Kernel Fisher Discriminant for Shape-based Classification in Epilepsy

S. Kodipaka [a] B.C.Vemuri [a,*] A.Rangarajan [a] C.M.Leonard [b] I.Schmallfuss [c] S.Eisenschenk [d]

[a]*Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL, 32611*
[b]*Department of Neuroscience, University of Florida, Gainesville, FL, 32611*
[c]*Department of Radiology, University of Florida, Gainesville, FL, 32611*
[d]*Department of Neurology, University of Florida, Gainesville, FL, 32611*

Abstract

In this paper, we present the application of Kernel Fisher Discriminant in the statistical analysis of shape deformations that indicate the hemispheric location of an epileptic focus. The scans of two classes of patients with epilepsy, those with a right and those with a left medial temporal lobe focus (RATL and LATL), as validated by clinical consensus and subsequent surgery, were compared to a set of age and sex matched healthy volunteers using both volume and shape based features. Shape-based features are derived from the displacement field characterizing the non-rigid deformation between the left and right hippocampi of a control or a patient as the case may be. Using the shape-based features, the results show a significant improvement in distinguishing between the controls and the rest (RATL and LATL) vis-a-vis volume-based features. Using a novel feature, namely, the normalized histogram of the 3D displacement field, we also achieved *significant improvement* over the volume-based feature in classifying the patients as belonging to either of the two classes LATL or RATL respectively. It should be noted that automated identification of hemispherical foci of epilepsy has not been previously reported.

Key words: Kernel Fisher discriminant, Shape deformations, Shape features, Shape classification, Epilepsy

1. Introduction

Statistical analysis of shape deformations, such as those likely to occur in epilepsy and other neurological disorders, necessitate both global and local parameter based characterization of the anatomical shape under study. The most popular method to achieve such a characterization has been size- and volume-based analysis [10,6]. However, this captures only one of the aspects necessary for complete analysis. A shape based description gives much more information, and may be combined with the former to help understand the anatomical structures better.

In this paper, we focus on developing an automated technique to aid in distinguishing between controls and patients with epilepsy and also to indicate the hemispheric location of an epileptic focus (right medial temporal lobe versus left medial temporal lobe) in the patients. *It should be noted that the research reported here does not attempt to determine the precise coordinates of the epileptic focus in the patients.* The key aspect is the use of local shape-based features as opposed to global volume information to achieve the task of automated classification.

1.1. *Literature Review*

In this paper, we use the non-rigid registration displacement field mapping the left and right hippocampus as the characterization of shape asymmetry. There are

---

* Corresponding author
  Email address: vemuri@cise.ufl.edu (B.C.Vemuri).

two goals here; (i) to separate controls from patients using the shape asymmetry characterized by displacement fields, and (ii) to separate patients with epileptic foci in either the left and right hemisphere (LATL versus RATL) using the displacement fields. Regardless of whether we are trying to separate the controls versus the patients or LATL from RATL, the problem of trying to classify subjects using a set of deformation field features remains as the fundamental problem.

In this study, we assume the existence of a training set which places us in the realm of supervised learning. That is, we begin with a set of labeled deformation fields where the labels reflect the dichotomies relevant to the study: controls versus patients and LATL versus RATL. Given this information, we can use classifiers such as support vector machines or an extension of the linear discriminant—the kernel Fisher discriminant for solving this problem. The classifier tries to reproduce the training set labels to the best of its ability. This is followed by an evaluation on the test set where the labels are unknown. The choice of features plays a crucial role in this process. In the following, we briefly survey relevant existing literature on pattern classification in brain MR image analysis, in order to clearly delineate previous work and point out the novel aspects of our work.

Gerig et al. [10] proposed the use of both volume measurements and shape based features – Mean Square Distance (MSD) – to detect group differences in hippocampal shape in schizophrenia. The local deformations are accounted for by first flipping one shape across the mid-sagittal plane, aligning the reference and the mirrored shape using the first ellipsoid and then by computing the MSD between the two surface shapes. The class differences are then accounted for by using SVM followed by performance evaluation using the leave-one-out technique. From the results reported, it was concluded that shape alone could not capture the class differences. This failure can be attributed either to weak shape features or the fact that the nature of the groups under study is such that shape alone cannot represent the entire class character.

Csernansky et al. [6] used high dimensional transformations of a brain template to compare the hippocampal volume and shape characteristics in schizophrenia and control subjects. The transformations are carried out in two steps namely, coarse registration using previously placed landmarks followed by local registration using fluid transformation. Optimal shape representations are obtained by computing the transformation vector fields from the triangulated graph of points superimposed on the hippocampus surface. Linear discriminant analysis is then used for classification based on the first six eigen vectors corresponding to the six largest eigenvalues. Log likelihood ratios were computed for all the subjects and Wilk's Lambda was used to test the statistical significance of group differences. On the other hand, volume estimation is done by measuring the volume enclosed by the transformed hippocampal surface. This was then analyzed using a two-way, repeated measures ANOVA, with diagnostic group and hemisphere as the factors. Note that there is an implicit Gaussian assumption on the training population of the shape differences when using the PCA to do the linear discriminant analysis – which may prove to be impractical.

An interesting implementation of the unsupervised clustering has been proposed by Duta et al. [8]. The idea is to automatically cluster the shape training set according to the PSS *(Procrustes sum of squares)* distance in the original shape space and then perform Procrustes analysis on each cluster for its prototype and information about shape variation. The mean shapes obtained are then used to detect object instances in the new image. The scheme has the scope of improvement by using an incremental learning process and needs to be tested for 3D anatomic structures. In a similar vein, Chui *et al.* [4] learn a mean 3D hippocampal shape with no bias from a set of unlabeled 3D hippocampal point-sets. The objective function used essentially minimizes the entropy of all the overlayed point-sets after they have been warped (using a thin-plate spline) to an emerging atlas. This work also needs to be tested in a classification setting. Also, since thin-plate splines are used, the warping into the atlas space is not guaranteed to be diffeomorphic.

Recently, Golland et al. [11,12,13] have investigated the use of a Support Vector Machine for shape-based classification in anatomical structures. In Golland et al. [12], distance transforms are used as the shape descriptor and shape differences are learned in the original high dimensional space using a Gaussian kernel while controlling the capacity (generalization error) of the classifier. The results of the scheme are demonstrated by finding the classification between schizophrenia patients and normal controls and an improvement from 63% to 73% has been shown when compared to using volume information alone. In their work, the segmented hippocampal shapes are registered to a reference using first and second order moments of distance field representation of the segmentations. This will lead to a bias toward the specific reference data. In more recent work, Golland et al. [13], describe a method for assessing the statistical significance of detected differences in medical image scans especially when the data (feature sets in our context) are very high dimensional and the number

of training samples are small. This work is based on a popular technique in statistics called permutation tests.

