Disease Classification with Hippocampal Shape Invariants

Boris Gutman[1] {bgutman@ucla.edu}, Yalin Wang[1,2], Jonathan Morra[1], Arthur Toga[1], and Paul Thompson[1]

[1] UCLA Laboratory of Neuro Imaging
[2] UCLA Department of Mathematics

Abstract. We present the first Support Vector Machine classification study using the feature space of shape invariants of hippocampal surfaces. Our shape invariants are based on rotationally invariant properties of spherical harmonics (SPH). A global conformal map is used for parameterization. Leave-one-out testing on 49 Alzheimer(AD) and 63 elderly control subjects yielded 75.5% sensitivity and 87.3% specificity with 82.1% correct overall.

1 Introduction

Numerous studies have explored the effect of Alzheimer's disease on brain structure. These generally undertake one of two goals: localizing AD's effect on brain anatomy or classifying subjects according to neuroanatomical differences between patients and controls. So far, the use of spherical harmonics in these studies has been entirely for the purpose of surface alignment and thus has belonged in the former category. To the best of the authors' knowledge, this is the first work to employ SPH-based shape invariants for classification.

Our general procedure may be summarized as follows: given a family of hippocampal surfaces, we map them to sphere in a rotation-preseving manner, in this case using the global conformal mapping. Next, we compute their spherical harmonic spectra and convert these spectra into an invariant shape desription. Finally, we apply the leave-one-out method with Support Vector Machines and feature selection to the shape description for validation.

The remainder of the paper follows this organization: in the first section we present some previous studies and discuss their relevance to the present work; in the second section, we describe how we extracted our hippocampal surfaces; in the third, we describe our exact means of feature generation; next, we give a synopsis of our use of SVM; in the fifth section we present empirical results; in the sixth we conclude the paper.

2 Previous Work

SPHARM has been the most common use of spherical harmonics in neuroimaging. Though the term is used increasingly loosely, it usually stands for a three-

P. Yushkevich and L. Wang (Eds.): MICCAI 2008 Workshop on Computational Anatomy and Physiology of the Hippocampus (CAPH'08), pp. 76–86, 2008.

step process, following any known spherical parameterization of a surface: estimating SPH coefficients of x, y, and z-components with a least-squares procedure, normalizing the orientation of the first-order ellipsoid and reconstructing the surface at regularly spaced points on the sphere [1, 2]. To refine such a coarse registration, Shen [2] also used a variant of the ICP algorithm to register hippocampal shapes in an AD study. Here, point correspondences were established by quickly shifting the surface signal on the sphere and minimizing the RMSD distance, while spatial alignment was done with a quaternion-based method. This method made use of a property of spherical harmonics that makes them unique among the possible L_2 bases on the sphere - they form a direct sum of orthogonal subspaces which are invariant to rotation. This fact is also key in our application of SPH in forming shape invariants.

In most AD classification studies to date, volume-based features, such as grey matter probability maps, were used with a Support Vector Machine classifier to predict diagnosis [3–5]. In the best ones, overall leave-one-out accuracy was 89-96 %. See this survey [6] for a good overview. More relevant to the current study, some recent works instead classified disease according to hippocampal shape-based features [7, 8], with pointwise displacements forming the feature set. [7] specifically used patch-averages of local displacement vectors projected onto the average normal. SVM was then applied to these local features to separate AD subjects from controls. Here, the best leave-one-out accuracy reached 94.9 %.

In a digression, we would like to say that several reviewers in the past have claimed that SVM and spherical harmonic representation have already been combined for classification. This, unfortunately, stems from a misunderstanding of what "spherical harmonic representation" means. In every classification paper on hippocampal shape (or any other neuroanatomical shape) we have found, spherical harmonics are simply used to align the surfaces before returning to a spatial representation for classification. We use a spectral feature space for classification. As an example, consider this Schizophrenia study [9]. Here, SPHARM was employed for surface alignment and vertex-wise displacement vectors for classification, after feature pruning with PCA. The best accuracy here was 77%.

