Multi-Object Analysis of Volume, Pose, and Shape Using Statistical Discrimination

Kevin Gorczowski, Martin Styner, *Member*, *IEEE*, Ja Yeon Jeong, J.S. Marron, Joseph Piven, Heather Cody Hazlett, Stephen M. Pizer, *Senior Member*, *IEEE*, and Guido Gerig, *Senior Member*, *IEEE*

Abstract—One goal of statistical shape analysis is the discrimination between two populations of objects. Whereas traditional shape analysis was mostly concerned with single objects, analysis of multi-object complexes presents new challenges related to alignment and pose. In this paper, we present a methodology for discriminant analysis of multiple objects represented by sampled medial manifolds. Non-euclidean metrics that describe geodesic distances between sets of sampled representations are used for alignment and discrimination. Our choice of discriminant method is the distance-weighted discriminant because of its generalization ability in high-dimensional, low sample size settings. Using an unbiased, soft discrimination score, we associate a statistical hypothesis test with the discrimination results. We explore the effectiveness of different choices of features as input to the discriminant analysis, using measures like volume, pose, shape, and the combination of pose and shape. Our method is applied to a longitudinal pediatric autism study with 10 subcortical brain structures in a population of 70 subjects. It is shown that the choices of type of global alignment and of intrinsic versus extrinsic shape features, the latter being sensitive to relative pose, are crucial factors for group discrimination and also for explaining the nature of shape change in terms of the application domain.

Index Terms—Shape, size and shape, shape analysis.

1 Introduction

Statistical shape modeling and analysis [1], [2], [3] is emerging as an important tool for understanding anatomical structures from medical images. Clinical applications favor a statistical shape modeling of multiobject sets rather than one of single structures outside of their multi-object context. Neuroimaging studies of mental illness and neurological disease, for example, are interested in describing group differences and changes due to neurodevelopment or neurodegeneration. These processes most likely affect multiple structures rather than a single one. An analysis of the structures jointly therefore should reveal more than studying them individually. Applications of multi-object analysis include segmentation and studying group differences. Litvin and Karl [4], for example, have proposed methodology for building a multiobject shape prior with application in 2D curve evolution segmentation. In this manuscript, we will focus on studying group differences in neuroimaging studies using discrimination analysis.

A fundamental difficulty in statistical shape modeling is the relatively small sample size, typically in the range of 20-50 samples in neuroimaging studies, compared to a high-dimensional feature space, commonly one to several orders of magnitude larger than the sample size. Given that we are describing the shape of several structures instead of a single one, the dimension of our feature space tends to be even higher. This difficulty must be considered when choosing among different methods for discrimination analysis [5]. We favor the distance-weighted discrimination (DWD) [6], which is similar to Support Vector Machines (SVMs), but it suffers less from data piling problems in high-dimensional low samples size (HDLSS) settings. Previous work in discriminating single anatomical objects has been done by Golland et al. [7] using distance transforms for shape features and SVM to discriminate populations. Yuschkevich et al. [8] also used SVM to discriminate 2D m-reps of corpora collosa.

Another context-specific choice is what features to use as input to the shape analysis. Most neurological studies focus solely on volume for the sake of simplicity [9], [10], [11], [12], [13], [14]. However, Styner et al. [15], [16] have shown that the shape of an object can be more useful in discriminating populations than volume for particular applications. In a multiobject setting, there may be an additional feature of interest: the relative pose of objects with respect to each other. A statistical description of multiobject pose variability was introduced in [17]. Since multiobject analysis of subcortical structures is novel, we have chosen to evaluate

- *K. Gorczowski, J.Y. Jeong, and S.M. Pizer are with the Department of Computer Science, University of North Carolina, CB 3175, Chapel Hill, NC 27599-3175.*
  E-mail: kgorcz@unc.edu, jjeong@email.unc.edu, smp@cs.unc.edu.
- *M. Styner is with the Departments of Computer Science and Psychiatry, University of North Carolina, CB 3175, Chapel Hill, NC 27599-3175.*
  E-mail: styner@cs.unc.edu.
- *J.S. Marron is with the Department of Statistics and Operations Research, University of North Carolina, CB 3260, Chapel Hill, NC 27599-3260.*
  E-mail: marron@email.unc.edu.
- *J. Piven and H.C. Hazlett are with the Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599.*
  E-mail: {jpiven, hcody}@med.unc.edu.
- *G. Gerig is with the Scientific Computing and Imaging Institute SCI, The University of Utah, Salt Lake City, UT 84112.*
  E-mail: gerig@sci.utah.edu.

Manuscript received 18 Oct. 2008; revised 19 Mar. 2009; accepted 2 Apr. 2009; published online 22 Apr. 2009.
Recommended for acceptance by A. Srivastava, J.N. Damon, I.L. Dryden, and I.H. Jermyn.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMISI-2008-10-0720.
Digital Object Identifier no. 10.1109/TPAMI.2009.92.

several different features, namely, volume, pose, shape, and the combination of pose and shape.

Several different geometric shape representations have been used to model anatomy, such as landmarks [18], dense collection of boundary points [19], or harmonic coefficients [20], [21]. Unlike the above explicit description, Tsai et al. [22] and Yang et al. [23] propose an implicit statistical object modeling by level sets with its inherent difficulties of topology preservation. Another shape analysis approach focuses on the analysis of spatial deformation maps [24], [25], [26], [27]. In this work, we employed explicit deformable shape modeling with a sampled medial mesh representation called m-rep, introduced by Pizer et al. [28]. Styner et al. [29] have compared the use of boundary and medial representations in the analysis of subcortical structures.

The work in this paper could be applied well to other shape descriptions, but we chose a medial description for several reasons. First, it gives a more intuitive representation of the interior of the object. The radius, which describes the distance from the medial axis to the boundary, serves as a localized measure related to the object's volume. This is particularly interesting for neuroimaging work because of the widespread use of volume data. Bouix et al. [30] studied hippocampi using the radius function defined on a flattened 2D medial sheet. Medial representations are also advantageous when attempting to describe certain nonlinear shape deformations such as bending and twisting [31]. Simple boundary representations are less suited to account for this type of variability. The sampled m-rep description is also relatively compact when compared to other shape representations. We can describe 10 subcortical structures using 210 medial atoms for a total of 1,890 features. While this is much higher than the number of data samples we typically have, it is less than the spherical harmonic representation that we have also computed and which uses about 10,000 features. The results of the study presented will show that the choice of a medial description was crucial to find relevant shape differences.

In summary, this paper presents a methodology for discriminant analysis on sets of objects. We choose the DWD method and feature sets of volume, pose, and shape. The latter is given by the sampled medial m-rep shape representation, requiring noneuclidean metrics to determine shape alignment and shape distance. The driving application is a longitudinal pediatric neuroimaging study.

2 Methods

In this section, we first present the motivating clinical data and discuss the methodology of the different features used in our discrimination analysis. These are the m-rep shape features and the local pose change features. We then summarize distance-weighted discrimination, along with the transformation of our raw data. Finally, our unbiased estimator for computing classification rates is presented.

