



Non-Euclidean classification of medically imaged objects via s-reps[☆]



Junpyo Hong^{a,*}, Jared Vicory^a, Jörn Schulz^d, Martin Styner^{a,c}, J.S. Marron^b,
Stephen M. Pizer^a

^a Department of Computer Science, University of North Carolina at Chapel Hill, USA

^b Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, USA

^c Department of Psychiatry, University of North Carolina at Chapel Hill, USA

^d Stavanger University Hospital, Norway

ARTICLE INFO

Article history:

Received 6 August 2014

Revised 7 January 2016

Accepted 28 January 2016

Available online 19 February 2016

Keywords:

Pattern classification

Statistical analysis

Shape analysis

ABSTRACT

Classifying medically imaged objects, e.g., into diseased and normal classes, has been one of the important goals in medical imaging. We propose a novel classification scheme that uses a skeletal representation to provide rich non-Euclidean geometric object properties. Our statistical method combines distance weighted discrimination (DWD) with a carefully chosen Euclideanization which takes full advantage of the geometry of the manifold on which these non-Euclidean geometric object properties (GOPs) live. Our method is evaluated via the task of classifying 3D hippocampi between schizophrenics and healthy controls. We address three central questions. 1) Does adding shape features increase discriminative power over the more standard classification based only on global volume? 2) If so, does our skeletal representation provide greater discriminative power than a conventional boundary point distribution model (PDM)? 3) Especially, is Euclideanization of non-Euclidean shape properties important in achieving high discriminative power? Measuring the capability of a method in terms of area under the receiver operator characteristic (ROC) curve, we show that our proposed method achieves strongly better classification than both the classification method based on global volume alone and the s-rep-based classification method without proper Euclideanization of non-Euclidean GOPs. We show classification using Euclideanized s-reps is also superior to classification using PDMs, whether the PDMs are first Euclideanized or not. We also show improved performance with Euclideanized boundary PDMs over non-linear boundary PDMs. This demonstrates the benefit that proper Euclideanization of non-Euclidean GOPs brings not only to s-rep-based classification but also to PDM-based classification.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Binary classification of objects of interest based on medical imaging has been a common objective (e.g., Kurtek et al., 2011; Gorczowski et al., 2010; Zhao et al., 2014). Researchers often wish to classify whether a subject has a disease or not based on geometric features of an anatomical structure from a medical image. Beyond simply providing a rule for classification is the desire to gain deeper scientific insights into phenomena underlying the disease.

These geometric features are often provided by shape representations and should be analyzed by statistical methods suitable for shapes. One of the most popular forms of shape representation is the *Point Distribution Model* (PDM) (e.g.,

Cootes et al., 1995; Styner et al., 2006; Davies et al., 2003). A boundary PDM is a tuple of boundary points on an object, with points corresponding across the training cases. Frequently, studies using PDMs capture shape variations through the statistical method of *Principal Component Analysis* (PCA) (Cootes et al., 1992; 1995), and classification is done using *Linear Discriminant Analysis* (LDA) or the *Support Vector Machine* (SVM) (Davies et al., 2003).

In this paper we investigate the possible improvements in classification that can arise from two modifications in the above method. The first is to statistically analyze the object representation data in the realization that, per (Kendall, 1984), even PDMs can be understood as lying on a curved manifold. We apply the method called *Principal Nested Spheres* (PNS) (Jung et al., 2012) for this purpose.

The second modification we consider is to augment the discrete positional features in a PDM by boundary directional features and object width features at discrete points. We show that this results in a more complicated curved manifold that can be statistically

[☆] This paper was recommended for publication by Dr. James Duncan.

* Corresponding author. Tel.: +1 9198020255.

E-mail address: jphong@cs.unc.edu (J. Hong).

analyzed by PNS. We recognize that there are other object representations and associated means of analysis that could be compared, but we leave those to future work.

The object representation we investigate that has *geometric object properties* (GOPs) that consist of not only positions but also directions and widths is the skeletal representation called the s-rep. We and others (Styner et al., 2004; Yushkevich and Zhang, 2013; Bouix et al., 2005) have found skeletal representations particularly effective for shape analysis. The discrete s-rep is a skeletal representation designed to combine tightness of fitting to the object segmentation with simplicity and stability of branching topology. The s-rep's directions lie on abstract spheres.

The method of analysis we propose is *distance weighted discrimination* (DWD) (Marron et al., 2007) on GOPs that are Euclideanized using PNS. We demonstrate that, both with PDMs and with s-reps, this statistical method produces more effective classification than those making less use of the geometry of the manifold in which the representation lies.

We apply our method to the problem of classifying 3D hippocampi as schizophrenic or healthy based on their GOPs. We have evaluated our method on a dataset that consists of 221 schizophrenic cases and 56 healthy control cases (McClure et al., 2013). In this application, we measured performances of methods by calculating area under the ROC curve (AUC). The results show that our proposed method on s-reps is superior, with non-overlapping confidence intervals, to

- the classification based on s-reps without Euclideanization.
- the classification based on volume, as is common in the neuroscience literature.
- the classification based on boundary PDMs with and without Euclideanization; also, the PDM-based classification with Euclideanization is shown superior to the PDM-based classification without Euclideanization.

This paper is organized as follows. Section 2 presents object representations and statistical methods used by others for classification as well as those used by us. Section 3 describes the hippocampi dataset. Section 4 describes our classification method. Section 5 presents the experimental analysis approach we have used. Section 6 gives the experimental results, and Section 7 discusses those results and draws conclusions.

2. Background

This section provides background information necessary to understand our method. We also briefly overview conventional shape representations, statistical analysis techniques, and classification methods.

2.1. Object model

At a high level there are two categories of object models that have been proposed for statistical analysis: continuous, parameterized models modulo parameterization (Kurtek et al., 2012; Jermyn et al., 2012; Bauer et al., 2010, 2012; Durrleman et al., 2014) and discrete models. Due to the discrete models' strengths in explicitly dealing with localized features, we focus on those models. Among the discrete models are those based on deformations of an atlas (Beg et al., 2005; Miller et al., 2002; Wang et al., 2007), those based on boundary PDMs (Cootes et al., 1995; Styner et al., 2006; Davies et al., 2003), and those based on skeletal models (Styner et al., 2004; Yushkevich and Zhang, 2013; Bouix et al., 2005; Schulz et al., 2013b). The PDM-based models have been the most popular. The skeletal models were designed to add local object width features and local directional features to those provided by PDMs.

We overview the two shape representations that we compare: PDM and s-rep. For each representation, we provide

- a brief descriptions of the representation.
- the dimensionality of the representation.
- the method used to capture modes of variation given a set of observations.

2.1.1. Point distribution model

A PDM is a point tuple for each object in a training set of example objects. In a boundary PDM each example object in the set has a set of enumerated points along its boundary, with points with corresponding index in each object chosen so as to be in correspondence across the training set. The training set is automatically aligned so that all the examples lie in the same coordinate system. Then, it models average shape by taking means on the positions over the set of example objects. It can also model allowed shape variation via a number of modes of variation.

