Morphometric Analysis of Brain Structures for Improved Discrimination

Li Shen[1], James Ford[1], Fillia Makedon[1], Yuhang Wang[1], Tilmann Steinberg[1],
Song Ye[1], and Andrew Saykin[2]

[1] DEVLAB, Computer Science, Dartmouth College, Hanover, NH 03755, USA
{li,jford,makedon,wyh,tilmann,yesong}@cs.dartmouth.edu,
[2] Psychiatry and Radiology, Dartmouth Medical School, Lebanon, NH 03756, USA
saykin@dartmouth.edu

Abstract. We perform discriminative analysis of brain structures using morphometric information. Spherical harmonics technique and point distribution model are used for shape description. Classification is performed using linear discriminants and support vector machines with several feature selection approaches. We consider both inclusion and exclusion of volume information in the discrimination. We perform extensive experimental studies by applying different combinations of techniques to hippocampal data in schizophrenia and achieve best jackknife classification accuracies of 95% (whole set) and 90% (right-handed males), respectively. Our results find that the left hippocampus is a better predictor than the right in the complete dataset, but that the right hippocampus is a stronger predictor than the left in the right-handed male subset. We also propose a new method for visualization of discriminative patterns.

1 Introduction

Classification via morphometric analysis is an important and challenging problem in medical image analysis. The goal is to identify morphometric abnormalities in a structure of interest that are associated with a particular condition to aid diagnosis and treatment. This paper studies classification techniques using shape and volume information. The techniques are demonstrated on MRI hippocampal data sets: (1) S_{all} of 35 schizophrenics and 21 controls; (2) S_{rhm} of 25 schizophrenics and 14 controls, all of whom are right-handed males from S_{all}.

There are several previous classification studies on hippocampal shape in schizophrenia. Csernansky et al. [1,2] studied hippocampal morphometry using an image-based deformation representation, and achieved jackknife accuracies of 80% and 74% through principal component analysis (PCA) and a linear discriminant. Golland, Timoner, et al. [3,4] conducted amygdala-hippocampus complex studies using distance transformation maps and displacement fields as shape descriptors, and achieved best accuracies of 77% and 87%, respectively, using support vector machines (SVMs). We studied hippocampal shape classification in [5] using a symmetric alignment model and binary images, and achieved 96% accuracy using only the second principal component after PCA.

Fig. 1. (a) Object surface and SPHARM reconstructions using coefficients up to degrees 1, 5 and 12. (b) Sample reconstructions: 2 controls and 2 patients.

The SPHARM description [6] is a parametric surface description using spherical harmonics as basis functions. It is a powerful surface modeling approach for arbitrarily shaped but simply-connected objects. It is suitable for surface comparison and can deal with protrusions and intrusions. Gerig, Styner, et al., have done numerous SPHARM studies for 3D medial shape (m-rep) modeling [7,8], model-based segmentation [9], and identifying statistical shape abnormalities of different neuroanatomical structures [7,8,10]; see [11] for a complete list. They have also used SPHARM in [12] for calculating hippocampal asymmetry, combined it with volume, and achieved 87% accuracy using SVM. Our previous study [13] closely followed this SPHARM model and combined it with PCA and Fisher's linear discriminant (FLD) in a hippocampal shape classification study and achieved 77% accuracy. This paper extends our previous work by integrating additional classification techniques and feature selection approaches as well as considering both shape and volume in order to obtain an improved classification. In addition, we introduce a new method for visualizing discriminative patterns.

2   Shape Description

We adopt the SPHARM expansion technique [6] to create a shape description for surfaces of 3D simply-connected volumetric objects (see Figure 1(a) for a hippocampal surface). Three steps are involved as follows.

(1) **Surface parameterization** creates a continuous and uniform mapping from the object surface to the surface of a unit sphere via an optimization for preserving area and minimizing angular distortions. The result is a bijective mapping between each point $\mathbf{v}$ on a surface and spherical coordinates θ and ϕ.

