Statistical Shape Analysis of Multi-Object Complexes

Kevin Gorczowski, Martin Styner, Ja-Yeon Jeong, J. S. Marron, Joseph Piven,
Heather Cody Hazlett, Stephen M. Pizer, Guido Gerig
University of North Carolina at Chapel Hill
{kgorcz,styner,gerig}@cs.unc.edu marron@stat.unc.edu

Abstract

An important goal of statistical shape analysis is the discrimination between populations of objects, exploring group differences in morphology not explained by standard volumetric analysis. Certain applications additionally require analysis of objects in their embedding context by joint statistical analysis of sets of interrelated objects. In this paper, we present a framework for discriminant analysis of populations of 3-D multi-object sets. In view of the driving medical applications, a skeletal object parametrization of shape is chosen since it naturally encodes thickening, bending and twisting. In a multi-object setting, we not only consider a joint analysis of sets of shapes but also must take into account differences in pose. Statistics on features of medial descriptions and pose parameters, which include rotational frames and distances, uses a Riemannian symmetric space instead of the standard Euclidean metric. Our choice of discriminant method is the distance weighted discriminant (DWD) because of its generalization ability in high dimensional, low sample size settings. Joint analysis of 10 subcortical brain structures in a pediatric autism study demonstrates that multi-object analysis of shape results in a better group discrimination than pose, and that the combination of pose and shape performs better than shape alone. Finally, given a discriminating axis of shape and pose, we can visualize the differences between the populations.

1. Introduction

Statistical shape modeling and analysis [7, 18] 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 neurolocal 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 both discrimination and segmentation. Litvin et al. [13], 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 [8]. We use the distance weighted discrimination (DWD) method [14], which is similar to Support Vector Machines (SVM) but 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. [10] using distance transforms for shape features and SVM to discriminate populations. Yushkevich et al. [24] also used SVM to discriminate 2-D 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. [20] 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. [2]. Since multi-object analysis of subcortical structures is novel, we have chosen to evaluate several different features, namely pose, shape, and the combination of pose and shape.

Several different geometric shape representations have been used to model anatomy, such as landmarks [1], dense collection of boundary points [4], or spherical harmonic decompositions [12]. Another shape variability approach focuses on the analysis of deformation maps [6, 5, 22].

1-4244-1180-7/07/$25.00 ©2007 IEEE

Whereas Yang et al. [23] 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 [16]. Styner et al. [21] 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. [3] studied hippocampi using the radius function defined on a flattened 2-D medial sheet. Medial representations are also advantageous when attempting to describe certain nonlinear shape deformations such as bending and twisting [19]. 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 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. 1a, is defined as $m = \{\mathbf{p}, r, \mathbf{U^{+1}}, \mathbf{U^{-1}}\} \in \mathcal{M}$, with $\mathcal{M} = \mathbb{R}^3 \times \mathbb{R}^+ \times \mathcal{S}^2 \times \mathcal{S}^2$.

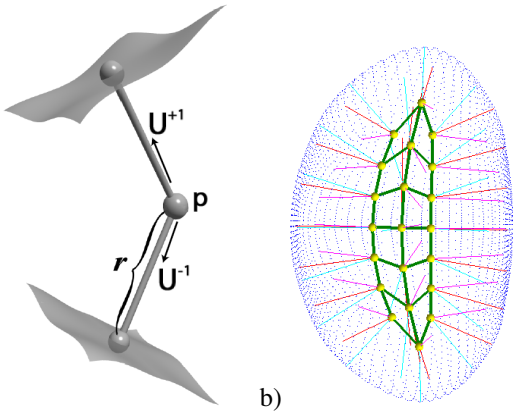

Figure 1. a) Medial atom: position ($\mathbf{p}$), radius (r), two normals to boundary ($\mathbf{U}$). b) M-rep figure composed of sheet of medial atoms. Implied boundary points displayed in blue.

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. [19] 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 while being penalized for becoming too irregular in its geometry [15].

2.2. 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. 2a 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. 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. 2c. 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 [11]. 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

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

where $\mathbf{R}(x)$ is the rotation of x to (1,0,0). For more details, see [9]. The distance between two m-reps is then the sum of geodesic distances between their corresponding atoms.