Consider a boundary PDM in the training set p with n boundary points. By scaling the entire point tuple such that the sum of squares of all the center-of-mass-relative point features has unit length, we can think of this as projection onto the unit hypersphere \mathbb{S}^{3n-4} . The dimensionality of $3n-4$ comes from the fact that we have used three degrees of freedom during alignment and one more degree of freedom in normalizing scale to unity. Therefore, as rigorously shown by (Kendall, 1984), a boundary PDM can be represented as a concatenation of this scaling factor and this normalized tuple of points; we can say that a boundary PDM abstractly lives on the manifold $\mathbb{R}^+ \times \mathbb{S}^{3n-4}$. The modes of variation are captured through a Principal Component Analysis (PCA)-like procedure. Although direct use of PCA is common, after the scaling Kendall's approach places the PDM on an abstract sphere. PCA is designed to analyze data on Euclidean space, so a variant of PCA that is designed to analyze data part of which is on a sphere is more appropriate (Kendall, 1984; Dryden and Mardia, 1998), though direct application of PCA to the non-scaled-normalized point features is more common.

Since the PDM in question represents points along the boundary, its PCA-like analysis provides no information about the object interior. Moreover, it does not directly represent local directional information or local object width information.

2.1.2. S-rep

A discrete s-rep is a skeletal discretization of the interior of the object. It consists of a grid of samples of the skeletal surface (which is an approximately medial surface) and, at each of these samples, vectors called spokes pointing from the skeletal surface to the object's boundary which are approximately normal to the boundary surface. These spokes explicitly capture local direction and local width information. Also, the spoke ends form a boundary PDM.

The number of these sample spokes is chosen to be the minimum to achieve a desired level of accuracy of each training object's boundary implied by the continuous s-rep interpolated from the discrete spokes by Vicory's work in (Tu et al., 2015c) as compared to the input object boundary from the image data. An example discrete s-rep of a hippocampus can be seen in Fig. 1.

For each case in the provided image data the initial set of s-reps are fitted by solving an optimization problem based on criteria including the following: no spokes are allowed to cross each other, grid sample points are approximately regularly spaced, spoke ends touch the object surface, spoke directions are approximately normal to the tangent object surface, and the 3-spoke assembly (magenta, red, and cyan in Fig. 1(a)) at each exterior skeletal point fits across the high curvature locus called the crest of the object (Koenderink, 1990).

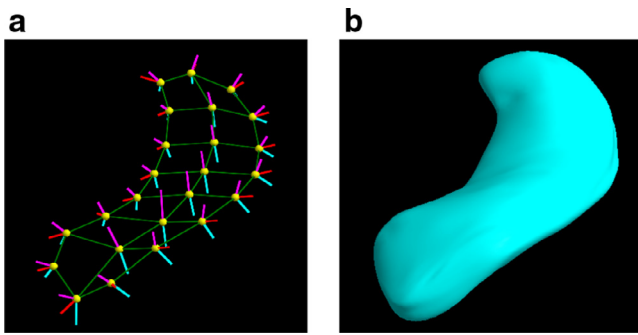


Fig. 1. (a) Skeletal model of a hippocampus s-rep; (b) solid model implied by that s-rep. Yellow spheres are sample points along the skeletal surface. Solid lines extending from these sample points are spoke vectors, which are approximately normal to the boundary surface. Interpolation of a discrete s-rep into a continuous skeleton with a continuous field of spokes forms a continuous s-rep whose spokes completely fill the interior of the object they are representing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Given such an initial set of s-reps, we want each spoke vector to be in correspondence across the training set. This is achieved through an iterative optimization process that involves repeating the following three steps for each iteration.

1. Extracting shape statistics of the current set of s-reps, i.e., mean shape and modes of variation.
2. Optimizing each case in the current set over modes of variation.
3. Extending or shortening each spoke to tighten the fit of the implied boundary to the boundary of the object data.

While this method provides repeatable models for a given training set of input boundaries as well as good correspondence of spokes across the training cases, separate work mentioned in Section 7 can lead to improved correspondence.

Consider a discrete s-rep \underline{s} with n spoke vectors and m grid sample points on the skeletal surface. The set of sample skeletal points forms a PDM that is aligned such that its center of gravity is at the origin. Additionally, this tuple of centered points is scaled by a factor making the sum of squared distances to the origin to be unity. Therefore, this PDM is described by a tuple of centered points that abstractly lives on the unit hypersphere \mathbb{S}^{3m-4} and an associated log-transformed scaling factor. The directional component of each spoke abstractly lives on the unit 2-sphere \mathbb{S}^2 , and the log-transformed associated length component of each spoke lives on the Euclidean space \mathbb{R}^1 . Thus, a single discrete s-rep abstractly lives on $\mathbb{R}^{n+1} \times \mathbb{S}^{3m-4} \times (\mathbb{S}^2)^n$. In our hippocampal dataset each discrete s-rep has 24 skeletal sample points and 66 spokes, putting the s-reps in our dataset on $\mathbb{R}^{67} \times \mathbb{S}^{68} \times (\mathbb{S}^2)^{66}$.

As described in detail in Section 2.2.3, modes of variation of s-reps are captured via *Composite Principal Nested Spheres (CPNS)* (Jung et al., 2010b), a PCA-like method used to analyze data some features of which do not live in a flat Euclidean space but on spheres. Here these features are the spoke directions present in an s-rep and the scaled tuple of skeletal points. Indeed, CPNS has been shown to be appropriate for analysis of PDMs, as well (Jung et al., 2010a). For more information on s-reps and CPNS, see (Pizer et al., 2012).

2.2. Statistical methods to capture data's variation

We provide brief descriptions of statistical analysis techniques used to capture underlying modes of variation of the input data. We first overview PCA, the conventional approach. Then, we briefly overview PNS analysis, a variant of PCA to analyze data that live on abstract spheres. Finally, we briefly describe CPNS, a statistical

analysis technique that is appropriate for analyzing the data that live on a Cartesian product of Euclidean space and hyperspheres.

2.2.1. Principal component analysis

Principal Component Analysis (PCA) has been an important statistical method for analyzing data. It provides a means of reducing the intrinsic dimension of data by capturing its major modes of variation. PCA has been widely used in the field of medical image analysis and computer vision because descriptions of objects of interest are often high dimensional, whereas the important variations can be quite low dimensional. Those modes of variation are often quite illuminating. PCA can be understood in terms of a forward or backward procedure. In a forward method you progressively build up the dimension of the approximating subspace being fitted to the data, whereas in a backward method you progressively reduce the dimension of the subspace being fitted to the data.

Both approaches yield the same result if the data lie on a Euclidean vector space. However, many shape features do not lie on a Euclidean space. The backward approach typically yields different results from the forward approach when applied to non-Euclidean data. As noted in (Damon and Marron, 2014), the backward approach is usually more appropriate to analyze those non-Euclidean features.