2.1 Motivation and Clinical Data

The driving clinical problem of this research is the need for a joint analysis of the set of 10 subcortical brain structures (see Fig. 2b). These structures include the left and right hemispheric hippocampus, amygdala, caudate, putamen, and pallide globe. The image data used in this paper are taken from an ongoing clinical longitudinal pediatric autism study [32]. This study includes autistic subjects (AUT) and typically developing, healthy controls (CONT) with baseline at age 2 and follow-up at age 4. For the study in this paper, we have selected 23 subjects from the autism group and 10 from the control group. For all of the autism subjects and 6 of the 10 controls, we have successful scans at age 2 and age 4. For the other four controls, we paired an age 2 scan of one subject with an age 4 scan of another unrelated subject. We also have four additional control age 2 scans that have no matching age 4 scan. This gives us a total of 70 samples: 46 autism and 24 control.

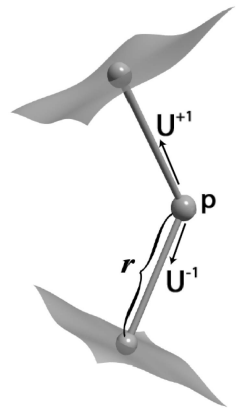


Fig. 1. Medial atom: position ($\mathbf{p}$), radius (r), and two normals to boundary ($\mathbf{U}$).

2.2 M-rep Shape Description

The m-rep shape description for a 3D object consists of a sheet of medial atoms, each of which is defined by a position, radius, and two unit-length normal vectors to the boundary (spokes). The radius represents the distance from the atom position to the corresponding point on the boundary of the object along the two normal vectors. The medial atom, as shown in Fig. 1, is defined as $m = \{\mathbf{p}, r, \mathbf{U^{+1}}, \mathbf{U^{-1}}\} \in \mathcal{M}$, with $\mathcal{M} = \mathbb{R}^3 \times \mathbb{R}^+ \times \mathcal{S}^2 \times \mathcal{S}^2$.

To obtain m-reps describing the subcortical structures in our study, we started with binary image segmentations from well-trained experts using semiautomated procedures.[1] We also needed an initial m-rep that would be deformed to fit the binary image. We constructed these initial medial models using the modeling scheme developed by Styner et al. [31] to determine the minimum sampling required for each model. Given a binary segmentation and initial model, the initial model is deformed through an optimization process such that the model best fits the image without becoming too irregular in its geometry [33]. This process is applied individually to each of the 10 anatomical objects using the Pablo tool [34], while the correspondence across samples is implicitly established by the deformation process on the template model. Figs. 2a and 2b show the medial atoms for a set of objects and the implied surfaces.

1. See https://www.ia.unc.edu/dev/tutorials for a detailed description of protocols and reliability results.

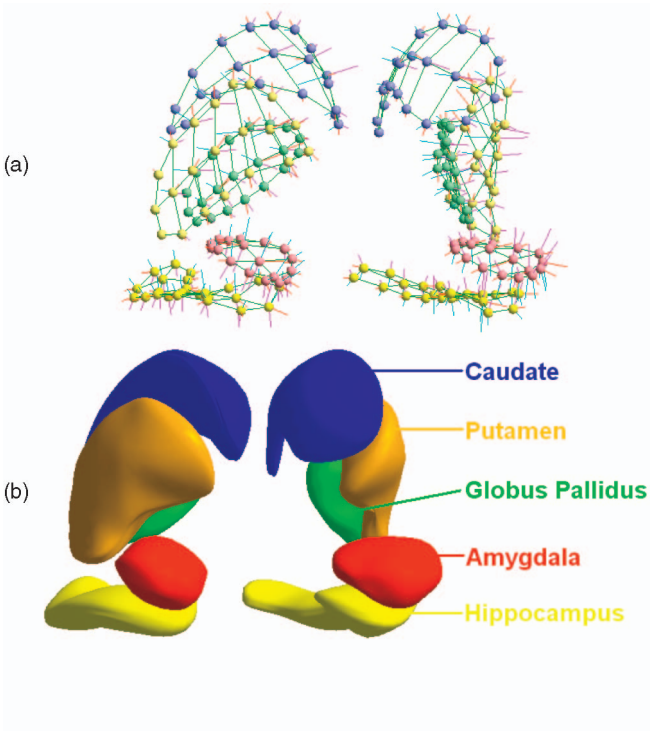


Fig. 2. M-reps of a multiobject complex. (a) Medial atoms. (b) Implied boundary surfaces of medial description.

2.3 Alignment and Pose

In a multi-object setting, it must be decided how to remove unimportant shape variability through alignment. We call aligning the object set as a whole, where transformations are applied to all objects jointly, a global alignment. As shown in Figs. 3a and 3b, after this global alignment, there are still local pose differences among the individual objects. In our case, we assumed that these single object pose differences were important because they represent the interobject changes within the multiobject set. Therefore, after the global alignment, we perform a second step referred to as the local alignment. In this step, we take the globally aligned object sets and align objects individually as would be done in a single-object setting. It is these local pose changes that we include as part of the overall variability of the objects. The resulting m-reps after the local alignment are what we refer to as pure shape and can be seen in Fig. 3c. For the purposes of this paper, the global alignment included translation and rotation. This accounted for any pose differences between the original images. The local alignment included translation, rotation, and scale to remove all remaining pose. When we

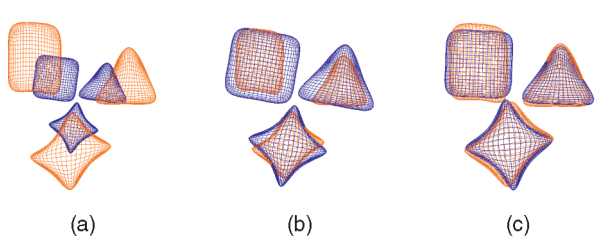


Fig. 3. Multi-object alignment. (a) Global translation and rotation. (b) Global translation, rotation, and scale. (c) Local translation, rotation, and scale after global translation and rotation.

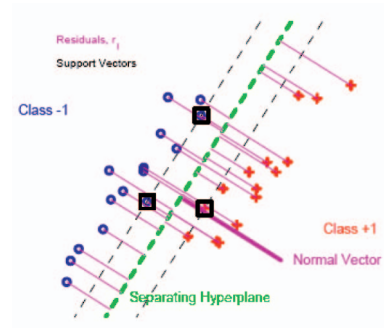


Fig. 4. Illustration of two-class discrimination with separating hyperplane and residuals. The support vectors that determine the discrimination direction computation in SVM are highlighted with an additional box marker. In DWD, all samples are included in the discrimination direction computation.

use the local pose changes as features for discriminant analysis, we have an 8D vector consisting of three elements for the translation, four for the orientation (stored as a quaternion), and one for the scale. After both global and local alignments have been finished, the final m-reps are in the mean pose position and used as the pure shape features.