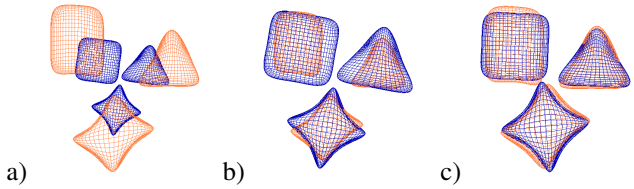

Figure 2. Multi-object alignment. a) Global translation and rotation. b) Global translation, rotation, and scale. c) Local translation, rotation, and scale after global translation and rotation.

2.3. 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. 3): 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

$$r_i = y_i(x_i' w + \beta) \qquad (2)$$

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. 4) 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.

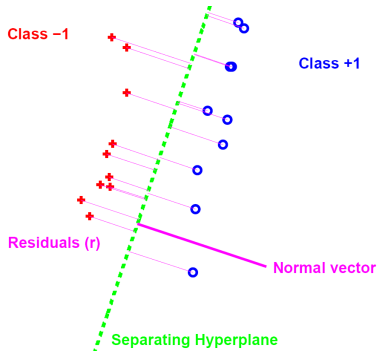

Figure 3. Illustration of two-class discrimination with separating hyperplane and residuals.

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.

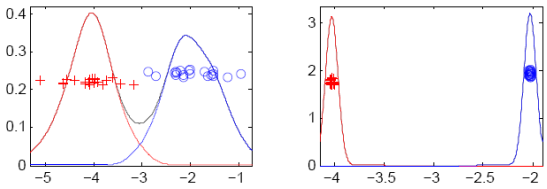

Figure 4. Left: Projection onto normal of optimal separating hyperplane. Right: Projection onto normal of separating hyperplane which exhibits data piling.

2.4. 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.

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. 5). 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 [9]. 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 $q = (w, \underline{\mathbf{v}})$ is defined as

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

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

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

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

$$Y = \frac{X - \overline{X}}{\sigma} \forall X \in \mathbb{R}^k \tag{5}$$

$$Y = \frac{\log\left(X - \overline{X}\right)}{\sigma} \forall X \notin \mathbb{R}^k \tag{6}$$

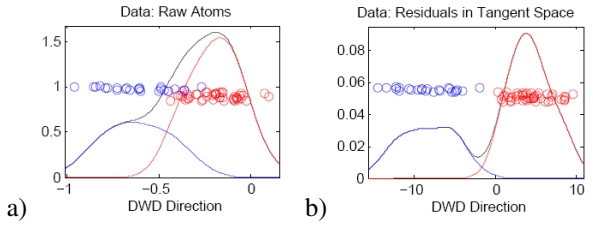

Figure 5. Separation of 70 multi-object m-reps into two populations given by DWD axis. a) Raw, nonlinear medial atom data. b) Atom data after projection into tangent space and subtraction of mean.

2.5. 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

$$s = x'w + \beta \tag{7}$$

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.

Algorithm 1 Training Set T Selection

$T = \varnothing$, $size = 0$, $i = 0$
while $size < n$ **do**
  if i mod 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

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.

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 three sections corresponding to the features used in the discriminant analysis: pose, shape, shape and pose combined. We then finish with some visualizations of the discriminating features.

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 images 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 both time points of a selected training subject for inclusion in the training set. Thus, the four control samples with only one time point 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. Shape Modeling

The process of fitting an m-rep to a binary image as described above was applied individually to each of the 10 anatomical objects using the Pablo tool [17]. The correspondence across samples is implicitly established by using a template model as the starting point of the deformation process. Fig. 6 shows medial atoms (a) and their implied surfaces (b) for the set of subcortical structures.

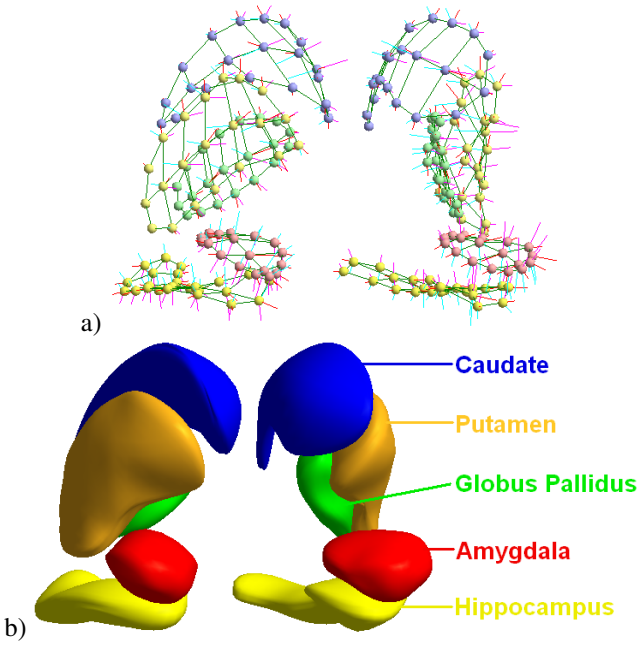

Figure 6. M-reps of a multi-object complex. a) Medial atoms. b) Implied boundary surfaces of medial description.

