

Discrimination Analysis using Multi-object Statistics of Shape and Pose

Kevin Gorczowski^a, Martin Styner^{a,b}, Ja Yeon Jeong^a, J. S. Marron^c, Joseph Piven^b, Heather Cody Hazlett^b, Stephen M. Pizer^a, Guido Gerig^{a,b}

^aDepartment of Computer Science, University of North Carolina at Chapel Hill

^bDepartment of Psychiatry, University of North Carolina at Chapel Hill

^cDepartment of Statistics and Operations Research, University of North Carolina at Chapel Hill

ABSTRACT

A main focus of statistical shape analysis is the description of variability of a population of geometric objects. In this paper, we present work towards modeling the shape and pose variability of sets of multiple objects. Principal geodesic analysis (PGA) is the extension of the standard technique of principal component analysis (PCA) into the nonlinear Riemannian symmetric space of pose and our medial m-rep shape description, a space in which use of PCA would be incorrect.

In this paper, we discuss the decoupling of pose and shape in multi-object sets using different normalization settings. Further, we introduce methods of describing the statistics of object pose and object shape, both separately and simultaneously using a novel extension of PGA. We demonstrate our methods in an application to a longitudinal pediatric autism study with object sets of 10 subcortical structures in a population of 47 subjects. The results show that global scale accounts for most of the major mode of variation across time. Furthermore, the PGA components and the corresponding distribution of different subject groups vary significantly depending on the choice of normalization, which illustrates the importance of global and local pose alignment in multi-object shape analysis. Finally, we present results of using distance weighted discrimination analysis (DWD) in an attempt to use pose and shape features to separate subjects according to diagnosis, as well as visualize discriminating differences.

Keywords: Statistical Shape Analysis, Statistics, Brain Morphometry

1. INTRODUCTION

Statistical shape modeling and analysis^{1,2} is emerging as an important tool for understanding anatomical structures from medical images. Clinical applications favor a statistical shape modeling of multi-object sets rather than one of single structures outside of their multi-object context. Neuroimaging studies of mental illness and neurodegeneration, 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 both discrimination and segmentation. Litvin et al.,³ for example, have proposed methodology for building a multi-object shape prior with application in 2D curve evolution segmentation.

A fundamental difficulty in statistical shape modeling is the relatively small sample size, typically in the range of 20 to 50 in neuroimaging studies. Given that we are describing the shape of several structures instead of a single one, the dimension of our feature space tends to be much higher than the number of data samples. This difficulty must be considered when choosing among the different methods for discrimination.⁴ We use the distance weighted discrimination (DWD), which is similar to Support Vector Machines (SVM), because 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.⁵ using distance transforms for shape features and SVM to discriminate populations. Yuschkevich et al.⁶ also used SVM to discriminate 2D m-reps of corpus collosa.

Another context-specific choice is what features to use as input to the discriminant method. Many neurological studies focus solely on volume for the sake of simplicity. However, Gerig et al.⁷ have shown that the shape of an object can be more useful in discriminating populations than volume for a particular application. In a multi-object setting, there may be an additional feature of interest: the relative pose of objects with respect to each other. A statistical description of multi-object pose variability was introduced in Bossa et al.⁸ Since multi-object analysis of subcortical structures is novel, we have chosen to evaluate 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,⁹ dense collection of boundary points,¹⁰ or spherical harmonic decompositions.¹¹ Another shape variability approach focuses on the analysis of deformation maps.¹²⁻¹⁵ Whereas Tsai et al.¹⁶ and Yang et al.¹⁷ describe statistical object modeling by level-sets, we use explicit deformable shape modeling with a sampled medial mesh representation called m-rep, introduced by Pizer et al.¹⁸ Styner et al.¹⁹ have compared the use of boundary and medial representations in the analysis of subcortical structures.

The work in this paper could be applied equally 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.²⁰ 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.²¹ Simple boundary representations struggle 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 1890 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.

This paper summarizes results of discriminant analysis on sets of objects. We choose the distance weighted discriminant (DWD) method and feature sets of volume, pose, and shape. The latter is given by the sampled medial m-rep shape representation. The driving application is a longitudinal pediatric neuroimaging study.

2. METHODS

In this section, we first discuss the methodology of the different features we use in our discrimination analysis. These are the m-rep shape features and the local pose change features. We then summarize the method of distance-weighted discrimination, along with the transformation of our raw data before inputting it into the DWD. Finally, we explain our method for building an unbiased estimator of the classification of untrained samples using DWD.