To align m-reps, we use a variation of the standard Procrustes method [35]. In a standard Procrustes alignment on a set of boundary points, the sum-of-squared euclidean distances between corresponding points are minimized. For our purposes, we instead minimize the geodesic distance between m-reps because they do not lie in a euclidean space [36]. The distance between two m-reps is then the sum of geodesic distances over all m-rep features $m = \{\mathbf{p}, r, \mathbf{U}^{+1}, \mathbf{U}^{-1}\}$. The geodesic distance $d(m_a, m_b)$ between two medial atoms m_a and m_b equals $\|\mathbf{p}_b - \mathbf{p}_a, \frac{r_b}{r_a}, \mathbf{R}(\mathbf{U}_b^{+1}) \cdot \mathbf{U}_a^{+1}, \mathbf{R}(\mathbf{U}_b^{-1}) \cdot \mathbf{U}_a^{-1}\|$, where $\mathbf{R}(x)$ is the rotation of x to $(1, 0, 0)$. The corresponding Procrustes error metric is $d(S_1, \ldots, S_N; M_1, \ldots, M_N) = \sum_{i=1}^{N} \sum_{j=1}^{i} d(S_i \cdot M_i, S_j \cdot M_j)^2$, where S_i, M_i are the ith transformation and the m-rep models. For more details, see [36]. This procedure results in the following steps:

1. Translations: First, the translational part of the alignment S_i is minimized once and for all by centering each m-rep model. That is, each model is translated so that the average of its medial atoms positions is the origin.
2. Rotations and Scalings: The ith m-rep model M_i is aligned to the mean of the remaining models, denoted as μ_i. The alignment is accomplished by a gradient descent algorithm on $SO(3) \times R^+$ to minimize $d(\mu_i, S_i \cdot M_i)^2$. The gradient is approximated numerically by a central differences scheme. This is done for each of the models.
3. Iterate: Step 2 is repeated until the error metric cannot be further minimized.

2.4 Distance-Weighted Discrimination

Discriminant analysis is concerned with finding the axis which best separates two populations. An optimization must be performed that somehow maximizes the distance between the discriminating axis and the data points while separating the two classes. It is formulated in a general way as follows (see Fig. 4): Given points x_i, class indicators $y_i \in \{+1, -1\}$,

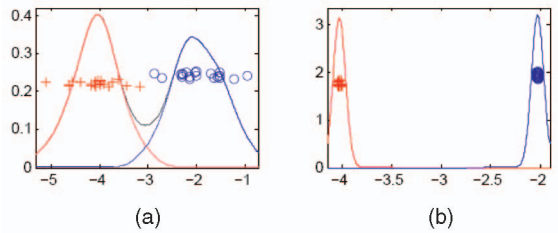


Fig. 5. (a) Projection onto normal of optimal separating hyperplane. (b) Projection onto normal of separating hyperplane which exhibits data piling.

and w the normal to the separating hyperplane, the distance or residual r from the points to the hyperplane is

$$r_i = y_i(w \cdot x_i + \beta), \tag{1}$$

where β determines the position of the hyperplane. One of the popular methods of discriminant analysis is SVMs. It attempts to maximize the minimum r_i. The main problem with this method is that it tends to use only a small subset of the population, those near the opposite class, to completely define the discriminating axis. It is manifested in the problem of "data piling" (see Fig. 5), where most of the samples from the same population group, when projected onto the normal of the discriminating axis, end up very close to each other. This leads to poor generalization performance when tested on new samples that were not included in the calculation of the discriminating axis: It is too specific to the samples from which it was computed.

Distance-weighted discrimination [6] is a method similar to SVM, but uses all sample points in the calculation of the discriminating axis.[2] It attempts to minimize the sum of the reciprocals of r_i. Thus, each point's contribution to the calculation is weighted inverse proportionally to the distance from that point to the opposite population. In this way, the DWD achieves a higher robustness when presented with new, untrained samples. This advantage is heightened further in the context of high-dimensional feature spaces with low sample sizes where it is best to use all information available from the low number of samples. A detailed discussion of DWD was published by Marron et al. [6].

DWD was specifically developed for the HDLSS. The shape analysis study presented in this paper is a good example of an HDLSS problem, as the dimensionality (1,890D data, joint m-rep shape description of 10 subcortical structures) is much larger than the number of samples (i.e., 70). Classical discrimination methodology based on Fisher Linear Discriminant is often not appropriate in such a setting due to data overfitting and is usually outperformed by SVM or DWD [6], [37].

2.5 Transformation of Raw Input Data

The m-rep shape description as well as the pose features contain rotational elements that are not part of a euclidean space. This can lead to reduced the performance of methods such as DWD that attempt to find a linear discriminant. Likewise, combining features with different units into one long feature vector can bias results toward features with larger variance. Finally, our data samples have unequal gender distributions within the two populations. We must first account for each of these issues before running DWD analysis.

2. Available at http://www.stat.unc.edu/faculty/marron/marron_software.html.

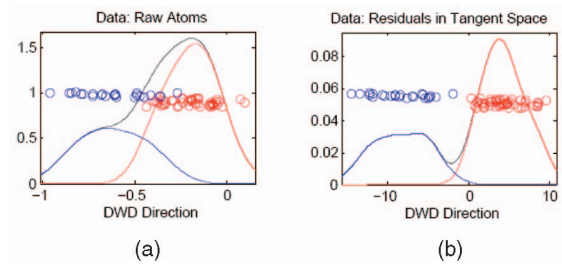


Fig. 6. Separation of 70 multiobject m-reps into two populations given by DWD axis. (a) Raw, nonlinear medial atom data. (b) Atom data after projection into tangent space and subtraction of mean.

While the application of DWD to nonlinear features may give a reasonable solution, we found through experimentation that the linearized form of the m-rep features gives a better discrimination result (see Fig. 6). To obtain a linear instance of our curvilinear m-rep and pose features, we project them into the tangent space at the geodesic mean point [36]. This involves taking the log map of each of the noneuclidean features. For the pose rotation, the log map of a unit-length quaternion $q = (w, \underline{\mathbf{v}})$ is defined as

$$\log q = \frac{\theta}{sin(\theta/2)} \cdot \underline{\mathbf{v}}, \quad \theta = 2arccos(w). \tag{2}$$

For the m-rep normal directions $\mathbf{U} = (x, y, z)$, the spherical log map is

$$\log \mathbf{U} = \left(x \cdot \frac{\theta}{sin(\theta)}, y \cdot \frac{\theta}{sin(\theta)} \right), \quad \theta = arccos(z). \tag{3}$$

For the pose scale and m-rep radius factors, it is a simple logarithmic transformation.

As a next step, we want to concatenate all linearized, log-map transformed m-rep features into a singe vector. In order to do so, we first must make them commensurate to avoid unwanted bias. For our purposes, we have chosen to normalize each feature by subtracting the mean and dividing by the standard deviation. This makes the weighting of points equal among separate features in the DWD calculation. So, for each feature, the final input to the DWD routine is of the form

$$Y = \frac{X - \overline{X}}{\sigma} \forall X \in \mathbb{R}^k, \tag{4}$$

$$Y = \frac{\log(X - \overline{X})}{\sigma} \forall X \notin \mathbb{R}^k. \tag{5}$$

The mean $\overline{X}$, however, is computed for each gender. By subtracting a gender-specific mean, we eliminate any disproportion in the gender sampling within our two populations. To build a gender-specific mean, we start with the lowest level of subcategories within our data and compute the mean of the samples in each. We then compute the mean of the subcategory means which are of the same gender. This gives us a gender-specific mean. Fig. 7 shows the process with subcategories according to the three criteria of gender, group, and time.

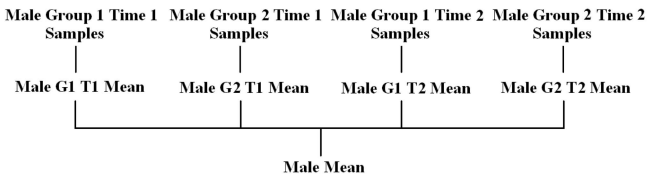


Fig. 7. Illustration of gender-specific mean calculation given gender, group, and time. The same process is applied to obtain a female mean.