1.2. *Overview of our algorithm*

In this paper, we demonstrate the application of the Kernel Fisher Discriminant (KFD) algorithm for shape based classification of hippocampal shapes in controls and epilepsy. In this context, we first have to learn the shape differences with in each of the three classes namely, controls, patients needing right anterior temporal lobectomy (RATL) and left anterior temporal lobectomy (LATL). Given a pair of sparse sets of data points (manually placed by an expert Neuro-scientist) corresponding to the boundaries of the left and right hippocampi of a subject, appropriate shape features are extracted by first fitting a model to the data sets using a deformable pedal surface [31]. The output of this fitting process is a smooth surface expressed using a triangular finite element representation. The nodal points of this mesh on the left and right hippocampus are then used as input to a similarity alignment scheme followed by a level-set non-rigid registration scheme. Rigid registration between the left and right hippocampus is achieved by using the Iterative Closest Point (ICP) algorithm [2]. The global scaling factor is accounted for by approximating the hippocampus shape using the smallest ellipsoid that encloses it and then equalizing the corresponding eigen values. The next step is to perform non-rigid registration between the left and right hippocampi which is done using the level-set registration method described in [36]. The local deformations obtained by non-rigid registration are then input to the Kernel Fisher classifier to capture the statistical difference between the classes considered. The analysis is done in two stages, namely, Controls vs. Rest and LATL vs. RATL. The validation is done using leave-one-out technique. The choice of Kernel Fisher Discriminant as the classifier has been motivated by the fact that it can separate the classes in a very high or infinite dimensional space using a linear classier and is simple to implement.

The Kernel Fisher Discriminant (KFD) training algorithm used in this work requires the solution to an eigen-value problem which can be solved efficiently using SVD decomposition. In contrast, using SVMs requires solving a much more difficult nonlinear optimization problem. Kernel Fisher Discriminant has shown results comparable to SVM in various other applications [19,20]. In this paper as well, we present a comparison of the performance of KFD with that of SVM.

1.3. *Organization of the Paper*

The rest of the paper is organized as follows: Section 2 briefly describes the deformable pedal surface model used for fitting a model to the given data points followed by the procedure of selection of shape based features. In Section 3, we summarize the Kernel Fisher Discriminant (KFD) method as a classifier. We introduce a new feature in Section 4.1, i.e. the 3D histogram of displacement vector field between left and right hippocampi, which yielded better performance than other features for the LATL vs. RATL case. Section 4.2 contains the experimental results and analysis, where KFD clearly out performs the other classifiers. This is followed by discussion and conclusions in Section 5.

2. Shape Extraction and Registration

In this section we will briefly discuss, the techniques used to segment the hippocampus from a human brain MRI followed by the methods employed for rigid and non-rigid registration of the corresponding hippocampi in any given subject. MR images used in this study were obtained using a 3-D MP-RAGE Turbo-FLASH protocol [21] with TR = 8.3-10 ms, TE = 3.3-4ms, flip angle = 10-20°, 24-25cm FOV, 130×256 or 192×256 (zero-filled to 256×256) matrix; 128 partitions of a 160mm thick slab in the third dimension, total acquisition time of 5 min. and 53 sec. to 6 min. 3 sec., spatial resolution of $(0.9 \times 0.9 \times 1.2)$ to $(1.0 \times 1.0 \times 1.25)$ in mm units. This sequence was drawn from the literature [21] and represents a good compromise between image acquisition time (the 3-D data set forms just one part of a one hour clinical exam), spatial resolution and contrast.

2.1. *Overview of the Deformable Pedal Surfaces*

In order to segment the hippocampi from the given image, we used the deformable pedal surface described in [31]. Pedal curves/surfaces are defined as the loci of the foot of the perpendiculars to the tangents of a fixed curves/surface from a fixed point called the pedal point [31]. A large class of shapes can be synthesized by varying the position of the pedal point which exhibits both global and local deformations. Physics-based control is introduced by using a snake to represent the position of this varying pedal point. Thus the model is called a "snake pedal" and allows for interactive manipulation through forces applied to the snake. The model also allows representation of global deformations such as bending and twisting without introducing addi-
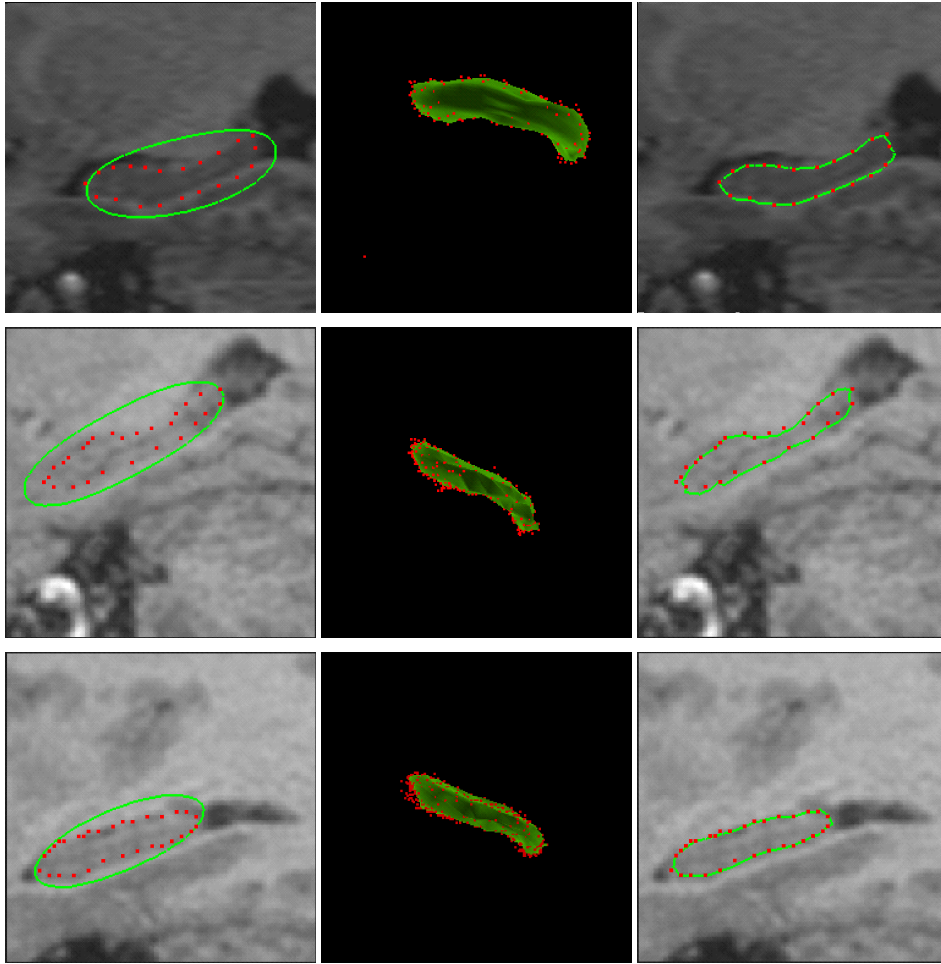
Figure 1. Model Fitting using the deformable pedal surface. First row: (MRI scan of a healthy volunteer superposed with boundary points placed by a neuro-scientist on a slice (dots) and a slice (white curve) of the initialized deformable pedal surface model (left), fitted model (center) followed by a slice of the fitted model superposed on the corresponding MR slice (right). Second row: MRI scan of a patient with left medial temporal lobe focus. Third row: MRI scan of a patient with right medial temporal lobe. Third column shows that fits are visually accurate.

tional parameters. To fit a model to a given set of data points in 2D/3D, a non-linear optimization scheme using Levenberg-Marquardt (LM) method in outer loop for estimating global parameters and the Alternating Direction Implicit (ADI) method in the inner loop for estimating the local parameters of the model is employed [31]. This fitting yields a finite element grid representation of the underlying shape in the data. The data here consisted of the MR images as well as the 3D points placed by a neuro-anatomist on the boundary of the hippocampus (shape of interest). Image-based as well as point-based forces were used in achieving the model fits to the data. In our fitting, the number of nodal points in the finite element (mesh) representation of the fitted surface shape is fixed to 21×40. For more details on this shape representation and fitting techniques, we re-

fer the reader to [31].