2.1. M-rep Shape Description

The m-rep shape description for a 3-D 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, seen in Fig. 1, is defined as $m = \{\mathbf{p}, r, \mathbf{U}^{+1}, \mathbf{U}^{-1}\} \in \mathcal{M}$, with $\mathcal{M} = \mathbb{R}^3 \times \mathbb{R}^+ \times \mathcal{S}^2 \times \mathcal{S}^2$.

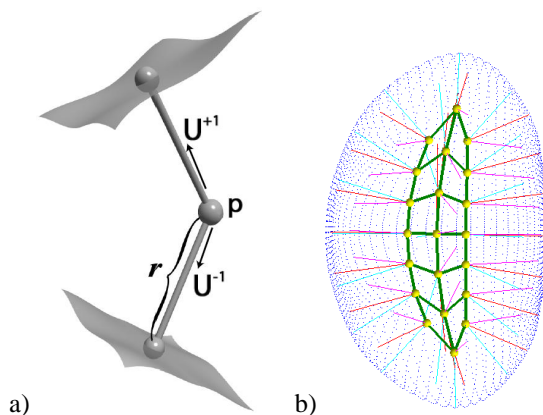


Figure 1. a) Medial atom: position (\mathbf{p}), radius (r), two normals to boundary (\mathbf{U}). b) M-rep figure composed of sheet of medial atoms. Implied boundary points displayed in blue.

To obtain m-reps describing subcortical structures, we started with binary image segmentations from well-trained experts using semi-automated procedures.* 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.²¹ 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.²² This process is applied individually to each of the 10 anatomical objects using the Pablo tool,²³ while the correspondence across samples is implicitly established by the deformation process on the template model. Fig. 2 shows the medial atoms for a set of objects (a) and the implied surfaces (b).

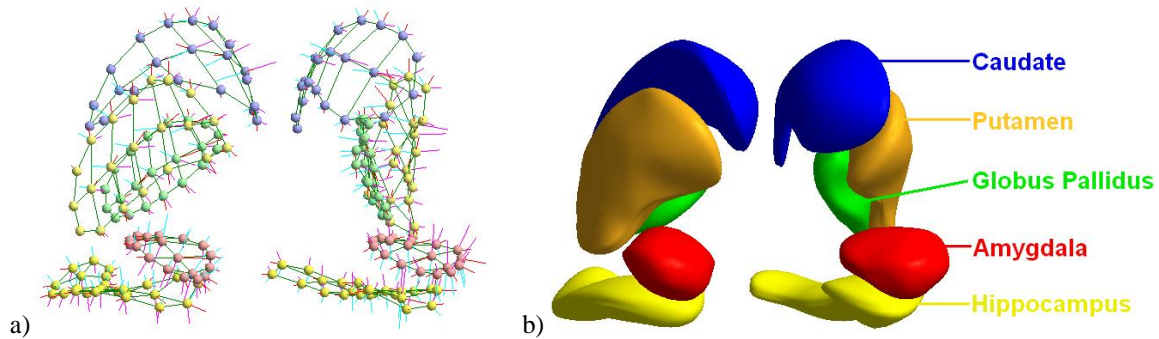


Figure 2. M-reps of a multi-object complex. a) Medial atoms. b) Implied boundary surfaces of medial description.

2.2. Estimating Variability of Multi-object Sets using Principal Geodesic Analysis

In linear space, variability of parameterized objects can be described by principle component analysis (PCA) of spherical harmonics¹¹ or point distribution models (PDM).¹⁰ However, a linear PCA cannot describe object rotations and the modeling cannot be extended to model points and normals. Non-linear modeling is achieved by principle geodesic analysis (PGA), developed by Fletcher et al.²⁴ PGA extends linear PCA into nonlinear space using “curved statistics” and is a natural generalization of PCA for data that are parameterized as curved manifolds. To recall, the intrinsic mean of a collection of points x_1, \dots, x_N on a Riemannian manifold M is the Fréchet mean $\mu = \operatorname{argmin} \sum_{i=1}^N d(x, x_i)^2$, where $d(\cdot, \cdot)$ denotes Riemannian distance on M . Whereas PCA in \mathbb{R}^3 generates linear subspaces that maximize the variance of projected data, geodesic manifolds are images of linear subspaces under the exponential map and are defined as the manifolds that maximize projected variance. Principle geodesics can be found by a recursive gradient descent or with an approximation by the log map and a linear PCA in the tangent space of the map (please see²⁴ for details). An important fact is that PGA can be used with parametrization schemes that include point locations, scale, and angle parameters.

