antialiased distance images were generated from the binary images according to [4]. We selected the first 96 of the 221 SG cases to control manual work as described in the following. Based on the distance images, we used the 96 cases of SG and all cases of CG to produce appropriate preliminary fits.

Two different models were used as initializations of the fitting procedure. The first initial model $\mathbf{m}_1$ was a CPNG backwards mean of 62 hippocampus fits presented in [6]. In addition, the second initial model $\mathbf{m}_2$ was derived from the CPNG backwards mean of manually adjusted fits of the control group. The initial models $\mathbf{m}_1$ and $\mathbf{m}_2$ were pre-aligned by translation and rotation, and fit to the hippocampi of CG and SG followed by an atom and spoke stage. As a results, two fittings corresponding to $\mathbf{m}_1$ and $\mathbf{m}_2$ are obtained

for each hippocampus. The fitting with the lowest objective function were selected for further processing. The objective function value is provided by the fitting software Pablo [5] and measures the goodness-of-fit of each s-rep model to the binary data.

The 96 SG and 56 CG fits were manually evaluated and adjusted when necessary. The adjusted fittings were refit by the second atom and spoke stage in order to minimize influence of the manual adjustment on the final fittings and to ensure that all spokes match the object boundary. Let $\tilde{A}_1$ be the set of 96 fits for SG and $\tilde{A}_2$ be the set of 56 fits for CG.

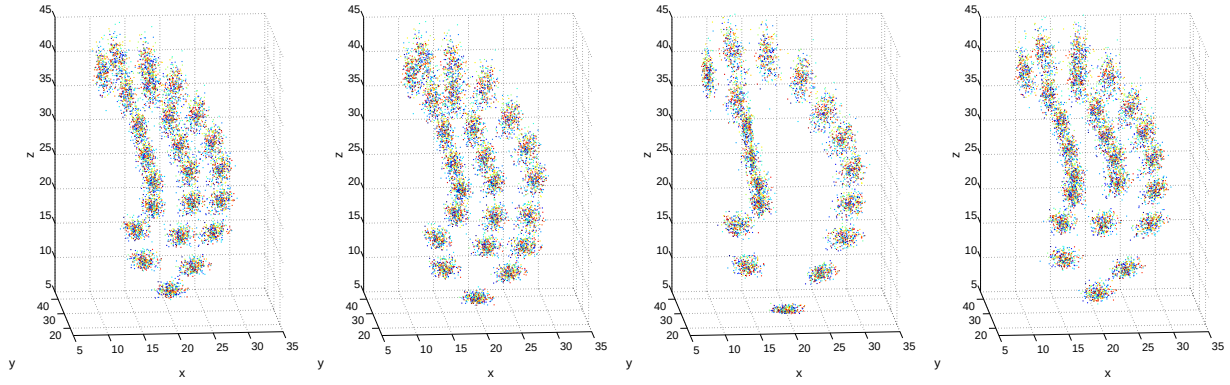
Correspondence across population is achieved by calculation of CPNG statistics. As a pre-processing step the obtained fittings must be aligned, otherwise the CPNG statistics would reflect undesirable rotational variations of the data. Therefore, the CPNG mean of the set union $\tilde{A}_1 \cup \tilde{A}_2$ was calculated. Afterwards, all fittings were translated and rotated to the mean by standard Procrustes alignment [1]. The alignment was based on the skeletal positions and not on the spoke ends, due to the CPNG analysis of the skeletal positions in a pre-shape space as described in Section 4 in the main article. Let $\bar{A}_1$ be the set of 96 aligned SG fits and $\bar{A}_2$ the set of 56 aligned CG fits. Finally, CPNG statistics were calculated for the s-rep populations $\bar{A}_1$, $\bar{A}_2$ and the pooled population $\bar{A}_1 \cup \bar{A}_2$.

2 Additional data analysis on fittings using a pooled shape distribution

The presented results in the main article are based on fittings obtained by the use of a pooled shape distribution during the CPNG stage (see Sections 7.1 in the main article). This section will present additional analyses and plots based on the same data.

2.1 Procrustes alignment of final fittings

Let $\tilde{A}$ be the obtained fittings of s-reps after the CPNG stage, final spoke stage and re-scaling into a world coordinate system as described in Section 7.1 in the main article. Figure 2 visualizes the skeletal positions and the spoke tail ends of $\tilde{A}$. Each spoke tail end is defined by the corresponding skeletal position, spoke direction and length.



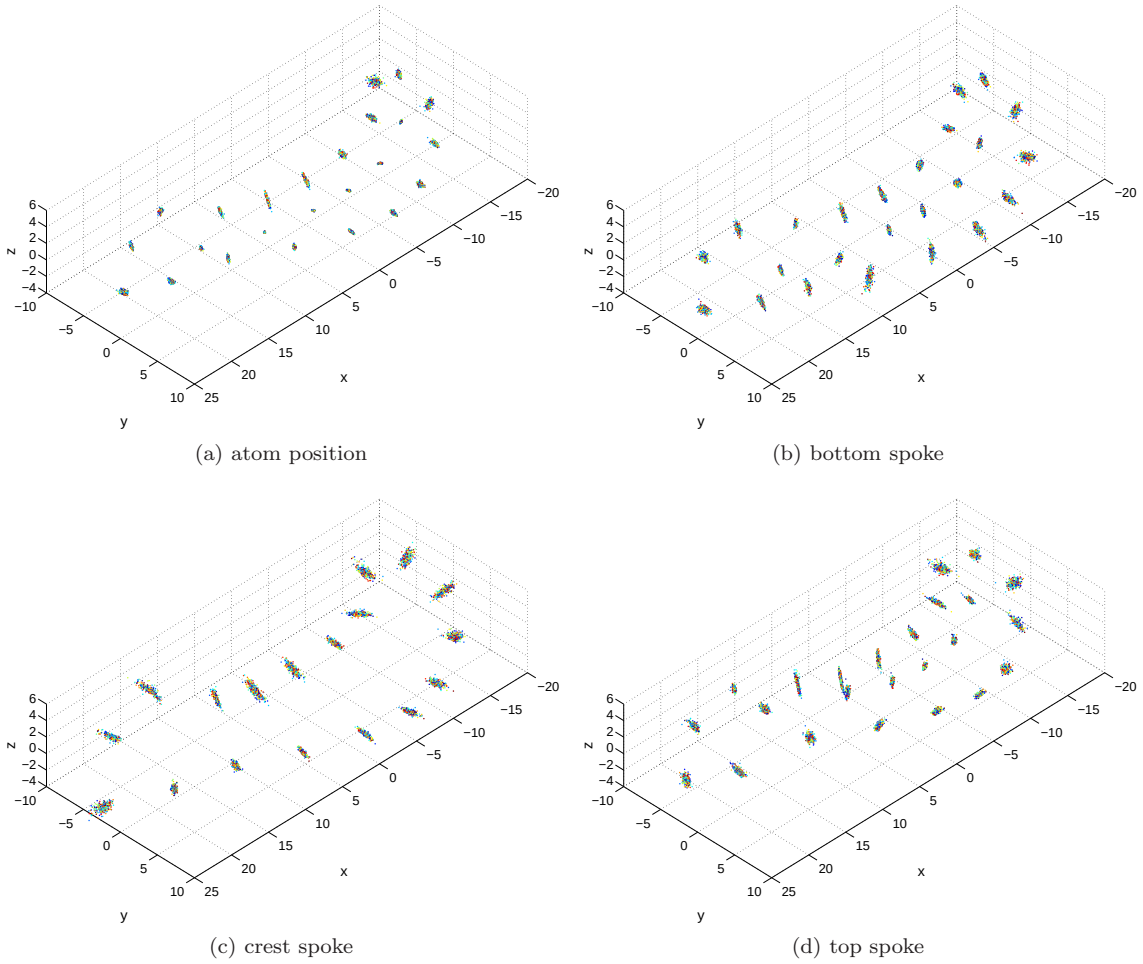
(a) atom position (b) bottom spoke (c) crest spoke (d) top spoke

Fig. 2: Final obtained s-rep fittings after the final spoke stage and re-scaling into a world coordinate system. Skeletal positions are depicted in (a). Bottom, crest and top spoke directions and lengths are depicted in (b-d) by the spoke tail ends based on the corresponding skeletal positions. The 277 fittings are represented by individual colors.

As discussed in Section 6.2.1 in the main article, an appropriate pre-processing of the data is required for a reasonable interpretation of the differences, e.g., between the latitude, longitude, x, y and z-coordinate using d^2. Let $\tilde{\mu}$ the overall backwards CPNG mean, estimated from the set union $\tilde{A}$ of obtained final fittings with

$$\tilde{A} = \tilde{A}_1 \bigcup \tilde{A}_2 = \{\tilde{\mathbf{s}}_{11}, \ldots, \tilde{\mathbf{s}}_{1N_1}, \tilde{\mathbf{s}}_{21}, \ldots, \tilde{\mathbf{s}}_{2N_2}\}.$$

The CPNG mean $\tilde{\mu}$ is translationally aligned by the subtraction of the mean of the locational components. In addition, the eigenvectors of the second moments about the center of the skeletal positions yields a rotational alignment to the x, y and z-axis. The translationally and rotationally aligned CPNG mean $\tilde{\mu}$ is called μ. Figure 3 depicts the translated, rotated and scaled s-reps of $\tilde{A}$ to μ using a standard Procrustes alignment [1], based on the skeletal positions of each s-rep $\tilde{\mathbf{s}} \in \tilde{A}$. The pre-processing removed undesirable variation from the data and enabled a meaningful interpretation for later analysis. This is highlighted by Figure 3 which shows considerable reduced variation compared to Figure 2.



(a) atom position (b) bottom spoke

(c) crest spoke (d) top spoke

Fig. 3: S-reps fittings are visualized after standard Procrustes alignment with translation, rotation and scaling based on the skeletal positions. The aligned skeletal positions are depicted in (a). Bottom, crest and top spoke directions and lengths are depicted in (b-d) by the spoke tail ends based on the corresponding skeletal positions. The 277 fittings are represented by individual colors.