2.2. *Overview of the Shape Registration*

Shape registration, in general, is required at both global and local levels. The global transformation deals with the estimation of the rigid parameters such as rotation, translation and scaling to align the coordinate systems of the two shapes under consideration. On the other hand, the local (registration) transformation deals with the problem of finding the (non-rigid transformation) displacement vector between the source and target shapes respectively.
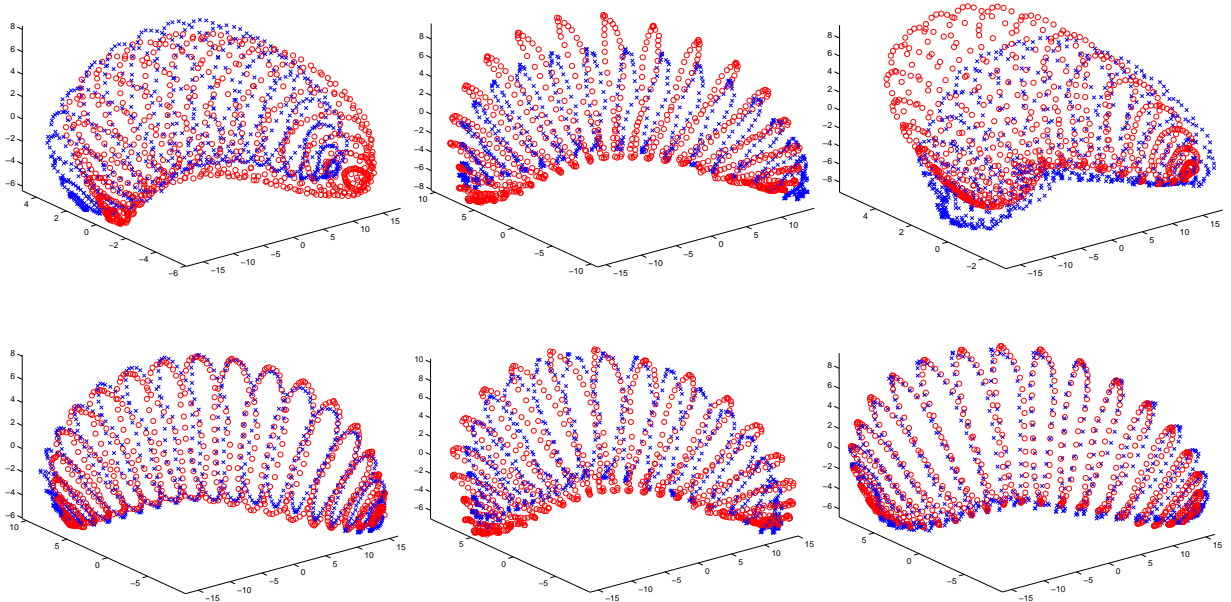
4

Figure 2. Each figure depicts the 3D points set from the deformable pedal surface fit before and after rigid registration of left (as crosses) and right (as circles) hippocampi using the ICP algorithm and scaling. Each column shows point sets for a healthy, LATL and RATL volunteers respectively; before registration (top row), after registration (bottom row).

2.2.1. *Rigid Registration*

In the present work we use the Iterative Closest Point (ICP) algorithm described in [2] to determine the rotation and translation between a subject's left and right hippocampus. One may use other more sophisticated techniques to achieve this goal. However, we were able to achieve sufficient accuracy with this technique. The method allows the determination of the rigid motion between the shapes, described by a point set in 3D, by minimizing the mean-squared distance and uses a quaternion representation of the rotations [2]. The choice of a rigid registration algorithm (ICP) is motivated by the fact that snake-pedal based model fitting yields a parametric description which does not provide a direct read of the correspondence between the left and right hippocampal surfaces. The corresponding left and right hippocampi of a subject may have a global scaling factor which is accounted for by approximating the shapes of the smallest ellipsoid that encloses the hippocampus and then equalizing their corresponding eigen values.

2.2.2. *Non-Rigid Registration*

The problem of finding the non-rigid registration can be formulated as a motion estimation task, in particular, estimation of the displacement field between the two given shapes. There are several techniques of varying sophistication for solving this problem and we re-

fer the reader to the tutorial [37] and the references therein. Before estimating the non-rigid registration, we first take the parametric surface representation of the hippocampi and convert them into an implicit (signed distance) representation. Then, we use a simple yet effective technique based on a level-set formulation described in [34] to estimate the non-rigid registration transformation(displacement field) that leads to the following governing equation:

$$\frac{dV}{dt} = [d_2(X) - d_1(V(X))]\mathcal{N} + \lambda \begin{pmatrix} \Delta u \\ \Delta v \\ \Delta w \end{pmatrix} \quad (1)$$

with the initial condition $V^{(t)} |_{t=0} = \overrightarrow{0}$

$$where, \quad \mathcal{N} = \frac{\nabla(G_\sigma * d_1(V(X)))}{(\| \nabla(G_\sigma * d_1(V(X))) \| + \alpha)} \quad (2)$$

where $V = (u, v, w)^T$ is the displacement vector at location X, the operation $V(X)$ stands for $(X - V)$, d_1 and d_2 denote the signed distance function representations of the source and target hippocampal shapes, Δ denotes the Laplacian operator, G_σ is Gaussian kernel and α is a small positive number serving as a stabilizing factor. The above differential equation can be solved by using the numerical implementation described in [32]. Once the non-rigid registration between the left and

5

the right hippocampi—in the form of the displacement field—is computed, it is further processed to extract features that can be used in a classification algorithm described in Section 4.1

2.2.3. *Feature Extraction Results*

We now present a set of experiments which demonstrate the hippocampal shape extraction using the deformable (snake) pedal surfaces in normal subjects, patients with right medial temporal lobe focus (RATL) and patients with left medial temporal lobe focus (LATL). The input consists of a sparse point set placed by a neuro-scientist on the MRI scans of brain to identify the hippocampal boundary. This is followed by an ROI (region of interest) selection wherein the deformable pedal surface model is initialized. Setting the parameters (such as bending, scaling) and using 3D forces, the desired fittings are achieved. All the fittings were done at interactive rates on an SGI machine.