PGA performs a compression of the multi-object shape variability to a small set of major eigenmodes of deformation. We assume that the first few modes describe most of the shape variability, the rest representing mostly noise. The quality of this compression can be evaluated with the criteria compactness, sensitivity and specificity as discussed in.²¹ As a preliminary test, we followed the standard procedure of projecting the multi-object sets into the shape space of the eigenmodes λ_i . This leads to a set of weights in the shape space that describe the deviation of individual shapes from the mean. In our case, each weight vector represents a multi-object shape set. We used these weights in exploring the different normalization environments below.

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 at once, a global alignment. As seen in Fig. 3a and b, after this global alignment there are still local pose differences among the individual objects. In our case, we assumed these single object pose differences were important because they represent the inter-object changes within the multi-object set. Therefore, after the global alignment, we perform a second step referred to as the local alignment.

*See <http://www.psychiatry.unc.edu/autismresearch/mri/roiprotocols.htm> for a detailed description of protocols and reliability results.

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 results 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 use the local pose changes as features for discriminant analysis, we have an 8-dimensional 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 are used as the pure shape features.

To align m-reps, we use a slight variation of the standard Procrustes method.²⁵ In a normal Procrustes alignment on a set of boundary points, the sum-of-squared distances between corresponding points is minimized. The standard Euclidean distance serves as the metric. For our purposes, we instead minimize the geodesic distance between m-reps because they do not lie within a Euclidean space. The geodesic distance $d(m_a, m_b)$ between two medial atoms m_a and m_b equals

$$\| \mathbf{p}_b - \mathbf{p}_a, \frac{r_b}{r_a}, \mathbf{R}(\mathbf{U}_b^{+1}) \cdot \mathbf{U}_a^{+1}, \mathbf{R}(\mathbf{U}_b^{-1}) \cdot \mathbf{U}_a^{-1} \| \quad (1)$$

where $\mathbf{R}(x)$ is the rotation of x to $(1,0,0)$. For more details, see.²⁴ The distance between two m-reps is then the sum of geodesic distances between their corresponding atoms.

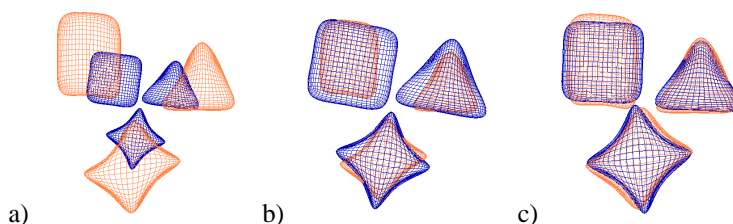


Figure 3. Multi-object alignment. a) Global translation and rotation. b) Global translation, rotation, and scale. c) Local translation, rotation, and scale after global translation and rotation.

Given the various options for alignment, we used PGA to assess the variability in the resulting aligned object sets. We started by applying PGA on objects globally aligned with and without scaling (see Figs. 3a,b). In order to compare the differences in the two global PGA's, we plotted the values of the first two major eigenmodes (λ_1, λ_2) of deformation (Fig. 4a,b). The arrows connect corresponding longitudinal pairs, which allows qualitative evaluation of the correlation between PGA modes and longitudinal changes.

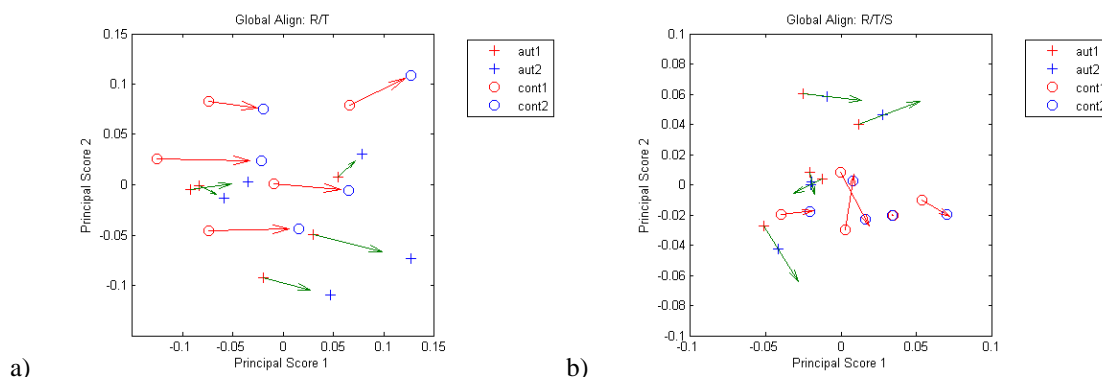


Figure 4. λ_1 vs λ_2 plot in pure global normalization settings. a) Global rotation and translation normalization; b) Global R/T/S normalization. The arrows indicate corresponding subject pairs. The range of (a) is larger than (b).

