A Study of Hippocampal Shape Difference Between Genders by Efficient Hypothesis Test and Discriminative Deformation

Luping Zhou[1], Richard Hartley[1,2], Paulette Lieby[2], Nick Barnes[2], Kaarin Anstey[3], Nicolas Cherbuin[3], and Perminder Sachdev[4]

[1] RSISE The Australian National University
[2] Vision Science and Technology Program, NICTA
[3] Centre for Mental Health Research, ANU
[4] Neuropsychiatric Institute, Prince of Wales Hospital, Sydney*

Abstract. Hypothesis testing is an important way to detect the statistical difference between two populations. In this paper, we use the Fisher permutation and bootstrap tests to differentiate hippocampal shape between genders. These methods are preferred to traditional hypothesis tests which impose assumptions on the distribution of the samples. An efficient algorithm is adopted to rapidly perform the *exact* tests. We extend this algorithm to multivariate data by projecting the original data onto an "informative direction" to generate a scalar test statistic. This "informative direction" is found to preserve the original discriminative information. This direction is further used in this paper to isolate the discriminative shape difference between classes from the individual variability, achieving a visualization of shape discrepancy.

1 Introduction

In this paper, we investigate methods of characterizing anatomical shapes as part of a clinical study. The purpose is to detect a statistically significant difference between two classes based on shape descriptors. In this paper we concentrate on classification for gender based on the hippocampus. Such analysis is typically used when attempting to discriminate between healthy and diseased populations, such as in the case of schizophrenia ([1]). We address two key problems in shape analysis: hypothesis testing and identifying the discriminative information.

The classification of the hippocampal shapes between genders is a difficult problem. Using common classifiers, such as kernel SVM and nonlinear discriminators, we could only generate a classification accuracy slightly better than chance. In this case, hypothesis testing becomes critical in answering whether the difference detected by the classifiers is a real difference or just random. Prevailing

methods for hypothesis testing in the existing medical applications make a distribution assumption for the test statistic, such as the t test or χ^2 test ([2,3,4,5]). A limitation of such methods is that a poor distribution model can result in unreliable testing results. For example, the Student's t test assumes a normal distribution of the population or a large sample size. However, this assumption is difficult to justify in many existing medical applications where the sample is small ([4,2,3]). Hence it is better to use hypothesis tests such as the Fisher permutation test and the bootstrap test, which make minimal assumptions about the sample distribution. However, the exact permutation and bootstrap tests are computationally intractable because of their formidably large test space. Generally practitioners only consider a small subset randomly drawn from the entire test space. This is known as the Monte-Carlo or the randomized permutation test.

Another issue to address is to identify the discriminative information between groups in order to understand the nature of the difference. One approach is to represent the shape as a Point Distribution Model (PDM) and analyze each landmark separately([3]). This method potentially ignores the correlations between the landmarks. Golland *et al.* in [2] recently proposed an approach introducing the concept of the "discriminative direction". This "direction" is a curved path with the property that when a shape deforms along its course, only the differences relevant for group discrimination appear.

This paper has two main contributions: (i) It applies the efficient permutation/bootstrap test proposed in [6] to the shape-determination problem; this involves adapting the method to multivariate data. The multivariate shape descriptors are "scalarized" by being projected onto a line. The projection direction is carefully selected using kernel Fisher discriminant analysis (KFDA) to preserve the discriminating information of the original data as much as possible. Compared with other approaches, our method is free of distribution assumptions and overcomes the problem of the overwhelmingly large search space involved in the exact permutation test. It searches the entire test space, resulting in a more accurate p-value estimation in trivial computational time. (ii) The "discriminative direction" is found by a different approach compared with [2]. In [2], the problem is formulated in terms of finding the movement direction in the original space that minimizes the movement in feature space parallel to a hyperplane determined using kernel SVM. Our method on the other hand formulates the problem from the perspective of movement in the feature space and links it to the inverse of the kernel problem ([7,8,9,10]) in machine learning. That is, when a shape descriptor moves along a direction found by KFDA in the feature space, the corresponding changes in the original space are computed as the preimages of the kernel mapping to form the course of the "discriminative direction". Thereby a more direct and conceptually simpler solution is achieved.