(2) **SPHARM expansion** expands the object surface into a complete set of SPHARM basis functions Y_l^m, where Y_l^m is the spherical harmonic of degree l and order m. The expansion takes the form: $\mathbf{v}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \mathbf{c}_l^m Y_l^m(\theta, \phi)$, where $\mathbf{c}_l^m = (c_{xl}^m, c_{yl}^m, c_{zl}^m)^T$. The coefficients $\mathbf{c}_l^m$ up to a user-desired degree can be estimated by solving a set of linear equations in a least-squares fashion. The object surface can be reconstructed using these coefficients, and using more coefficients leads to a more detailed reconstruction. See Figure 1(a) for an example.

(3) **SPHARM normalization** creates a shape descriptor (i.e., excluding translation, rotation, and scaling) from a normalized set of SPHARM coefficients,

which are comparable across objects. *Rotation invariance* is achieved by aligning the degree 1 ellipsoid. *Scaling invariance* is achieved by normalizing the object volume. Ignoring the degree 0 coefficient results in *translation invariance*. Figure 1(b) shows four reconstruction examples after normalization.

Using a uniform icosahedron subdivision of spherical surfaces (see [9]), we convert the SPHARM coefficients to a dual landmark representation, which is more intuitive. We use icosahedron subdivision level 3, resulting in $n = 642$ landmarks for each object and $3n = 1926$ feature elements. Clearly, we have many more dimensions than training objects. PCA [14] is applied to reduce dimensionality to make classification feasible. This involves eigenanalysis of the covariance matrix $\mathbf{\Sigma}$ of the data: $\mathbf{\Sigma P} = \mathbf{DP}$, where the columns of $\mathbf{P}$ hold eigenvectors, and the diagonal matrix $\mathbf{D}$ holds eigenvalues of $\mathbf{\Sigma}$. The eigenvectors in $\mathbf{P}$ can be ordered decreasingly according to respective eigenvalues, which are proportional to the variance explained by each eigenvector. Now any shape $\mathbf{x}$ in the data can be obtained using $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb}$, where $\mathbf{b}$ is a vector containing the components of $\mathbf{x}$ in basis $\mathbf{P}$, which are called *principal components*. Since eigenvectors are orthogonal, $\mathbf{b}$ can be obtained using $\mathbf{b} = \mathbf{P}^T(\mathbf{x} - \bar{\mathbf{x}})$.

Given a dataset of m objects, the first $m-1$ principal components are enough to capture all the data variance. Thus, $\mathbf{b}$ becomes an $m-1$ element vector, which can be thought of a more compact representation of the shape $\mathbf{x}$. This model is a point distribution model (PDM) [9]. We apply PDM to each hippocampal data set to obtain a $\mathbf{b}$ (referred to as a *feature vector* hereafter) for each shape.

3 Classifiers

We examine several variants of Fisher's linear discriminants (FLDs) and linear support vector machines (SVMs) in our study.

FLD projects a training set (consisting of c classes) onto $c - 1$ dimensions such that the ratio of between-class and within-class variability is maximized, which occurs when the FLD projection places different classes into distinct and tight clumps. See [14] for how to calculate this optimal projection $\mathbf{W}_{\text{opt}}$. We have only two classes, and the FLD basis $\mathbf{W}_{\text{opt}}$ becomes a column vector $\mathbf{w}$. New feature vectors can be compared to the training set, and thus classified, by projecting them onto $\mathbf{w}$. We choose four approaches to perform classification in the FLD space: (1) FLD-BM, (2) FLD-1NN, (3) FLD-3NN, and (4) FLD-NM.

FLD-BM assumes a normal distribution in the FLD space for each class. Using a Bayesian model (BM), the certainty that a test subject could be explained by each class's distribution can be calculated based on the training set (see [13] for details). FLD-BM assigns a new object to the class corresponding to the largest posterior probability computed by the BM. **FLD-1NN** and **FLD-3NN** are two k nearest neighbour (kNN) classifiers with $k = 1$ and $k = 3$ respectively. A kNN classifier assigns a new object to the most common class in the k nearest labelled training objects. **FLD-NM** is a nearest mean (NM) classifier, which assigns a new object to the class having the nearest mean.