2.2 Visualization of generated permutations

The distribution of $P = 1000$ permuted sample means $\hat{\nu}_{1l}$ for SG and $\hat{\nu}_{2l}$ for CG (see Section 6.2.2 in the main article) is visualized in Figure 4, $l = 1, \ldots, P$. The permuted sample means are depicted by the projections of the scaled CPNG scores matrix Z_{Comp} of $\{\hat{\nu}_{1l}, \hat{\nu}_{2l} \mid l = 1, \ldots, P\}$ (see Section 4 in the main article) onto the distance-weighted discrimination (DWD) direction and the first three orthogonal directions to the DWD direction as described in Marron et al. [3] and Qiao et al. [7]. Red circles depict permuted SG means and blue circles permuted CG means. The larger variance of CG is due to the unbalanced group size (SG contains 221 cases and CG 56 cases). The observed Gaussian distributions indicate appropriate permutation sets.

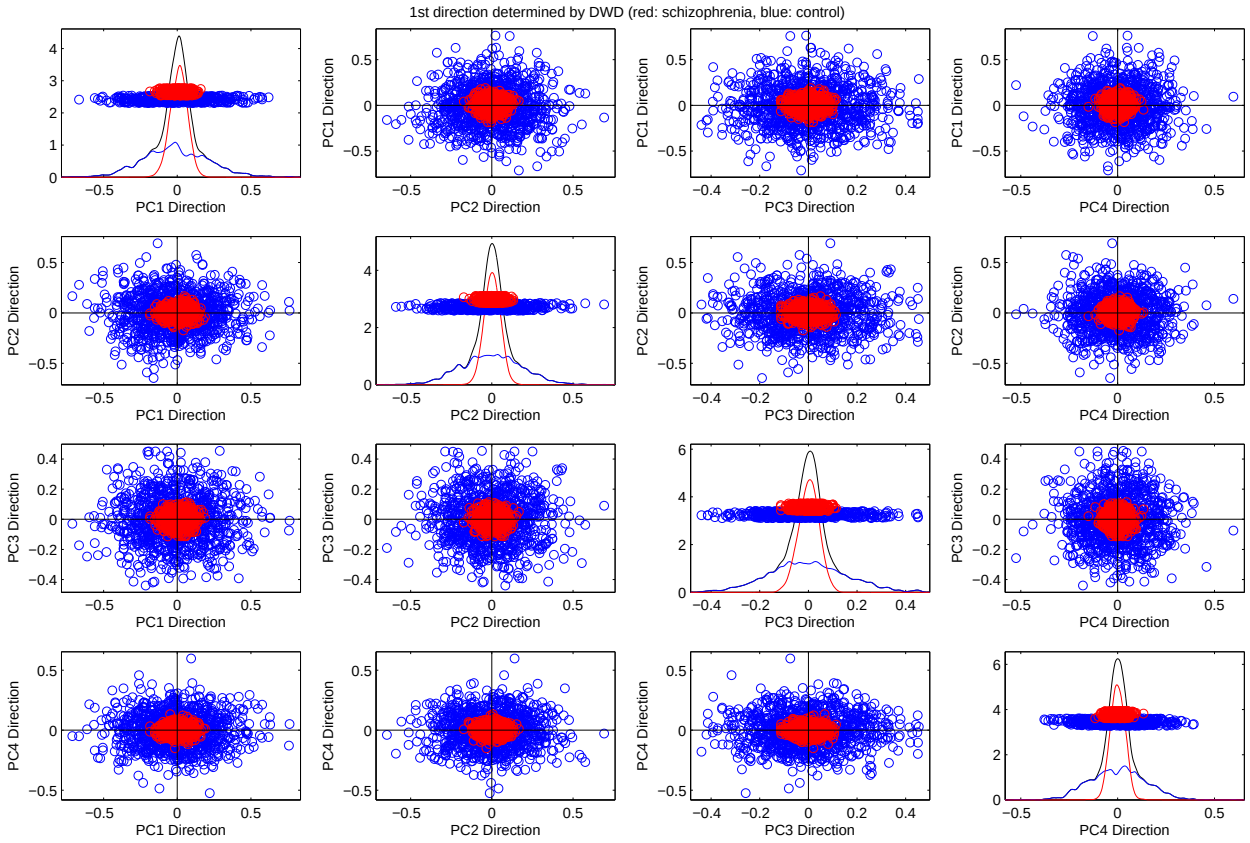


Fig. 4: Scatter plots and jitterplots (diagonal) with KDE are showing the distribution of permuted sample means projected on the DWD direction and the first three orthogonal directions to the DWD direction. Additionally, the KDE of the pooled distribution of SG and CG is shown in the jitterplots. Red circles depict permuted SG means and blue circles permuted CG means.

2.3 DiProPerm results using a MD test statistic and a DWD projection direction

Figure 5 visualizes the DiProPerm test reported in Table 1 in Section 7.2 in the main article using a mean difference (MD) test statistic and DWD as the projection direction. The DiProPerm test is based on the evaluation of the scaled CPNG scores matrix Z_{Comp} as described in Section 4 in the main article. The DiProPerm test is a global test and the hypothesis of identical mean between the two populations was rejected given a significance level $\alpha = 0.05$.

2.4 ROC analysis compared to feature-by-feature test results using distance measure d^2 and PP1

This section evaluates the performance of the feature-by-feature test by Receiver Operating Characteristic (ROC) curves. The ROC analysis gives a curve lying in $[0,1]\times[0,1]$, which quantifies the amount of "overlap" of each GOP between the samples of the two populations. The ROC curve resulting from the observed data is visualized by a red line in the following plots. In addition, for each permutation a ROC curve is generated, represented by a blue line, which results in an envelope under the null distribution. In the following, each envelope is visualized by the first $1,000$ of the $30,000$ permutations. A ROC curve of the observed data close to the boundary of this envelope indicates a significant feature. The comparison is done using the distance measure d^2 and the standard pre-processing of the data as described in Section 6.2.1 in the main article. The GOPs that represent latitude and longitude of the spoke direction are normalized corresponding to the mean shift as explained in Section 6.2.4 in the main article.

The feature-by-feature test results are reported in Figures 7 and 8 in Section 7.3 in the main article. Several GOPs were tested as statistically significant including the global scaling factor $|U_{0K}| = 2.7627$ given

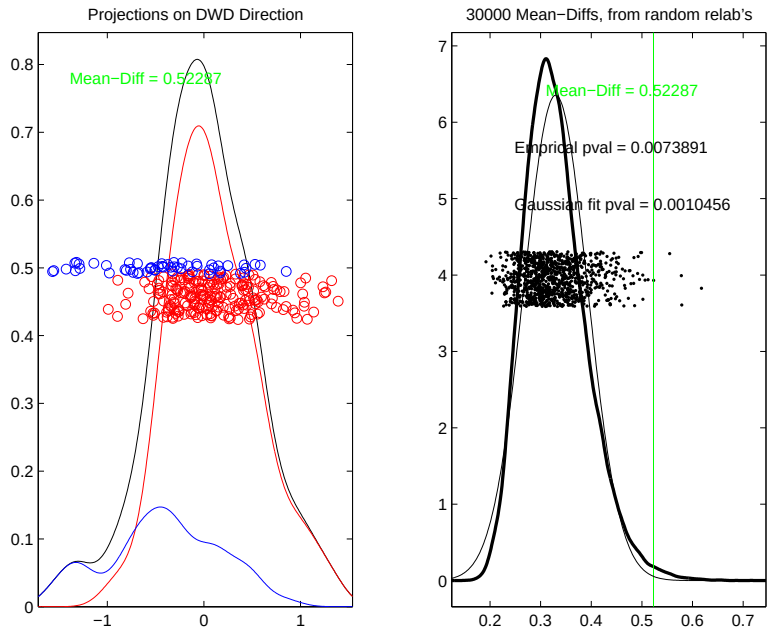


Fig. 5: The DiProPerm hypothesis test of mean differences based on the scaled CPNG scores matrix Z_{Comp} of the final fittings after pre-processing by PP1. DiProPerm is a two sample mean hypothesis test. The left plot shows a jitterplot by the projection of the data on the DWD direction together with the kernel density estimates (KDEs) of the distribution of SG (red circles), CG (blue circles) and the set union SG∪CG. The right plot shows a jitterplot of the mean differences of the 30,000 permutations, a KDE of the distribution of the MD test statistic in addition to the MD between the observed population SG and CG (green line).

a corrected threshold $\lambda = 2.2917$. Figure 6 depicts the ROC curve for the global scaling factor (red) together with the envelope (blue) obtained from the permutations. A major part of the red curve is located close to the boundary of the envelope. Thus, Figure 6 indicates a significant GOP in agreement with the obtained feature-by-feature test result.

The area under the curve (AUC) value is a simple numerical summary which is useful for a comparison of several ROC curves, e.g., a comparison of the ROC curves between the figures below.

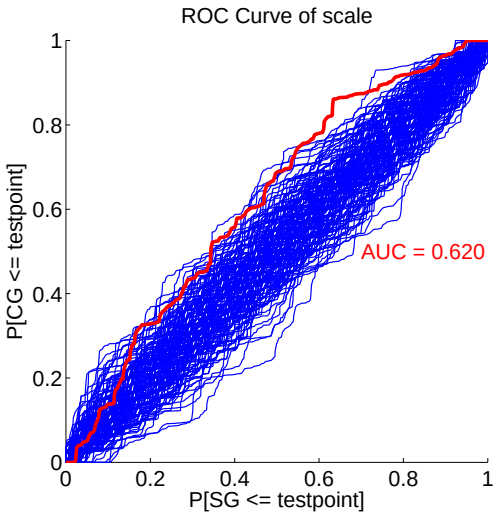
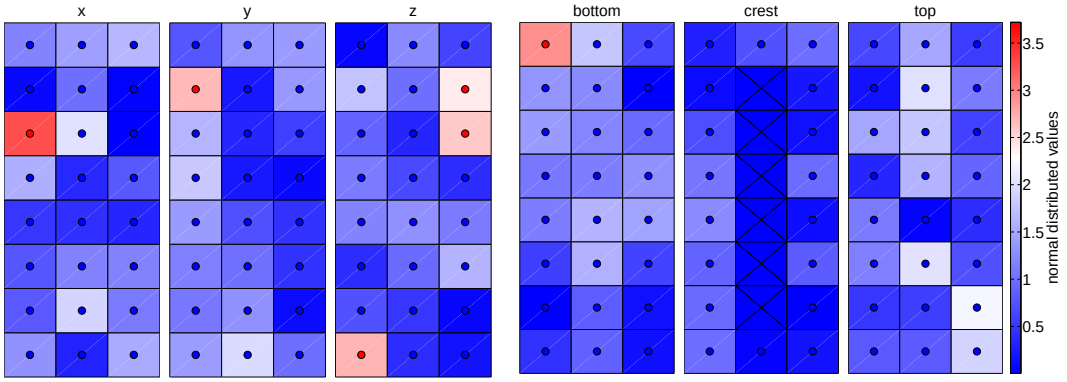
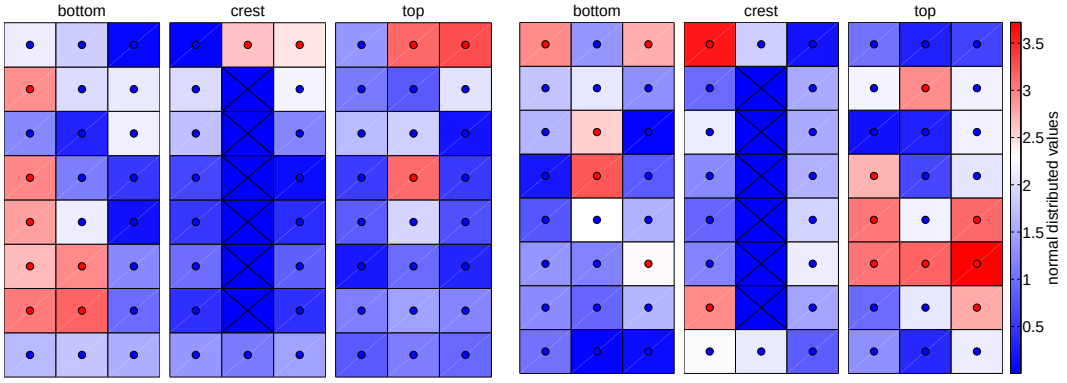