2   Method

In this paper, 3D hippocampal shapes are represented by the spherical harmonics expansion (SPHARM) [11,5,3]. A shape descriptor vector is composed of $3(l+1)^2$

spherical harmonic coefficients $(c_{x0}^0, c_{y0}^0, c_{z0}^0, \cdots, c_{xn}^{-m}, \cdots, c_{zn}^m)^\top$, where l is the degree, m the order and c the coefficient. Here hippocampal shape descriptors for two classes (men and women) are compared in order to determine if the shape vectors for the two classes are significantly different.

2.1   Review of the Efficient Permutation and Bootstrap Tests in [6]

Suppose we have m observations $x_1, x_2, \ldots, x_m$ belonging to class A (for instance males) and $n - m$ observations $x_{m+1}, \ldots, x_n$ belonging to class B (for instance females). A (scalar) statistic t_i (for instance height) is measured and the difference between the mean value of the statistic for classes A and B is found; thus

$$t = \frac{1}{m} \sum_{i=1}^m t_i - \frac{1}{n-m} \sum_{i=m+1}^n t_i \ .$$

Suppose that this results in a higher value for class A than class B. The question is whether this difference is significant. Under the null-hypothesis, the value of this test statistic t should not be significantly different from zero.

For the Fisher permutation test, we randomly repartition the n samples into two classes A_r and B_r of sizes m and $n - m$, resulting in a test statistic t_r^* defined as before. The permutation test seeks the fraction p of times among all such partitions that the original statistic t is greater than t_r^*. The boostrap test differs from the permutation test in that the two classes A_r and B_r are drawn from the original observations with replacement.

The exact permutation/bootstrap test has an extremely large test space. The size of the entire test space, N, is C_n^m for the permutation test and n^n for the bootstrap test. Even when n is around 100, the exhaustive enumeration of all possible test statistic values is computationally intractable.

A more efficient method is proposed in [6] for the exact test. Instead of using the exhaustive enumeration, it expands p as an infinite Fourier series. The method is elaborated as follows. In [6], p is defined as $N^{-1} \sum_{r=1}^N H(t_r^* - t)$, where $H(x)$ is a step function: 0 if $x < 0$, $1/2$ if $x = 0$, and 1 otherwise. Assuming that H is a periodic function (with period exceeding the largest value of $t_r^* - t$), taking its infinite Fourier series and exchanging the order of the finite and infinite sums, p becomes a convergent series

$$p = \frac{1}{2} + \frac{2}{\pi} \Im \left(\sum_{k=1}^\infty \frac{\Psi(2k-1) \exp(-i(2k-1)t)}{2k-1} \right) \tag{1}$$

where $\Im(z)$ is the imaginary part of z and $\Psi(2k-1) = N^{-1} \sum_{r=1}^N \exp(ikt_r^*)$. For details, refer to [6].

The advantage of this procedure is that the infinite sum converges rapidly when N is large. More significantly, when the test statistic is a linear combination of the observed data, the computation of $\Psi(k)$ may be carried out very efficiently, as shown in [6]. Note that this method converges to an exact p value without explicitly enumerating all the resamplings of the data.

2.2   Extension of the Method in [6] to Multivariate Data

The method in [6] is applicable to scalar data only. This can be clearly seen from the definition of p where the step function, $H(x)$, is only valid for scalars. We solve this problem by "scalarizing" the multivariate data. All data are projected onto an "informative" direction which can maximally preserve the discriminative information in the original data. The direction $\mathbf{w}$ is realized by maximizing the Fisher linear discriminant criterion (FLDA): $\mathcal{J}(\mathbf{w}) = (\mathbf{w}^\top \mathbf{S}_B \mathbf{w})/(\mathbf{w}^\top \mathbf{S}_W \mathbf{w})$ where $\mathbf{w}$ is the projection vector, $\mathbf{S}_B$ is the between-class scatter matrix, and $\mathbf{S}_W$ is the within-class scatter matrix. Hence the resulting projections are gathered compactly within classes, and separated between classes.