A linear classifier corresponds to a decision hyperplane separating different classes. The margin is defined as the distance from the decision hyperplane to the closest training set exemplar. The aim in training a linear SVM is to find the separating hyperplane with the largest margin; the expectation is that the larger the margin, the better the generalization of the classifier. We employ the publicly available *OSU SVM Classifier Matlab Toolbox version 3.00* [15]. We denote a linear SVM classifier with the cost of the constraint violation x as **SVM-Cx**. **SVM-C1**, **SVM-C10**, **SVM-C100** are applied in our experiments. We refer the readers to [14,15] for more technical and implementation details.

4 Experimental Studies

This work, unlike our previous work [13], uses PCA applied to all data in a single step, rather than constructing a new basis for each jackknife trial based on individual training sets. This is a simpler approach that should minimize representation errors. Classification is performed on feature vectors after PCA, and a jackknife approach [14] is used for accuracy estimation. Two alternatives for feature selection are considered: (1) use varying numbers of principal components to test classification; (2) use varying numbers of significant features (SFs).

SFs are based on the observation that selecting an appropriate subset of features often improves classification accuracy. In our study, features are principal components (PCs), and we feel that some PCs are more *important* than others for classification, but not necessarily matching the ordering of the variance amounts they explain. To rank the effectiveness of features, we obtain a p-value for each feature via a two-sample t-test. A lower p-value implies stronger group difference statistically and corresponds to a more *significant* feature. We hypothesize that more significant features can help more in classification.

We investigate three feature selection schemes in experiments: (1) **PC**: use the first n PCs; (2) **SF**: use the first n SFs, where the SF ordering is determined by running t-test on each feature using all the objects; (3) **SF-J**: use the first n SFs, as in **SF**, but with the t-test applied only to each jackknife training set.

We examine two hippocampal data sets[1]: (1) $\mathbf{S_{all}}$ of 35 schizophrenics and 21 controls, and (2) $\mathbf{S_{rhm}}$ of 25 schizophrenics and 14 controls, all of whom are right-handed males from $\mathbf{S_{all}}$. In each set, left and right hippocampi are studied separately. We use $\mathbf{S}_X^Y$ to denote the set of Y ($\in$ {left, right}) hippocampi in $\mathbf{S}_X$, where $X \in$ {all, rhm}. Thus, we have 4 data sets: $\mathbf{S}_{all}^{left}$, $\mathbf{S}_{all}^{right}$, $\mathbf{S}_{rhm}^{left}$ and $\mathbf{S}_{rhm}^{right}$. We have also 7 classifiers and 3 feature selection schemes. Our experiments include every combination, but due to space limitations we present a few typical examples in detail. Figure 2 shows the experimental results of FLD-BM and SVM-C10 on $\mathbf{S}_{all}^{left}$ and $\mathbf{S}_{rhm}^{right}$. The following observations are true for all the experiments: **(1)** The **SF** results show a nearly perfect classification for each classifier in the best case; however, in this case feature selection introduces

---

[1] All the hippocampi are manually segmented from the MRI scans using the Brains software developed by Iowa MHCRC Image Processing Lab.
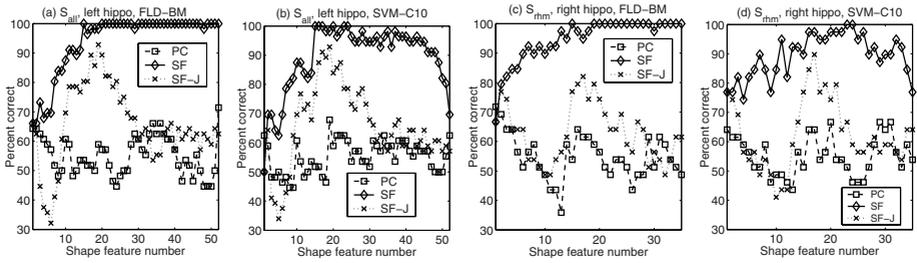
Fig. 2. Jackknife classification: shape feature only, different feature selections. (a) FLD-BM on $\mathbf{S}_{all}^{left}$. (b) SVM-C10 on $\mathbf{S}_{all}^{left}$. (c) FLD-BM on $\mathbf{S}_{rhm}^{right}$. (d) SVM-C10 on $\mathbf{S}_{rhm}^{right}$.
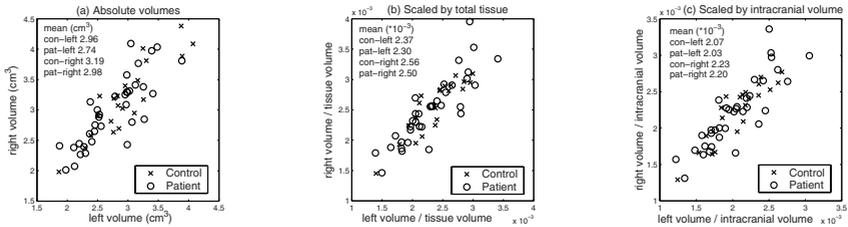
**Fig. 3.** Volume information for $\mathbf{S}_{all}$: (a) V0, (b) V1, (c) V2.