The λ_1 axis in the PGA without global scaling normalization seems to characterize mainly differences between age 2 and 4 (Fig. 4a), as indicated by the parallel alignment of the connecting arrows to the λ_1 axis. After scaling normalization (Fig. 4b) no coherent alignment of the arrows is visible. This suggests that the main effect of longitudinal change is reflected in the scaling normalization and thus the overall size of the object sets. Also, corresponding longitudinal pairs cluster quite well in the plot including scaling normalization. Thus we can be reasonably confident that by including scaling normalization, we have accounted for the major differences between ages. This allows us to use samples from all ages in our discrimination analysis.

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. 5): given points x_i , class indicators $y_i \in \{+1, -1\}$, and w the normal to the separating hyperplane, the distance or residual, r , from the points to the hyperplane is

$$r_i = y_i(x'_i w + \beta) \quad (2)$$

where β determines the position of the hyperplane. One of the popular methods of discriminant analysis is Support Vector Machines (SVM). It attempts to maximize the minimum r_i . The main problem with this method is 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. 6) 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.

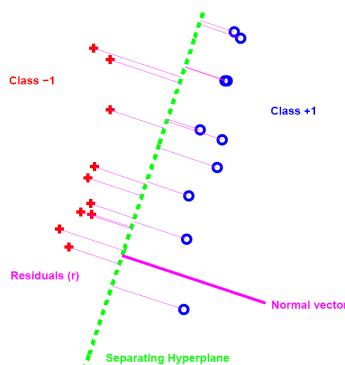


Figure 5. Illustration of two-class discrimination with separating hyperplane and residuals. The samples marked with boxes would act as the support vectors in SVM, whereas all samples are included in DWD.

Distance weighted discrimination is a method similar to SVM, but uses all sample points in the calculation of the discriminating axis. It attempts to minimize the sum of the reciprocals of r_i . Through this, each point’s contribution to the calculation is weighted 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.

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 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 towards features with larger variance. Therefore, we must first account for each of these issues before running DWD analysis on m-rep and pose features.

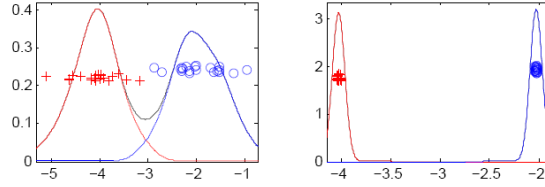


Figure 6. Left: Projection onto normal of optimal separating hyperplane. Right: Projection onto normal of separating hyperplane which exhibits data piling.

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. 7). 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.²⁴ This involves taking the log map of each of the non-Euclidean features. For the pose rotation, the log map of a unit-length quaternion $q = (w, \mathbf{v})$ is defined as

$$\log q = \frac{\theta}{\sin(\theta/2)} \cdot \mathbf{v} \quad , \quad \theta = 2\arccos(w) \quad (3)$$

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

$$\log \mathbf{U} = \left(x \cdot \frac{\theta}{\sin(\theta)}, y \cdot \frac{\theta}{\sin(\theta)} \right) \quad , \quad \theta = \arccos(z) \quad (4)$$

For the pose scale and m-rep radius factors it is just the logarithm function.

To concatenate features of differing units, 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

$$Y = \frac{X - \bar{X}}{\sigma} \forall X \in \mathbb{R}^k \quad (5)$$

$$Y = \frac{\log(X - \bar{X})}{\sigma} \forall X \notin \mathbb{R}^k \quad (6)$$

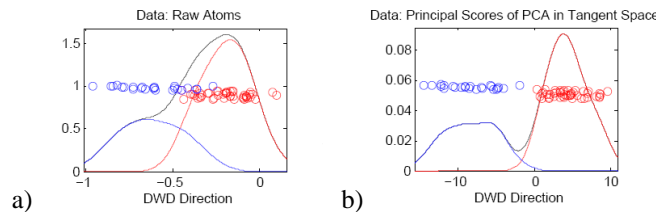


Figure 7. Separation of 70 multi-object m-reps into two populations given by DWD axis. a) Raw, nonlinear medial atom data. b) Atom data after projection into tangent space and subtraction of mean.

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 one-dimensional 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

$$s = x'w + \beta \quad (7)$$

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 strategy for selecting training samples (see Algorithm 1). We would alternately choose a random autism or a random control sample. With this sample from one group, we chose the sample from the other group that was a best match according to age and gender to the subject from the first group. This gave us one sample from each group. Since our data is longitudinal, we always included each sample's counterpart across time. From here, the process was repeated but starting with a random sample from the opposite group than in the previous iteration. After several iterations, we would have a training set with an equal number of samples from each group.