Fig. 6: The ROC curve of the global scaling factor (red) is visualized together with the envelope (blue) obtained from the permutations.

Figure 7 below is identical to Figure 8 in the main article and shows the magnitude of significance of each GOP using the difference measure d^2. In order to simplify the visualization all standard normal values $U_{0k}, k = 1, \ldots, K$ are presented in absolute values. The color map is non-linear defined from blue to white to red. The corrected threshold $\lambda = 2.2917$ defines the color white, blue and red visualize non-significant and significant values, respectively. Blocks which show a white color have U_{0k} around the threshold λ. The blue small circles inside each block mark whether a U_{0k} is less than or equal to the threshold λ. Red small circles mark if an U_{0k} is greater than the threshold λ and therewith statistical significant.



(a) skeletal position (b) spoke length



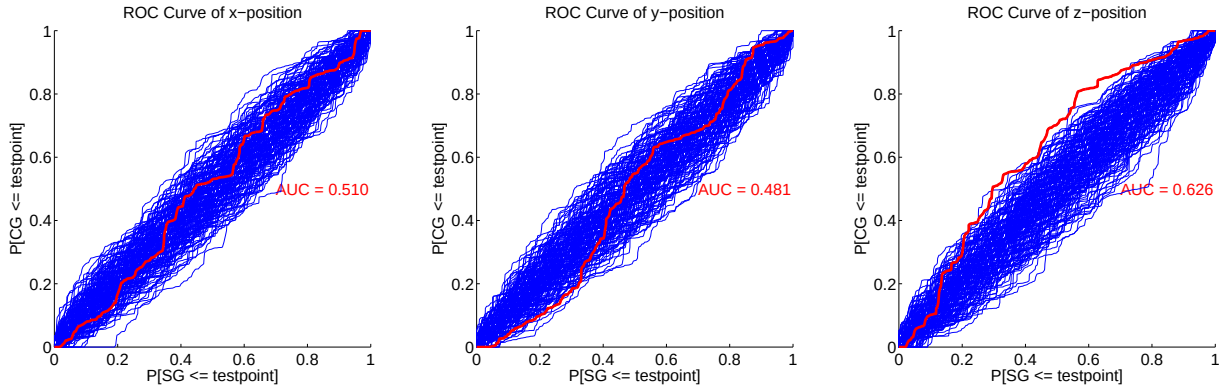
(c) latitude spoke direction (d) longitude spoke direction

Fig. 7: Colored significant map of U_{0k} using difference measure d^2 with a corrected threshold $\lambda = 2.2917$. Each box corresponds to a GOP. The color map on the left side is non-linear and has a range from blue (not significant) to white (λ) to red (significant). The circle inside each box marks whether an U_{0k} is less or equal than the threshold λ (symbolized by blue) or if an U_{0k} is greater than the threshold λ (symbolized by red).

The results are presented on the basis of the 3×8 skeletal sheet such as the 24 skeletal x-positions in Figure 7a. The skeletal sheet is numbered from bottom to top and from left to right, i.e., atom 1 correspond to the the left bottom block, atom 8 to the left top block, atom 9 to the middle bottom block, atom 16 to the middle top block, atom 17 to the right bottom block and finally, atom 24 correspond to the right top block. In the following, we compare results for selected GOPs from Figure 7 with the ROC analysis.

Figure 8 visualizes the ROC curve of the skeletal x, y and z-position of atom 22. Figure 7a indicates the z-position of atom 22 as statistically significant. The x and y-position are not statistically significant whereas the x-position shows a lower value than the y-position of atom 22. These results are reflected in Figure 8 by the ROC analysis. The ROC curve for the x-position of atom 22 is located close to the center of the envelope, the ROC curve for the y-position is located closer to the boundary of the envelope in some regions whereas the ROC curve for the z-position is close to the boundary in major parts of the envelope.

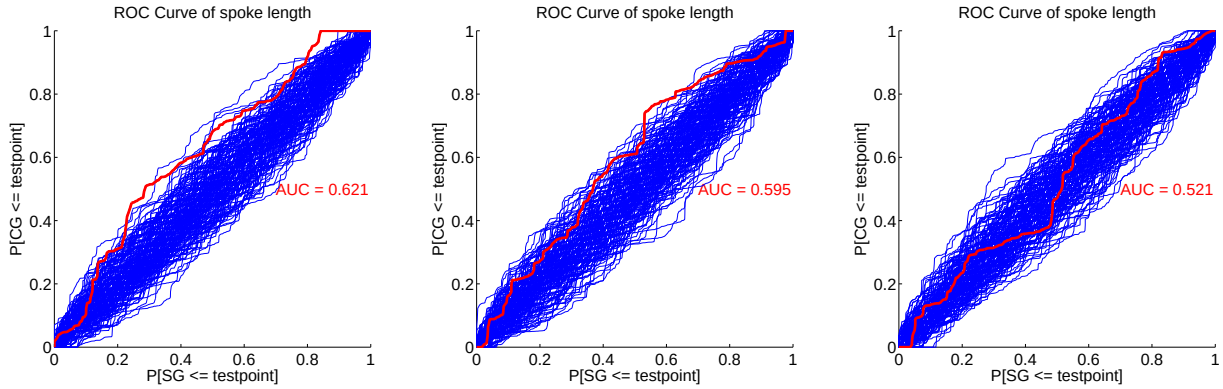
Figure 9 visualizes the ROC curve of the bottom spoke lengths of atom 8, 16 and 24. Figure 7b indicates the bottom spoke length of atom 8 as statistically significant whereas the bottom lengths of atom 16 and 24 are not significant. Furthermore, atom 24 shows a lower value than atom 16. These observations are reflected in the ROC analysis and the AUC values in Figure 9. The ROC curve in Figure 9a is located closer to



(a) atom 22 (x-position) (b) atom 22 (y-position) (c) atom 22 (z-position)

Fig. 8: ROC curves are visualized for (a) the x-position, (b) the y-position and (c) the z-position of atom 22 from the skeletal 3×8 sheet. The blue lines depict the ROC curves from the permutations and define an envelope. The red line depicts the ROC curve between the observed samples of two populations.

the boundary of the envelope than the ROC curve in Figure 9b, and again more than the ROC curve in Figure 9c.



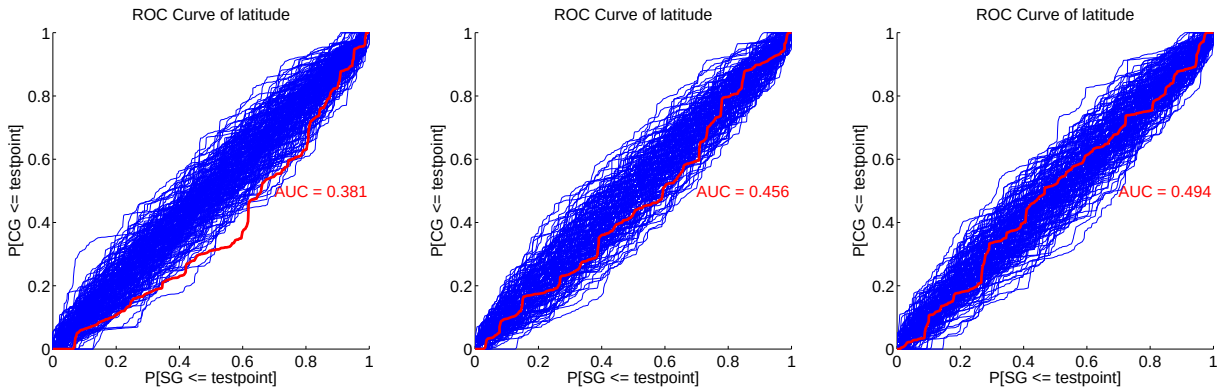
(a) atom 8 (bottom) (b) atom 16 (bottom) (c) atom 24 (bottom)

Fig. 9: ROC curves are visualized for the spoke lengths for (a) atom 8, (b) atom 16 and (c) atom 24 on the bottom side of the skeletal 3×8 sheet. The blue lines depict the ROC curves from the permutations and define an envelope. The red line depicts the ROC curve between the observed samples of two populations.

Figure 10 visualizes the ROC curve of the latitude spoke directions of atom 3 on the bottom, crest and top of skeletal sheet. Figure 7c indicates the latitude spoke direction of atom 3 on the bottom of the skeletal sheet as statistically significant whereas the latitude spoke direction on the crest and top are not significant. The box color of the top latitude spoke direction of atom 3 reflects a smaller value than the crest latitude spoke direction of atom 3. As above, all observations are reflected by the corresponding ROC curves in Figure 10.