some bias, as test subjects are included in the selection process. However, it is interesting to see that a feature subset does exist that supports nearly perfect classification. **(2)** The **SF-J** results always outperform the **PC** for each classifier in terms of the best case. The improvements range from 3% to 28% for all the cases. In **SF-J**, the classes are not separated well if there are insufficient features, while using too many introduces extra noise. **(3)** The performances of FLD-BM, FLD-1NN, FLD-3NN and FLD-NM are similar, and so are those of SVM-C10 and SVM-C100. However, SVM-C1 underperforms SVM-C10, which indicates the cost of constraint violation needs to be set appropriately in SVMs.

We observe that **SF-J** is an effective and unbiased feature selection scheme. Thus, in the rest of the study, each classification is performed using this scheme.

4.1   Including Volume

We conduct additional experiments by adding volume to the discriminative analysis to see if this can help improve the classification. We consider three types of volume information: (1) the absolute volume **V0**, (2) the volume scaled by the total tissue volume **V1**, and (3) the volume scaled by the total intracranial volume **V2**. Figure 3 shows these volumes just for $\mathbf{S}_{all}$, and $\mathbf{S}_{rhm}$ is a subset of $\mathbf{S}_{all}$. The distributions of these three types of volumes are different from each other. However, none seems able to distinguish patients from controls well.

We use volume as an additional feature and combine shape and volume together in the classification. To treat them equally, we scale all the volumes in
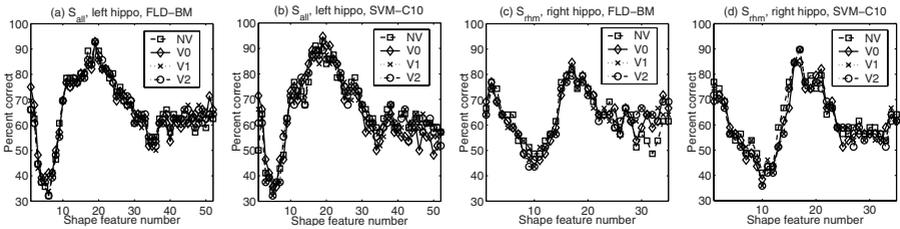
Fig. 4. Jackknife classification: including volumes, SF-J feature selection. (a) FLD-BM on $\mathbf{S}_{all}^{left}$. (b) SVM-C10 on $\mathbf{S}_{all}^{left}$. (c) FLD-BM on $\mathbf{S}_{rhm}^{right}$. (d) SVM-C10 on $\mathbf{S}_{rhm}^{right}$.
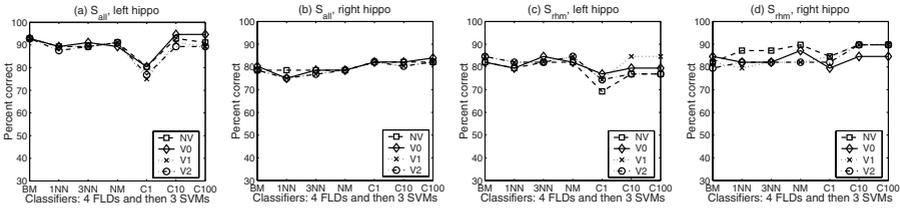
**Fig. 5.** Best jackknife accuracies using different classifiers and SF-J feature selection.

each data set so that the volume variance becomes the same as the total shape variance. We examine all combinations of classifiers and volume information (**V0, V1, V2,** and no volume (**NV**)) on each of the four datasets using SF-J feature selection. The results do not change much after adding any type of volume information. Figure 4 shows some detailed results.