This experiment was then run many times resulting in different training sets. After several runs, all of the data samples were included in the test set at least a few times. From the results of these experiments, we could then build an unbiased estimate of each sample's classification. For each sample, we computed it's mean DWD score over those runs of the experiment for which it was in the test set. In this way, we calculate a classification for a sample only when the discriminating axis was computed without any knowledge of that sample. The box plots in the following sections are of these unbiased mean DWD scores.

Algorithm 1 Training Set T Selection

```

 $T = \emptyset$ ,  $size = 0$ ,  $i = 0$ 
while  $size < n$  do
  if  $i \bmod 2 = 0$  then
     $s$  = random sample from autism group
     $t$  = closest matching sample to  $s$  from control group
  else
     $s$  = random sample from control group
     $t$  = closest matching sample to  $s$  from autism group
  end if
   $s'$  = counterpart of  $s$  across time
   $t'$  = counterpart of  $t$  across time
   $T = T \cup \{s, s', t, t'\}$ 
   $size = size + 4$ 
   $i = i + 1$ 
end while

```

3. RESULTS

In this section, we describe our data set and the results of our leave-many-out experiment. We have divided the results into four sections corresponding to the features used in the discriminant analysis: volume, pose, shape, shape and pose combined. We then finish with some visualizations of the discriminating differences.

3.1. Motivation and Clinical Data

The driving clinical problem of this research is the need for a joint analysis of the set of subcortical brain structures, over and above that of individual structures. The image data used in this paper is taken from an ongoing clinical longitudinal pediatric autism study. 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 results shown here, 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 4 controls, we paired an age 2 scan of one subject with an age 4 scan of another unrelated subject. We also have 4 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.

In the classifier experiments, we always included all details of a selected training subject. Thus, the four unpaired control samples were always left out of the training set. From our specific data, we chose a training set, in the manner described above, consisting of 32 out of the 70 available samples. Thus, it included 16 samples from the control group and 16 from the autism 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 for at least a few runs; the minimum number of runs in the test set for any sample turned out to be 4. From these, we could calculate an unbiased mean DWD score.

3.2. 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 10-dimensional feature space to serve as the basis for the discriminating axis. 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. 8a shows that there is a clear difference between the mean of the autism group and control group. This difference is statistically significant with $p < 0.0001$. As another measure of the discriminatory performance of the volume, Table 1 shows the average percentage (71%) of the 38 test samples that were correctly classified over the 100 runs of the leave-many-out experiment.

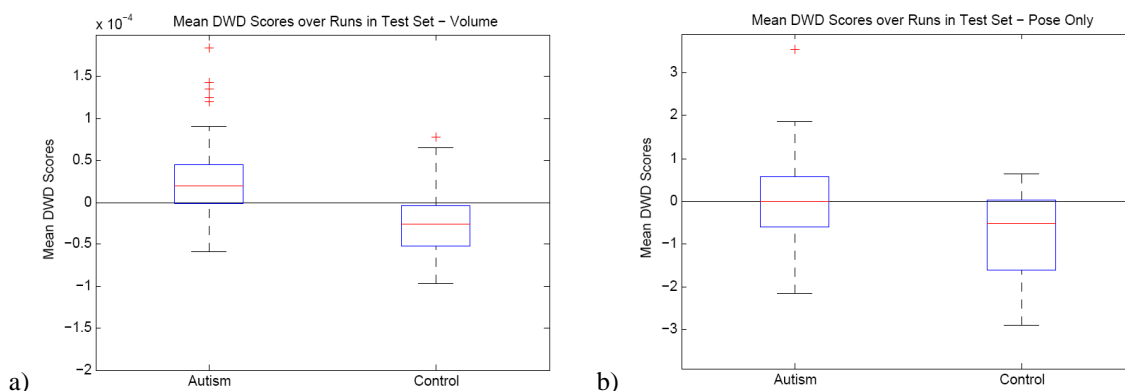


Figure 8. Box plot (median, 25 and 75 percentiles, min/max) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, less than zero classified as control. a) Volume features, $p < 0.0001$. b) Pose features, $p = 0.001$.

3.3. Pose

The next step was to explore the significance of local pose changes. For each sample, these features totaled 70: three for translation, three for rotation, and one for uniform scale across 10 objects. The raw features were transformed as described above, thus reducing the quaternion representing the rotation to a three-dimensional vector. The significance in the volume discrimination led us to believe that the pose, which includes uniform object scale factors, would also show significance. This is the case for the mean DWD scores, $p = 0.001$ (Fig. 8b), but the test sample classification accuracy was considerably lower than volume as Table 1 illustrates. Also, there were 24 individual runs in which the classification accuracy was below 50%, 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 to the DWD calculation because the same experiment run with only the scale factors gave an average classification rate of 67% as opposed to 56% using the entire pose. From these results, we conclude that the pose does include some relevant information for discrimination but that most of it is likely in the uniform scale factors.