The results are organized as follows: in Fig. (1), left to right, the images show a slice of an MR brain scan in which the boundary of the hippocampus was identified by a neuro-scientist via sparsely placed points on the shape boundary. The image also shows a slice of the initialized deformable (snake) pedal surface model seen as a white curve. The middle frame shows the fitted deformable pedal surface in 3D. The last frame of the first row depicts a slice of the fitted model superposed on the corresponding slice (which is the same as the first frame of this row) of the MRI scan. The second row and third row depict the same organization as the first row but the results are those obtained for the LATL and RATL patient scans. As can be seen from the figures, a visually accurate fit is achieved in all the three cases. The superimposed mesh of 21×40 on the model gives a new point set of cardinality 840×3 (each point is in 3D).

We present the rigid and non-rigid registration results obtained on the point sets obtained by the model fitting of corresponding left and right hippocampi of a healthy volunteer/patient. The process of rigid registration begins by taking a subject as the reference and registering all the left and right hippocampi to the left and right hippocampi of the reference using the ICP algorithm proposed in [2]. This is done because the data for different subjects (controls/patients) might have been acquired in different coordinate systems. Thus in order to compare all the subjects, it is necessary to maintain a common coordinate frame. This is followed by estimating the rotation and translation between the corresponding left and right hippocampi using the ICP algorithm [2]. The

next step is to remove the global scaling factor among a pair of hippocampi using the method discussed earlier. The results of this entire process are shown in Fig. (2): each top frame shows the initial position of the left hippocampus with respect to the right hippocampus before the rigid registration and the bottom frame depicts the two after the registration using the estimated rotation, translation and scaling parameters.

The above procedure leads us to the next step of finding the non-rigid registration using the level-set formulation [31]. This requires the signed distance images which are formed using the Fast Marching Method [1,27]. The lattice for which the signed-distance is defined with respect to the given shape, is of dimension $128 \times 128 \times 128$ with a voxel distance of 0.2 (this value was selected empirically). To ensure that the given shape is connected, we interpolate the point set using Splines. Using the signed distance images so obtained, the local deformation is found by using the equation (1). The various parameters used can be summarized as follows: $\sigma = 1.2$ for smoothing the source image after applying displacement field, $\sigma = 1$ for smoothing the displacement field at each iteration, α as the stabilizing factor has value 0.01 and $\lambda = 5$. The results obtained are shown in Fig. 3 for a healthy (CTRL), LATL and RATL volunteer respectively. As can be seen, the difference between the source and target image after the rigid registration is almost fully eliminated using the local deformation field.

3. The Kernel Fisher Discriminant Algorithm

The classification problem can be approached in two ways, namely *supervised* and *unsupervised*, with the discriminant function being either linear or non-linear. The classical approach begins with the optimal Bayes classifier by assuming the normal distribution for the classes and when linear discriminant analysis is applied, leads to the Fisher algorithm. The Fisher approach [9] is based on projecting d-dimensional data onto a line with the hope that the projections are well separated by class. Thus, the line is oriented to maximize this class separation.

However, the features in the input space may not possess sufficient discriminatory power for separation of class via linear projection techniques. This problem can be tackled by mapping the input data into a very high dimensional space and using a linear classifier in this new feature space; thereby giving an implicit non-linear classification in the input space. This is the basic idea behind the Kernel Fisher Discriminant algorithm.
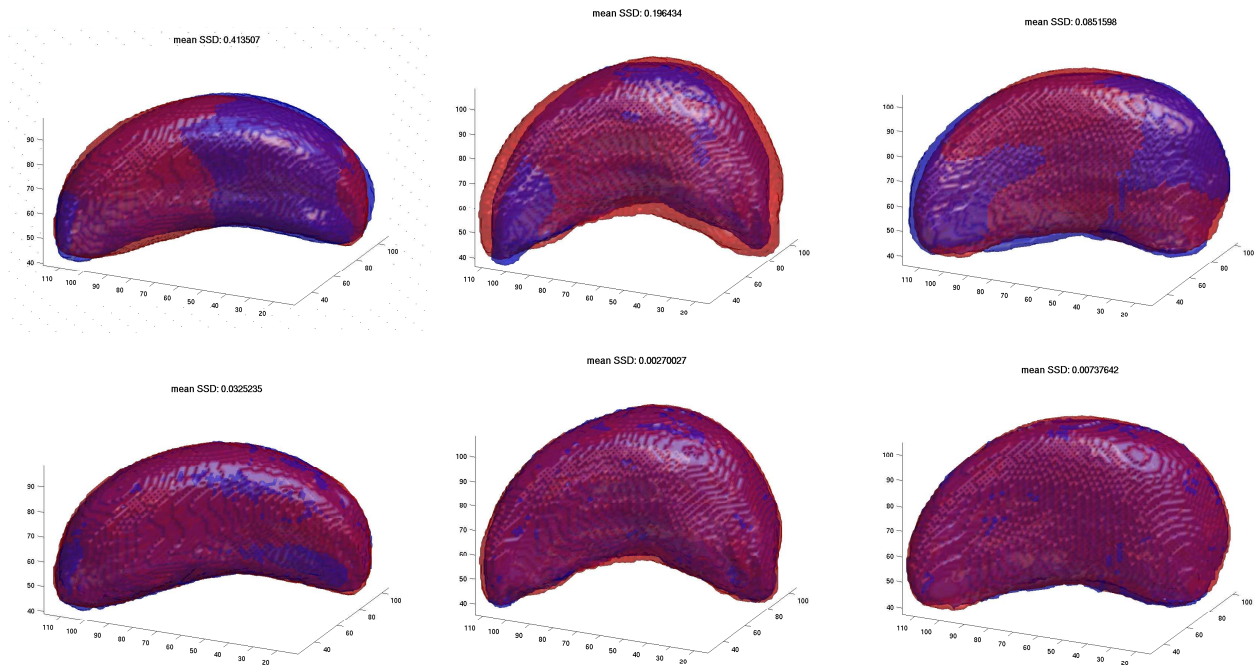
Figure 3. Non-rigid registration of left (in blue) and right (in red) hippocampi using a level set formulation. Point sets for a healthy, LATL and RATL volunteers respectively. First column: After rigid registration. Second column: After non-rigid registration we can see complete visual alignment and the reduction in the sum of squared difference (SSD)
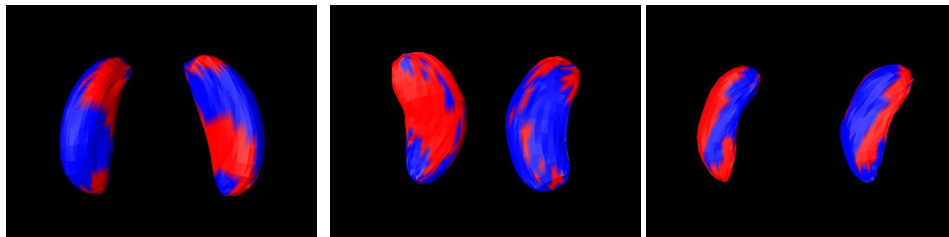

Figure 4. Sign of displacement field for different subjects of type CTRL, LATL and RATL respectively. Positive displacement is color coded with RED and the negative with BLUE on the hippocampal surface.