Actually we use a nonlinear version, known as kernel Fisher discriminant analysis (KFDA) [12], to seek the "informative" direction, because the real-world data is often not linearly separable. KFDA maps the data nonlinearly into another space of higher dimension, $\mathcal{F}$, and performs FLDA in $\mathcal{F}$ instead. Let $\mathbf{x}$ ($\mathbf{x} \in \mathcal{R}^d$) be a sample, and $\Phi : \mathcal{R}^d \to \mathcal{F}$ be the nonlinear mapping. In $\mathcal{F}$ we may write the discriminant criterion as $\mathcal{J}(\mathbf{w}) = (\mathbf{w}^\top \mathbf{S}_B^\Phi \mathbf{w})/(\mathbf{w}^\top \mathbf{S}_W^\Phi \mathbf{w})$ where $\mathbf{S}_B^\Phi$ and $\mathbf{S}_W^\Phi$ are defined in terms of values $\Phi(\mathbf{x})$ in $\mathcal{F}$.

The mapping Φ can be very complex. Hence KFDA uses the kernel trick to avoid computing it explicitly. Consider that $\mathbf{w}$ lies in a space spanned by the data samples: $\mathbf{w} = \sum_i \alpha_i \Phi(\mathbf{x}_i)$. The objective function $\mathcal{J}(\mathbf{w})$ may be written in terms of the coefficients α_i and the inner product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. Instead of explicitly computing Φ, we define a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$; the function Φ is then implicitly defined by the relationship $k(\mathbf{x}_i, \mathbf{x}_j) \equiv \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. The RBF (radial basis function) kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ is used in this paper. With this kernel, the optimal values of α_i may be computed to maximize $\mathcal{J}(\mathbf{w})$. Then any data vector $\mathbf{x}$ may be scalarized by projection onto the discriminative direction $\mathbf{w}$ according to the formula $\mathbf{w}^\top \Phi(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i k(\mathbf{x}_i, \mathbf{x})$.

Note that the projection direction $\mathbf{w}$ in KFDA is a non-linear discrimination axis in the original shape space. Moreover, the projection direction is a combination of the features in the shape space (linear in FLDA, and nonlinear in KFDA). The combination weights reveal the correlations of multiple variables with regard to the discrimination.

2.3   Discriminative Deformation

The shape difference needs to be located. We address this issue by deforming a particular male hippocampus to a female one and vice versa. This deformation reveals the discriminative differences between classes, excluding any differences due to the individual variability within a class. This is called a "discriminative deformation" in this paper, and is obtained by deforming a shape along the "discriminative direction" (see Section 1). This paper seeks the "discriminative direction" by utilizing the nonlinear mapping $\Phi : \mathcal{R}^d \to \mathcal{F}$ provided by KFDA and computing the inverse of the kernel mapping. In short, the "discriminative direction" is formed by simply computing the preimages of a movement parallel

to the direction $\mathbf{w}$ (see Section 2.2) in $\mathcal{F}$ for a particular shape descriptor $\Phi(\mathbf{x})$. In this way, we avoid solving the complex optimization problem in [2].

In $\mathcal{F}$, if $\Phi(\mathbf{x})$ moves along the direction $\mathbf{w}$ for a step of λ, it arrives at a new position: $\Phi(\mathbf{x}) + \lambda\mathbf{w}$, where $\mathbf{w}$ is a unit vector. If the preimage of the new position in $\mathcal{R}$ is estimated as $\mathbf{z}$, the error $\rho = \|\Phi(\mathbf{x}) + \lambda\mathbf{w} - \Phi(\mathbf{z})\|^2$ should be minimized. This optimization problem is equivalent to maximizing $\langle\Phi(\mathbf{x}) + \lambda\mathbf{w}, \Phi(\mathbf{z})\rangle$ when a Gaussian RBF kernel is used. Notice that for an RBF kernel there is a relationship between the input space distance $d_{ij}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ and the feature space distance $\tilde{d}_{ij}^2 = \|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|^2$, namely $d_{ij}^2 = -2\sigma^2 \log((2 - \tilde{d}_{ij}^2)/2)$. Making a simplifying assumption that the preimage of $\Phi(\mathbf{x}) + \lambda\mathbf{w}$ does exist ($\rho = 0$), we obtain a closed form solution for $\mathbf{z}$, namely