3.4. Shape

Fig. 9 shows the results of using only the m-rep shape features for the DWD calculations. Once again, the mean DWD scores for the test samples were significantly different ($p < 0.0001$, Fig. 9a). Also, the classification accuracy of shape was better than pose with an average correctness rate of 60%. More encouraging was the much lower number (6) of runs with

less than 50% accuracy when compared with pose. The DWD methodology proved its usefulness and stability in high dimensional low sample size settings here because the m-rep shape features for the 10 subcortical structures number about 2000 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 best the performance of the pose features.

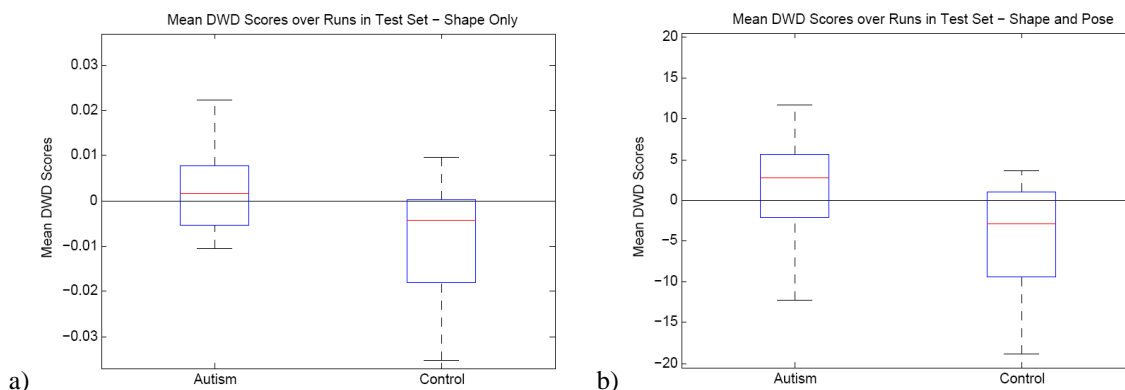


Figure 9. Box plot (median, 25 and 75 percentiles, min/max) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, less than zero classified as control. a) Shape features, $p < 0.0001$. b) Shape and pose features combined, $p < 0.0001$.

3.5. 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 multi-object complex. The differences between the mean DWD scores remained significant ($p < 0.0001$). The results shown in Table 1 reflect this in that the average classification rate (64%) was higher using both pose and shape than either individually. Likewise, there were no individual runs in which the classification accuracy was below 50%.

3.6. Shape and Scale: M-rep Radii

One of the ongoing questions with a discretely sampled m-rep shape description is the issue of correspondence. Currently, the correspondence is implicitly defined in the process of fitting a template model to each binary image. However, it is not clear that features such as the spokes and atom positions are sufficiently stable to give a meaningful correspondence across samples. 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 with a better correspondence.

To explore this question, we chose to run our same analysis on only the radius feature of each medial atom without any alignment. This is because the radius is less sensitive to a noisy correspondence than some of the other medial atom features; 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 including both shape and scale information in the form of local width. This makes it a more intrinsic measure than the other medial atom features. Compared to the other features in

Table 1. Classification Accuracies for Test Samples over 100 Runs of Leave-Many-Out Experiment

Feature	Mean	Std. Dev.
Volume	71 %	± 9 %
Pose	56 %	± 10 %
Shape	60 %	± 7 %
Shape and Pose	64 %	± 7 %
Shape and Scale (Radii Only)	75 %	± 7 %

Table 1, the mean classification accuracy over 100 runs of the leave-many-out experiment is best when using only the atom radii at 75%. Fig. 10 shows the mean DWD scores using only the radii. While the radii give the best classification rate of any features 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. 10 shows that the overlap between the populations is not drastically better than with the other features. However, the jump in classification accuracy from using all 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.

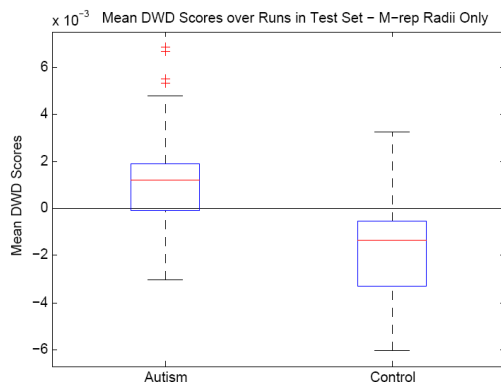


Figure 10. M-rep radii features: Box plot (median, 25 and 75 percentiles, min/max) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, less than zero classified as control. $p < 0.0001$.