Figure 5. Direction of displacement field for different subjects of type CTRL, LATL and RATL respectively. Direction components in (x, y, z) are color coded with Red, Green, Blue components on the hippocampal surface.

Let ϕ be a non-linear mapping to some feature space $\mathcal{F}$. Given two sets of features $\{\chi_1, \chi_2\}$, the vector $\underline{w} \in \mathcal{F}$, along which separation is optimal in the new feature space, can be found by maximizing the *Fisher criterion*:

$$J(\underline{w}) = \frac{\underline{w}^T S_B^\phi \underline{w}}{\underline{w}^T S_W^\phi \underline{w}} \qquad (3)$$

With the class means as $m_i^\phi = \frac{1}{l_i} \sum_{j=1}^{l_i} \phi(x_j^i)$ and l_i being the number of samples in χ_i, the *between-class* and *within-class* covariance matrices S_B^ϕ and S_W^ϕ can be defined as follows

$$S_B^\phi = (m_1^\phi - m_2^\phi)(m_1^\phi - m_2^\phi)^T, \qquad (4)$$

$$S_W^\phi = \sum_{i=1,2} \sum_{x \in \chi_i} (\phi(x) - m_i^\phi)(\phi(x) - m_i^\phi)^T \qquad (5)$$

Eq. (3) can be solved by formulating it in terms of inner-products $(\phi(x) \cdot \phi(y))$ of the training patterns [19]. Use of Mercer kernels in KFD and SVM has given a new approach to the problem of pattern recognition. The wide variety of kernels allows various non-linear classifiers. However, a crucial step is the choice of the kernel function which is entirely data dependent.

As explained in [19] using the kernel theory, Eq. (3) can be rewritten as

$$J(\alpha) = \frac{\alpha^T M \alpha}{\alpha^T N \alpha} \qquad (6)$$

$$N \stackrel{\text{def}}{=} \sum_{j=1,2} K_j (I - 1_{l_j}) K_j^T \qquad (7)$$

$$M \stackrel{\text{def}}{=} (M_1 - M_2)(M_1 - M_2)^T \qquad (8)$$

$$(M_j)_i \stackrel{\text{def}}{=} \frac{1}{l_j} \sum_{p=1}^{l_j} (K_j)_{ip} \qquad (9)$$

where α is the set of coefficients corresponding to the training patterns such that $\underline{w} = \sum_{i=1}^{l} \alpha_i \phi(x_i)$. With inner products replaced by the kernel function, we get Eqs. (8,9,7) in which, K_j is a $l \times l_j$ matrix with $(K_j)_{nm} \stackrel{\text{def}}{=} k(x_n, x_m^j)$; K_j is the kernel matrix for class j, I is the identity matrix and 1_{l_j} is the matrix with all entries $1/l_j$.

The optimum direction of projection can be found by taking the leading eigenvector of $N^{-1}M$. This approach is called the *Kernel Fisher Discriminant* (KFD) [19]. The projections of a new vector $\underline{x}$ onto $\underline{w}$ can be obtained by

$$(\underline{w} \cdot \phi(x)) = \sum_{i=1}^{l} \alpha_i k(x_i, x) \qquad (10)$$

The proposed setting is ill-posed due to the estimation of l-dimensional covariance structures from l samples which can cause the matrix N to not be positive definite. The problem can be solved by adding a positive multiple of the identity matrix to N [19] such that $N_\lambda = N + \lambda I$. A good choice for λ can be the least non-zero singular value of N.

4. Experimental Results and Validation

In this section, we present the experimental results obtained by testing the performance of Kernel Fisher Discriminant(KFD) against Linear Fisher Discriminant (LFD) [9] and Support Vector Machines (SVM) with both Polynomial (PLY) and Gaussian Radial Basis (RBF) Kernels for the statistical analysis of shape deformations that indicate the hemispheric location of an epileptic focus. We used libSVM [3] to test the performance of SVM. The scans of two classes of epilepsy patients, those with a right and those with a left medial temporal lobe focus (RATL and LATL), as validated by clinical consensus and subsequent surgery, are compared to a set of age and sex matched healthy volunteers using both volume and shape based features. *The data set consists of 23 healthy volunteers referred to as controls (CTRL), 15 LATL and 16 RATL patients*. For each healthy volunteer/patient, we are given the point sets placed by a neuro-scientist on the MRI scan of the left and right hippocampi. These point sets then undergo three stages of processing namely, model fitting using the Snake Pedal Model [31], shape registration (both rigid and non-rigid) and finally classification to recognize the group. Shape based features are derived from the local displacement field between the left and right hippocampi of a subject.

4.1. *Shape Feature Selection*

After the process of rigid and non-rigid registration mentioned above, we are ready to select the appropriate features and a classifier to capture the shape differences in the classes considered.

With volume based study being popular in this particular application [10,6], our objective is to compare it with the shape-based features in learning the shape differences between the three groups. Hence, we use both volume and shape features for comparison purposes. The feature vector obtained from volume information is two dimensional, with L/R and (L-R)/(L+R) as the components, where L and R are the volumes of the left and right hippocampi respectively.

Shape-based features are derived from the local displacement field between the left and right hippocampi

of a healthy volunteer/patient for the fitted point sets. *The features used in the shape-based classification are: (i) sign of the displacement vector at each grid/mesh point on the hippocampal shape, (2) direction of the displacement vector at each grid/mesh point on the hippocampal shape and (3) the normalized 3D histogram of the entire displacement field.* The sign of the displacement is defined as follows: given the displacement vector for a point on the zero-set of the implicit representation of the source shape, determine the cube in which the displaced point falls in the implicit source shape. Depending on the sign of the vertices (since each vertex was assigned a +/- sign while forming the implicit representation i.e., a distance field/image) of the enclosing cube, assign a sign to the magnitude of the displacement. From left to right, in Fig. (4), we depict a visualization of this feature for a CTRL, LATL and RATL respectively.

The direction vector is obtained by finding the unit vector corresponding to the displacement vector at each point on the zero set. Fig. (5) depicts a visualization of this feature for a CTRL, LATL and RATL subject from left to right. The feature vector for the sign of displacement is of length 762 while the direction vector is of length 762×3 (each point has an (x, y, z) component of displacement). These numbers are arrived at from the fact that there are 840 points on the zero set and the first and last row of the 21×40 mesh represents the north and south pole as described in [31].

To reduce the dimensionality of the feature vectors, principal component analysis (PCA) was applied and only a subset of leading eigen-vectors corresponding to the non-zero eigen values was used in the classifier. The choice of which eigen basis to use was determined via scanning for the best test performance in the classification.

The third feature we used was the joint histogram of the (x, y, z) components of the 3D displacement vector field. Prior to construction of the 3D histogram feature, the displacement vector field was de-noised by reconstructing it in a reduced eigen basis. Once again, the choice of the number of basis was determined by scanning for the best test performance in the classification. As seen subsequently, the joint 3D histogram of the components of the 3D displacement vector field proved to be the best of all the features in discriminating the LATL vs. RATL and showed comparable performance in the controls vs. the rest to the results when using other shape features.