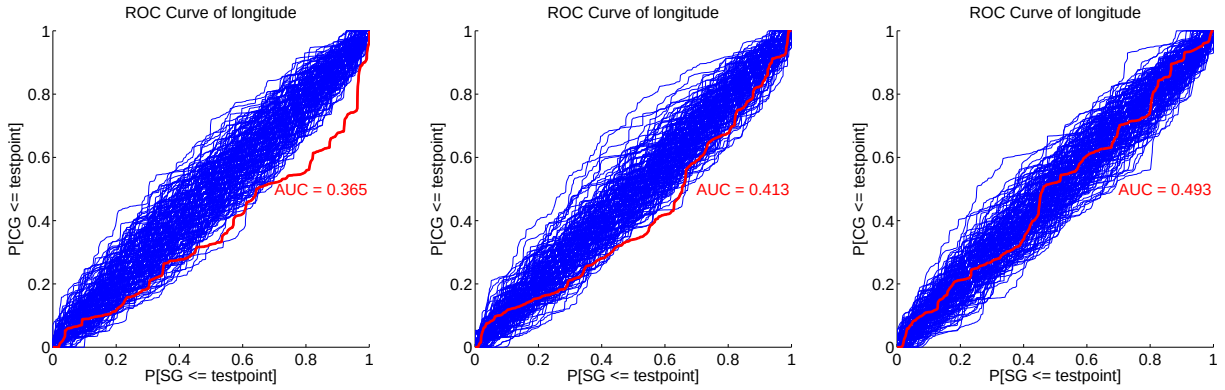
Finally, Figure 11 visualizes the ROC curve of the longitude spoke direction on the crest of atom 8, 16 and 24. Figure 7d indicates a statistically significant longitude spoke direction of atom 8 on the crest of the skeletal sheet whereas the longitude spoke direction on the crest of atom 16 and 24 are not significant. The color for atom 24 reflects a considerably smaller value than for atom 16. A comparison with Figure 11 confirms these observations. The ROC curve in Figure 11a is mostly located outside or close to the boundary of the envelope wheres the ROC curve of Figure 11c is close to the center of the envelope.

The observations described in this section verify the correctness of the feature-by-feature test results on the basis of selected GOPs. The ROC visualization of all 271 GOPs described by the distance measure d^2 was omitted for the purpose of clarity of this article .



(a) atom 6 (bottom) (b) atom 6 (crest) (c) atom 6 (top)

Fig. 10: ROC curves are visualized for the spoke latitude directions for atom 3 on (a) the bottom, (b) the crest and (c) the top of the skeletal 3×8 sheet. The blue lines depict the ROC curves from the permutations and define an envelope. The red line depicts the ROC curve between the observed samples of two populations.



(a) atom 8 (crest) (b) atom 16 (crest) (c) atom 24 (crest)

Fig. 11: ROC curves are visualized for the spoke longitude directions for (a) atom 8, (b) atom 16 and (c) atom 24 on the crest of the skeletal 3×8 sheet. The blue lines depict the ROC curves from the permutations and define an envelope. The red line depicts the ROC curve between the observed samples of two populations.

2.5 Test results for the unsigned difference measure d^1

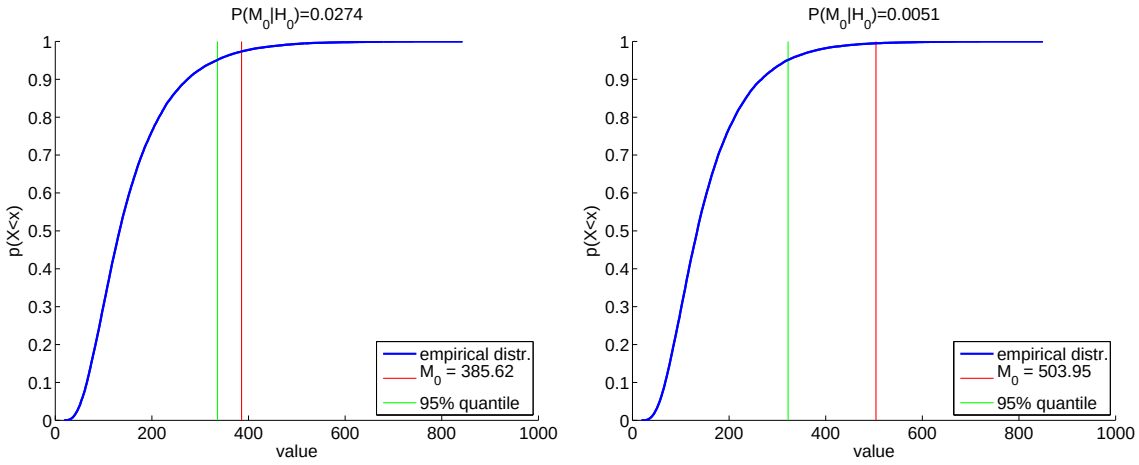
This section reports hypothesis test results using distances measure d^1 as described in Section 1.3. Results are based on the pre-processing methods PP1 and PP2 as described in Section 7.2 in the main article.

2.5.1 Global test results using d^1

Figure 12 shows the global test results for difference measures d^1 using PP1 and PP2. The global hypothesis of equal sample means is rejected and a statistical significant difference between the shape distribution of SG and CG is established ($p = 0.0274$ for PP1 and $p = 0.0051$ for PP2 with $p = P(M_0|H_0)$). These results correspond to the results using d^2 ($p = 0.0109$ for PP1 and $p = 0.0029$ for PP2) as presented in Section 7.2 in the main article. The larger p-values for d^1 are due to less information is being used for the unsigned differences, because the correlation structure between the GOPs was removed after the applied mapping to a full multivariate Gaussian as described in Section 1.3. Thus, results presented in the main article are quantified by the conservative test results in this section.

2.5.2 Single GOP test results using d^1

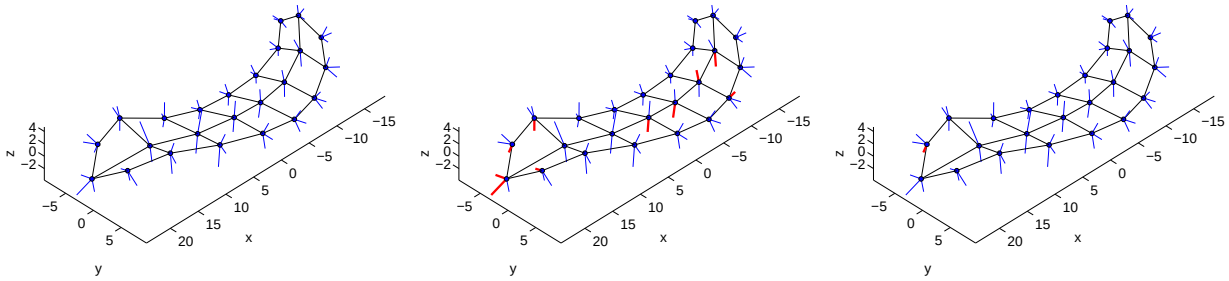
Figures 13 and 14 visualize the feature-by-feature test results for the difference measure d^1 using PP1. Recall that each discrete slabular s-rep is organized into 24 atoms by a 3×8 grid. Thereby, the measure d^1 (see



(a) PP1 (b) PP2

Fig. 12: Global test results using PP1 in (a) and PP2 in (b). The empirical distribution of $M_l, l = 1, \ldots, 30,000$ is shown together with M_0 and the 95% quantile of the empirical distribution.

Section 1.3) results in 157 GOPs with 24 GOPs corresponding to the skeletal position of each atom, 66 GOPs for the spoke directions (bottom, crest and top), 66 GOPs for the spoke lengths (bottom, crest and top) and 1 GOP for the global scaling factor. Figure 14 shows the magnitude of significance as described for Figure 7 in Section 2.4. The corrected threshold from the feature-by-feature test is $\lambda = 2.5532$.

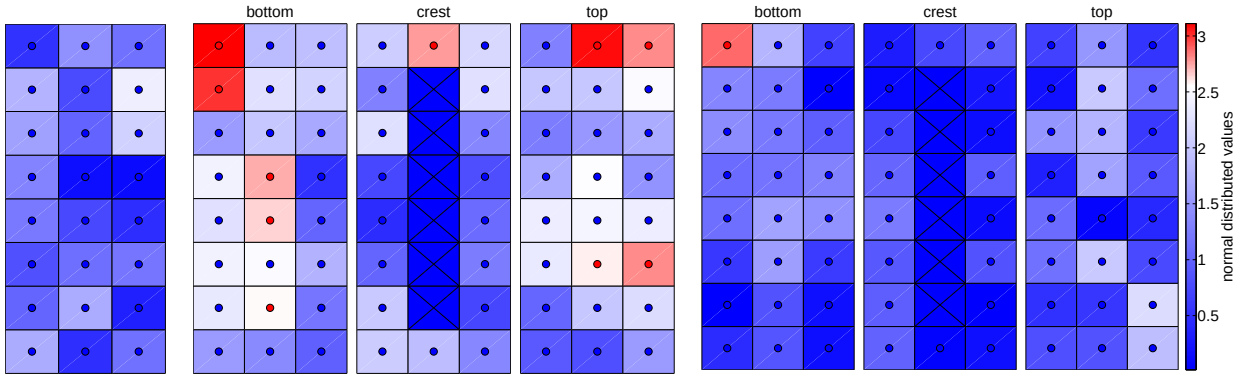


(a) skeletal position (b) spoke direction (c) spoke length

Fig. 13: Significant GOPs using PP1 and difference measure d^1 based on the 3×8 skeletal sheet of the SG CPNG mean. Test results are shown in (a) for the skeletal positions, in (b) for the spoke directions and in (c) for the spoke lengths. No skeletal position is statistically significant where non-significant skeletal positions are marked by small blue circles and significant skeletal positions are marked by large red circles. Similar, non-significant spoke directions and lengths are marked by small blue lines whereas significant spoke directions and lengths are marked by wide red lines.

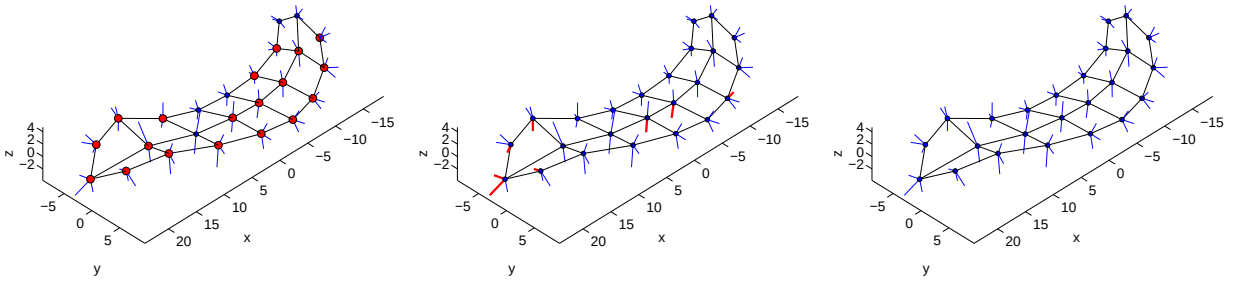
Figures 13 and 14 show several statistically significant GOPs. No skeletal position but one spoke length and 10 spoke directions are statistically significant. Moreover, the global scaling factor τ between SG and CG was found statistically significant by the GOP $|U_{0K}| = 2.7704$.