3.3. Pose

The first step of our experiment was to explore the significance of local pose changes. For each sample, these features totalled 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 knowledge that there are volume differences between autistic and typical brains led us to believe that the pose, which includes uniform object scale factors, might show significance on its own. This is the case for the mean DWD scores, p=0.001. Fig. 7 shows the distributions of the mean DWD scores for each group. The red line is the median, the top and bottom of the blue box are the 25^{th} and 75^{th} percentiles, the extensions to the box are the minimum and maximum excluding outliers, and red plus signs are outliers defined as being more than 1.5 times the interquartile range from the box. Although the p-value for the mean DWD scores shows significance, the test sample classification accuracy, which is the percentage of the 38 test samples correctly classified as autism or control, was poor (54%, see Table 1). Also, 24 out of the 100 individual runs had a classification accuracy 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%. 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.

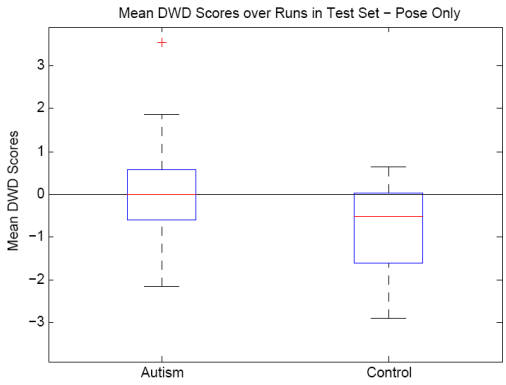

Figure 7. Pose features: Box plot (median, 25 and 75 percentiles, min/max) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, less than zero classified as control. p=0.001.

Table 1. Classification Accuracies for Test Samples over 100 Runs of Leave-Many-Out Experiment and P-values of Mean Classification Scores

Feature	Mean	Std. Dev.	P-value
Pose	56 %	± 10 %	0.001
Shape	60 %	± 7 %	< 0.0001
Shape and Pose	64 %	± 7 %	< 0.0001

3.4. Shape

Fig. 8 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. 8). 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. Here the DWD methodology proved its usefulness and stability in high dimensional low sample size settings because the m-rep shape features for the 10 subcortical structures number about 2000 in total, whereas the pose had 70. Even in this high dimensional space, the DWD still generalized well enough to best the performance of the pose features.

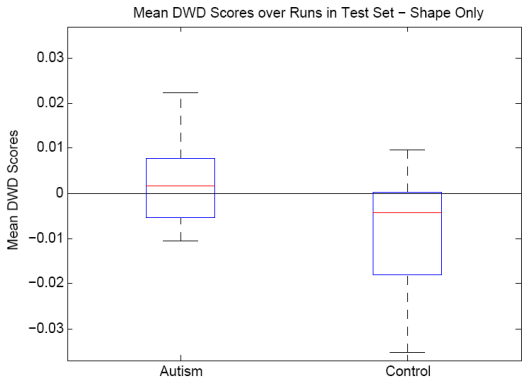

Figure 8. Shape features: Box plot (median, 25 and 75 percentiles, min/max) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, less than zero classified as control. p<0.0001.

3.5. Shape and Pose

Finally, we concatenated the pose and shape parameters into one feature vector. This gave us the most complete description of the variability of the multi-object complex. 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%. The differences between the mean DWD scores remained significant (p<0.0001).

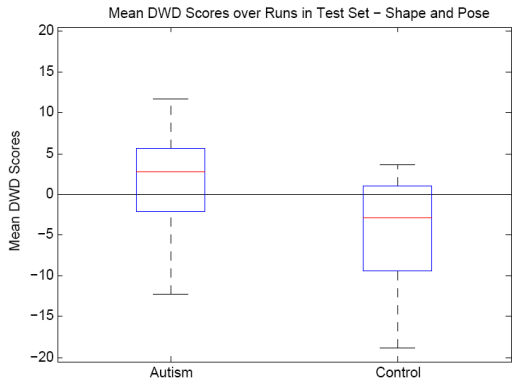

Figure 9. Shape and pose features combined: Box plot (median, 25 and 75 percentiles, min/max) of mean DWD scores of each group over those runs in which the samples were in the test set. Greater than zero classified as autism, less than zero classified as control. p<0.0001.