Davies et al. have done yet another Schizophrenia study using the minimal distance length approach to statistically align hippocampal parameterizations in [10]. For classification, Linear Discriminant Analysis (LDA) is used to find the discriminant vector in the feature space for distinguishing diseased subjects from controls. The work is compared to SPHARM, with both approaches yielding a Students t-statistic for the group difference of less than 2.3 along the discriminant vector. The authors claim that an SVM classifier on this feature space yielded practically the same results. s

An interesting study by Gorczowski, Styner, et al. [11]recently appeared on classification using multi-object complexes. Their approach is advantageous to ours in that it takes into account the relative position of several subcortical structures with respect to each other, while we can only combine several shape invariants from every structure individually. This study, however, acknowledged that the classification results are improved when pose is eliminated from the fea-

ture space and only structure-intrinsic features (here, radii of m-reps developed by Gerig, Styner and co-workers) are used. Though the validation method was more robust than ours, the accuracy is inferior: 75%.

Since the number of features in these studies is much larger than in the present paper, a very robust feature selection is required as a preprocessing step before SVM can give reasonable results. The usual and evidently quite reliable means of doing this is the recursive feature elimination (RFE), as in [3, 4, 7]. Here, an SVM model is iteratively trained and at each step the weakest (least-weighted) feature is removed. This is repeated until the classification rate or the particular cost function of the SVM model stop increasing. Though this method is SVM-centric and well-suited for the problem, it is more expensive than our simple feature selection technique. Using the fact that our feature space is already an order or two of magnitude smaller (N = 510), we have simply used a Students t-statistic threshold. This nave feature selection technique may explain why our accuracy is inferior to some of the best results in the studies we mention here.

A unifying aspect of the studies above is their emphasis on locally-based features: in each case a feature corresponds to either a voxel or a point on the surface. While this facilitates visualization, it may not take full advantage of some pattern a shape exhibits globally. For this reason we have chosen a feature set in which each feature represents a unique aspect of the entire hippocampal shape rather than an individual point. Thus, each feature may represent shape variation in multiple locations, increasing its discriminative power. A further advantage of our feature set is that it does not require registration. Since each feature vector is entirely intrinsic to the shape, we require no pre-alignment or point correspondence.

[12] perhaps comes closest to our approach in that it uses two non-local shape features to classify Schizophrenia subjects and controls. This study does indeed use a spherical harmonic representation, specifically SPHARM, to align the left and flipped right amygdala-hippocampal surfaces for each subject. However, once the shapes are aligned, the study again returns to a simple spatial measure (not spherical harmonics or any features derived from them) to classify the shapes. Two asymmetry measures, volume difference and mean square distance (MSD) with the volume normalized, are used in an SVM classification. A good accuracy of 87 % is achieved.

3 Surface Extraction

Initially, structural MRI images are automatically converted into binary hippocampal masks with the help of the recent Auto Context Model (ACM) [13]. ACM uses a few hand-traced ROI examples as a training set for AdaBoost to create a voxel-level classification function. We then convert the masks to a signed distance function and apply topology-constrained mean curvature flow following Han's TGDM algorithm [14]. We ensure that our initialization has the topology of a closed genus-zero surface, thus guaranteeing a surface of correct topology upon extraction. However, since we use Marching Cubes to generate a triangle

mesh from the level set function, our resulting mesh has far more triangles than the level of detail justifies, many of which are highly acute. Applying our spherical parameterization algorithm directly to this mesh would take a long time with suboptimal results. To resolve this issue we apply QEM remeshing [15] to simplify all meshes to 5000 triangles, and subdivide the result to 20000 triangles while applying local smoothing. The first step greatly reduces subsequent computation time (MC meshes tend to have between 40000 and 80000 triangles), while the second ensures triangle niceness, important for achieving the best parameterization. Lastly, a visual check is done on each mesh to ensure that the original masks correspond to a hippocampal shape.

4 Generation of Shape Invariants

Following our surface extraction from MRI data, we proceed to generate an invariant description of each shape. Our generation consists of four steps: (1) spherical conformal parameterization following [16],(2) sampling the mesh at regularly spaced spherical coordinates, (3) computing SPH coefficients of each surface with the help of a spherical fft [17], (4) computing shape invariants from SPH coefficients.