2.6 Unbiased Classification Using Leave-Many-Out Experiments

To test the performance of the DWD, we chose to implement a leave-many-out, cross-validation experiment. We first divided our data samples into a training set and a testing set. The discriminating axis was computed using the training set. Each sample from the test set was then projected onto the DWD axis with the resulting 1D projected value serving as the classification score (hence, known as the DWD score). The DWD method produces both a discriminating axis and a threshold β. The threshold value is the amount by which the training data, after projected onto the DWD, must be shifted such that zero becomes the best dividing point between populations. Therefore, given a DWD axis w and a test sample feature vector x, the DWD score becomes $s = w \cdot x + \beta$. The discrete classification into one of the diagnosis groups is then simply the sign of the DWD score.

In order to make the training set unbiased, we used the following selection strategy (see Algorithm 1): We alternately choose a single random autism or control sample $g1$ over both timepoints. With this sample from one group, we chose the sample $g2$ from the other group that was the closest match according to the subject's age and was of the same gender as the subject from the first group. This results in samples $g1$ and $g2$, one from each group at the same timepoint. Since our data are longitudinal, i.e., we have two data sets for each subject, one at age 2 and one at age 4, we also included for each of the samples their corresponding data $g1'$, $g2'$ at the other timepoint. This process was repeated but starting with a random sample from the opposite group than in the previous iteration.

Algorithm 1. Training Set T Selection
$T = \emptyset, size = 0, i = 0$
while $size < n$ **do**
  if i mod $2 = 0$ **then**
    $g1 =$ random sample from autism group
    $g2 =$ closest matching sample to $g1$ from control group
  else
    $g1 =$ random sample from control group
    $g2 =$ closest matching sample to $g1$ from autism group
  end if
  $g1' =$ corresponding, other sample of $g1$
  $g2' =$ corresponding, other sample of $g2$
  $T = T \cup \{g1, g1', g2, g2'\}$
  $size = size + 4$
  $i = i + 1$
end while

After several runs, all of the data samples are included in the testing set of at least in a few runs. From the results of these experiments, an unbiased estimate of each sample's classification is built by computing its mean DWD score over only those runs for which it was in the test set. In this way, we calculate a classification for a sample only when the discriminating axis is computed without any knowledge of that sample. The box plots in the following sections show these unbiased mean DWD scores.

As mentioned before, the study presented in this paper is based on a total of 70 samples with 46 autism and 24 control with information at ages 2 and 4 including four unpaired control samples, which were always left out of the training set. We chose the training set consisting of 32 out of the 70 available samples with 16 samples in each group. The remaining 38 samples served as the test set. The experiment was then run 100 times. The number of runs was chosen heuristically such that each of the 70 samples was included in the test set at least in a few runs; the minimum number of runs in the test set for any sample was 4. From test set results, we then calculated an unbiased mean DWD score. As the two groups are not equally represented in the test set (30 autistic cases and 8 controls), we assess the discrimination accuracy of an individual test by averaging the discrimination accuracy of the two subgroups. Without averaging of the subgroups, a classifier that would always guess "autism" would result in a classification accuracy of $30/38 = 78.9\%$. Our average discrimination accuracy results for such a simple classifier in 50 percent classification accuracy.

3 Results

In this section, we describe the results of our discrimination-based shape analysis experiments divided into five sections corresponding to the analyzed features: volume, pose, shape, shape and pose, and m-rep radii. All statistical p-values reported in this section are computed via standard, nonparametric permutation tests.

3.1 Discrimination by Diagnosis

3.1.1 Volume

Because of its prevalence in neuroimaging studies, we first assessed the ability of object volumes to discriminate between the autism and control groups. The volumes were computed from the implied surface boundary of the m-reps. The 10 subcortical structures gave us a 10D feature space for the discriminant analysis. We computed the mean DWD score for each sample over the runs in which that sample was in the test set. This gives us an unbiased average classification score for each sample. Fig. 8 and Table 1 show that there is a clear, significant difference ($p < 0.001$) between the median score of the autism group and control group with an average classification performance of 74 percent.

3.1.2 Pose

The next step was to explore the significance of local pose changes. For each sample, there were 70 features (10 objects $\times$ 3 translation, 3 rotation, 1 uniform scale) transformed to the log-map. Similarly to the volume discrimination, the pose analysis, which includes uniform object scale factors, showed significantly different ($\mathrm{p} = 0.03$) mean DWD scores (Fig. 9).

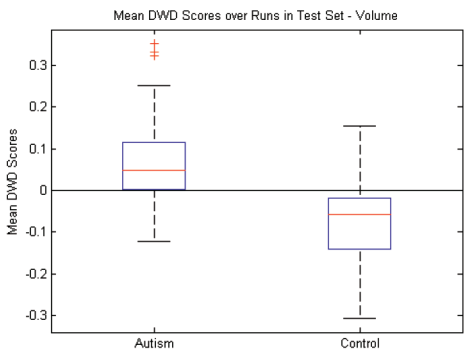


Fig. 8. Volume features: Box plot (median, 25 and 75 percentiles, minimum/maximum) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, and less than zero classified as control. $p < 0.001$.

TABLE 1
Classification Accuracy over 100 Runs

Feature	Mean	Std. Dev.
Volume	74 %	± 8 %
Pose	57 %	± 9 %
Pose (Scale Only)	64 %	± 7 %
Shape	56 %	± 8 %
Shape and Pose	55 %	± 8 %
Shape and Scale (Radii Only)	76 %	± 7 %

However, the test sample classification accuracy was considerably lower than volume as illustrated in Table 1. There were also 20 individual runs in which the classification accuracy was at or below 50 percent, a result that would be outperformed by a random coin flip. The translation and rotation components of the pose seemed to be adding mostly noise and instability to the DWD calculation because the same experiment run with only the scale factors gave an average classification rate of 64 percent and $p = 0.002$ as opposed to 54 percent and $p = 0.1$ using the translations and rotations. From these results, we conclude that the pose does include some relevant information for discrimination but it is likely in the uniform scale factors. The classification rate of volume (74 percent) compared with the scale factors (64 percent) is illustrative of shape-based uniform scaling factors not capturing full information of volumetric measurements.

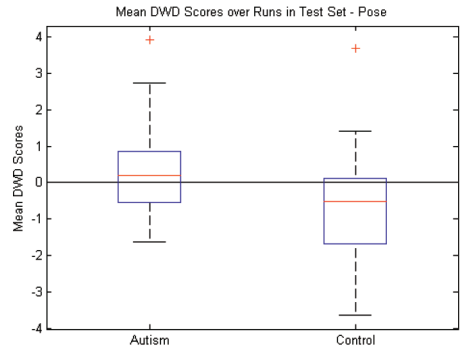


Fig. 9. Pose features: Box plot (median, 25 and 75 percentiles, minimum/maximum) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, and less than zero classified as control. $p = 0.03$.