$$\mathbf{z} = \frac{(2 - \tilde{d}^2(\phi(\mathbf{x}) + \lambda\mathbf{w}, \Phi(\mathbf{x})))\mathbf{x} + \lambda\sum_{i=1}^{l}\alpha_i(2 - \tilde{d}^2(\phi(\mathbf{x}) + \lambda\mathbf{w}, \Phi(\mathbf{x}_i)))\mathbf{x}_i}{2 - \tilde{d}^2(\phi(\mathbf{x}) + \lambda\mathbf{w}, \Phi(\mathbf{x})) + \lambda\sum_{i=1}^{l}\alpha_i(2 - \tilde{d}^2(\phi(\mathbf{x}) + \lambda\mathbf{w}, \Phi(\mathbf{x}_i)))},$$

(2)

where $\tilde{d}^2(\Phi(\mathbf{x}_i), \Phi(\mathbf{x}) + \lambda\mathbf{w})$ is $2 + \lambda^2\alpha^\top K\alpha + 2\lambda\alpha^\top K_x - 2k(x_i, x) - 2\lambda\alpha^\top K_{x_i}$. Given n training data, K is the $n \times n$ kernel matrix, $K_{ij} = k(x_i, x_j)$. Further, α denotes $(\alpha_1, \cdots, \alpha_n)^\top$, K_x denotes $(k(x, x_1), \cdots, k(x, x_n))^\top$, and K_{x_i} denotes $(k(x_i, x_1), \cdots, k(x_i, x_n))^\top$.

3   Results

A database of 353 left hippocampi of healthy individuals, including 180 males and 173 females, is used. The hippocampi have been hand-traced from MRI scans which are drawn from the 60-64 year old age cohort from the first wave of a longitudinal project, PATH through Life, from the Centre for Mental Health Research, the Australian National University. The whole data set is partitioned into a training data set of size 150 (used to train KFDA), and a test data set of size 203. All shapes are represented using SPHARM coefficients with degree 5.
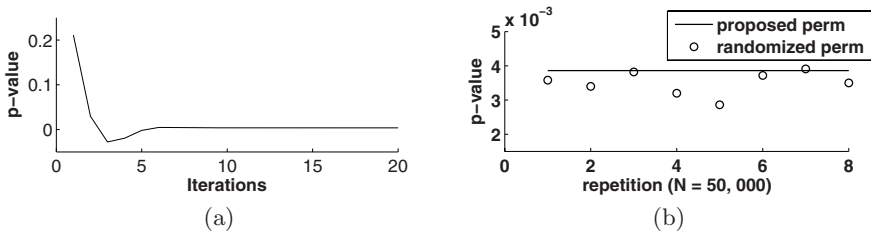

(a)   (b)

Fig. 1. (a) Convergence of the p-value calculated by the proposed permutation test. (b) In the randomized permutation test ($N = 50,000$), the p-value fluctuates around the value estimated by the proposed method when repeating the test with random resampling.

Hypothesis test.   In our hypothesis test, the null hypothesis is defined by *a lack of significant statistical difference between the shapes of genders*. Two aspects are tested: the convergence speed and the discriminating ability. The convergence speed of our method is compared with the randomized permutation test. To facilitate a fair comparison, the convergence of the p-value is tested on the same KFDA projections in the randomized version. KFDA uses a RBF kernel. The parameters σ and μ for KFDA are tuned via a 5-fold cross validation. The results are obtained on a computer with a 3.0GHz CPU and 2.0GB RAM. In the randomized test, the p-value fluctuates around 0.0035 in eight repetitions of 50,000 random resamplings (Fig. 1 (b)). It takes 490 seconds for the randomized test to test only 10^{-57} of the entire test space. However, the p-value calculated by our methods takes only 10 iterations (2.28 seconds) in the permutation test and 36 iterations (0.05 seconds) in the bootstrap test to converge to 0.0039. Due to the large size of our test space, Fig. 1 (a) indicates that the p-value converges well without the "spike" effect mentioned in [6].