Figure 5 shows the best accuracies achieved by different classifiers using **SF-J** feature selection approach. SVM-C10 performs the best, with 95% accuracy for $\mathbf{S}_{all}^{left}$ using shape and **V0**, 82% for $\mathbf{S}_{all}^{right}$ using shape, 84% for $\mathbf{S}_{rhm}^{left}$ using shape and **V1**, 90% for $\mathbf{S}_{rhm}^{right}$ using shape. All other classifiers except SVM-C1 perform closely to SVM-C10. Another observation is that the left hippocampi predict better in $\mathbf{S}_{all}$ while the right ones predict better in $\mathbf{S}_{rhm}$. This suggests that gender and handedness may affect hippocampal shape changes in schizophrenia.

The 95% accuracy achieved for $\mathbf{S}_{all}^{left}$ greatly outperforms our previous result [13] and is competitive with other hippocampal studies [1,2,3,4,5,12] using different methods on similar types of data.

4.2   Visualization of Discriminative Patterns

Based on the PCA and FLD framework, we introduce a new method for visualizing discriminative patterns. Applying PCA and FLD as detailed above to a shape set, we get a discriminative value v for each shape $\mathbf{x}$: $v = \mathbf{x}_\delta^T * B_{pca} * B_{fld} = \mathbf{x}_\delta^T * W$, where $\mathbf{x}_\delta = \mathbf{x} - \bar{\mathbf{x}}$ is the deformation of $\mathbf{x}$ from the mean shape $\bar{\mathbf{x}}$, B_{pca} consists of a subset of eigenvectors, depending on which PCs are selected, and B_{fld} is the corresponding FLD basis. Thus W is a column vector that weights the contribution of each deformation element in $\mathbf{x}_\delta$ to v. Given a landmark location l, we use
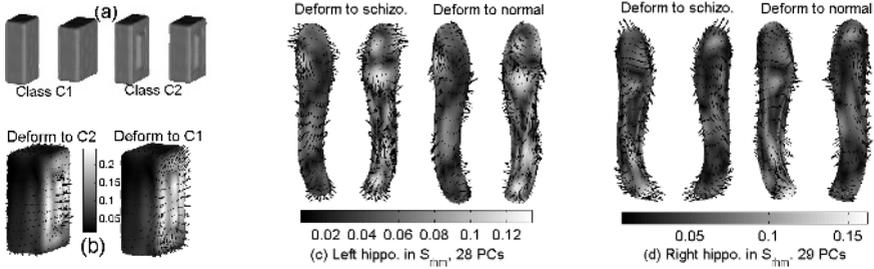
Fig. 6. (a) Sample synthetic surfaces. (b-d) Visualization of discriminative patterns on mean surfaces: (b) synthetic data set, (c) $\mathbf{S}_{rhm}^{left}$, and (d) $\mathbf{S}_{rhm}^{right}$. In (b), the first and second plots show the mappings of the weight vector and its negation, suggesting deformations towards C2 and C1, respectively. In each of (c-d), the weight vectors are mapped onto the first two views and indicate the directions towards a more schizophrenic shape, while their negation are onto the last two views showing the directions towards a more normal shape. The magnitude of local weight vector is coded in color. Light color indicates more discriminative power while dark color less.

$\mathbf{x}_\delta(l)$ to denote the vector containing deformation fields associated with l in $\mathbf{x}_\delta$, and $W_\delta(l)$ the vector of the corresponding weights in W. Thus the contribution made by l can be calculated as $C(l) = \mathbf{x}_\delta(l)^T * W(l)$. Based on this formula, we have two observations: (1) A large magnitude of $W(l)$ indicates that location l has discriminative power, since even small local deformations will have a noticeable effect on the overall classification. (2) Assume Class A has more positive discriminative values v's than Class B. The vector $W(l)$ actually indicates the local deformation direction towards Class A. The reason is that the local contribution $C(l) = \mathbf{x}_\delta(l)^T * W(l)$ is maximized if the local deformation $\mathbf{x}_\delta(l)$ shares the same direction as $W(l)$, which makes the value v more positive. In contrast, $-W(l)$ indicates the deformation direction towards Class B.