We now formulate the conformal mapping problem. Define triangle mesh $M \equiv (K, g)$, where K is a simplicial complex and $g : |K| \to \mathbb{R}^3$ defines the shape of the mesh in space. Let $i, j, \{i, j\} \in K$ be vertices and their corresponding edge. Then, given $\boldsymbol{f}, \boldsymbol{g} : M \to \mathbb{R}^3$, we define the inner product on M by

$$< \boldsymbol{f}, \boldsymbol{g} >= \frac{1}{2} \sum_{m \in \{x,y,z\}} \sum_{\{i,j\} \in K} k_{i,j}(f_m(i) - f_m(j))(g_m(i) - g_m(j)) \qquad (1)$$

and the energy of $\boldsymbol{f}$ by

$$E(\boldsymbol{f}) =< \boldsymbol{f}, \boldsymbol{f} > \qquad (2)$$

It is well known [16] that setting the string energy weights $k_{i,j}$ to $\frac{1}{2}(\cot(\alpha) + \cot(\beta))$, where α and β are angles opposite $\{i, j\}$, yields the equation for harmonic energy, whose minimization leads to a conformal map in the genus-zero case. The cotangent weights are precisely the cause of numerical instability from highly acute triangles. Imposing the additional constraints

(a) $|\boldsymbol{f}|^2 = 1$

(b) $\frac{\int_M \boldsymbol{f} dA}{\int_M dA} = 0$ (Center of mass constraint)

ensures that the mapping is to $\mathbb{S}^2$.

We reformulate this minimization problem in order to apply the faster conjugate gradient method. Since we may no longer check the constraints at every iteration of naive gradient descent, the energy to minimize becomes

$$E(\boldsymbol{f}, \lambda, \nu) = \frac{1}{2} E_{harmonic}(\boldsymbol{f}) + \lambda E_{sphere}(\boldsymbol{f}) + \frac{\nu}{2} E_{center}(\boldsymbol{f}) \qquad (3)$$

$$= \frac{1}{2}\sum_{i\neq j} k_{i,j}|\boldsymbol{f}_i - \boldsymbol{f}_j|^2 + \lambda[\sum_{i\in K}|\boldsymbol{f}_i|^2 - 1] + \frac{\nu}{2}|\sum_{i\in K}\boldsymbol{f}_i A_i|^2$$

where λ, ν are Lagrange multipliers and $\boldsymbol{f}_i$ are the spherical mesh veriteces. Computation time for a 20000 triangle HP mesh requires on average about 10 minutes with the conjugate gradient method and roughly 45 minutes with the gradient descent method on a 2.41 GHz, 1 Gb RAM Gateway PC.

Following the spherical parameterization, we sample the inverse map at $\theta_j = \pi(2j+1)$, $\phi_k = 2\pi(k)/2B$, $0 \leq j,k < 2B$. For this study, we set the bandwidth B to 256. Although in general, spherical sampling requires $O(mn)$ operations, where m and n are the numbers of triangles of the original and the new mesh, taking advantage of the regularity of our samples reduces this number to $O(m+n)$. In conjunction with fast spherical harmonic transforms, this means that we can find SPH coefficients as quickly as the SPHARM least squares approximation, while reaching much higher bandwidth.

Once sampled, our meshes are put through the spherical fft algorithm of [17] to get the SPH coefficients of x, y and z components. We now give a brief review of spherical harmonics.

Spherical harmonics are functions $f : \mathbb{S}^2 \rightarrow \mathbb{C}$ which are simultaneously eigenfunctions of the Laplace-Beltrami and the angular momentum operators; they are expressed explicitly as

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4(l+m)!}} P_l^m(cos\theta)e^{im\phi} \tag{4}$$

for degree and order $m, l \in \mathbb{Z}$, $|m| \leq l$, where $P_l^m(x)$ is the associated Legendre polynomial. Spherical harmonics form a countable orthonormal basis for square-integrable functions on the sphere. A projection of a function $f \in L_2(\mathbb{S}^2)$ onto this basis yields the SPH coefficients $\widehat{f}(l,m) =< f, Y_l^m >$, where $< f,g >$ is the usual L_2 inner product.