Figures 15 and 16 are identical to both previous figures except for the use of PP2 instead of PP1. Several skeletal positions are statistically significant in contrast to Figures 13a and 14a with no statistically significant skeletal position. The volume difference between the two populations is reflected by the skeletal positions using d^1 and PP2. Thus, Figures 15a and 16a show rather significant differences from a global deformation than from local deformations. Figures 14c and 16c show only small differences, which reveals that the global volume information is described by scaling of the skeletal grid. The spoke lengths are designed to capture only local differences whereas the skeletal position captures global scale differences. Similar results between spoke directions are expected because of the scaling invariance of $u_{ij} \in S^2$.



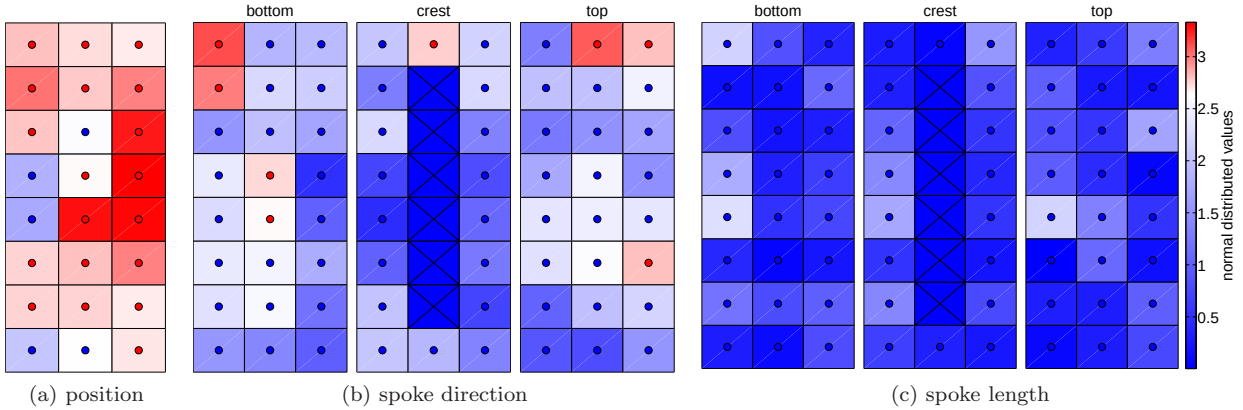
(a) position (b) spoke direction (c) spoke length

Fig. 14: Colored significant map of U_{0k} with a corrected threshold $\lambda = 2.5532$ using PP1 and difference measure d^1. Each box represents a GOP which correspond to a skeletal atom. The color map on the left side is non-linear and has a range from blue (not significant) to white (λ) to red (significant). The circle inside each box marks whether an U_{0k} is less or equal than the threshold λ (symbolized by blue) or if an U_{0k} is greater than the threshold λ (symbolized by red).



(a) skeletal position (b) spoke direction (c) spoke length

Fig. 15: As Figure 13, now based on PP2 and difference measure d^1.



(a) position (b) spoke direction (c) spoke length

Fig. 16: As Figure 14, now based on PP2 and difference measure d^1 with a corrected threshold $\lambda = 2.6368$.

A comparison of the results in this section with Section 7.3 in the main article leads to very similar observations and conclusions. Thereby, the results in the main article are quantified by the conservative test results presented in this section which not use the correlation structure between the GOPs (see Section 1.3). This is reflected by less significant GOPs, in particular for the spoke directions.

Using difference measure d^2 a significant volume difference was observed in the x and y-directions but not in the z-direction for the aligned hippocampi. Thus, we could obtain additional information using d^2 compared to d^1.

2.6 Asymptotic behavior of the global test for the two difference measures d^1 and d^2

This section will study the asymptotic behavior of the global test (described in Section 6.2 in the main article) for an increasing permutation size using PP1. The reported empirical p-values are 0.0274 for d^1 and 0.0109 for d^2 using $30,000$ permutations and given a significance level of $\alpha = 0.05$.

We have randomly selected subsets of $P = 500, 1000, 1500, 2000, 2500, \ldots, 29500$ from the set of 30000 permutations and applied the proposed testing procedure of Section 6.2 in the main article. Figure 17 visualizes the results and indicates a stabilization of the p-value from the global test after around $10,000$ permutations. Surprisingly, we observe a p-value equal to zero for a very small permutation size. This section will show the Mahalanobis space as the cause on the basis of distance measure d^1.

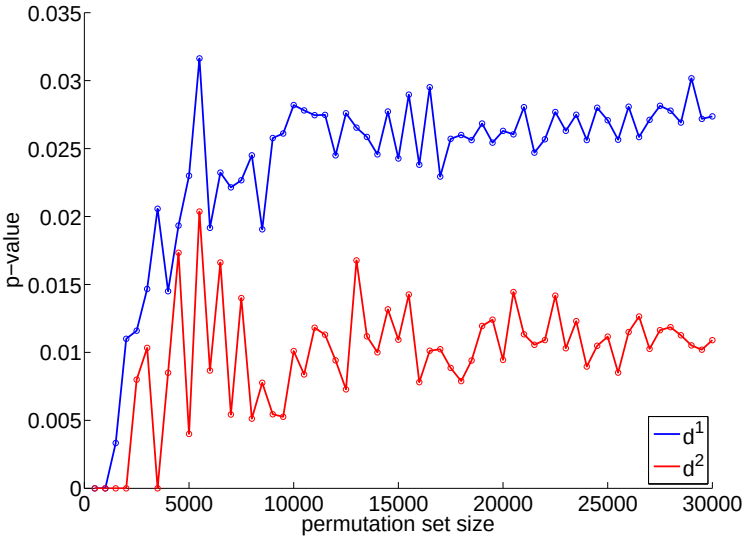


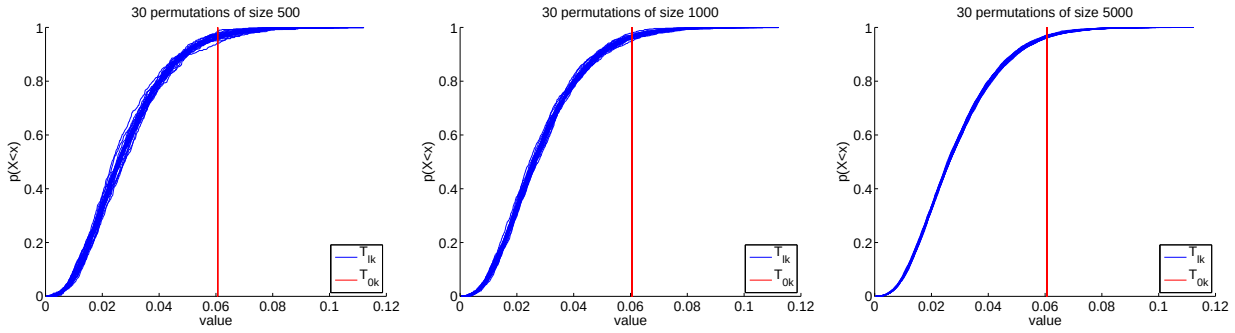
Fig. 17: The p-values are plotted against the number of permutations using difference measures d^1 and d^2. 30000 permutation were generated. The hypothesis test was calculated on randomly chosen subsets with $500, 1000, 1500, 2000, 2500, \ldots, 29500$ permutations.

In order to elaborate the convergence behavior of d^1, we have generated 30 random permutation sets with 500, 1000 and 5000 permutations for each permutation set. Afterwards, we applied the proposed testing procedure of Section 6.2 in the main article.

First, we calculated the difference measure $T_l = d^1(\mathbf{t}_{1l}, \mathbf{t}_{2l})$ (see Section 6.2.5 in the main article) between the s-reps $\mathbf{t}_{1l}$ and $\mathbf{t}_{2l}$, $l = 1, \ldots, P$ where P is the number of permutations. Each blue line in Figure 18 shows the cumulative empirical distribution for the chosen element $k = 22$ from the 157 dimensional GOP d^1-difference vector T_l. The selected element describes the atom position 22 from the 3×8 skeletal grid. Each plot contains 30 cumulative empirical distributions (blue lines) corresponding to each permutation set. We observe a higher variance of the envelope for a smaller permutation set size. $T_0 = d^1(\mathbf{t}_1, \mathbf{t}_2)$ is identical for all 30 permutation sets.

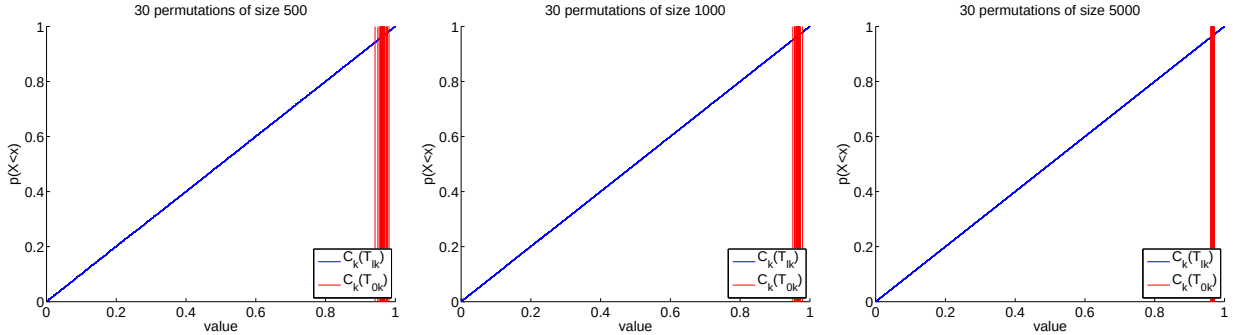
Afterwards, we estimated the empirical cumulative functions C_k for $k = 1, \ldots, K$ partial tests following to Section 6.2.5 in the main article. As a result, we obtained for each GOP difference a p-value $C_k(T_{lk})$, and $C_k(T_{0k})$ respectively. The cumulative empirical distribution of the calculated p-values are depicted in Figure 19. The p-values of the 30 permutations sets have by construction a uniform distribution. Therefore, no variance is visible between the blue line in Figures 19a-19c. However, we observe a larger variance of the red line for smaller permutation set size. The cumulative function C_k bases on the empirical distribution which shows larger variation for a smaller permutation set size in Figure 18. Therefore, the observed larger variance between $C_k(T_{0k})$ (red line) can be expected.