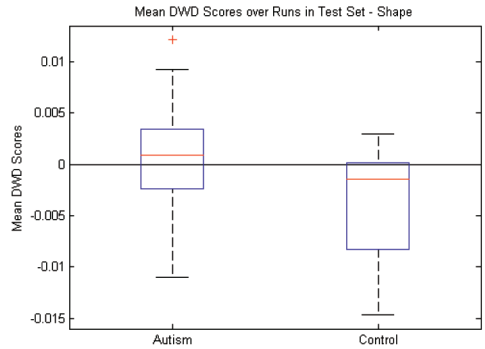


Fig. 10. Shape features: Box plot (median, 25 and 75 percentiles, minimum/maximum) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, and less than zero classified as control. $p = 0.01$.

3.1.3 Shape

Fig. 10 shows the results of using only the m-rep shape features for the DWD calculations. As with volume and pose, the mean DWD scores for the test samples were significantly different ($p = 0.01$, Fig. 10). The classification accuracy of shape was equal to that of pose at 56 percent. The DWD methodology proved its usefulness and stability in high-dimensional low sample size settings because the shape features numbered about 2,000 in total, whereas the volume and pose were 10 and 70, respectively. Even in this high-dimensional space, the DWD still generalized well enough to equal the performance of the pose features.

3.1.4 Shape and Pose

Finally, the high classification accuracy of the volumes compared with both the shape and pose features led us to combine the latter two. This gave us the most complete description of the variability of the multiobject complex. The differences between the mean DWD scores were not significant ($p = 0.38$, Fig. 11). Once again, the pose features seem to be mostly noise since combining them with shape produced a nonsignificant result while shape alone was significant. The classification accuracy (55 percent) was similar to the shape and pose features individually, as shown in Table 1.

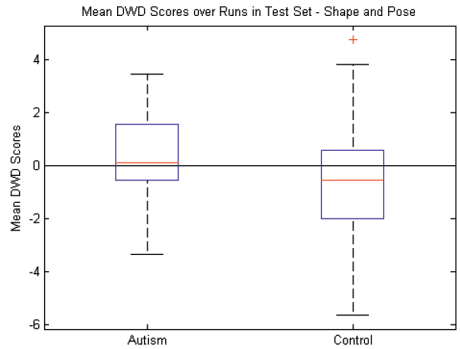


Fig. 11. Shape and pose features combined: Box plot (median, 25 and 75 percentiles, minimum/maximum) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, and less than zero classified as control. $p = 0.38$.

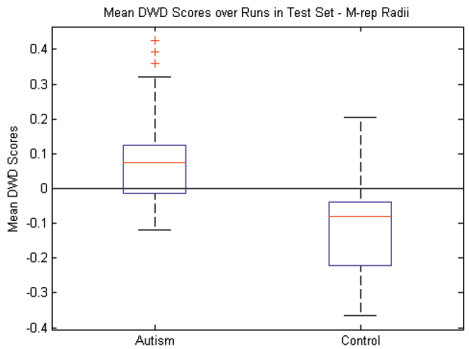


Fig. 12. M-rep radii features: Box plot (median, 25 and 75 percentiles, minimum/maximum) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, and less than zero classified as control. $p < 0.001$.

3.1.5 Shape and Scale: M-rep Radii

One of the ongoing research questions with a discretely sampled m-rep shape description is whether information about spokes and atom positions are sufficiently stable based on the method to establish correspondence. Out of this arises the question of whether statistics on these features, which we used as part of our shape analysis above, could be tighter by focusing on the medial radii features that are considered more stable. A small change in the atom position or spoke directions will generally not cause a large change in the radius. It is also invariant to any object-level translation and rotation, while encoding both local shape and scale information. This makes it a more intrinsic measure than the other medial atom features. Compared to the other features in Table 1, the mean classification accuracy is the best when using only the atom radii at 76 percent. Fig. 12 shows the mean DWD scores using only the radii. While the radii give the best classification rate of any feature we studied, it is difficult to say outright that they are the best features to use. First, it may be our specific application that lends itself to a local measure that is related to volume like the radii. It is known and evidenced above that there are volume differences between autistic and typically developing brains. Also, Fig. 12 shows that the overlap between the populations is not drastically better than with volume. However, the jump in classification accuracy from using all of the medial atom features to only the radii suggests that there is a certain amount of noise in the other features which ends up being correlated to the detriment of the DWD calculation. At the same time, witnessing the radii outperform volume as a discriminating feature adds validation to our general choice of the medial shape representation.

3.2 Evaluation of Bias

To verify that the mean classification scores were unbiased, we employed random, normally distributed input data with the same random number seeding and the exact same training and testing sets. The random data were generated with the same mean, variance, and dimension as our actual data. The p-value of the mean DWD scores was 0.22 and the classification accuracy was 49 percent.

3.3 Discrimination by Age

To complete our analysis, we ran the same experiments as above, with precisely the same gender-corrected data, but discriminating according to age instead of diagnosis. Not surprisingly, most features show improved performance when discriminating by time (Table 2) rather than by diagnosis (Table 1), as a size differences across time are expected to be larger. Especially noteworthy is the pose, which does marginally better than the volume, m-rep radii, and scale factors. Our previous work [38] has shown that the first principal mode of our pooled data (both time points) aligned very closely with age encoding mainly a large global scaling effect. This age-related change mainly affects the local translation parameters, which helps to explain why the entire local pose performs slightly better than scale or volume alone. Our conclusion is that since pose does well at discriminating by age, removing pose for the analysis of diagnosis groups leaves us with discriminating differences that are not heavily influenced by having pooled data across time and sufficient sample size. However, a direct method of correcting for age would be preferable.

TABLE 2
Classification Accuracies for Discrimination by Age

Feature	Mean	Std. Dev.
Volume	72 %	± 5 %
Pose	74 %	± 6 %
Pose (Scale Only)	72 %	± 6 %
Shape	64 %	± 6 %
Shape and Pose	72 %	± 6 %
Shape and Scale (Radii Only)	70 %	± 6 %

3.4 Visualization

To visualize the changes in shape along the DWD direction, we start with the mean m-rep of the autism group, which is deformed along the unit-length DWD toward the control group. The distance along the DWD direction by which the autism mean is deformed is defined as the distance between the mean of each group's projections onto the DWD direction. The final m-rep which has been deformed by this full distance is then used to represent the control group.

For robustness, we chose to use the mean DWD direction over all the runs instead of using a single run from the leave-many-out experiment. Fig. 13a shows colormaps of surface distances between the two object sets representing the diagnosis groups using shape only. The measurement is the distance from the autism to the control group. In this

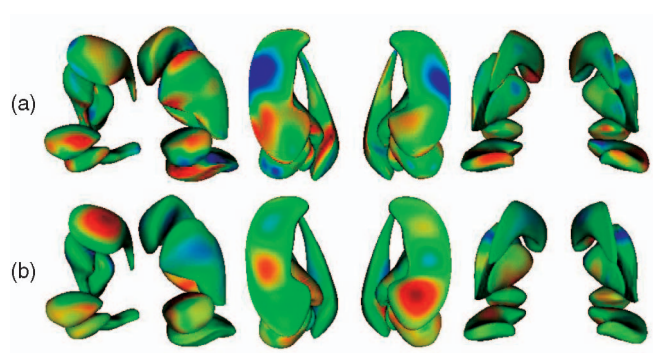


Fig. 13. Colormap of surface distances from autism mean m-rep to deformed m-rep along DWD direction using (a) shape only and (b) m-rep radii only. Red = inward distance, green = zero distance, and blue = outward distance.