A key property of spherical harmonics is their behavior under a shift on the sphere. Given an element of the rotation group $R \in SO(3)$, a rotated spherical harmonic is expressed as

$$Y_l^m(R^{-1}\omega) = \sum_{n=-l}^{l} Y_l^n(\omega) D_{m,n}^l(R) \tag{5}$$

where $D^l(R)$ are the irreducible representations of $SO(3)$ (see [18] for a possible definition). This property has already been used in medical imaging application [2]. The implication is that the effect of a rotational shift of spherical harmonics on the sphere can be expressed completely by the spherical harmonics of the same order. Thus, given two scalar functions $f, h : \mathbb{S}^2 \rightarrow \mathbb{R}$, where $h(\omega) = f(R^{-1}\omega)$, the L_2-norms of their within-order components are equal:

$$\sum_{|m|<l} ||\widehat{f}(l,m)||^2 = ||Proj_{Span\{Y_l^{-l}, Y_l^{-l+1}, \ldots, Y_l^l\}} f||_2^2 = \tag{6}$$

$$||Proj_{Span\{Y_l^{-l},Y_l^{-l+1},\ldots,Y_l^l\}}h||_2^2 = \sum_{|m|<l} ||\widehat{h}(l,m)||^2$$

Before we present our shape invariants, we should note that our spherical parameterization preserves rotation in the following sense. Suppose $M' = (K, g')$ is a mesh where $g' = R \circ g$, and $\boldsymbol{f}, \boldsymbol{f'} : \mathbb{S}^2 \rightarrow M, M' \in \mathbb{R}^3$ are two inverse conformal mappings. Then $\boldsymbol{f'}(\omega) = R \circ \boldsymbol{f}(R^{-1}\omega)$, $\omega \in \mathbb{S}^2$. This property is necessary for achieving invariance and not shared by some of the faster spherical parameterizations, in particular those which affix an artificial north pole [19].

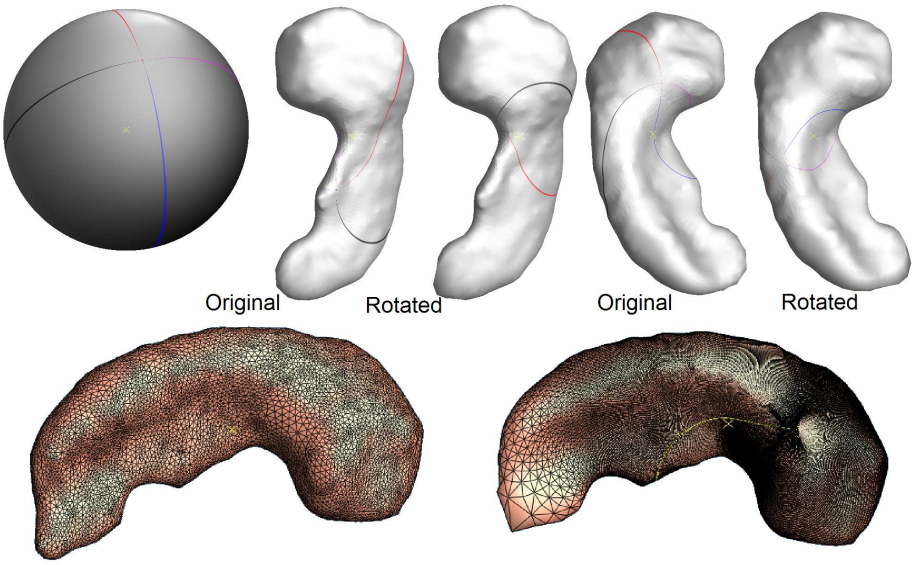

Fig. 1. Rotation and Undersampling: the bottom row shows the original 20K triangle HP surface (left) and its sampled version (right). Note the undersampling. The top row shows the effect of a random rotation. The same surface was rotated by $\alpha = 151.8$ $\beta = 75.6$ $\gamma = 259.5$ reparameterized and resampled. Note the slightly different parameterization of the undersampled region.