3.6. 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%.

3.7. 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.

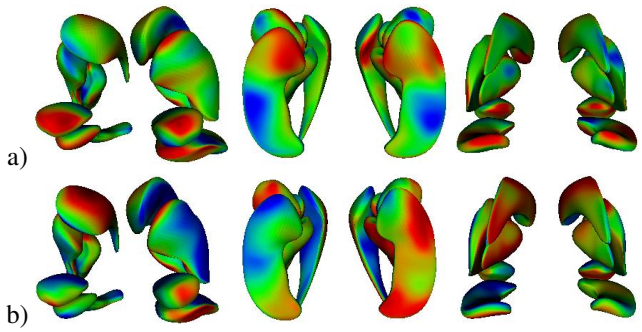

Figure 10. Colormap of surface distances from autism mean m-rep to deformed m-rep along DWD direction using a) shape only, b) shape and pose. Red, green, and blue are inward distance, zero distance, and outward distance respectively.

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. 10 shows colormaps of surface distances between the two multi-object sets representing each of the two diagnosis groups. 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 Fig. 10a, 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. 10a) 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.

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 multi-object complexes 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 choose 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 experiments 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 individu-

ally, 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. The purpose is not to achieve highly accurate classifications (they are diagnosed by psychiatrists using non-imaging scores), but to study morphologic changes as part of a better understanding of phenotypes and biological mechanisms. The reported group differences have to be interpreted with caution since there is a mismatched gender ratio between the patient and control groups; an appropriate correction scheme is in development. In the future, we would like 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. The MRI datasets and expert segmentations were funded by NIH RO1 MH61696 and NIMH MH64580.

References

[1] F. Bookstein. Shape and the information in medical images: A decade of the morphometric synthesis. In *MMBIA*, 1996.

[2] 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)*, page 59, 2006.

[3] S. Bouix, J. C. Pruessner, D. L. Collins, and K. Siddiqi. Hippocampal shape analysis using medial surfaces. *NeuroImage*, 25:1077–1089, 2005.

[4] 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*, pages 38–59, 1995.

[5] 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*, 95:11406–11411, September 1998.

[6] 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*, 20:88–97, Jan./Feb 1996.

[7] I. Dryden and K. Mardia. Multivariate shape analysis. *Sankhya*, 55:460–480, 1993.

[8] I. Dryden and K. Mardia. *Statistical Shape Analysis*. Wiley, 1998.

[9] 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*, volume 23, pages 995–1005, 2004.

[10] P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9:69–86, 2005.

[11] C. Goodall. Procrustes methods in the staistical analysis of shape. *Journal of the Royal Statistical Society*, 53(2):285–339, 1991.

[12] A. Kelemen, G. Székely, and G. Gerig. Elastic model-based segmentation of 3d neuroradiological data sets. In *IEEE Trans. Med. Imaging*, volume 18, pages 828–839, 1999.

[13] A. Litvin and W. C. Karl. Coupled shape distribution-based segmentation of multiple objects. In *Information Processing in Medical Imaging*, pages 345–356, 2005.

[14] J. Marron and M. Todd. Distance weighted discrimination. Technical report 1339, Operations Research and Industrial Engineering, Cornell University, 2002.

[15] 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):85–106, 2003.

[16] 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*, volume 18, pages 851–865, 1999.

[17] 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 Physiscs Journal*, 32(5):1335–1345, May 2005.

[18] C. G. Small. *The statistical theory of shape*. Springer, 1996.

[19] 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(3):207–220, Sept 2003.

[20] 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(13):4872–4877, March 2005.

[21] M. Styner, J. A. Lieberman, D. Pantazis, and G. Gerig. Boundary and medial shape analysis of the hippocampus in schizophrenia. In *MedIA*, pages 197–203, 2004.

[22] 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*, 404:190–193, 2000.

[23] 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(8):940–948, Aug 2004.

[24] P. Yushkevich, S. M. Pizer, S. Joshi, and J. Marron. Intuitive, localized analysis of shape variability. In *Information Processing in Medical Imaging*, pages 402–408, 2001.