Subsequently, we calculated standard normal distributed variables from the uniformly distributed p-values by the inverse cumulative normal distribution function as described in the previous Section 1.3. Figure 20 visualizes the calculated standard normal distributed variables U_{lk} (blue) and U_{0k} (red). The blue and red lines show a larger variance for smaller permutation set size. However, the mean of T_{0k}, $C_k(T_{0k})$ and U_{0k} is similar for different permutation set size.



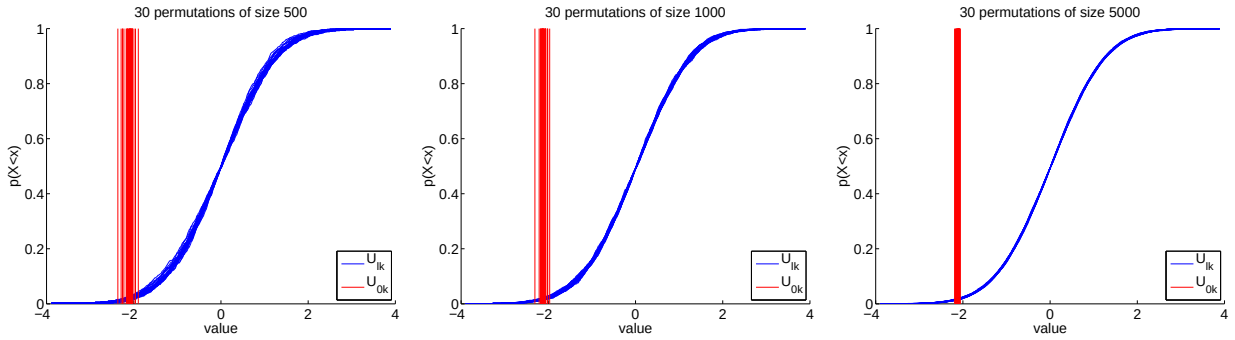
(a) 500 (b) 1000 (c) 5000

Fig. 18: The cumulative empirical distributions of GOP differences are depicted for a selected GOP using difference measure d^1. Each plot visualizes 30 random permutation sets of sizes 500, 1000 and 5000 (corresponding to 30 blue lines in each plot). The selected GOP is the atom position 22.



(a) 500 (b) 1000 (c) 5000

Fig. 19: The cumulative empirical distributions of the p-values $C_k(T_{lk})$ (blue) are depicted together with $C_k(T_{0k})$ (red) using difference measure d^1. Each plot visualizes 30 random permutation sets of sizes 500, 1000 and 5000. The selected GOP is the atom position 22.

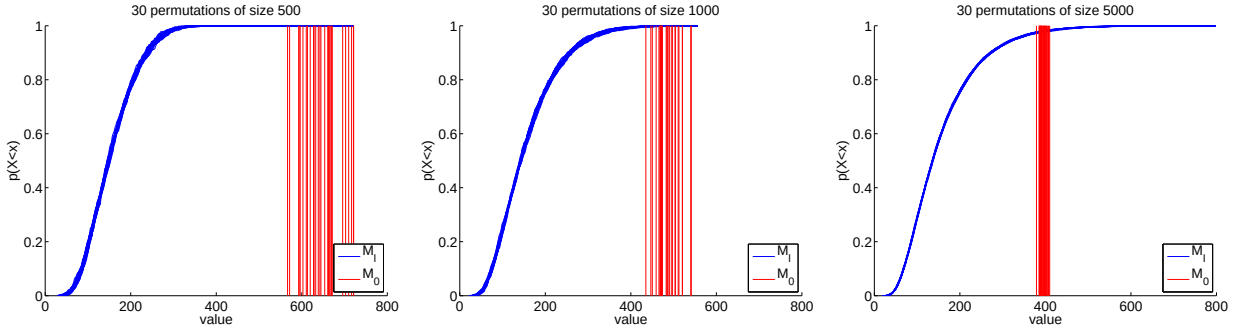


(a) 500 (b) 1000 (c) 5000

Fig. 20: The cumulative empirical distributions of the standard normal variables U_{lk} (blue) are visualized together with U_{0k} (red) using difference measure d^1. Each plot visualizes 30 random permutation sets of sizes 500, 1000 and 5000. The selected GOP is the atom position 22.

Finally, the p-values of the global tests were obtained by the estimation of the covariance matrix $\hat{\Sigma}_U$ from U_{lk} and the Mahalanobis distance as a combining function (see Section 6.2.6 in the main article). For each permutation $l = 1, \ldots, P$, we obtained the Mahalanobis distance M_l in addition to M_0 between the two populations SG and CG. Figure 21 shows the Mahalanobis distance for the three different permutation

set sizes. A smaller permutation set size strongly increase the variance of M_0. In addition, the blue curves indicate a smaller slope for higher permutation set size. In contrast to the previous figures, we observe a change in the mean value of M_0 with a larger value for smaller permutation set size. As a result, $p(M_0)$ is 0 (see equation (9) in the main article) using a small permutation set size such as 500 because $H(M_l, M_0) = 0$ for all $l = 1, \ldots, P$.



(a) 500 (b) 1000 (c) 5000

Fig. 21: Cumulative empirical distributions of Mahalanobis distances M_l (blue) are visualized together with M_0 (red) using difference measure d^1. Each plot visualizes 30 random permutation sets of sizes 500, 1000 and 5000. The selected GOP is the atom position 22.

Figures 18 to 21 and additional simulations on the covariance matrix found the covariance matrix as the reason for the convergence behavior in Figure 17. The Mahalanobis distance combines all GOPs to a corrected global test by the covariance matrix $\hat{\Sigma}_U$. A smaller permutation set size increases the magnitude of the elements of the covariance matrix, i.e., leads to a larger variance between the matrix elements of $\hat{\Sigma}_U$. As a result, the covariance matrix assigns different weights to the GOPs by the Mahalanobis distance.

Therefore, we recommend a permutation set size greater than $10,000$ for the proposed global hypothesis test. The study of an alternative combining functions for the global hypothesis test is left for future research.

3 Data analysis on an alternative group of final fittings

Besides the obtained final fittings using a joint shape distribution during the CPNG stage as described in Section 7.1 in the main article, we have generated a second group of final fittings derived from CPNG stages using a pooled shape distribution (FG1), two individual shape distributions (FG2) and two individual interchanged shape distributions (FG3). Interchanged shape distributions means the use of the estimated individual CG shape distribution for the re-fitting of the SG population during the CPNG stage, and the individual SG shape distribution for the re-fitting of the CG population. In each CPNG stage, the obtained backward mean was translational and rotational aligned to the data, i.e, the alignment of the CPNG backward mean of

1. $\bar{A}_1 \cup \bar{A}_2$ to the 221 and 56 CG cases for FG1,
2. $\bar{A}_1$ to the 221 SG cases and of $\bar{A}_2$ to the 56 CG cases for FG2,
3. $\bar{A}_2$ to the 221 SG cases and of $\bar{A}_1$ to the 56 CG cases for FG3.

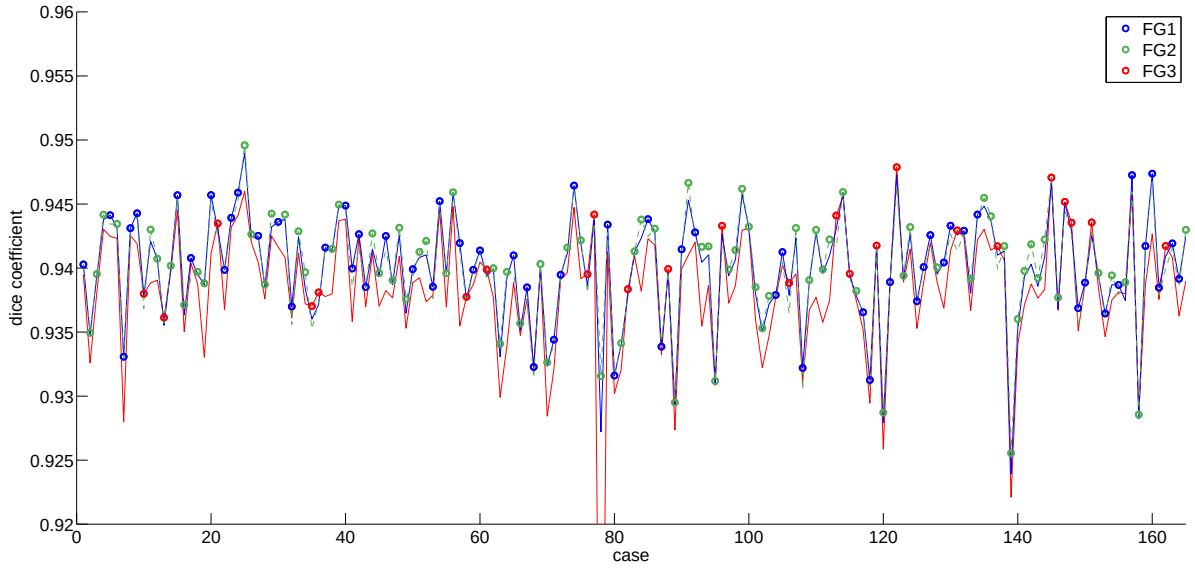
Afterwards, the means were optimized inside the CPNG shape space with an additional final spoke stage (see Section 5 in the main article). As a result, we obtained three fittings for each hippocampus. We chose the fitting with the largest Dice similarity coefficient. The Dice coefficient is a measure of the volume and was calculated between the original binary image B_1 and the binary image B_2 generated from each fitting. The coefficient is defined by

$$d_{vol}(B_1, B_2) = 2\frac{|B_1 \cap B_2|}{|B_1| + |B_2|} \tag{4}$$

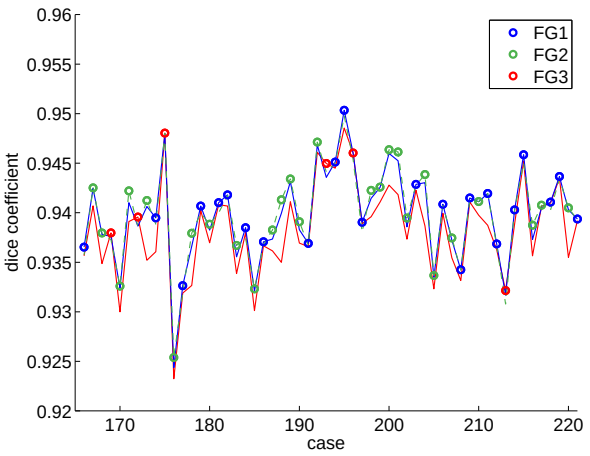
where $|\cdot|$ denotes the number of voxels that describe hippocampal tissue. Figure 22 shows the Dice coefficients of SG and CG for all three fitting types. Accordingly, the second group of final SG fittings consist of 84

fittings from FG1, 107 fittings from FG2 and 30 fittings from FG3. The second group of final CG fittings consist of 18 fittings from FG1, 21 fittings from FG2 and 17 fittings from FG3.