Table 1. Comparison of the p-value in different hypothesis tests

Test 1:	Test 2: FLDA		Test 3: KFDA	
Randomized	Permutation	Bootstrap	Permutation	Bootstrap
0.0339	0.0651	0.0645	0.0039	0.0039

The operation of projection may cause information loss. Hence we need to test the discriminating ability of our proposed methods. Three experiments have been conducted. Test 1 is independent of projection: a randomized permutation test is used whose test statistic is the Euclidean distance between the mean shape descriptors of two randomly selected groups. Test 2 and Test 3 seek the informative projection direction by FLDA and KFDA respectively. The p values in the hypothesis tests are summarized in Table 1. The smaller p value in Test 3 suggests that KFDA better preserves the discriminating information than both FLDA in Test 2 and the Euclidean distance in Test 1. Further experiments, not discussed here, show that KFDA achieves a better separation for the mean projections of the male and female groups.

The hypothesis test is conducted on normalized data, i.e. the hippocampi are normalized for volume. Indeed, it has often been reported that hippocampal volumes differ for the sexes ([13]). Our findings confirm other results that show male and female hippocampal shapes differ ([14]).

Discriminative shape deformation. A synthetic data set consisting of two concentric circles is used to test the discriminative direction method we derived (Fig 2). The circles are projected onto a direction **w** found by KFDA in the feature space (Fig 2 (a)). When a particular point p moves along **w** in the feature space, the corresponding moving trace, i.e. the discriminative direction at point p in the original space, is recovered in Fig 2 (b). It also can be seen that our method works consistently at different points. Our method is further applied
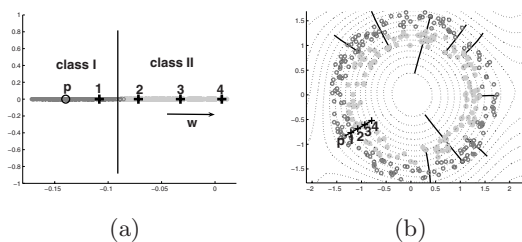
Fig. 2. Discriminative direction of two synthetic classes (dark and gray dots): (a) point p moving along $\mathbf{w}$ in the feature space; (b) the corresponding motion in the original space, i.e. the discriminative direction at p. Nine further examples are shown.
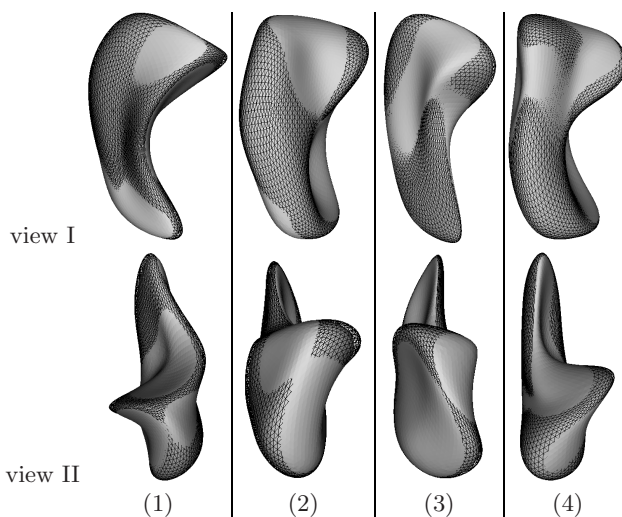

Fig. 3. Four left male hippocampi (each one in a column) are deformed to be female-like and viewed from two perspectives. The deformed female-like shape (in solid) and the original male shape (in mesh) are overlapped for comparison. The solid part is only visible when it protrudes the mesh, and vice versa. A female hippocampus is more likely to be thicker in the middle part of the hippocampal head than a male hippocampus.

on real hippocampi to expose the nature of shape changes between genders. Fig 3 gives four examples of left male hippocampi before and after the discriminative deformation. The original male shape and the deformed female-like shape are overlapped for comparison. It can be seen that a female hippocampus is more likely to be thicker in the middle part of the hippocampal head than a male hippocampus. This is supported by the findings in [14]. Note that each deformed hippocampus resembles its original shape since the deformation only changes the class features while maintaining the individual information.