4.2. *Classification Results*

The classification is done in two stages, namely, first we distinguish between controls and patients with epilepsy and then determine the hemispheric location of an epileptic focus in the patients. In each of the Figures (6-9), we plot the Gaussian distribution of feature projections for each class on the discriminant direction for LFD and KFD so as to analyze separability.

4.2.1. *Classification of Healthy Volunteers vs. Patients with Epilepsy*

We begin our analysis by distinguishing between healthy volunteers and the rest (RATL and LATL). Fig. (6) *shows the classification obtained between the two classes using volume-based information for the linear and non-linear classifiers.* The plots show the projection of the data sets on the separating line (either in input space or projected space). As can be seen from the plots, *the two classes cannot be separated using a volume (only) feature based study.* Even the non-linear classifier is unable to separate them out as their characteristics are mixed in nature. The results are further validated using leave-one-out test and have been summarized in Table (1). The average training accuracy achieved is nearly 76% while the maximum test set accuracy is only 79.56%. We empirically explored the best parameters for KFD and SVM, as listed in Table (1).

The next stage of the analysis is based on shape features for distinguishing the two classes. The sign of displacement vector is of length 762 while the number of samples we are dealing with is 54. Since the shape based features are high dimensional, we have used an optimal number of principal components of the samples for all the classification methods. In each classification method, we searched for the optimal number of leading principal components giving us the best testing performance.

We begin with the linear classifier and the principal components of the sign of displacement as the feature vector. Analysis using leave-one-out as shown in Table (2) also shows an improvement over using only volume information. The next step is to use KFD as the non-linear classifier and compare it with SVM. The degree of polynomial kernel and radius of Gaussian radial basis kernel were determined experimentally. The best classification results for each method in Table (2) were obtained for the leading (cumulative) 54, 30, 54, 54 and 23 principal components respectively. Table (2) depicts the leave-one-out training and testing results using the
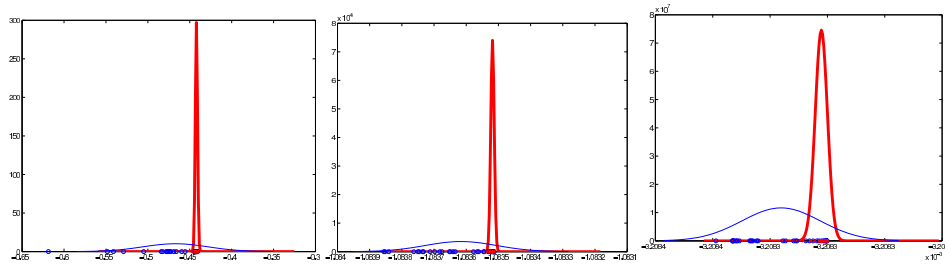
Figure 6. Classification results for Controls (thick line) vs. Rest using volume based feature. Classification using LFD (left), classification with KFD using polynomial basis of degree 3 (center), classification with KFD using RBF (radius=30) (right).

	LFD	SVM with Poly. basis (degree=7)	SVM with RBF (radius=0.05)	KFD with Poly. basis (degree=3)	KFD with RBF (radius=30)
Training	72.26%	79.59%	79.63%	77.81%	77.53%
Testing	72.22%	77.78%	**79.56**%	77.78%	75.93%

Table 1
Controls vs. Rest, Feature: Volume (Controls=23, Rest=31)

	LFD	SVM with Poly. basis (degree=1)	SVM with RBF (radius=400)	KFD with Poly. basis (degree=1)	KFD with RBF (radius=1800)
Training	93.26%	100%	100%	100%	100%
Testing	88.89%	94.44%	87.04%	94.44%	**96.30**%

Table 2
Controls vs. Rest, Feature: Sign of Displacement (Controls=23, Rest=31)
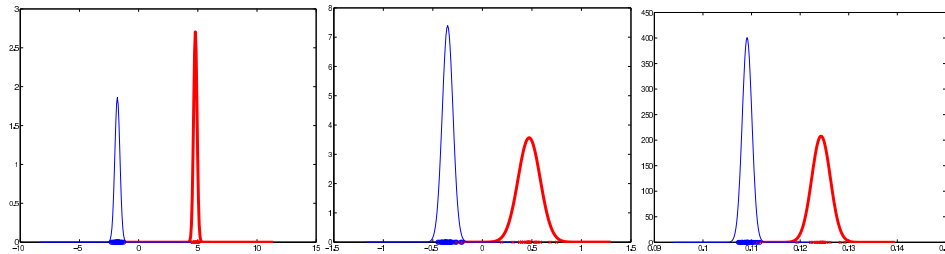

Figure 7. Classification results for Controls (thick line) vs. Rest using Direction of Displacement as feature: LFD (left), KFD with Poly. basis (degree=2) (center), KFD with radial basis (radius=1200) (right).

	LFD	SVM with Poly. basis (degree=2)	SVM with RBF (radius=500)	KFD with Poly. basis (degree=2)	KFD with RBF (radius=1200)
Training	94.55%	100%	100%	100%	99.83%
Testing	88.89%	**96.30**%	94.44%	**96.30**%	94.44%

Table 3
Controls vs. Rest, Feature: Direction of Displacement (Controls=23, Rest=31)

	LFD	SVM with Poly. basis (degree=6)	SVM with RBF (radius=0.2)	KFD with Poly. basis (degree=2)	KFD with RBF (radius=40)
Training	83.30%	91.79%	85.19%	98.15%	98.18%
Testing	83.30%	88.89%	85.19%	88.89%	**90.74**%

Table 4
Controls vs. Rest, Feature: 3D Histogram of Displacement (Controls=23, Rest=31)

different classifiers.

Table (3) summarizes the results obtained with direction vector of the displacement field as the feature vector. As can be seen in Fig. (7), the classification improves significantly. It shows a considerable improvement (over LFD) in the training classification with the non-linear classifier. Note that this conclusion is arrived at from the fact that the probability of error in classification is less for KFD than for LFD. For the five different classifiers, the best testing results were obtained for 54, 21, 54, 23 and 46 leading (cumulative) principal components respectively.

The 3D Histogram of the displacement vector field is our next feature of choice for the classification. Table (4) shows that this novel histogram feature performed comparably to the sign and direction of displacement features. For the five different classifiers in Table (4), the best testing results were obtained for 27, 23, 21, 31 and 23 leading (cumulative) principal components respectively. Number of bins in the histogram based feature for these results were $(9 \times 9 \times 9)$, $(8 \times 8 \times 8)$, $(7 \times 7 \times 7)$, $(4 \times 4 \times 4)$ and $(4 \times 4 \times 4)$ respectively. Note that number of bins is an important parameter in the classifier performance for this feature. However, by using a density estimator e.g., Parzen window estimator, we can circumvent this free parameter estimation issue. On the average, performance of KFD was better than other classifiers. The best accuracy achieved being 96.3% for sign and direction of displacement vector field features with KFD and SVM. *Evidently, all the shape based features show a considerable improvement over using the volume information.*

4.2.2. *Classification within Epileptics (LATL vs. RATL)*

The next task is to identify if the hemispheric location of an epileptic focus is in RATL or LATL. We begin the experiments with the volume feature. As can be seen in Fig. (8), volume is not able to distinguish between RATL and LATL using either LFD or KFD. Leave-one-out performance, as shown in Table (5), also shows low training and test set accuracy and does not improve much with the use of the non-linear classifiers.