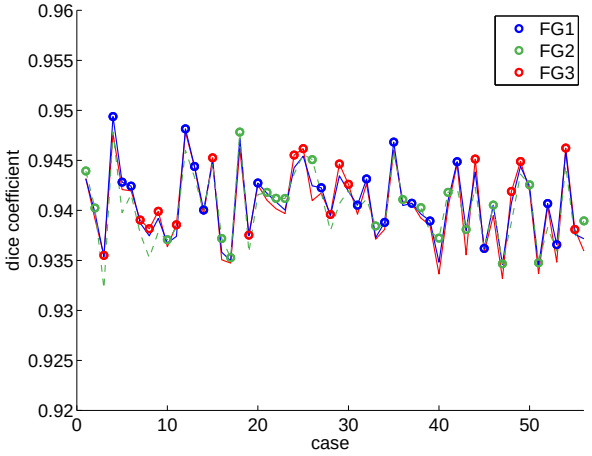
Figure 22 shows also an average volume overlap of 94% for both groups which indicates accurate fittings. We observe an outlier for case 73 of SG for FG3 due to a poor fitting result. The variance of the Dice coefficient is small for both groups. Nevertheless, a larger variance inside SG can be observed. Moreover, we can observe that FG1 and FG2 leads to a comparable Dice coefficient. The Dice coefficient of FG3 is inferior to FG1 and FG2 for SG but comparable for CG. There are two reasons for this observation. First, schizophrenia is a heterogeneous disease and also contains hippocampi variations between healthy patients. Therefore, the interchanged shape distribution from the schizophrenia cases can also describe the control cases. Second, both populations have an unbalanced size with a higher number of schizophrenics.



(a) SG (cases 1-165)



(b) SG (cases 166-221)

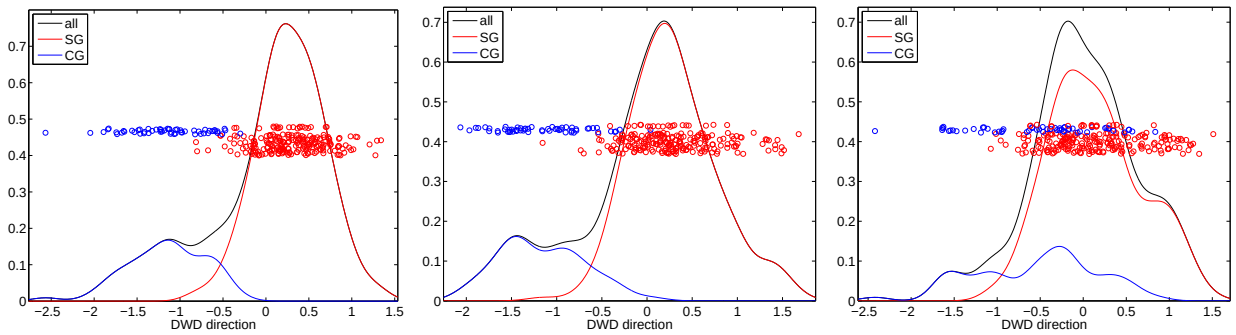


(c) CG

Fig. 22: Dice coefficient between the final fittings for (a-b) SG and (c) CG. The coefficient is depicted for the three types of obtained fittings by using a pooled shape distribution (FG1), two individual distributions (FG2) and two interchanged individual distributions (FG3) during the CPNG stage. The maximal Dice coefficient is depicted by a circle for each case colored by the corresponding class. The solid and dashed lines connect all points of the corresponding classes and depict the variance. SG shows larger variance than CG in correspondence with the heterogeneous character of the schizophrenia disease.

In addition to Figure 5 in the main article, Figure 23 shows the distribution of of SG and CG fittings obtained from (a) two individual distributions during the CPNG stage, (b) two interchanged individual

distributions and (c) of SG and CG fittings selected by the Dice criteria. The distributions are visualized by the projections of the CPNG score matrix Z_{Comp} on the DWD direction. Figures 23a and 23b show high separation properties between SG and CG. In contrast, a difference between the populations is not very strongly visible in Figure 23c which visualizes the second group of final fittings. The group is a compromise between independent fittings and a small bias as discussed in Section 7.1 in the main article.



(a) individual (b) interchanged (c) selected

Fig. 23: Jitterplot and KDEs show the distribution of SG and CG fittings projected onto the DWD direction. SG and CG fittings are obtained by using (a) two individual distributions during the CPNG stage, (b) two interchanged individual distributions during the CPNG stage and (c) by a selection of the final fittings using the Dice criteria. Additionally, the KDE of the pooled distribution of SG and CG is shown (all). A difference between the populations is visible for (a) and (b) but not very strong in (c).

The obtained second group of final fittings were used to test each of the hypotheses

$$H_0 : \{\mu_1 = \mu_2\} \text{ versus } H_1 \ : \ \{\mu_1 > \mu_2\} \qquad \text{(one-sided)} \tag{5}$$

for a one-sided test in case the difference measure is unsigned (e.g., d^1) and

$$H_0 : \{\mu_1 = \mu_2\} \text{ versus } H_1 \ : \ \{\mu_1 \neq \mu_2\} \qquad \text{(two-sided)} \tag{6}$$

for a two-sided test in case the difference measure is signed (e.g., d^2). The hypotheses are tested by the proposed global and feature-by-feature test in Section 6.2 in the main article at a significance level of $\alpha = 0.05$.

3.1 Global test results

Table 1 shows the global test results for the difference measures d^1 and d^2 for the two different pre-processing methods. Both difference measures rejected the hypothesis of equal population means and established a statistical significant difference between the two populations. In addition, DiProPerm results are reported in Table 1. All reported values are consistent with the results obtained from fittings using a pooled shape distribution, see Table 1 in the main article and Section 2.5 above. We observe an overall improved p-value in Table 1, particularly for the difference measure d^2. Thus, the second group of final fittings reveals an improved separation of the two populations, schizophrenics and controls.

3.2 Single GOP test results

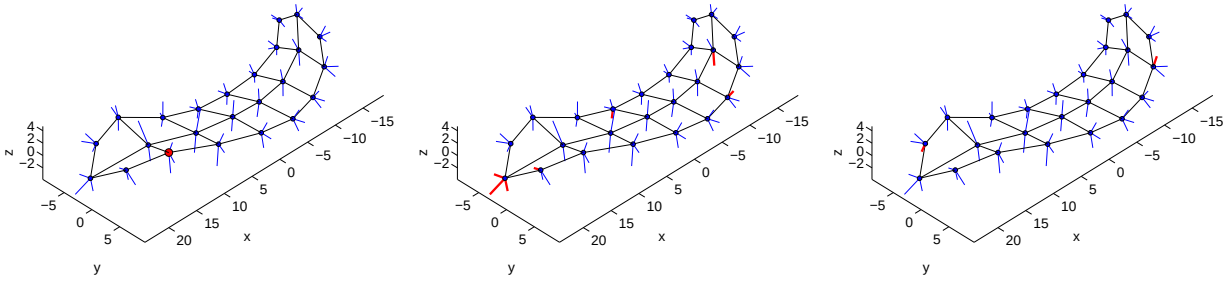
This section presents feature-by-feature test results for the two distance measures d^1 and d^2 using PP1. We have left out additional results for PP2 because neither additional informations nor conclusions would be added to this section.

Figures 24 and 25 visualize the feature-by-feature test results for the difference measure d^1 and correspond to Figures 13 and 14 above. The corrected threshold is $\lambda = 2.5632$. The measure d^1 results in 157 GOPs with

Table 1: Empirical p-value results using difference measures d^1 and d^2 for the proposed global hypothesis test in comparison with results obtained by DiProPerm. Two different pre-processing steps were applied: (PP1) Full Procrustes alignment with scaling. (PP2) Full Procrustes alignment without scaling. Three different projection directions were used for DiProPerm.

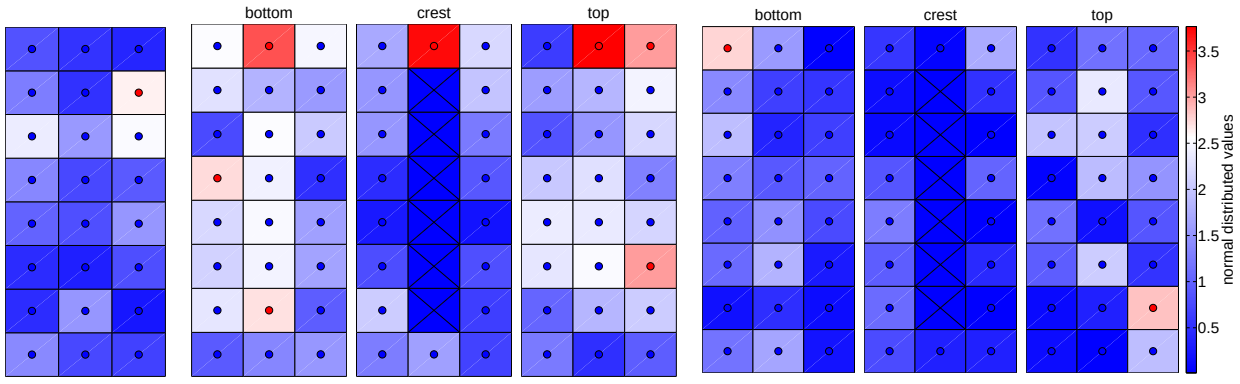
method	empirical p-value PP1	PP2
Mahalanobis distance		
difference measure d^1	0.0245	0.0043
difference measure d^2	0.0013	0.0009
DiProPerm using MD-statistic		
DWD direction vector	0.0018	0.0011
SVM direction vector	0.0039	0.0051

24 GOPs corresponding to the skeletal position of each atom, 66 GOPs for the spoke directions (bottom, crest and top), 66 GOPs for the spoke lengths (bottom, crest and top) and one GOP for the global scaling factor. Figure 24 and 25 show statistically significant GOPs. One skeletal position, two spoke lengths and 7 spoke directions are statistically significant compared to Figure 13 above where no skeletal position but one spoke length and 10 spoke directions are statistically significant. Moreover, the global scaling factor τ between SG and CG was found statistically significant.