We can map W onto the mean surface to show significant discriminative regions and even deformation directions towards a certain class. Figure 6(a) shows shape examples based on two classes of 14 synthetic rectangular surfaces each, with bumps centered on one face in the second class. Perfect discrimination is achieved after PCA and FLD using 3 PCs. Figure 6(b) shows the mapping of W onto the mean surface, which clearly matches our intuition. On the hippocampal data, perfect discrimination is achieved after PCA and FLD using 28 PCs and 29 PCs for left and right cases in $\mathbf{S}_{rhm}$, respectively. Figure 6(c-d) shows the mapping of W to surface locations. Mapping results show that discriminative patterns appear in the both anterior and posterior regions of the left hippocampus and the posterior region of the right one.

5   Conclusion

Our proposed classification framework for brain structures employs SPHARM surface modeling and PDM for shape description. Several FLD and linear SVM

variants together with several feature selection schemes are used to perform classification. We demonstrate our techniques in hippocampal shape analysis in schizophrenia. We test our whole data set and also a subset controlled for gender and handedness (i.e., right-handed males), and consider both inclusion and exclusion of volume information. Exhaustive experimentation is performed on each combination of techniques, and reveals that the proposed **SF-J** feature selection technique works effectively with most classifiers and improves the jackknife accuracy significantly. We achieve the best jackknife accuracies of 95% for the whole set and 90% for right-handed males, competitive with the best results in similar studies using different techniques on similar types of data. Our result suggests the left hippocampus being a stronger predictor in the whole set while the right one stronger over right-handed males. We also present an effective method for visualizing discriminative patterns based on the PCA and FLD framework.

References

1. Csernansky, J.G., Joshi, S., et al.: Hippocampal morphometry in schizophrenia by high dimensional brain mapping. Proc. Nat. Acad. Sci. USA **95** (1998) 11406–11
2. Csernansky, J.G., Wang, L., et al.: Hippocampal deformities in schizophrenia characterized by high dimensional brain mapping. Am J Psy. **159** (2002) 2000–6
3. Golland, P., Grimson, W.E.L., et al.: Small sample size learning for shape analysis of anatomical structures. In: MICCAI'00, LNCS 1935. (2000) 72–82
4. Timoner, S.J., Golland, P., Kikinis, R., Shenton, M.E., et al.: Performance issues in shape classification. In: MICCAI'02, LNCS 2488. (2002) 355–362
5. Saykin, A.J., Flashman, L.A., et al.: Principal components analysis of hippocampal shape in schizophrenia. In: Int. Congress on Schizophrenia Research. (2003)
6. Brechbühler, C., Gerig, G., Kubler, O.: Parametrization of closed surfaces for 3D shape description. Computer Vision and Image Understanding **61** (1995) 154–170
7. Styner, M., Gerig, G., Pizer, S., Joshi, S.: Automatic and robust computation of 3D medial models incorporating object variability. Int. J. Computer Vision (2003)
8. Styner, M., Lieberman, J., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophreni. In: MICCAI'03. (2003)
9. Kelemen, A., Szekely, G., Gerig, G.: Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Trans. on Medical Imaging **18** (1999) 828–839
10. Gerig, G., Styner, M., et al.: Hippocampal shape alterations in schizophrenia: Results of a new methodology. In: 11th Bi. W. Workshop on Schizophrenia. (2002)
11. Gerig, G.: Selected Publications. http://www.cs.unc.edu/~gerig/pub.html (2003)
12. Gerig, G., Styner, M.: Shape versus size: Improved understanding of the morphology of brain structures. In: MICCAI'01, LNCS 2208. (2001) 24–32
13. Shen, L., Ford, J., Makedon, F., Saykin, A.: Hippocampal shape analysis: Surface-based representation and classification. In: SPIE Medical Imaging, Proc. 5032. (2003) 253–264
14. Duda, R.O., Hart, P.E., et al.: Pattern Classification (2nd ed). Wiley, NY (2000)
15. Ma, J., Zhao, Y., Ahalt, S.: OSU SVM Classifier Matlab Toolbox (ver 3.00). http://eewww.eng.ohio-state.edu/~maj/osu_svm/ (2002)