Finally, SPH-based shape invariants are defined as:

$$s(l) = \sum_{i\in\{x,y,z\}} \sum_{|m|<l} ||\widehat{f}_i(l,m)||^2 \tag{7}$$

For clarification, we now give a sketch of a proof of $s(l)$s' rotational invariance, omitting some details given in [20]. With last paragraph's notation in mind, consider $s'(l) = \sum_{i\in\{x,y,z\}} \sum_{|m|<l} ||\widehat{f}'_i(l,m)||^2$, and rotation matrix elements $R_{i,j}, i, j \in \{x, y, z\}$. Then,

$$s'(l) = \sum_{j\in\{x,y,z\}} \sum_{|m|<l} ||\langle \sum_{i\in\{x,y,z\}} R_{j,i} f_i(R^{-1}\omega), Y_l^m(\omega)\rangle||^2 = \tag{8}$$
$$\sum_{j\in\{x,y,z\}} \sum_{|m|<l} || \sum_{i\in\{x,y,z\}} R_{j,i}\langle f_i(R^{-1}\omega), Y_l^m(\omega)\rangle||^2$$
$$= \sum_{j\in\{x,y,z\}} \sum_{|m|<l} ||\langle f_j(R^{-1}\omega), Y_l^m(\omega)\rangle||^2.$$

Note the use of the rotation-preserving property of our conformal map above. Now applying (6) to the last line above, we see that

$$s'(l) = \sum_{j\in\{x,y,z\}} \sum_{|m|<l} ||\langle f_j(\omega), Y_l^m(\omega)\rangle||^2 = \sum_{j\in\{x,y,z\}} \sum_{|m|<l} ||\widehat{f}_j(l,m)||^2 = s(l). \tag{9}$$

A recent review claimed that the Euclidean norm of each vector SPH coefficient by itself is rotation-invariant, due to Parseval's theorem. (5) shows that this is certainly not the case. See [18] for an in-depth description of the relationship between the rotation group and spherical harmonics.

By setting the zero-order coefficient to zero we achieve translational invariance. Essentially, the l-th shape invariant is the L^2 norm of the Euclidean distance from the surface to the average value $(\widehat{f}_x(0,0), \widehat{f}_y(0,0), \widehat{f}_z(0,0))$ of the spherical map, projected onto the l-th order subspace.

For all its advantages, the conformal map has one significant shortcoming – its large area distortion. We illustrate this in figure 1: regions of extreme Gaussian curvature which protrude are mapped to very small regions on $\mathbb{S}^2$ and suffer from undersampling. This can potentially cause our shape description to lose its invariance in practice. Figures 1 and 2 illustrate the effect a random rotation of a surface has on its invariants. In the first hundred orders, error is within 2 %. More importantly, the greatest error of the invariants selected for SVM classification (see section 6) is within 0.5%.

5 Support Vector Machine Classification

SVM [21] seeks an optimally separating hyperplane to distinguish two classes within a feature space. Given $\{\boldsymbol{x}_i, c_i\}_{i=1}^n$ data points and their classes c_i, linear SVM minimizes $||\boldsymbol{w}||^2 + C\sum_{i=1}^n \xi_i$ constrained by $c_i(\boldsymbol{w}\boldsymbol{.}\boldsymbol{x}_i - b) \geq 1 - \xi_i$, where ξ_i are the slack variables, measuring the degree of a data point's misclassification, and $\boldsymbol{w}$ are the weights defining the hyperplane. A datum's class is determined by the sign of the SVM score $\boldsymbol{w}\boldsymbol{.}\boldsymbol{x}_i - b$. In this study, we used Joachims' svmPerf package, described in [22].

Shape invariants form our feature space. Since we have a left and a right hippocampus, each hemisphere contributes $B-1$ features for a total of 510. Though far smaller than the initial sets of locally-based models, this is still too large to train a good model given our number of subjects. Feature selection is needed.

Feature selection is a problem encountered in many SVM classification studies, and a wide array of literature on the subject exists [23]. While the most used selection method in AD studies seems to be optimal thresholding of the SVM weights $\boldsymbol{w}$ with cross-validation within the training set [3], also known as REF, for now we have chosen a simple t-statistic thresholding $\frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$.