4   Conclusions

Using the Fisher permutation and bootstrap tests applied to a KFDA projection of the shape vector, it was demonstrated that the difference between male and female hippocampal shapes is significant at the 99.6% level. The KFDA test outperforms a randomized test (Test 1) based on difference between shape-cluster distance (96.6%), and is very much faster.

The shape discrepancy revealed by our experiments is global for two reasons: (i) The shapes are represented by SPHARM coefficients, which are global features. (ii) Discriminative deformation is able to illustrate the localized shape changes as demonstrated in [2]. However, as our large data set is population representative, the significant inter-individual hippocampal shape variation renders localization of class difference difficult. Future work will be to investigate discrimination for pathology vs normal using local features like PDM.

References

1. Styner, M., Lieberman, J., Pantazis, D., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophrenia. Medical Image Analysis 8(3), 197–2003 (2004)
2. Golland, P., Grimson, W.E., Shenton, M.E., Kikinis, R.: Detection and analysis of statistical differences in anatomical shape. Medical Image Analysis 9(1), 69–85 (2005)
3. Shen, L., Ford, J., Makedon, F., Saykin, A.: Hippocampal shape analysis surface-based representation and classification. In: Proceedings of SPIE-Medical Imaging, pp. 253–264 (2003)
4. Gerig, G., Styner, M., Shenton, M.E., Lieberman, J.: Shape versus size: Improved understanding of the morphology of brain structures. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 24–32. Springer, Heidelberg (2001)
5. Gerig, G., Styner, M., Jones, D., Weinberger, D., Lieberman, J.: Shape analysis of brain ventricles using spharm. In: Proceedings of Workshop on Mathematical Methods in Biomedical Image Analysis MMBIA, pp. 171–178 (2001)
6. Gill, P.M.: Efficient calculation of p-values in linear-statistic permutation significance tests. Journal of Statistical Computation and Simulation 77(1), 55–61 (2007)
7. Schoelkopf, B., Mika, S., Smola, A., Raetsch, G., Mueller, K-R.: Kernel PCA pattern reconstruction via approximate pre-images. In: Proceedings Eighth International Conference Artificial Neural Networks, pp. 147–152 (1998)
8. Mika, S., Schoelkopf, B., Smola, A.J., Mueller, K-R., Scholz, M., Raetsch, G.: Kernel PCA and de-noising in feature spaces. In: Proceedings of Advances in Neural Information Processing Systems, pp. 536–542 (1999)
9. Kwok, J.T., Tsang, I.W.: The pre-image problem in kernel methods. IEEE Transactions on Neural Networks 15(6), 1517–1525 (2004)
10. Rathi, Y., Dambreville, S., Tannenbaum, A.: Statistical shape analysis using kernel PCA. In: Proceedings of SPIE Electronic Imaging (2006) pp. 425–432 (2006)
11. Kelemen, A., Szekely, G., Gerig, G.: Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Transactions on Medical Imaging 18(10), 828–839 (1999)

12. Mika, S., Raetsch, G., Weston, J., Schoelkopf, B., Mueller, K-R.: Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, pp. 41–48. IEEE Computer Society Press, Los Alamitos (1999)
13. Maller, J.J., Réglade-Meslin, C., Anstey, K.J., Sachdev, P.: Sex and symmetry differences in hippocampal volumetrics: Before and beyond the opening of the crus of the fornix. Hippocampus 16, 80–90 (2006)
14. Bouix, S., Pruessner, J.C., Collins, D.L., Siddiqi, K.: Hippocampal shape analysis using medial surfaces. NeuroImage 25, 1077–1089 (2005)