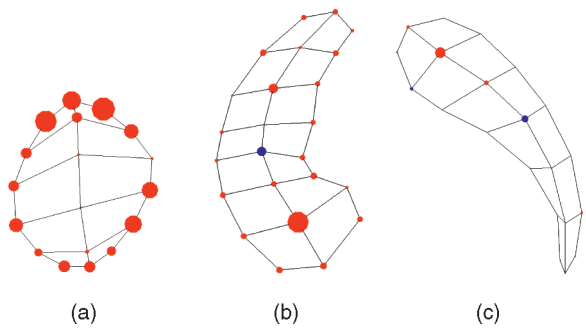


Fig. 14. Visualization of radii change from controls to autism in (a) right amygdala, (b) right hippocampus, and (c) right caudate. The size of the ball at an atom position is proportional in size to the log of the control's radius minus the log of the autism's radius. Red is an increase in radius from control to autism and blue is a decrease.

figure, we see that the amygdala and hippocampus undergo strong shape changes between the groups relative to the other three structures. There is a distinct inward deformation of the hippocampus tail as well as an outward change in the midsection. A large portion of each amygdala also presents a difference in shape.

Given that the m-rep radii performed best as a discriminating feature, we also wanted to visualize these differences. We used the same procedure as above to obtain a representative set of radii for each diagnosis group. Fig. 13b shows the surface distance between m-reps of each group with only the radii modified. The strongest individual radii changes appear to be in the hippocampus and the caudate. To assess the overall radii differences between groups, we calculated $\triangle r = \log(r_c/r_a)$ for all atoms with r_c the radius of the control group and r_a the radius of the autism group. Fig. 14 shows spheres plotted at the mean atom positions with size and color illustrating $\triangle r$. We see a clear increase in local widths when deforming from controls to autism at almost all positions across the three structures. Both the hippocampus and caudate show increases in the posterior head with slight decreases in the body section. These overall radii decreases are supported by Table 3 which lists the percent volumetric change from autism to control. All structures show a larger volume in the autism group.

TABLE 3
Volume Percent Change from Autism to Control

Object	$\%\triangle_2$	$\%\triangle_4$
Left Amygdala	-19.85	-15.41
Right Amygdala	-21.05	-18.19
Left Caudate	-14.38	-3.84
Right Caudate	-13.96	-5.49
Left Hippocampus	-10.92	-7.98
Right Hippocampus	-10.29	-6.69
Left Globus Pallidus	-10.19	-16.17
Right Globus Pallidus	-2.57	-8.27
Left Putamen	-6.43	-3.66
Right Putamen	-4.54	-2.24

$\%\triangle_A = \frac{\overline{Vol}_{cont} - \overline{Vol}_{aut}}{\overline{Vol}_{cont}} \times 100$, *at age* A.

4 Discussion and Conclusion

This research demonstrates work in progress toward shape analysis and group discrimination of multi-object complexes. Traditionally, shape analysis is concerned with single objects following a well-developed mathematical framework employing linear alignment. In a multi-object setting, linear alignment will remove global translation, rotation, and scale, but will not account for relative object pose. A joint analysis of only globally aligned sets of shapes will therefore include residual pose differences. Here, we discuss and explore the various options for global and local alignment of sets of shapes. We propose an initial global alignment to map each data set into a common coordinate frame. This step is followed by a local alignment for each object, and the alignment parameters are kept as pose information. Joint shape analysis will thus use *pure* shape features not affected by residual pose. Our features are mapped into Riemannian symmetric space, the appropriate choice for pose, and medial m-rep features that include rotational frames and positive reals. It is then straightforward to perform separate analyses for pose, shape, or pose and shape. Alternatively to the stepwise feature selection procedure, one could also use an automatic feature selection approach. The focus here though is mainly on understanding the interaction of pose and shape features in the multiobject settings.

Our results show that, in this specific application, pose features do not give statistically significant discrimination. Shape features also did not show significance, except when isolating a particular feature of the m-rep shape description, namely, the radius measure, which combines locality and local scale. Although sampled medial representations use a lower number of features than densely sampled surfaces, we still face the problem of HDLSS. In typical shape discrimination applications, two populations of less than 50 samples are represented by thousands of features. This problem is even more pronounced with sets of objects. For classification, we applied the DWD method, which is designed to be robust for HDLSS data analysis problems. Unbiased analysis by repeated leave-many-out experiments resulted in classification rates and significance values. Other classifiers suited for the HDLSS setting could be chosen as well, and we make no claim that the best classification method was chosen here.

The driving application is a pediatric autism study with autistic and normal children imaged at 2 and 4 years of age. We focus on a joint analysis of five left and right subcortical structures represented as sampled medial representations. Note that the relatively low classification rates can be explained by the nature of the underlying clinical problem. Morphologic neurodevelopmental phenotypes are often reflected by only subtle differences. Discrimination rates in psychiatric pathology are commonly low with considerable overlap between normal anatomy and pathology. Near perfect discrimination rates would be very suspicious of overfitting. Furthermore, our results fit well with the current literature on autism [32]. Prior shape analysis results in the hippocampus [39] correlate well with our findings (enlargement in the head and tail sections and reduction in the body section).

In the future, we will explore multivariate classification by selection of a best-separating subspace rather than a single axis. Further, we will develop a technique to explore the covariance structure of sets of shapes. This will help clinicians to explore links between morphological changes and underlying biological processes.

ACKNOWLEDGMENTS

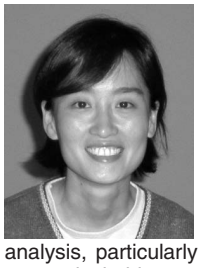
This research is supported by the NIH NIBIB grant P01 EB002779, the NIH Conte Center MH064065, and the UNC Neurodevelopmental Research Core NDRC. The MRI images of infants and expert manual segmentations are funded by NIH RO1 MH61696 and NIMH MH64580.