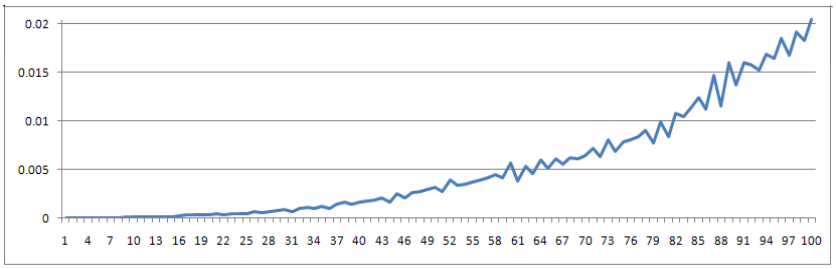

Fig. 2. Effect of undersampling on invariance. We show the relative error of invariants corresponding to the surface in fig. 1: $\frac{|s(l)-s'(l)|}{s(l)}$ vs. l, where l is the order of the invariant.

6 Experimental Results

Our data set consisted of 112 1.5T T1-weighted MRI scan images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, with 49 AD patients and 63 controls, age and gender-matched (mean age: 76.14, 76.76, p = .609). We ascertained the discriminative power of our shape description with the leave-one-out test. For each training set $\{\boldsymbol{x}_i, c_i\}_{i=1}^{n-1}$, we selected a feature if its t-statistic exceeded a threshold t_{min}. After testing a few subjects, we noticed that the best overall accuracy is achieved with $6.7 \leq t_{min} \leq 6.9$, and set it globally to 6.8. This yielded between 6 and 14 features, depending on which subject was left out. All selected features $s(l)$ were of order $37 \leq l \leq 58$. Our margin/error coefficient C was set to 1000. All features were normalized w.r.t. standard deviation (differently for each left out subject) and translated so that $min(x) = -max(x)$. The transformation was saved and applied to the remaining subject. The result was 75.5% sensitivity and 87.3% specificity for a total correct rate of 82.1% (AD is considered positive).

By comparison, hippocampal volume gave 67.3/76.2 % sensitivity/specificity in a leave-one-out test, with 72.3% correct overall. To combine our best features into one measure, we ran SVM on the entire data set with the same t_{min} and C and obtained each subject's SVM score. In regression, SVM score correlated slightly better with Mini-Mental State Examination (MMSE) and Clinical Dementia Rating Sum of Boxes (CDR) scores than volume: for MMSE, $R^2_{vol} = .253$, $R^2_{SVM} = .291$; for CDR, $R^2_{vol} = .276$, $R^2_{SVM} = .295$, $p < .001$ for all.

Since all our selected discriminating features came from the right hippocampus, consistent with a locally-based study on this data [13], we ran the same tests using only right HP volume. We found it is a worse predictor than combined volume in all cases.

As a measure of how much new information is contained in our shape description compared to volume, we ran a linear regression on combined volume and SVM score. The two quantities are at best very weakly correlated, as shown in figure 3.

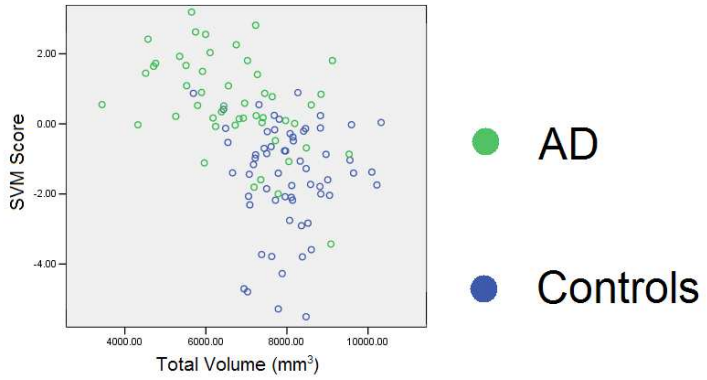

Fig. 3. Scatter of SVM score and total hippocampal volume. Regression $R^2 = .16$

7 Conclusion and Future Work

We have presented an alternative means of disease classification based on neuroanatomical shapes. Our method looks for a shape's global patterns rather than locally based variation in order to discriminate two groups of subjects. In experiments, our method's accuracy was on par with volume and surface displacement-based models when compared to studies of comparable size. This method may well be useful in complementing existing classification methods in selecting biomarkers for the purpose of early detection, etc. Though several existing studies have had better overall accuracy results, we believe that our features contain information that is unavailable through sptatially-based approaches. Thus, we would like to see this feature set combined with some existing ones for an improved classification accuracy.