3.7. Evaluation of Bias

To verify that the mean classification scores were unbiased, we ran our same experiments using random, normally distributed input data. We used the same random number seeding and the exact same training and testing sets. The random data was generated with the same mean, variance, and dimension as our actual shape data. The p-value of the mean DWD scores for this case was 0.22 and the average classification accuracy was 49%.

3.8. Visualization

To visualize the changes in shape along the DWD direction, we start with the mean m-rep of the autism group. Then, we deform the autism mean m-rep along the unit-length DWD which points 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 line. The final m-rep which has been deformed 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. We also used all 70 samples to compute the distance between the projected group means, which gives us the distance to deform along the DWD direction, and all 46 autism samples for the autism mean m-rep. Fig. 11 shows colormaps of surface distances between the two multi-object sets representing each of the two diagnosis groups using shape only. The measurement is the distance when starting from the autism group and deforming towards the control group. Red, green, and blue coloring denote inward, zero, and outward deformations respectively. In this image, 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 (seen in far right of Fig. 11a) as well as an outward change in the midsection. A good portion of each amygdala is covered with some non-green coloring denoting a change in shape as well.

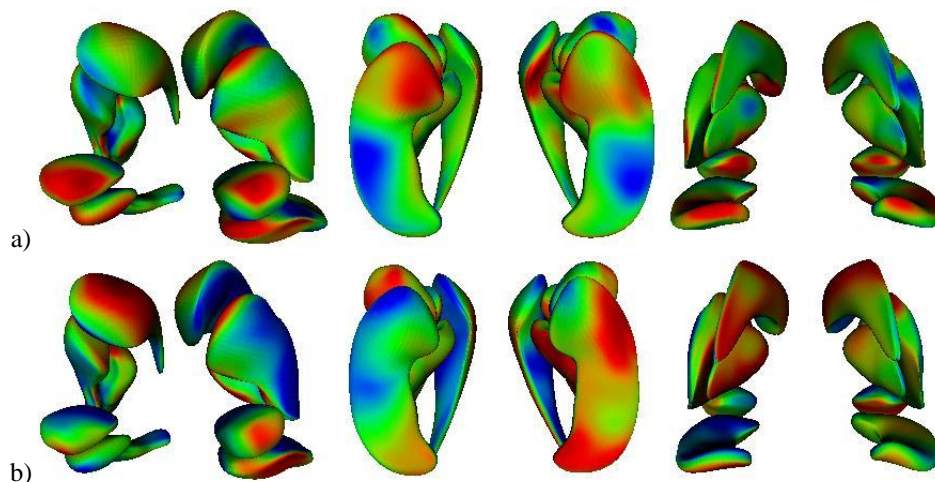


Figure 11. Colormap of surface distances from autism mean m-rep to deformed m-rep along DWD direction using a) shape only, b) shape and pose. Red, green, and blue are inward distance, zero distance, and outward distance respectively.

4. CONCLUSION

This research demonstrates work in progress towards shape analysis and group discrimination of multi-object complexes. Traditionally, shape analysis is mostly concerned with representation and statistical analysis of single objects, mostly following a well developed mathematical framework that proposes linear alignment and subsequent statistical analysis of corresponding features.

In a multi-object setting, the alignment step has to be reconsidered. Linear alignment of a population of sets of objects will remove global translation, rotation and scale, but will not account for relative object pose variability. A joint analysis of only globally aligned sets of shapes will therefore include these residual pose differences into the statistical shape model. Here, we discuss and explore the various options for global and local alignment of sets of shapes. We propose an initial global alignment with rotation and translation to map each dataset into a common coordinate frame. This step is followed by a local alignment of each object individually, but the alignment parameters translation, rotation and scale are kept as pose parameter vectors. Shape analysis of the joint set of objects will therefore use *pure* shape features not affected by any residual pose differences. Features are mapped into Riemannian symmetric space, the appropriate choice for medial atom features that include rotational frames and positive reals, and are ready for statistical analysis. The same technique can be applied to the vectors of the joint pose parameters of the multi-object complexes. It is then straightforward to chose pose, shape, or pose and shape as features for group discrimination.