The analysis is followed by using shape based information for distinguishing between the two classes as explained in the Controls vs. Rest case. LFD, SVMs and KFDs are tested for the shape based features. Table (6) shows the quantitative results obtained by leave-one-out test. For the five different classifiers, the best testing results were obtained for 22, 21, 30, 22 and 27 leading (cumulative) principal components respectively. Results obtained using direction vector as the feature vector are

shown in Table (7). It can be seen that the performance of this feature vector is poorer than the sign of displacement feature. For the five different classifiers, the best testing results were obtained for 31, 28, 31, 29 and 31 leading (cumulative) principal components respectively.

The 3D histogram feature of the displacement vector field yielded far better results as seen in in Table (8). Fig. (9) shows that the classes are well separated upon using this feature. Same was the case with all the shape based features in general. For the five different classifiers, the results were obtained for 23, 24, 31, 25 and 22 leading (cumulative) principal components. The number of histogram bins for these results were $(7 \times 7 \times 7)$, $(12 \times 12 \times 12)$, $(12 \times 12 \times 12)$, $(5 \times 5 \times 5)$ and $(10 \times 10 \times 10)$ in each of the directions (x, y, z) respectively. Over all, though SVM performed better with sign and direction of displacement vector field features, KFD with RBF basis using the 3D joint histogram of the components of the displacement vector field has yielded the best result, i.e. 90.32% accuracy, for the LATL vs. RATL classification. In this case, only one LATL and two RATL patients were misclassified.

Of the 31 patients analyzed using our shape classifier, we found that 26% were subjected to the potentially harmful invasive procedure involving intra-cranial electrode placement which can potentially lead to complications such as development of subdural or epidural hematomas. *Our shape-based classifier correctly diagnosed the hemisphere of epileptic focus in all the 8 patients who underwent intra-cranial electrode placement.* Therefore, all these patients could have avoided this potentially harmful procedure and the associated morbidity. We anticipate that a large percentage of patients with intractable epilepsy will benefit from the use of our shape analysis tool when applied to a larger population sample.

5. Discussion and Conclusion

From the results section, it is evident that we are able to distinguish between controls and patients with a better accuracy as compared to distinguishing between RATL and LATL. This can be due to various reasons. The shape differences among the pathologies may be highly correlated, hence making it difficult to separate them out. Also the number of data samples for the patients with pathology is quite small which hinders a sufficient representation of the population. This can also be seen in high training accuracy but lower test accuracy among the patients. Thus, it is not conservative to say that we have much to do in this context i.e., use
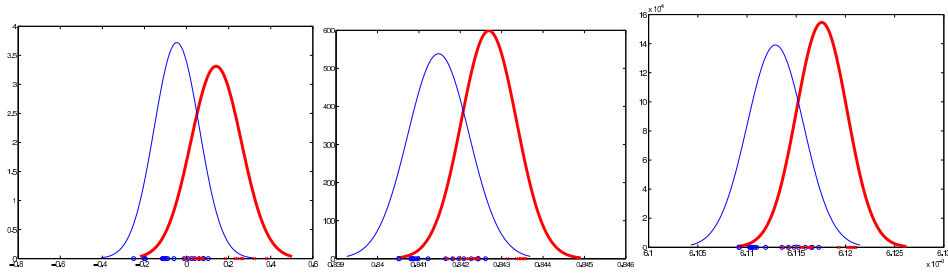
Figure 8. Classification results for LATL (thick line) vs. RATL using volume based feature. Classification using LFD (left). classification with KFD using polynomial basis of degree 1 (center), classification with KFD using RBF (radius=1200) (right).

	LFD	SVM with Poly. basis (degree=3)	SVM with RBF (radius=3)	KFD with Poly. basis (degree=1)	KFD with RBF (radius=1200)
Training	79.78%	83.25%	80.65%	79.78%	80.11%
Testing	77.42%	**80.65**%	**80.65**%	77.42%	77.42%

Table 5
LATL vs. RATL, Feature: Volume (LATL=15, RATL=16)

	LFD	SVM with Poly. basis (degree=2)	SVM with RBF (radius=400)	KFD with Poly. basis (degree=1)	KFD with RBF (radius=1800)
Training	93.30%	100%	100%	100%	98.92%
Testing	77.42%	**80.65**%	64.52%	77.42%	74.19%

Table 6
LATL vs. RATL, Feature: Sign of displacement (LATL=15, RATL=16)

	LFD	SVM with Poly. basis (degree=1)	SVM with RBF (radius=1000)	KFD with Poly. basis (degree=1)	KFD with RBF (radius=500)
Training	84.52%	100%	96.77%	100%	76.13%
Testing	64.52%	**74.19**%	61.29%	70.97%	70.97%

Table 7
LATL vs. RATL, Feature: Direction of Displacement (LATL=15, RATL=16)
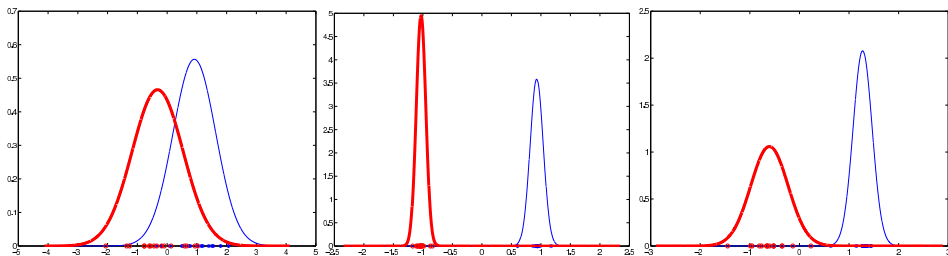

Figure 9. Classification results for LATL (thick line) vs. RATL using 3D histogram of the displacement vector as feature vector; (left) classification using LFD, classification using KFD with Poly. basis (degree=1) (center), (right) classification with KFD using RBF (r=1).

	LFD	SVM with Poly. basis (degree=15)	SVM with RBF (radius=0.01)	KFD with Poly. basis (degree=1)	KFD with RBF (radius=1)
Training	77.42%	91.29%	100%	100%	99.68%
Testing	77.42%	83.87%	80.65%	87.10%	**90.32**%

Table 8
LATL vs. RATL, Feature: 3D Histogram of Displacement (LATL=15, RATL=16)

larger number of training samples for controls as well as LATL and RATL. Also, the choice of kernel is crucial for achieving good generalization. This issue requires much more empirical testing and validation in order to determine the best kernel for the task. The choice of feature vector is key to achieving good training and generalization performance. As explained earlier, the principal components were introduced for the dimensionality reduction of the sign and direction of displacement vector fields. In the case of the 3D histogram feature, they were used for denoising the displacement vector field prior to the construction of the 3D histogram.