Forward PCA increases dimension by adding the component that captures the most remaining variance; at each iteration a component that best describes the data and that is orthogonal to previous components is added to form a new best fitting manifold so that the current manifold is the best fitting submanifold of the data in the original dimension. The principal component scores are found by projecting all the data onto the found submanifold.

In contrast, the backward view of PCA progressively reduces the intrinsic dimension of the manifold by removing the component of the least variance from all the data points; at the beginning of each iteration the data is projected onto the submanifold found in the previous iteration, and then the best fitting submanifold is found by minimizing the sum of squared distances of all the projected data.

2.2.2. Principal nested spheres

Principal Nested Spheres (PNS) analysis is a special case of backward PCA on hyperspheres. PNS progressively reduces intrinsic dimension by finding the best fitting subsphere \mathbb{S}^{k-1} that is nested in the current hypersphere \mathbb{S}^k . At each iteration, the data points are first projected onto the subsphere found in the previous iteration; then the fitting is done by minimizing the sum of squared geodesic distances of all the projected data points to the subsphere. Over the training cases PNS will yield a tuple of signed geodesic distances to the best fitting subsphere for each dimension-reduction iteration. As long as the commonly satisfied criterion that the projected data points are much closer to the fitted subsphere than to the poles of that subsphere holds, these signed geodesic distances provide an appropriate Euclideanized form of their spherical counterparts. The final result of PNS yields Euclideanized variables and a set of geodesic polar systems that provide a means of transformations between the original space and Euclideanized space and vice versa. The dimension 0 point in feature space produced at the end of this iteration is called the backwards mean. (Jung et al., 2012) provides more information on the method.

2.2.3. Composite principal nested sphere analysis

Suppose the data of interest live on a Cartesian product of a Euclidean vector space and hyperspheres. Such an instance includes any model described by a combination of GOPs involving PDMs, lengths, directions, and scaling. In this case, PNS is applied independently to each GOP that lives on a hypersphere. As noted in the previous subsection, each application of PNS on spherical

GOPs produce their Euclidean counterparts. Then we apply conventional PCA on the matrix of Euclideanized values concatenated with the already Euclidean components. To make the components appropriately commensurate (Jung et al., 2010b) when analyzing shape variations of s-reps, we multiply each Euclideanized value derived from a PDM by the geometric mean of the scale factors in the training population, and we multiply each Euclideanized value derived from a direction by the geometric mean of its associated length.

2.3. Classification methods

We briefly describe two binary classification methods: SVM and DWD. We concentrate on linear classification methods because this framework is easier for scientists to gain insights from studying features. We especially pay attention to the separating direction vector in the feature space pointing from one class to the other. Large entries in this vector indicate that the corresponding feature is relevant. A good separating direction provides additional information and insight into the data by visualizing the trends between the classes by linearly interpolating and synthesizing the data in the original feature space along the direction.

2.3.1. Support vector machine

SVM (Cortes and Vapnik, 1995) is a binary classification method that yields a separation direction in the feature space. SVM then classifies a new example by thresholding the scalar value of the projection of its feature tuple onto this direction.

2.3.2. Distance weighted discrimination

DWD is a classification method similar to SVM but which is more robust to noise and limited sample size. Like SVM, DWD takes in two classes of data and yields a separating direction that can be used to classify new data points through projection and thresholding. Unlike SVM, the separating direction computed by DWD is influenced by all points in the data set. A full description of DWD can be found in (Marron et al., 2007).

3. Materials

In this work, we study the problem of classifying hippocampi as schizophrenic or healthy. We have chosen to use the s-rep to represent hippocampi; we will show later in Section 6 that rich geometric features such as directions provided by the s-rep proved to be important discriminating features between schizophrenic hippocampi and healthy hippocampi. In the original study, 238 schizophrenics and 56 healthy controls were enrolled. High resolution Magnetic Resonance Imaging (MRI) scans (multi-site SPGR T1 weighted imaging on 1.5 T scanner at $0.9375 \times 0.9375 \times 1.5 \text{ mm}^3$ voxel resolution) were performed on the subjects. The MRI scans were rigidly aligned to a common coordinate system prior to the segmentation to account for variations in sensor field of view and magnetic field. The hippocampi were automatically segmented from the aligned MRI scans. Then segmented hippocampi were positionally and rotationally aligned. In the data provided, the hippocampi had been normalized in volume with the original volumes provided as a separate scaling feature. Details on the original MRI hippocampi dataset can be found in (McClure et al., 2013), and those on the segmentation method can be found in (Gouttard et al., 2007).

We have chosen to analyze the shape of the left hippocampus in this study because that was the data available. The choice of left versus right hippocampus would not affect the finding as there is no biological correlation between the sidedness of the hippocampus and schizophrenia. Moreover, records of the left hippocampus were

not available for 17 patients from the schizophrenia group. Therefore, the dataset consists of 221 schizophrenia cases and 56 control cases.

A set of s-reps fitted to these MRIs were provided to us. S-reps were fitted using shape statistics drawn from the set where both schizophrenic cases and control cases were pooled together. Detailed description of the actual s-rep fitting procedure can be found in (Schulz et al., 2013b; Merck et al., 2008).

4. Method

The novelty of our classification method comes from the fact that we recognize that some GOPs are not Euclidean and that we appropriately take that into account during classification. Our classification method works as follows.

1. Applying PNS to Euclideanize GOPs that live on a sphere and commensurating those features to millimeters.
2. Learning the separation direction from these features concatenated with the originally Euclidean features in the training data using DWD.
3. Computing the function that maps from values projected onto the separation direction to the probability of belonging to the schizophrenic group based on Bayes' Theorem (Fig. 2).
4. Classifying each case in the test set based on the probabilities computed using the function from the previous step.

In this particular classification problem, positive examples are s-reps from the schizophrenic group and negative examples are s-reps from the control group. In the following subsections, we provide detailed description for each step.

4.1. Euclideanization of s-reps and basis of the transformation between s-rep space and Euclidean space

As we have noted in Section 2.1.2, a discrete s-rep has some spherical GOPs, i.e., each spoke's direction and the PDM formed by its skeletal sample points. We apply PNS separately to each spherical GOP, producing corresponding Euclideanized variables. This is consistent with the shape statistics used in fitting, namely modes of variation calculated using PNS.

We considered both great subspheres and small subspheres at each iteration of PNS to Euclideanize spherical GOPs of the representation. Hypothesis testing was performed to decide which subspheres to use at each iteration of PNS. Supplementary material of (Jung et al., 2012) provides details on the hypothesis testing. Along with the Euclideanized variables, PNS yields a polar system to be used as the basis of a transformation between the original s-rep space and the corresponding Euclidean space, in both directions.

We concatenate the already Euclidean and Euclideanized variables and scale each so that they are commensurate. These variables form the feature space on which classifiers are trained and tested. We denote these concatenated variables as the composite data matrix.

4.2. Learning separating direction

The composite matrix computed via PNS is the input to DWD. DWD learns a feature space separating direction between the two classes, i.e., the schizophrenic and the control group, via the training set of discrete s-reps Euclideanized as described in the previous section.