(a) skeletal position (b) spoke direction (c) spoke length

Fig. 24: As Figure 13, now based on PP1, difference measure d^1 and the alternative group final fittings.



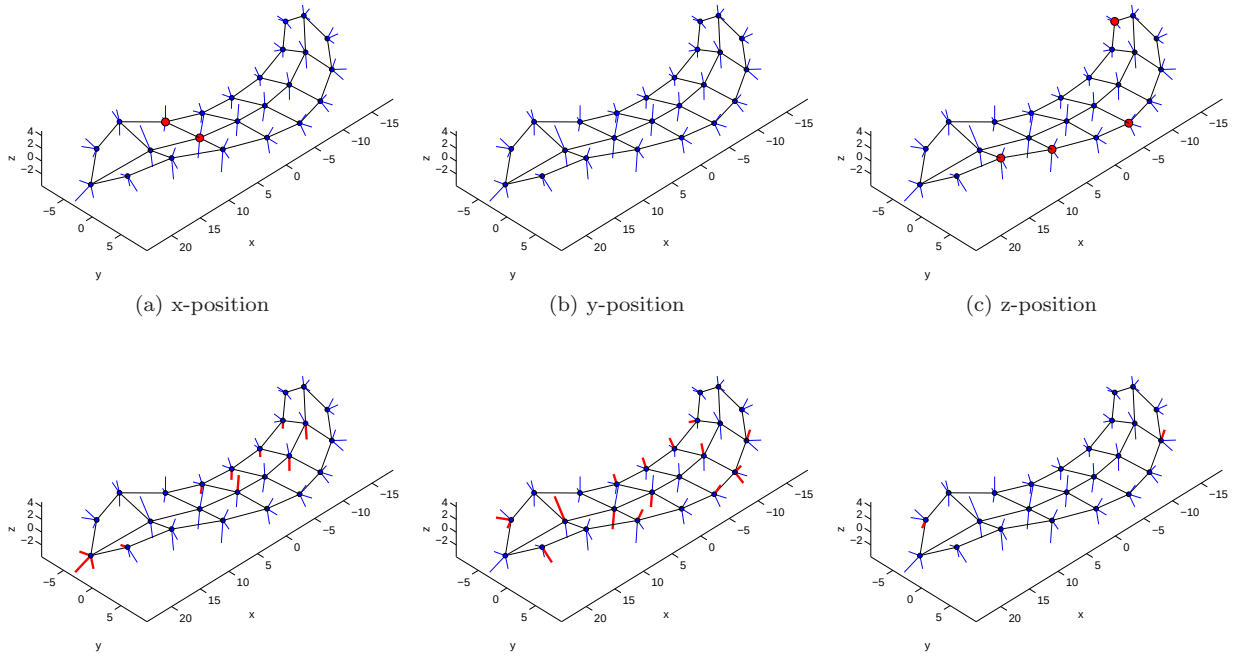
(a) position (b) spoke direction (c) spoke length

Fig. 25: As Figure 14, now based on PP1, difference measure d^1 and the alternative group final fittings with a corrected threshold $\lambda = 2.5632$.

Figure 25 shows the magnitude of significance as described for Figure 7 in the previous Section 2.4. The corrected threshold from the feature-by-feature test is $\lambda = 2.5632$. The GOP $|U_{0K}| = 2.7388$ is statistically significant where the index K corresponds to the global scale factor τ. A comparison of Figure 25 with

Figure 14 above shows a very similar pattern between the colored significant maps except the pattern between the bottom spoke directions. In the previous Figure 14, we observe two significant atoms 7 and 8 (top right of the skeletal sheet) and two significant atoms 12 and 13 (center middle) which are not significant in Figure 25. A detailed interpretation of this observation is left as an open question for the future. However, the second group of the final s-reps reflects tighter fittings based on the Dice coefficient. Therefore, the two populations are better separated, which decreases noise artifacts and yields a larger threshold $\lambda = 2.5632$ compared to $\lambda = 2.5532$ in Section 2.5.2.

Figures 26 and 27 visualize the feature-by-feature test results for the difference measure d^2 and correspond to Figures 7 and 8 in the main article. The measure d^2 results in 271 GOPs with 72 GOPs corresponding to the skeletal position of each atom (x, y and z-position), 66 GOPs for the latitude spoke directions (bottom, crest and top), 66 GOPs for the longitude spoke directions (bottom, crest and top), 66 GOPs for the spoke lengths (bottom, crest and top) and one GOP for the global scaling factor. The corrected threshold is $\lambda = 2.5214$. Figure 26 and 27 show statistically significant GOPs. Two skeletal x-positions, no y-position, 4 z-positions, one bottom, no crest and one top spoke lengths, 7 bottom, one crest and three top latitude spoke directions, 5 bottom, two crest and 9 top latitude spoke directions are statistically significant. Moreover, the GOP $|U_{0K}|$ is 2.7198 and is statistically significant where the index K corresponds to the global scale factor τ.



(a) x-position (b) y-position (c) z-position

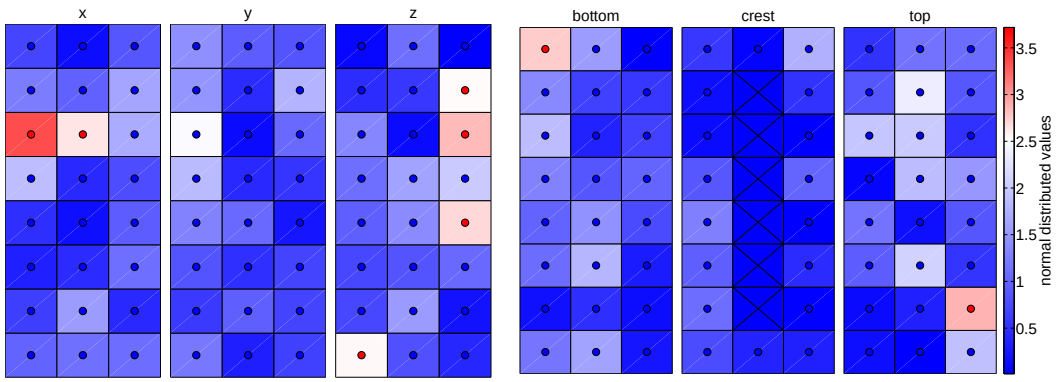
(d) latitude direction (e) longitude direction (f) spoke length

Fig. 26: As Figure 13, now based on PP1, difference measure d^2 and the alternative group final fittings.

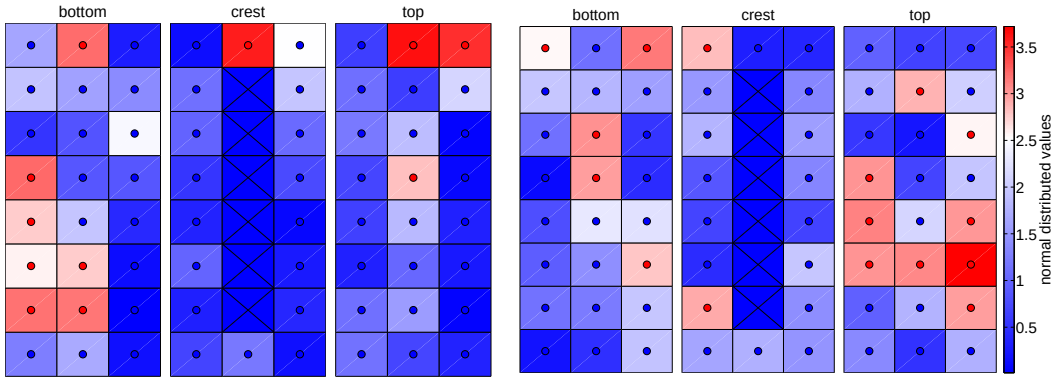
As before, a comparison of Figure 27 with Figure 8 in the main article shows a very similar pattern between the colored significant maps. The lower color intensity for several boxes in Figure 27 is due to a larger threshold $\lambda = 2.5214$ compared to $\lambda = 2.2917$ in the main article.

3.3 Conclusion

The additional data analysis by the second group of final fittings in this section confirms the results and conclusions of the main article and Section 2 above. The global test results establish smaller p-values compared to the results from the first group of final fittings. This indicates a better separation of the two populations by the second group of final fittings. The feature-by-feature test show similar patterns between the colored significant maps and demonstrate therewith the sensitivity of the proposed test in the case of less separated fittings.



(a) skeletal position (b) spoke length



(c) latitude spoke direction (d) longitude spoke direction

Fig. 27: As Figure 14, now based on PP1, difference measure d^2 and the alternative group final fittings with a corrected threshold $\lambda = 2.5214$.

References

1. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. John Wiley & Sons, Chichester (1998)
2. Jung, S., Dryden, I.L., Marron, J.S.: Analysis of principal nested spheres. Biometrika **99**(3), 551–568 (2012)
3. Marron, J.S., Todd, M.J., Ahn, J.: Distance weighted discrimination. J. Amer. Statist. Assoc. **102**(480), 1267–1271 (2007)
4. Niethammer, M., Juttukonda, M.R., Pizer, S.M., Saboo, R.R.: Anti-aliasing slice-segmented medical images via Laplacian of curvature flow. In preparation (2013)
5. Nitrc: S-rep fitting, statistics, and segmentation. http://www.nitrc.org/projects/sreps (2013)
6. Pizer, S.M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J.N., Huckemann, S., Marron, J.S.: Nested sphere statistics of skeletal models. In: Innovations for Shape Analysis: Models and Algorithms, Lecture Notes in Comput. Sci., pp. 93–115. Springer (2013)
7. Qiao, X., Zhang, H.H., Liu, Y., Todd, M.J., Marron, J.S.: Weighted distance weighted discrimination and its asymptotic properties. J. Amer. Statist. Assoc. **105**(489), 401–414 (2010)