REFERENCES

[1] D. Thomson, *On Growth and Form,* second ed. Cambridge Univ. Press, 1942.
[2] I. Dryden and K. Mardia, "Multivariate Shape Analysis," *Sankhya,* vol. 55, pp. 460-480, 1993.
[3] C.G. Small, *The Statistical Theory of Shape.* Springer, 1996.
[4] A. Litvin and W.C. Karl, "Coupled Shape Distribution-Based Segmentation of Multiple Objects," *Proc. Int'l Conf. Information Processing in Medical Imaging,* pp. 345-356, 2005.
[5] I. Dryden and K. Mardia, *Statistical Shape Analysis.* Wiley, 1998.
[6] J. Marron, M. Todd, and J. Ahn, "Distance-Weighted Discrimination," *J. Am. Statistical Assoc.,* vol. 102, no. 480, pp. 1267-1271, Jan. 2007.
[7] P. Golland, W.E.L. Grimson, M.E. Shenton, and R. Kikinis, "Detection and Analysis of Statistical Differences in Anatomical Shape," *Medical Image Analysis,* vol. 9, pp. 69-86, 2005.
[8] P. Yushkevich, S.M. Pizer, S. Joshi, and J. Marron, "Intuitive, Localized Analysis of Shape Variability," *Proc. Int'l Conf. Information Processing in Medical Imaging,* pp. 402-408, 2001.
[9] R. McCarley, C. Wible, M. Frumin, Y. Hirayasu, J. Levitt, I. Fischer, and M. Shenton, "MRI Anatomy of Schizophrenia," *Biological Psychiatry,* vol. 45, pp. 1099-1119, 1999.
[10] J. Levitt, R. McCarley, C. Dickey, M. Voglmaier, M. Niznikiewicz, L. Seidman, Y. Hirayasu, A. Ciszewski, R. Kikinis, F. Jolesz, and M. Shenton, "MRI Study of Caudate Nucleus Volume and Its Cognitive Correlates in Neuroleptic-Naive Patients with Schizotypal Personality Disorder," *Am. J. Psychiatry,* vol. 159, no. 7, pp. 1190-1197, 2002.
[11] R.S.E. Keefe, L.J. Seidman, B.K. Christensen, R.M. Hamer, T. Sharma, M.M. Sitskoorn, S.L. Rock, S. Woolson, M. Tohen, G.D. Tollefson, T.M. Sanger, and J.A. Lieberman, "Long-Term Neurocognitive Effects of Olanzapine or Low-Dose Haloperidol in First-Episode Psychosis," *Biological Psychiatry,* vol. 59, no. 2, pp. 97-105, 2006.
[12] J. Juranek, P.A. Filipek, G.R. Berenji, C. Modahl, K. Osann, and M.A. Spence, "Association between Amygdala Volume and Anxiety Level: Magnetic Resonance Imaging (MRI) Study in Autistic Children," *J. Child Neurology,* vol. 21, no. 12, pp. 1051-1058, Dec. 2006.
[13] M. Langen, S. Durston, W.G. Staal, S.J.M.C. Palmen, and H. van Engeland, "Caudate Nucleus Is Enlarged in High-Functioning Medication-Naive Subjects with Autism," *Biological Psychiatry,* vol. 62, no. 3, pp. 262-266, 2007.
[14] J. Munson, G. Dawson, R. Abbott, S. Faja, S.J. Webb, S.D. Friedman, D. Shaw, A. Artru, and S.R. Dager, "Amygdalar Volume and Behavioral Development in Autism," *Archives of General Psychiatry,* vol. 63, no. 6, pp. 686-693, June 2006.
[15] M. Styner, A. Lieberman, R.K. McClure, D.R. Weinberger, D.W. Jones, and G. Gerig, "Morphometric Analysis of Lateral Ventricles in Schizophrenia and Health Controls Regarding Genetic and Disease-Specific Factors," *Proc. Nat'l Academy of Sciences USA,* vol. 102, no. 13, pp. 4872-4877, Mar. 2005.
[16] G. Gerig, M. Styner, M. Shenton, and J. Lieberman, "Shape versus Size: Improved Understanding of the Morphology of Brain Structures," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention,* pp. 24-32, 2001.
[17] M.N. Bossa and S. Olmos, "Statistical Model of Similarity Transformations: Building a Multi-Object Pose," *Proc. 2006 IEEE Conf. Computer Vision and Pattern Recognition Workshop,* p. 59, 2006.
[18] F. Bookstein, "Shape and the Information in Medical Images: A Decade of the Morphometric Synthesis," *Proc. Workshop Math. Methods in Biomedical Image Analysis,* 1996.
[19] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, pp. 38-59, 1995.
[20] A. Kelemen, G. Székely, and G. Gerig, "Elastic Model-Based Segmentation of 3D Neuroradiological Data Sets," *IEEE Trans. Medical Imaging,* vol. 18, no. 10, pp. 828-839, Oct. 1999.
[21] L. Staib and J. Duncan, "Model-Based Deformable Surface Finding for Medical Images," *IEEE Trans. Medical Imaging,* vol. 15, no. 5, pp. 1-12, Oct. 1996.
[22] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky, "Shape-Based Approach to Curve Evolution for Segmentation of Medical Imagery," *IEEE Trans. Medical Imaging,* vol. 22, no. 2, pp. 137-154, Feb. 2003.
[23] J. Yang, L.H. Staib, and J.S. Duncan, "Neighbor-Constrained Segmentation with Level Set Based 3D Deformable Models," *IEEE Trans. Medical Imaging,* vol. 23, no. 8, pp. 940-948, Aug. 2004.
[24] C. Davatzikos, M. Vaillant, S. Resnick, J. Prince, S. Letovsky, and R. Bryan, "A Computerized Method for Morphological Analysis of the Corpus Callosum," *J. Computer Assisted Tomography,* vol. 20, pp. 88-97, Jan./Feb. 1996.
[25] J. Csernansky, S. Joshi, L. Wang, J. Haller, M. Gado, J. Miller, U. Grenander, and M. Miller, "Hippocampal Morphometry in Schizophrenia via High Dimensional Brain Mapping," *Proc. Nat'l Academy of Sciences USA,* vol. 95, pp. 11406-11411, Sept. 1998.
[26] P. Thompson, M. Mega, and A. Toga, "Disease-Specific Brain Atlases," *Brain Mapping: The Disorders,* Academic Press, 2000.
[27] P. Thompson, J. Giedd, R. Woods, D. MacDonald, A. Evans, and A. Toga, "Growth Patterns in the Developing Brain Detected by Using Continuum Mechanical Tensor Maps," *Nature,* vol. 404, pp. 190-193, 2000.
[28] S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney, "Segmentation, Registration, and Measurement of Shape Variation via Image Object Shape," *IEEE Trans. Medical Imaging,* vol. 18, no. 10, pp. 851-865, Oct. 1999.
[29] M. Styner, J.A. Lieberman, D. Pantazis, and G. Gerig, "Boundary and Medial Shape Analysis of the Hippocampus in Schizophrenia," *Medical Image Analysis,* vol. 8, no. 3, pp. 197-203, 2004.
[30] S. Bouix, J.C. Pruessner, D.L. Collins, and K. Siddiqi, "Hippocampal Shape Analysis Using Medial Surfaces," *NeuroImage,* vol. 25, pp. 1077-1089, 2005.
[31] M. Styner, G. Gerig, J. Lieberman, D. Jones, and D. Weinberger, "Statistical Shape Analysis of Neuroanatomical Structures Based on Medial Models," *Medical Image Analysis,* vol. 7, no. 3, pp. 207-220, Sept. 2003.
[32] H.C. Hazlett, M.D. Poe, G. Gerig, R.G. Smith, and J. Piven, "Cortical Gray and White Brain Tissue Volume in Adolescents and Adults with Autism," *Biological Psychiatry,* vol. 59, no. 1, pp. 1-6, Jan. 2006.
[33] S. Pizer, T. Fletcher, Y. Fridman, D. Fritsch, A. Gash, J. Glotzer, S. Joshi, A. Thall, G. Tracton, P. Yushkevich, and E. Chaney, "Deformable m-reps for 3d Medical Image Segmentation," *Int'l J. Computer Vision,* vol. 55, no. 2, pp. 85-106, 2003.
[34] S.M. Pizer, P.T. Fletcher, S. Joshi, A.G. Gash, J. Stough, A. Thall, G. Tracton, and E.L. Chaney, "A Method and Software for Segmentation of Anatomic Object Ensembles by Deformable m-reps," *Medical Physics J.,* vol. 32, no. 5, pp. 1335-1345, May 2005.
[35] C. Goodall, "Procrustes Methods in the Statistical Analysis of Shape," *J. Royal Statistical Soc.,* vol. 53, no. 2, pp. 285-339, 1991.
[36] P. Fletcher, C. Lu, S. Pizer, and S. Joshi, "Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape," *IEEE Trans. Medical Imaging,* vol. 23, no. 8, pp. 995-1005, Aug. 2004.
[37] S. Sen, M. Foskey, J. Marron, and M. Styner, "Support Vector Machine for Data on Manifolds: An Application to Image Analysis," *Proc. Fifth IEEE Int'l Symp. Biomedical Imaging: From Nano to Macro,* pp. 1195-1198, Apr. 2008.
[38] M. Styner, K. Gorczowski, T. Fletcher, J. Jeong, S.M. Pizer, and G. Gerig, "Multi-Object Statistics Using Principal Geodesic Analysis in a Longitudinal Pediatric Study," *Lecture Notes in Computer Science,* G.-Z. Yang, T. Jiang, D. Shen, L. Gu, and J. Yang, eds., pp. 1-8, Springer, Aug. 2006.
[39] S.R. Dager, L. Wang, S.D. Friedman, D.W. Shaw, J.N. Constantino, A.A. Artru, G. Dawson, and J.G. Csernansky, "Shape Mapping of the Hippocampus in Young Children with Autism Spectrum Disorder," *Am. J. Neuroradiology,* vol. 28, pp. 672-677, 2007.