4.3. Computing the function that maps from projected feature values to the probability of schizophrenia

Given a separation direction and a case with an unknown class label, our objective is to compute that case's probability

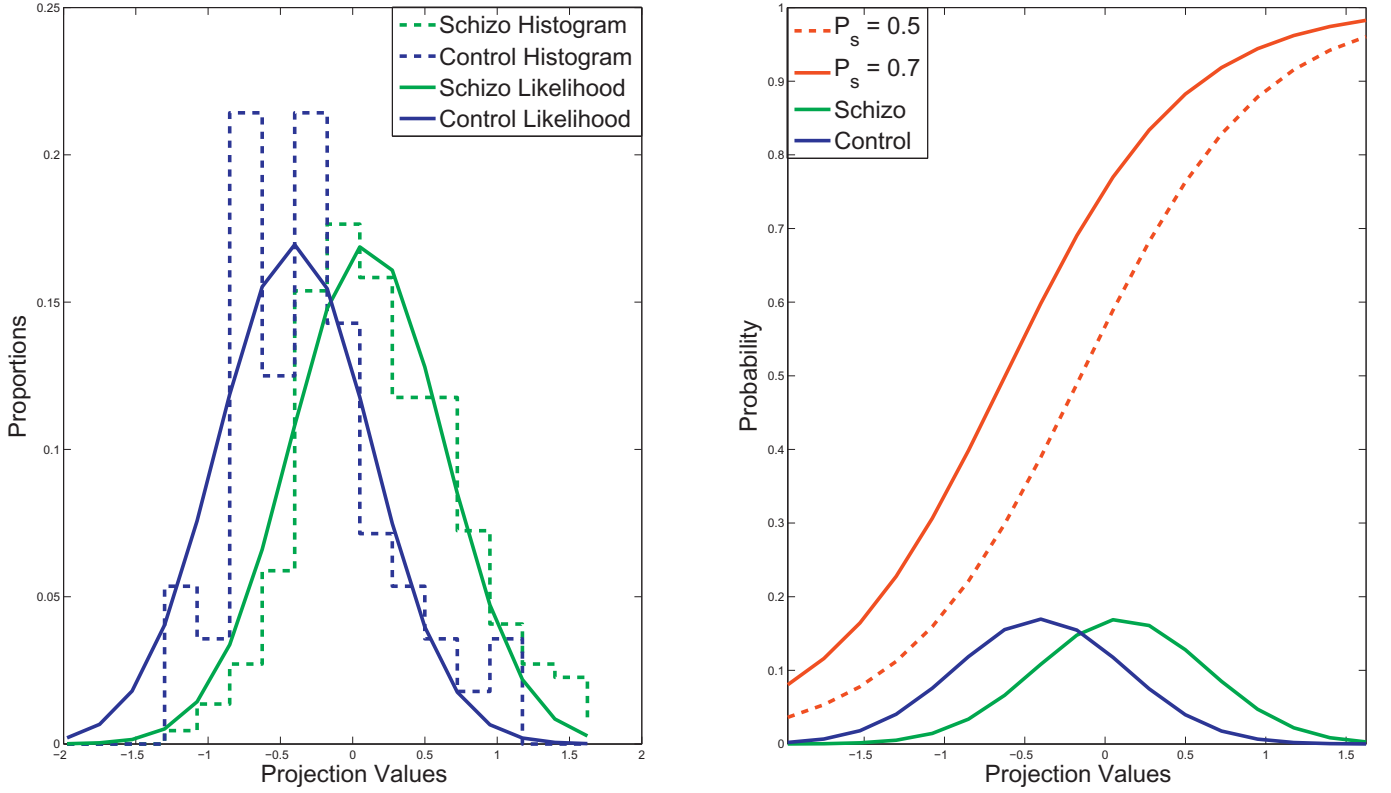


Fig. 2. Visualizations of (left) the class likelihoods and (right) the probability mapping function overlaid on top of the distributions. The empirical histogram of the scalar projection of the control cases in the training set onto the separation direction is plotted in the blue dotted lines; then the Gaussian probability distribution for the controls is plotted in the blue solid curve. The histogram for the schizophrenic class is plotted in the green dotted lines, and the corresponding Gaussian probability distribution for the schizophrenic class is plotted in the solid green curve. The function on the right that maps from the scalar projection onto the direction to the probability of being schizophrenic is plotted as solid and dashed curves respectively for two different values of $p(\text{schizo})$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

of belonging to the schizophrenic group. Using Bayes' Theorem, we can express this probability in terms of a prior and a likelihood of each class. We derive likelihood probabilities, i.e., the probability distributions of each class, given the s -rep features, by forming a pair of histograms (Fig. 2) each describing statistics of a class.

Using the trained polar system, we first transform the s -rep of interest into a point in feature space. Let d_X be the scalar value resulting from projecting that data point X onto the separation direction; let $\{d_{\text{schizo}}\}$ be projection values of positive training examples, and let $\{d_{\text{control}}\}$ be projection values of negative training examples. We form a pair empirical histograms of d_{schizo} and d_{control} . By treating d_{schizo} and d_{control} as random variables, we derive a probability distribution for each class from the respective histograms. The F-test failed to reject the null hypothesis that the two distributions are Gaussian with a common standard deviation. We therefore computed the sample means of the respective histograms and the unbiased least square estimate of their pooled variance. These were used to fit Gaussians forming the class likelihood probability distributions.

With these two distributions, $p(d_X|\text{schizo})$ and $p(d_X|\text{control})$, we can infer a class label of an unknown case if the projection value of that case d_X is given. It can be formulated by using Bayes' Theorem as follows.

By Bayes' theorem,

$$p(\text{schizo}|d_X) = \frac{p(\text{schizo})p(d_X|\text{schizo})}{p(\text{schizo})p(d_X|\text{schizo}) + (1 - p(\text{schizo}))p(d_X|\text{control})} \quad (1)$$

This can be reduced to

$$p(\text{schizo}|d_X) = \frac{p(\text{schizo})}{p(\text{schizo})(1 - R(d_X)) + R(d_X)} \quad (2)$$

where

$$R(d_X) = \exp\left[-\frac{1}{2}\left\{\left(\frac{d_X - \mu_{\text{control}}}{\sigma}\right)^2 - \left(\frac{d_X - \mu_{\text{schizo}}}{\sigma}\right)^2\right\}\right] \quad (3)$$

where

$$\sigma^2 = \frac{(n_{\text{schizo}} - 1)\sigma_{\text{schizo}}^2 + (n_{\text{control}} - 1)\sigma_{\text{control}}^2}{(n_{\text{schizo}} - 1) + (n_{\text{control}} - 1)} \quad (4)$$

where n_{schizo} denotes the number of observations for the schizophrenic observations, σ_{schizo} denotes the standard deviation of the scalar projections onto the direction for the schizophrenic observations, and similarly for the controls with n_{control} and σ_{control} .