Preliminary results indicate that shape features prove to be quite promising for discriminating between controls and epileptics. However, the same shape features are relatively less successful in inter-hemispheric discrimination between subjects of pathology. Kernel Fisher Discriminant out performed other classifiers for the LATL vs. RATL case and was on par with SVM, for the epileptics vs. healthy volunteers classification. We used a new feature namely, the 3D joint histogram of the components of the displacement vector field and found a significant improvement over the volume-based feature and the two shape features used previously, in the LATL vs. RATL classification. We expect to improve the classification performance in this area by i) increasing the number of patient studies, ii) better feature selection and pruning and iii) improving the classifier.

6. Acknowldgements

References

[1] J.A. Baerentzen, On The Implementation of Fast Marching Methods for 3D Lattices, *Technical Report IMM* **13** (2001)

[2] P.J. Besl and N.D. McKay, A Method of Registration of 3-D Shapes, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **14(2)** (1992) 239–255

[3] C. Chang and C. Lin, LIBSVM: a Library for Support Vector Machines (2001)

[4] H. Chui, A. Rangarajan, J. Zhang and C.M. Leonard, Unsupervised Learning of an Atlas from Unlabeled Point-sets, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **26(2)** (2004), 160–173

[5] C. Cortes and V. Vapnik, Support Vector Networks, *Machine Learning* **20** (1995) 273–297

[6] J.G. Csernansky, S. Joshi, L. Wang, J.W. Haller, M. Gado, J.P. Miller, U. Grenander and M.I. Miller, Hippocampal Morphometry in Schizophrenia by High Dimensional Brain Mapping, *Neurobiology* **95(10)** (1998) 11406-11411

[7] D. DeCoste, M.C. Burl, A. Hopkins and N.S. Lewis, Support Vector Machines and Kernel Fisher Discriminants:A Case Study using Electronic Nose Data, *Workshop on Mining Scientific Datasets*(2001)

[8] N. Duta, M. Sonka and A.K. Jain, Learning Shape Models from Examples using Automatic Shape Clustering Procrustes Analysis, *Proc. Info. Processing in Medical Imaging* **1613** (1999) 370–375

[9] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics, **7** (1936) 111–132

[10] G. Gerig, M. Styner, M.E. Shenton and J.A. Lieberman, Shape vs. Size: Improved Understanding of the Morphology of Brain Structures, *MICCAI* (2001) 24–32

[11] P. Golland, W. Eric, L. Grimson and R. Kikinis, Statistical Shape Analysis Using Fixed Topology Skeletons: Corpus Callosum Study, *Proc. Info. Processing in Medical Imaging* **1613** (1999) 382–387

[12] P. Golland, W. Eric, M.E. Shenton and R. Kikinis, Small Sample Size Learning for Shape Analysis of Anatomical Structures, *Proc. MICCAI* LNCS **1935** (2000) 72–82

[13] P. Golland, W. Eric, L. Grimson, M.E. Shenton and R. Kikinis, Deformation Analysis for Shape Based Classification, *Proc. IPMI* LNCS **2082** (2001) 517–530

[14] Y. Guo, Snake Pedals: Active Geometric Models for Shape Modeling and Recovery, *PhD Dissertation, University of Florida* (1998)

[15] M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active Contour Models, *International Journal of Computer Vision* **1** (1987) 321–331

[16] R. Kimmel, A. Amir and A.M. Bruckstein, Finding Shortest Path on Surfaces using Level Set Propagation, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17(6)** (1995) 635–640

[17] S.H. Lai and B.C. Vemuri, Robust and Efficient Computation of Optical Flow, *IEEE Symposium on Cvs. , Miami, Florida* (1995) 455–460

[18] R.E. Lynch, J.R. Rich and D.H. Thomas, Tensor Product Analysis of Alternating Direction Implicit Methods, *J. Soc. Industrial and applied Math.* **13(4)** (1965) 995–1007

[19] S. Mika, G. Ratsch and J. Weston, Fisher Discriminant Analysis with Kernels, *Neural Networks for Signal Processing* **IX** (1999) 41–48

[20] S. Mika, A. Smola and B. Schölkopf, An improved Training Algorithm for Kernel Fisher Discriminant, *Proceedings AISTATS* (2001) 98–104

[21] J. P. I. Mugler and J. R. Brookeman, Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP-RAGE). *Magnetic Resonance in Medicine*, **15**, (1990) 152-157

[22] W.H. Press, S.S. Teukolsky, W.T. Vetterling and B.P. Flannery, Numerical Recipes in C, *Cambridge Univ. Press, Cambridge* (1992)

[23] E. Rouy and A. Tourin, A Viscosity Solutions Approach to Shape-From-Shading, *SIAM J. Num. Anal* **29(3)** (1992) 867–884

[24] S. Saitoh, Theory of Reproducing Kernels and its Application, *Longman Scientific & Technical, Harlow, England* (1988)

[25] B. Schölkopf, C. Burges and A. Smola, Advances in Kernel Methods-Support Vector Learning, *MIT Press, Cambridge MA* (1999)

[26] B. Schölkopf, A. Smola and K,-R. Müller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, *Neural Computation* **10** (1998) 1299–1319

[27] J.A. Sethian, Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Material Science, *Cambridge University Press, Cambridge* (1999)

[28] C. Studholme, D.L.G. Hill and D.J. Hawkes, Automated 3D Registration of MR and CT Images in the Head, *Medical Image Analysis* **1(2)** (1996) 163–175

[29] R. Szeliski and J. Coughlan, Hierarchical Spline-Based Image Registration, *IEEE Conf. on Computer Vision and Pattern Recognition* (1994) 194–201

[30] V. Vapnik, The Nature of Statistical Learning Theory, *Springer Verlag, New York* (1995)

[31] B.C. Vemuri and Y. Guo, Compact and Versatile Geometric Models with Physics-Based control, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22(5)** (2000) 445–459

[32] B.C. Vemuri, S. Huang, S. Sahni, C.M. Leonard, C. Mohr, R. Gilmore and J. Fitzsimmons, An Efficient Motion Estimator with Application to Medical Image Registration, *Medical Image Analysis, Oxford University Press* **2(1)** (1998) 79–98

[33] B.C. Vemuri and A. Radisavljevic, Multiresolution Stochastic Hybrid Shape Models with Fractal Priors, *ACM Trans. Graphics* (1994) 177–207

[34] B.C. Vemuri, J. Ye, Y. Chen and C. Leonard, A Level-Set based Approach to Image Registration, *Workshop on Mathematical Methods in Biomedical Image Analysis June 11-12* (2000) 86–93

[35] P. Viola and W.M. Wells III, Alignment by Maximization of Mutual Information, *International Journal of Computer Vision* **24(2)** (1997) 137–154

[36] J. Ye, A Level-Set Based Approach to Image Registration, *M.S. Thesis, University of Florida* (2000)

[37] L. Younes, Deformations, Warping and Object Comparisons - A tutorial, *Presented at ECCV* (2000)