Kevin Gorczowski received the BS degree in computer science from the University of Illinois at Urbana-Champaign in 2005 and the MS degree in computer science from the University of North Carolina at Chapel Hill in 2007. He is currently a software engineer at Morphormics, Inc., in Chapel Hill.

Martin Styner is an assistant professor in the Department of Computer Science with a joint appointment in the Department of Psychiatry at the University of North Carolina at Chapel Hill (UNC). He is the codirector of the UNC Neuro Image Research and Analysis Laboratory and the associate director of the Developmental Neuroimaging Core in the Carolina Institute for Developmental Disabilities at UNC. His main field of research is in medical image processing and analysis. He has an extensive background in anatomical structure and tissue segmentation, morphometry using shape analysis, modeling and atlas building, diffusion tensor image analysis, as well as intra- and intermodality registration. His current research focuses on structural- and connectivity-based brain analysis within the developing brain in humans, primates, and rodents. He is a member of the IEEE.



Ja Yeon Jeong received the BS degree in mathematics from Pohang University of Science and Technology, Korea, in 1998, the MS degree in mathematics from the University of Nebraska-Lincoln in 2001, and the MS degree in 2007 from the University of North Carolina at Chapel Hill, where she is currently working toward the PhD degree in computer science. She is currently employed as a product analyst at Morphormics, Inc. Her research interests lie in medical image analysis, particularly in segmentation and statistical shape analysis of anatomical objects.

J.S. Marron received the PhD degree in statistics in 1982 from the University of California, Los Angeles, where he is currently the Amos Hawley professor of statistics and operations research. Furthermore, he is a professor of biostatistics, an adjunct professor of computer science, and a research member of the Lineberger Comprehensive Cancer Center at the University of North Carolina at Chapel Hill.

Joseph Piven received the MD degree from the University of Maryland in 1981. He completed training in general and child and adolescent psychiatry at the Johns Hopkins Hospital in Baltimore. He continued in research training in the genetics of neurobehavioral disorders, during a postdoctoral Merck Fellowship at Johns Hopkins. He joined the Faculty of the Department of Psychiatry at the University of Iowa from 1990 to 1999. He is currently the Sarah Graham Kenan Professor of Psychiatry, Pediatrics, and Psychology at the University of North Carolina at Chapel Hill, and the director of the Carolina Institute for Developmental Disabilities, home of the NICHD-funded Developmental Disabilities Research Center, TEACCH Program, UCEDD, and LEND Programs. He directs an NIH-funded Postdoctoral Research Training Program in Neurodevelopmental Disorders at UNC, and is the director of an Autism Center of Excellence Network study of brain development in infants at risk for autism. He is a founding editor of the *Journal of Neurodevelopmental Disorders*. His research is focused on the pathogenesis of autism-related disorders and includes studies on the molecular genetics of autism, the neuropsychological basis of autism and the broad autism phenotype, and magnetic resonance imaging of early brain development.

Heather Cody Hazlett received the PhD degree from the School of Psychology at the University of Georgia, where she studied under Dr. George Hynd with a focus on child neuropsychology. She completed her APA predoctoral psychology internship at Children's Hospital, Harvard Medical School. She did her postdoctoral training at the University of North Carolina (UNC), working with Dr. Joseph Piven and Dr. Stephen Hooper. She is currently an assistant professor in the Department of Psychiatry in the UNC School of Medicine. Her primary area of research involves neuroimaging studies of neurodevelopment disorders such as autism, Fragile X syndrome, and Down syndrome.

Stephen M. Pizer received the bachelor's degree in applied mathematics from Brown University in 1963, and the PhD degree in computer science from Harvard University in 1967. From 1962 to 1973, he was a member of the Physics Research Laboratory at Massachusetts General Hospital, and since 1967, he has been on the faculty of the University of North Carolina at Chapel Hill (UNC). A Kenan professor of computer science, radiation oncology, radiology, and biomedical engineering at UNC, he founded and has led UNC's multidepartmental Medical Image Display & Analysis Group. He is on the faculty of Computer Science's Graphics and Image Laboratory, and is a cofounder and the vice president of Science of Morphormics, Inc. His research, centered since 1962 on medical image processing and display, presently focuses on image and object shape analysis, probability on shape and appearance models, and segmentation. Other research directions have included interactive 3D graphics, human vision, image quality analysis, contrast enhancement, and image restoration. He is the author of a book on medial representations, two books on numerical computing, three proceedings in computer graphics and image analysis, and more than 270 published chapters and journal and proceedings articles. For many years, he was an associate editor of the *IEEE Transactions on Medical Imaging*. He is a senior member of the IEEE and a member of the IEEE Computer Society.

Guido Gerig received the PhD degree from the Swiss Federal Institute of Technology, ETH Zurich, Switzerland, in 1987. He joined the faculty at the University of North Carolina Chapel Hill as a Taylor Grandy professor in August 1998, with a joint appointment in the Departments of Computer Science and Psychiatry. In 2008, he accepted a new faculty position in the School of Computing and the Scientific Computing and Imaging Institute (SCI) at the University of Utah, with adjunct appointments in Biomedical Engineering and Psychiatry. He serves as the director of the Utah Center for Neuroimage Analysis (UCNIA). His research interests include development of methodologies for spatiotemporal and longitudinal analysis of volumetric image data with focus on modeling of the early developing brain, representation and analysis of 3D shape, techniques for representation, and statistical analysis of white matter tracts using diffusion-weighted imaging. Method developments are driven by challenging clinical applications that include research in schizophrenia, autism, multiple sclerosis, infants at risk for mental illness, and studies of aging populations. He is a member of the editorial board of the journal *Medical Image Analysis* (Elsevier), a member of the board of the Medical Image Computing and Computer Assisted Intervention Society, and regularly serves in various roles at major international conferences. He is a senior member of the IEEE and a member of the IEEE Computer Society.