In summary, we end up with the function mapping from projection value d_X along the separation direction and $p(\text{schizo})$ to $p(\text{schizo}|d_X)$. $p(\text{control}|d_X)$ is the complement of $p(\text{schizo}|d_X)$. Not only does this probability communicate intuitively to a user how certain a classification of a new case is but also its basis on parameterized probability distributions allows stable predictions in the tails of the distribution. Fig. 2 illustrates how the mapping from d_X to $p(\text{schizo}|d_X)$ varies for different values of $p(\text{schizo})$.

4.4. Classification based on probability produced by the mapping function

We decide the class label of an unknown case given projected value d_X and the prior $p(\text{schizo})$ by comparing $p(\text{schizo}|d_X)$

and $p(\text{control}|d_X)$. In particular, we study how $p(\text{schizo}|d_X)$ and $p(\text{control}|d_X)$ varies as we vary the prior $p(\text{schizo})$.

5. Experimental analysis

We first compare the performance of our method against classification based on global volume and against classification based on non-Euclideanized s-reps.

To evaluate each method, we use repeated 4-fold cross validation so that we do not introduce bias in the testing procedure. We first randomly partition the positive example set (schizophrenic group) into 4 roughly equal size subsets and likewise with the negative example sets (control group). We set aside one of the subsets from each class for validation and used the remaining subsets to collect statistics necessary for the classification method; this process is repeated so that every pair of quarters over both classes is used for validation.

A conventional way to compare classification methods is via ROCs, and in particular via the area under ROC. However, in data such as ours arising from cross validation the standard methods for comparing ROCs are not applicable because the data from different tries of the cross validation are not statistically independent. Instead, we directly compute true positive rate and true negative rate by varying the prior $p(\text{schizo})$ from 0 to 1. These two curves can be used to form an ROC (Fig. 3). The area under this curve tells us classification performance averaged over the range of prior probability.

We have conducted 625 rounds of these cross validations yielding 10,000 pairs of true positive rate and true negative rate against the prior. We pool these pairs of curves over 10,000 cross validation rounds to yield a single ROC (Fig. 3). We then compute area under this final ROC (AUC_0). We report that value in Table 1.

In addition, we computed confidence intervals at the 95% level for all the methods given 10,000 AUCs for each method. To do this, we can think of AUC_0 as corresponding to its index k among the sorted 10,000 individual AUCs. Under the conservative estimate that these individual AUCs are drawn randomly with replacement from a uniform distribution over the interval $[0,1]$, $k = AUC_0 \times N$. From this uniform distribution, we can estimate the confidence interval using order statistics (Gentle, 2009; Jones, 2009).

Order statistics $U_1 \leq U_2 \leq \dots \leq U_N$ are drawn independently from the distribution $Uniform(0, 1)$. Under this assumption the k th order statistic, U_k , follows the beta distribution $\beta(k, N+1-k)$. The mean and variance of $\beta(a, b)$ are $a/(a+b)$ and $\frac{ab}{(a+b)^2(a+b+1)}$. Therefore, U_k has expected value of $\frac{k}{N+1}$ and variance of $\frac{k(N+1-k)}{(N+1)^2(N+2)}$.

Because our sample size of 10,000 is sufficiently large, we can approximate the beta distribution by a normal distribution. In that case, the expected value of U_k is approximately AUC_0 , and the variance is approximately $\frac{k(N-k)}{N^3} = \frac{AUC_0(1-AUC_0)}{N}$. Thus the standard deviation of U_k is $\sqrt{\frac{AUC_0(1-AUC_0)}{N}}$.

With this approximation we computed each method's 95% confidence interval. These intervals are reported in Table 1.

5.1. S-rep based method compared to boundary PDM-based methods

The boundary PDM is a common approach to represent a shape; boundary PDMs represent a shape via a collection of points along the object's boundary. We wish to compare the qualities of classification when hippocampal shapes are represented by s-reps vs. boundary PDMs to see if the rich geometric information provided by s-reps does increase discriminative power over classification based on boundary information (Fig. 4).

In order to make a fair comparison between boundary PDMs and s-reps, we need boundary PDMs that can be compared directly

to s-reps. Recall that s-reps are a collection of spoke vectors pointing from skeletal sample points to the object's surface and that s-reps are fitted such that the spoke vectors are in approximate correspondence across all cases in the training population; we form boundary PDMs from these spoke end points. We will refer these boundary PDMs as srep-PDMs.

In order to make comparisons for PDMs not based on s-reps, we also create PDMs by the conventional method based on spherical harmonics. We used a standard software pipeline (Styner et al., 2006) to create boundary PDM with 4002 points. The two cases in schizophrenic groups produced badly formed PDMs, so we removed those two cases for this analysis. We will refer these boundary PDMs as spharm-PDMs.

Once the points are in correspondence, we classify in two different ways. First, we applied our DWD-based method directly to the point coordinate features. Second, in order to understand advantages of the Euclideanization on that type of the shape data, we applied PNS to the point tuples to yield Euclideanized features as well as a commensurated scale, and then we applied our DWD-based method to these features. The same cross validation strategy used with s-reps was applied to each of these methods. For each method, in Table 1 we report the AUC as well as confidence intervals. While these confidence intervals are valid, their not overlapping does not strictly indicate statistical significance, as these confidence intervals can be made as small as desired by carrying out arbitrarily many cross validation trials. However, since to our knowledge the statistics literature fails to have a satisfactorily powerful test for the significance of findings from the cross validation experiments, we resort to reporting these confidence intervals.

6. Results

Table 1 reports the performances of all the aforementioned methods in terms of the average AUC and its associated confidence interval. First, all of the methods show improvement over random guessing with non-overlapping confidence intervals. Second, there is no overlap among the confidence intervals for the best performing classification method based on s-reps, best performing classification method based on PDMs, and the method based on volume alone. That is, s-rep-based classification with Euclideanization is superior to all the other methods.

For s-reps, s-rep-PDMs, and spharm-PDMs, classification using Euclideanization is superior to that without Euclideanization. For the boundary PDMs derived from spherical harmonics, the confidence intervals in respect to the improvement in performance from Euclideanization do overlap.

With Euclideanization both forms of model yield similar if not better classification than the common approach in the literature in hippocampal classification of using volume alone. The Euclideanization is so important for s-rep-based classification that without it even classification based on volume alone is superior.

Table 1
Table of averaged AUC of ROCs, confidence interval corresponding to 95% level of the aforementioned methods and random guessing.