In the future we would like to improve our method by experimenting with other surface-intrinsic scalar maps such as mean curvature and conformal factor, area and angle-preserving parameterizations and refined feature selection. Lastly, we intend to create a visualization of the discriminant features in our next publication.

References

1. Gerig, G., et al.: Shape analysis of brain ventricles using spharm. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01). (2001)
2. Shen, L., et al.: Morphometric analysis of hippocampal shape in mild cognitive impairment: An imaging genetics study. In: IEEE 7th International Conference Bioinformatics and Bioengineering. (2007)
3. Vemuri, P., et al.: Alzheimer's disease diagnosis in individual subjects using structural mr images: Validation studies. Neuroimage **39**(3) (2008) 1186–97
4. Kloppel, S., et al.: Automatic classification of mr scans in alzheimer's disease. Brain **131**(3) (2008) 681–9
5. Boitino, C., et al.: Volumetric mri measurements can differentiate alzheimer's disease, mild congintive impairment and nomral aging. International Psychogeriatrics **14** (2002) 59–72
6. Silverman, D., Thompson, P.: Structural and functional neuromaging: Focusing on mild cognitive impairment. Applied Neurology **2** (2006) 10–22
7. Li, S., et al.: Hippocampal shape analysis of alzheimer disease based on machine learning methods. American Journal of Neuroradiology **28** (2007) 1339–45
8. Kim, J., et al.: Morphometry of the hippocampus based on a deformable model and support vector machines. Artificial Intelligence in Medicine, Springer LNCS **3581** (2005) 353–362
9. Shen, L., et al.: Hippocampal shape analysis: surface-based representation and classification. In: SPIE. (2003)
10. Davies, R., et al.: Shape discrimination in the hippocampus using an mdl model. In: IPMI. (2003)
11. Gorczowski, K., et al.: Discrimination analysis using multi-object statistics of shape and pose. In: SPIE. (2007)
12. Shenton, M., et al.: Amygdala-hippocampal shape differences in schizophrenia: the application of 3d shape models to volumetric mr data. Psychiatry Research: Neuroimaging **115** (2002) 15–35
13. Morra, J., et al.: Mapping hippocampal degeneration in 400 subjects with a novel automated segmentation approach. In: IEEE International Symposium on Biomedical Imaging (ISBI 2008). (2008)
14. Han, X., et al.: A topology-preserving level set method for geometric deformable models. IEEE Trans. on Pattern Analysis and Machine Intelligence **25**(6) (2003) 755–769
15. Garland, M., Heckbert, P.: Surface simplification using quadric error metrics. In: SIGGRAPH. (2005)
16. Gu, X., Wang, Y., et al.: Genus zero surface conformal mapping and its application to brain surface mapping. IEEE Transactions on Medical Imaging **23**(8) (2004) 949
17. Healy, D.and Rockmore, D., et al.: Ffts for the 2-sphere-improvements and variations. J. Fourier Anal. Applicat. **9**(4) (2003) 341–385
18. Talman, J.: Special functions: A group-theoretic approach. W.A.Benjamin, Inc. (1968)
19. Haker, S., et al.: Conformal surface parameterization for texture mapping. IEEE Transactions on Visualization and Computer Graphics **6**(2) (2000) 181–189
20. Kazhdan, M., et al.: Rotation invariant spherical harmonic representation of 3d shape descriptors. In: Symposium on Geometry Processing. (2003)
21. Cotes, C., Vapnik, V.: Support-vector networks. Machine Learning **20** (1995)

22. Joachims, T.: Training linear svms in linear time. In: KDD. (2006)
23. Guyon, I., et al.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3** (2003) 1157–1182