Although sampled medial representations use a lower number of features than densely sampled surfaces, we still face the HDLSS problem (high dimensionality low sample size). This problem is even more pronounced with the analysis of object sets, resulting in a feature space dimensionality which is magnitudes larger than the number of samples. In typical applications similar to the one described here, two populations of 25 samples are each represented by 2000 features to provide a sufficiently detailed representation for 10 3-D objects. For classification, we applied the distance-weighted discrimination (DWD) method, which is a variant of support vector machine discrimination but is designed to be robust for HDLSS data analysis problems. Unbiased statistical analysis by repeated leave-many-out analysis finally results in classification rates and significance values (p-values).

The driving application is a pediatric autism study with autistic and typically developing 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 after model fitting. The combination of pose and shape performed better than either did individually, and shape analysis was better than analysis of pose. Please note the discrepancy of relatively low classification rates in the presences of highly significant population differences. This might possibly be explained by the nature of the underlying clinical problem. Morphologic phenotypes in neurodevelopmental disorders are often reflected by only subtle differences and increased heterogeneity. In the future, we will explore multi-variate classification by selection of a best-separating

subspace rather than a single axis. Further, we will have to develop a technique to explore the covariance structure of sets of shapes in order to explain their interrelationship. This will help clinicians to explore links between morphological changes and underlying biological processes.

5. ACKNOWLEDGEMENTS

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

REFERENCES

1. I. Dryden and K. Mardia, "Multivariate shape analysis," *Sankhya* **55**, pp. 460–480, 1993.
2. C. G. Small, *The statistical theory of shape*, Springer, 1996.
3. A. Litvin and W. C. Karl, "Coupled shape distribution-based segmentation of multiple objects," in *Information Processing in Medical Imaging*, pp. 345–356, 2005.
4. I. Dryden and K. Mardia, *Statistical Shape Analysis*, Wiley, 1998.
5. P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis, "Detection and analysis of statistical differences in anatomical shape," *Medical Image Analysis* **9**, pp. 69–86, 2005.
6. P. Yushkevich, S. M. Pizer, S. Joshi, and J. Marron, "Intuitive, localized analysis of shape variability," in *Information Processing in Medical Imaging*, pp. 402–408, 2001.
7. 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. of the National Academy of Sciences* **102**, pp. 4872–4877, March 2005.
8. M. N. Bossa and S. Olmos, "Statistical model of similarity transformations: Building a multi-object pose," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, p. 59, 2006.
9. F. Bookstein, "Shape and the information in medical images: A decade of the morphometric synthesis," in *MMBIA*, 1996.
10. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," in *Computer Vision and Image Understanding*, pp. 38–59, 1995.
11. A. Kelemen, G. Székely, and G. Gerig, "Elastic model-based segmentation of 3d neuroradiological data sets," in *IEEE Trans. Med. Imaging*, **18**, pp. 828–839, 1999.
12. C. Davatzikos, M. Vaillant, S. Resnick, J. Prince, S. Letovsky, and R. Bryan, "A computerized method for morphological analysis of the corpus callosum," *J. of Comp. Assisted Tomography* **20**, pp. 88–97, Jan./Feb 1996.
13. 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. Natl. Acad. Sci. USA* **95**, pp. 11406–11411, September 1998.
14. P. Thompson, M. Mega, and A. Toga, *Brain Mapping: The Disorders*, ch. Disease-Specific Brain Atlases. Academic Press, 2000.
15. 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* **404**, pp. 190–193, 2000.
16. 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 Transactions on Medical Imaging* **22**, pp. 137–154, Feb 2003.
17. J. Yang, L. H. Staib, and J. S. Duncan, "Neighbor-constrained segmentation with level set based 3d deformable models," *IEEE Transactions on Medical Imaging* **23**, pp. 940–948, Aug 2004.
18. S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape," in *IEEE Trans. Med. Imaging*, **18**, pp. 851–865, 1999.
19. M. Styner, J. A. Lieberman, D. Pantazis, and G. Gerig, "Boundary and medial shape analysis of the hippocampus in schizophrenia," in *MedIA*, pp. 197–203, 2004.
20. S. Bouix, J. C. Pruessner, D. L. Collins, and K. Siddiqi, "Hippocampal shape analysis using medial surfaces," *NeuroImage* **25**, pp. 1077–1089, 2005.
21. M. Styner, G. Gerig, J. Lieberman, D. Jones, and D. Weinberger, "Statistical shape analysis of neuroanatomical structures based on medial models," *Medical Image Analysis (MEDIA)* **7**, pp. 207–220, Sept 2003.
22. 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," *International Journal of Computer Vision IJCV* **55**(2), pp. 85–106, 2003.
23. 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 Journal* **32**, pp. 1335–1345, May 2005.
24. P. Fletcher, C. Lu, S. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," in *IEEE Transactions on Medical Imaging*, **23**, pp. 995–1005, 2004.
25. C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society* **53**(2), pp. 285–339, 1991.