Methods	AUC	Confidence intervals
s-reps + PNS + DWD	0.6457	[0.6363, 0.6551]
s-reps + DWD	0.5617	[0.5520, 0.5715]
boundary srep-PDMs + PNS + DWD	0.5981	[0.5885, 0.6077]
boundary srep-PDMs + DWD	0.5769	[0.5672, 0.5866]
boundary spharm-PDMs + PNS + DWD	0.5750	[0.5653, 0.5847]
boundary spharm-PDMs + DWD	0.5734	[0.5638, 0.5831]
volume + DWD	0.5754	[0.5657, 0.5851]
Random guessing	0.5000	[0.4902, 0.5098]

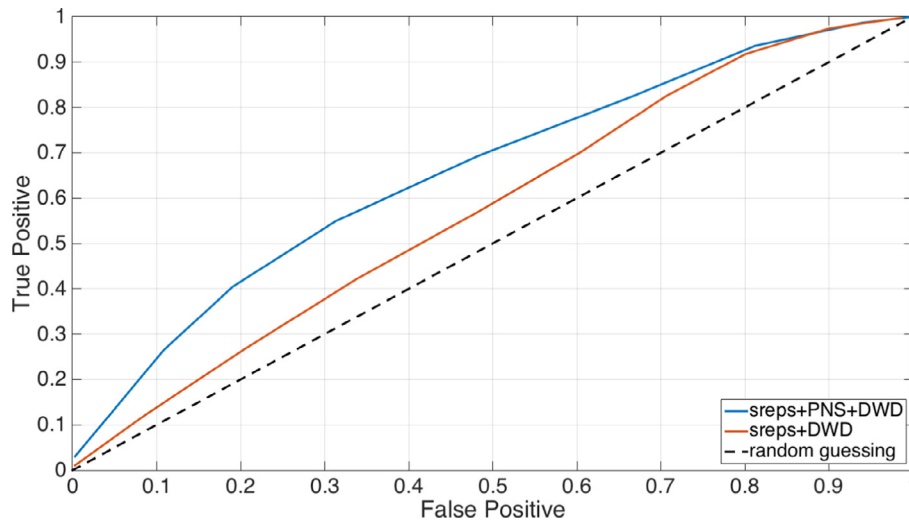


Fig. 3. The ROCs for s-rep based classification methods with and without PNS based Euclideanization. The classification method of s-reps without Euclideanization of spherical GOPs in s-reps yields AUC of 0.5617. Our proposed method that uses s-reps as the object representation and uses DWD as the classification method with Euclideanization of s-rep's spherical GOPs via PNS yields the AUC of 0.6550.

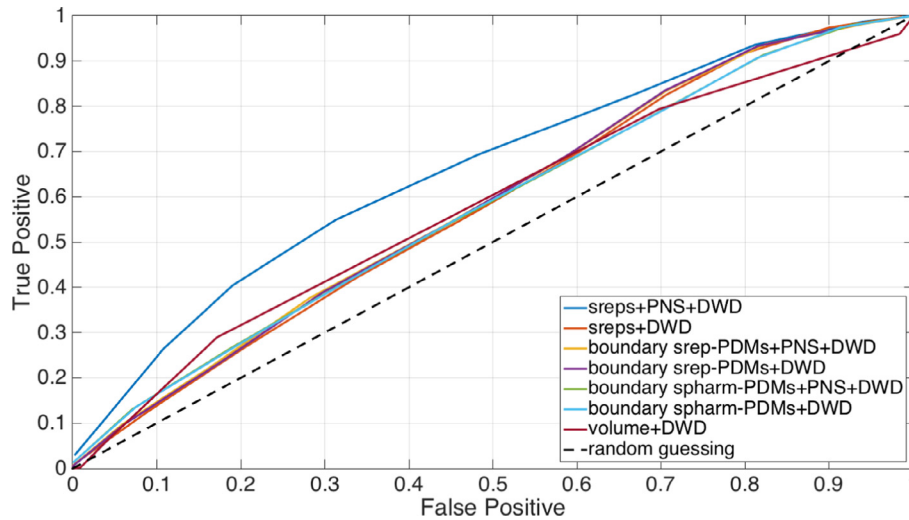


Fig. 4. The ROCs for aforementioned classification methods with and without PNS based Euclideanization. Our proposed method that uses s-reps as the object representation and uses DWD as the classification method with Euclideanization of s-rep's spherical GOPs via PNS yields the AUC of 0.6550.

7. Discussion and conclusion

In this paper, we have presented a novel classification method that recognizes that rich geometric information is provided by s-reps and that that information does not live in Euclidean space. We have shown improvement in classification performance when all of the GOPs of either s-reps or boundary PDMs derived from s-reps are Euclideanized via PNS analysis. Indeed, since shape is essentially non-Euclidean, it is not surprising that trying to analyze the geometrically rich s-rep models without Euclideanization notably harms the performance. We believe that the advantages of Euclideanization to shape classification is the primary message of this paper.

We have not seen significant advantage to using Euclideanization on boundary spharm-PDMs. One possible cause is the number of points in spharm-PDMs; there are a total of 4002 points for each case in spharm-PDMs whereas there are only a total of 66 points for each case in srep-PDMs. This significantly increases the dimensionality of the sphere in which the shape representation resides, so the curved manifold can be more well approximated by a flat space.

We have also shown that s-rep-based classification does provide an advantage over traditional volume based classification of hippocampi under schizophrenia; we therefore claim that shape descriptions add additional discriminative power. We have also shown improvement in classification accuracy when using s-reps over boundary PDMs assuming both are appropriately Euclideanized; we conclude that local object directions and local object width add discriminative power.

We chose this classification between schizophrenics and controls as our target problem partly because the discriminability of these shapes was not previously studied and also because its low level of classification accuracy could be expected to particularly strongly illustrate the effects of object representations and statistical analysis methods. It remains to be seen how strongly this effect applies with shape classes that are more easily discriminated, i.e., for classifications that are clinically useful.

Our method yields a separating direction through the pooled backwards mean in the feature space of the Euclideanized s-reps. Each point on this vector can be used to generate an s-rep using the polar system. Viewing the sequence of the s-reps as an

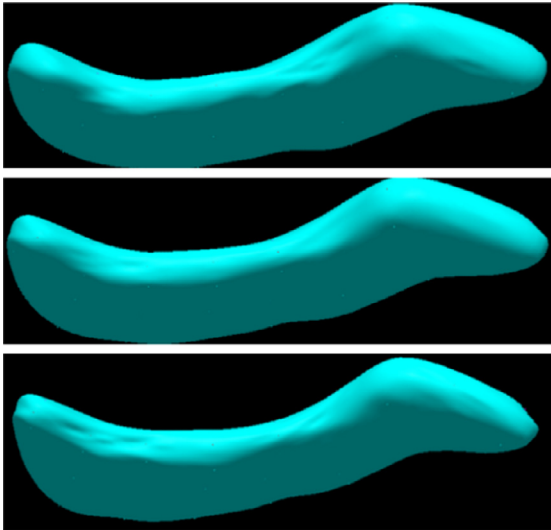


Fig. 5. Selected frames from the sequence of the s-reps while walking along the separation direction through the pooled backwards mean from the schizophrenic class to the control class. Viewing the sequence as a looping movie makes the local shape changes between the two classes more noticeable.

animation yields understanding of the shape changes between the two classes. Fig. 5 shows selected frames from the sequence. Our group's paper on hypothesis testing on shapes using PNS-Euclideanization (Schulz et al., 2013a) analyzes the discriminability between these two classes of hippocampi locality by locality and GOP by GOP.

The experiments described in this paper were done on a single data set of 277 hippocampus s-reps. These s-reps were fitted, as described in Section 2.1.2, using statistics computed from the entire dataset. This introduces a bias in classification evaluation because, when the data is partitioned into training and testing groups during cross validation, the s-rep models in the testing group have their fits affected by not only the training data but also the testing data. Unfortunately, the cost required to correct this bias by recomputing statistics and s-reps in every iteration is prohibitive, so the bias could not be removed. Instead, we choose to examine the effect of this bias has on a single partitioning of the cross validation.

For that partitioning, we fit s-reps to the training hippocampi and testing hippocampi using statistics computed using the training data only; this reflects the procedure that would be used when applying a trained classifier to previously unseen data. Using these unbiased s-rep fits, we performed the experiment described in Section 4 only on that partitioning. For the method on the classification of s-reps with Euclideanization, the unbiased analysis yields an AUC for this partition of 0.600. The analysis on the same partition using the original biased s-reps yields an AUC of 0.591. The difference is about 0.2 times the standard deviation of the AUCs across partitions. While this result comes from the only one partition, this suggests that there are negligible effects of the bias from the model fitting.

There are still some further questions to be investigated.

- To see if our results extend to other anatomic objects and diseases, we would like to apply the method on different application problems, e.g., classification of Alzheimer patients or of infants at high risk of autism based on shapes of the neuroanatomical structures. We are also interested to see classification quality when there are multiple structures involved, e.g., hippocampus and caudate.

- In Euclideanizing a spoke direction using PNS, we apply PNS separately because we are making the naive assumption that each direction is independent. However, because object surface is continuous and smooth, each direction is highly correlated to its neighbors. We would like to produce a Euclideanization method that reflects this correlation. Also, others are suggesting methods for statistical analysis directly on the curved shape-feature space manifold (Benjamin Eltzner and Huckemann, 2015; Sommer, 2015), and it would be interesting to evaluate classification methods using these ideas.
- As previously mentioned in Section 2.1.2 the method we used to achieve spoke correspondence in s-reps across the training set could be improved. In separate work, reported in (Tu et al., 2015b) and in (Tu et al., 2015a) under review, we created a method to improve the correspondence by spoke shifting on each training case, so as to minimize an entropy measure. This entropy measure reflects both shape probability distribution tightness and uniformity of coverage of the spokes in each training case. The shape probability distribution used is derived from the same PNS approach used in this paper. The correspondence was shown to be improved in a set of lateral ventricles and in a subset of the hippocampi used in this paper. It would be interesting to see whether classification of hippocampus could be improved using these correspondence improved models. Finally, (Tu et al., 2015c) also showed improved PDM correspondence when using the spoke tips as the PDM as compared to a PDM derived from spherical harmonics and then improved in correspondence by the entropy-based method of (Cates et al., 2006). This further justifies our decision to use the s-rep derived PDM instead of PDMs derived from spherical harmonics in the classification study reported in this paper.
- Other work is in progress comparing different statistical methods against DWD. It would be interesting to see how DWD for our purpose compares to other statistical methods such as Support Vector Machine, Difference of the Means, and Random Forests.
- It would be interesting to measure the relative power of classification via other shape representations that have been used in the anatomic shape analysis literature, including but not limited to parameterized surface representations used in (Kurtek et al., 2012; Jermyn et al., 2012; Bauer et al., 2010, 2012; Durrleman et al., 2014), deformation fields used in (Lancaster et al., 2003), Villalon-Reina et al., the spherical harmonic coefficients used in (Gerig et al., 2001), spherical wavelet coefficients used in (Nain et al., 2007), and atlas deformation representations such as LD-DMM momentum (Beg et al., 2005; Miller et al., 2002; Wang et al., 2007).
- Whereas this paper compares the classification performances of shapes, we are preparing another work on comparison of probability distribution estimation on shapes as we vary the representation and whether Euclideanization is used, as well as one focusing on probability distribution estimations on shape change using Euclideanization.

Acknowledgment

We thank Eli Lilly User Initiated Information Technology for supporting the previous study (McClure et al., 2013) in which the original MRI dataset of the schizophrenic patients was collected. The work (McClure et al., 2013) was supported by the Eli Lilly User Initiated Information Technology Grant no. PCG TR:033107/F1D-MC-X252, NIH Roadmap Grant no. U54 EB005149-01, National Alliance for Medical Image Computing, HD 03110, P30 HD003110.

Appendix

As noted in Section 7, we visualize the hippocampal shape difference between the schizophrenics and the healthy controls by linearly interpolating points in feature space; these points are interpolated along the separation vector that points from the positive class (schizophrenics) to the negative class (controls) passing through the mean of all the training cases in the Euclideanized feature space. We generate a sequence of s-reps from these interpolated points. We create an animation using these s-reps that loops back and forth three times in two different views, i.e., axial view and coronal view. We strongly recommend interested readers to take a look at the full sequence in the supplementary data.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.media.2016.01.007](https://doi.org/10.1016/j.media.2016.01.007).

References

- Bauer, M., Harms, P., Michor, P.W., 2010. Sobolev metrics on shape space of surfaces. arXiv preprint arXiv:1009.3616.
- Bauer, M., Harms, P., Michor, P.W., 2012. Almost local metrics on shape space of hypersurfaces in n-space. *SIAM J. Imaging Sci.* 5 (1), 244–310.
- Beg, M.F., Miller, M.I., Trounev, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* 61 (2), 139–157.
- Benjamin Eltzner, S.J., Huckemann, S., 2015. Dimension reduction on polyspheres with an application to skeletal representations.
- Bouix, S., Pruessner, J.C., Louis Collins, D., Siddiqi, K., 2005. Hippocampal shape analysis using medial surfaces. *Neuroimage* 25 (4), 1077–1089.
- Cates, J., Meyer, M., Fletcher, T., Whitaker, R., 2006. Entropy-based particle systems for shape correspondence. In: 1st MICCAI Workshop on Mathematical Foundations of Computational Anatomy: Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability, pp. 90–99.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1992. Training models of shape from sets of examples. In: *BMVC92*. Springer, pp. 9–18.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models—their training and application. *Comput. Vis. Image Underst.* 61 (1), 38–59.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297. doi:10.1007/BF00994018.
- Damon, J., Marron, J.S., 2014. Backwards principal component analysis and principal nested relations. *J. Math. Imaging Vis.* 50 (1–2), 107–114.
- Davies, R.H., Twining, C.J., Allen, P.D., Cootes, T.F., Taylor, C.J., 2003. Shape discrimination in the hippocampus using an MDL model. *Information processing in medical imaging : proceedings of the ... conference* 18, 38–50.
- Dryden, I.L., Mardia, K.V., 1998, 4. Statistical shape analysis. John Wiley & Sons New York.
- Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trounev, A., 2014. Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage* 101, 35–49.
- Gentle, J.E., 2009, Vol. 308. Computational statistics. Springer.
- Gerig, G., Styner, M., Jones, D., Weinberger, D., Lieberman, J., 2001. Shape analysis of brain ventricles using spharm. In: *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on*. IEEE, pp. 171–178.
- Gorcuzowski, K., Styner, M., Jeong, J.Y., Marron, J.S., Piven, J., Hazlett, H.C., Pizer, S.M., Gerig, G., 2010. Multi-object analysis of volume, pose, and shape using statistical discrimination. *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (4). doi:10.1109/TPAMI.2009.92.652–61
- Gouttard, S., Styner, M., Joshi, S., Smith, R.G., Cody Hazlett, H., Gerig, G., 2007. Subcortical structure segmentation using probabilistic atlas priors 6512.65122J–65122J–11 URL: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1299623>, doi: 10.1117/12.708626.
- Jermyn, I.H., Kurtek, S., Klassen, E., Srivastava, A., 2012. Elastic shape matching of parameterized surfaces using square root normal fields. In: *Computer Vision—ECCV 2012*. Springer, pp. 804–817.
- Jones, M., 2009. Kumaraswamy's distribution: a beta-type distribution with some tractability advantages. *Stat. Methodol.* 6 (1), 70–81.
- Jung, S., Dryden, I.L., Marron, J.S., 2012. Analysis of principal nested spheres. *Biometrika* 99 (3), 551–568. doi:10.1093/biomet/ass022. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3635703&tool=pmcentrez&rendertype=abstract>.
- Jung, S., Liu, X., Marron, J., Pizer, S.M., 2010a. Generalized pca via the backward stepwise approach in image analysis. In: *Brain, Body and Machine*. Springer, pp. 111–123.
- Jung, S., Marron, J.S., Pizer, S.M., 2010b. A backward generalization of pca for image analysis. In: *International Symposium on the Occasion of the 25th Anniversary of McGill University Centre for Intelligent Machines*, pp. 111–124.
- Kendall, D.G., 1984. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* 16 (2), 81–121.
- Koenderink, J.J., 1990, Vol. 2. Solid shape. Cambridge Univ Press.
- Kurtek, S., Klassen, E., Ding, Z., Avison, M.J., Srivastava, A., 2011. Parameterization-invariant shape statistics and probabilistic classification of anatomical surfaces. In: *Information Processing in Medical Imaging*. Springer, pp. 147–158.
- Kurtek, S., Klassen, E., Gore, J.C., Ding, Z., Srivastava, A., 2012. Elastic geodesic paths in shape space of parameterized surfaces. *Pattern Anal. Mach. Intell. IEEE Trans.* 34 (9), 1717–1730.
- Lancaster, J.L., Kochunov, P.V., Thompson, P.M., Toga, A.W., Fox, P.T., 2003. Asymmetry of the brain surface from deformation field analysis. *Hum. Brain Mapp.* 19 (2), 79–89.
- Marron, J.S., Todd, M.J., Ahn, J., 2007. Distance-weighted discrimination. *J. Am. Stat. Assoc.* 102 (480), 1267–1271. doi:10.1198/016214507000001120.
- McClure, R.K., Styner, M., Maltbie, E., Lieberman, J.A., Gouttard, S., Gerig, G., Shi, X., Zhu, H., 2013. Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type. *Psychiatry Res.* 211 (1), 1–10. doi:10.1016/j.psychres.2012.07.001.
- Merck, D., Tracton, G., Saboo, R., Levy, J., Chaney, E., Pizer, S., Joshi, S., 2008. Training models of anatomic shape variability. *Med. Phys.* 35 (8), 3584–3596. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2809709&tool=pmcentrez&rendertype=abstract>.
- Miller, M.I., Trounev, A., Younes, L., 2002. On the metrics and euler-lagrange equations of computational anatomy. *Annu. Rev. Biomed. Eng.* 4 (1), 375–405.
- Nain, D., Haker, S., Bobick, A., Tannenbaum, A., 2007. Multiscale 3-d shape representation and segmentation using spherical wavelets. *Med. Imaging IEEE Trans.* 26 (4), 598–618.
- Pizer, S. M., Jung, S., Goswami, D., Vicory, J., Chaudhuri, R., Damon, J. N., Huckemann, S., Marron, J. S., 2012. Nested Sphere Statistics of Skeletal Models, 1–21.
- Schulz, J., Pizer, S. M., Marron, J., Godtliebsen, F., 2013a. Nonlinear hypothesis testing of geometrical object properties of shapes applied to hippocampi.
- Schulz, J., Pizer, S. M., Marron, J. S., Godtliebsen, F., 2013b. Nonlinear Hypothesis Testing of Geometrical Object Properties of Shapes Applied to Hippocampi.
- Sommer, S., 2015. Anisotropic distributions on manifolds template estimation and most probable paths.
- Styner, M., Lieberman, J.A., Pantazis, D., Gerig, G., 2004. Boundary and medial shape analysis of the hippocampus in schizophrenia. *Med. Image Anal.* 8 (3), 197–203.
- Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Levitt, J.J., Shenton, M.E., Gerig, G., 2006. Framework for the statistical shape analysis of brain structures using spharm-pdm. *Insight J.* (1071) 242–250.
- Tu, L., Styner, M., Vicory, J., Elhabian, S., Paniagua, B., Prieto, J. C., Yang, D., Whitaker, R., Styner, M., Pizer, S. M., 2015a. Skeletal shape correspondence through entropy.
- Tu, L., Styner, M., Vicory, J., Paniagua, B., Prieto, J.C., Yang, D., Pizer, S.M., 2015b. Skeletal shape correspondence via entropy minimization. In: *SPIE Medical Imaging*. International Society for Optics and Photonics.94130U–94130U.
- Tu, L., Vicory, J., Elhabian, S., Paniagua, B., Prieto, J.C., Whitaker, R., Styner, M., Pizer, S.M., 2015c. Entropy-based correspondence improvement of interpolated skeletal models.
- Villalon-Reina, J., Prasad, G., Joshi, S. H., Jalbrzikowski, M., Toga, A., Bearden, C., Tompson, P. M., Statistical analysis of maximum density path deformation fields in white matter tracts.
- Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J.C., Csernansky, J.G., Miller, M.I., 2007. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type. *Med. Imaging IEEE Trans.* 26 (4), 462–470.
- Yushkevich, P.A., Zhang, H.G., 2013. Deformable modeling using a 3d boundary representation with quadratic constraints on the branching structure of the blum skeleton. In: *Proceedings of the 23rd International Conference on Information Processing in Medical Imaging*. Springer-Verlag, Berlin, Heidelberg, pp. 280–291. doi:10.1007/978-3-642-38868-2_24.
- Zhao, Q., Okada, K., Rosenbaum, K., Kehoe, L., Zand, D.J., Sze, R., Summar, M., Linguraru, M.G., 2014. Digital facial dysmorphismology for genetic screening: Hierarchical constrained local model using {ICA}. *Med. Image Anal.* 18 (5), 699–710. <http://dx.doi.org/10.1016/j.media.2014.04.002>.