

Kernel-Based Morphometry of Diffusion Tensor Images

Madhura Ingalhalikar, Parmeshwar Khurd, and Ragini Verma

Abstract Voxel-based group-wise statistical analysis of diffusion tensor imaging (DTI) is an integral component in most population-based neuroimaging studies such as those studying brain development during infancy or aging, or those investigating structural differences between healthy and diseased populations. The majority of studies using DTI limit themselves by testing only certain properties of the tensor that mainly include anisotropy and diffusivity. However, the pathology under study may affect other aspects like the orientation information provided by the tensors. Therefore, for detecting subtle pathological changes it is important to perform group-wise testing on the whole tensor, which encompasses the changes in anisotropy, diffusivity and orientation. This is rendered challenging by the fact that conventional linear statistics cannot be applied to tensors. Moreover, the pathology over the population is unknown and could be non-linear, further complicating the group-based statistical analysis. This chapter gives a perspective on performing voxel-wise morphometry of tensor data using kernel-based approach. The method is referred as Kernel-based morphometry (KBM) as it models the tensor distribution using kernel principal component analysis (kPCA), which linearizes the data in high dimensional space. Subsequently a Hotelling T^2 test is performed on the high dimensional kernelized data to determine statistical group differences. We apply this method on simulated and real datasets and show that KBM can effectively identify the underlying tensorial distribution. Thus it can potentially elucidate pathology-induced population differences, thereby establishing a kernelized full tensor framework for population studies.

M. Ingalhalikar (✉) • P. Khurd • R. Verma
Department of Radiology, Section of Biomedical Image Analysis, University of Pennsylvania,
Philadelphia, PA 19104, USA
e-mail: Madhura.Ingalhalikar@uphs.upenn.edu; Ragini.Verma@uphs.upenn.edu

1 Introduction

Diffusion tensor imaging (DTI) is an important MRI modality for studying white matter connectivity and organization non-invasively [6, 32]. There exists a wide variety of applications of DTI that include investigating neurological and psychiatric disorders [2, 21, 25], structure-function relationships [11, 20, 38], evaluation of brain connections and creating the structural brain connectome [8, 19], and as a tool to assist in computer guided surgery and treatment planning [13, 29].

Clinical investigation of pathology-induced changes requires a group-based statistical analysis of DT images which can identify regional differences between controls and patients. Voxel-based morphometry (VBM) [4] is a popular form of statistical analysis and has been widely adopted by the neuroimaging community. VBM has an advantage over region-of-interest (ROI) based analysis as it does not require an a priori hypothesis regarding regions affected by disease. Conventionally, VBM on DTI images is performed via analyzing the scalar maps of anisotropy and/or diffusivity that are computed from the tensor and then spatially normalized to a common template. A statistical voxel-wise p-value map is then computed from these scalar maps with the application of standard tests (e.g. t-tests) for statistical inference. Commonly used scalar images for such analyses include the fractional anisotropy (FA) and the apparent diffusion coefficient (ADC) [28, 30, 37]. The main disadvantage of these methods is that they do not use the complete information available in the DTI dataset, but rather make an a priori assumption that group differences will affect only a particular aspect of the tensors which is usually quantified through scalar indices like FA. Moreover, the combined results from different scalar maps may be difficult to interpret as they can potentially contain spatially overlapping patterns. Tensors provide both shape information (in the form of eigenvalues) that are captured in scalar maps of anisotropy and diffusivity measures computed from the tensor data and underlying fiber orientation information in the form of eigenvectors. However, voxel-wise statistical analysis of the tensors is complicated as the tensor is a 3×3 positive definite symmetric matrix located at each voxel and has an underlying non-linear manifold structure [3, 14, 31]. Furthermore, disease-induced changes between the subjects may not be linear adding to the complexity of the group-based statistical analysis.

Non-scalar features [34, 35, 42] such as the principal eigen-directions (PD) of the tensors have also been used in some analyses. In [34] a Bipolar-Watson model was introduced for analysis of PD's. This model takes into account the symmetric nature of the tensor while performing the statistical analysis and assumes the underlying distribution to be a Wishart distribution. However, in the regions where white matter fibers cross, the tensors are oblate in nature and therefore applying statistics to PD's may demonstrate ambiguous results.

A few methods have attempted on analyzing tensors in the VBM type setting. For example, work by Li et al. performed tensor regression analysis [27] where a linear regression model was assumed that imposed distributional assumptions on the tensors under consideration. In some other works, [15, 26] tensors were

characterized using Riemannian symmetric spaces. In [3] a simple and efficient Riemannian framework based on Log-Euclidean (LE) transform was introduced. Such methods rely upon the assumption that the tensors around a given voxel from various subjects belong to a principal geodesic sub-manifold and that these tensors obey a normal distribution on that sub-manifold. The basic principle of these methods is sound, namely that statistical analysis of tensors must be restricted to the appropriate manifold of positive definite symmetric tensors, which is known to be a cone embedded in \mathfrak{R}^6 . However, there is no guarantee that the representations of the tensors on this sub-manifold will have normal distributions since the pathology imposes its own structure and the tensors measured at a given voxel, from n subjects, typically lie on a much more restricted submanifold of the space of symmetric positive definite matrices. A novel approach was suggested by Verma et al., where a manifold learning method (Isomap) was employed for tensor analysis [40]. The focus of this work was on learning embeddings (or features) parameterizing the underlying manifold structure of the tensors. The learned features belonged to a low-dimensional linear manifold parameterizing the higher-dimensional tensor manifold and were subsequently used for group-wise statistical analysis. In general, manifold learning approaches [10] may be used to estimate the embedding of the manifold that represents the tensor measurements fairly well, however, depending on the number of samples used to learn the underlying manifold structure, it may not always be possible to determine the structure or validate its correctness or they may fail to estimate the probability distribution (non-Gaussian) on the (flattened) manifold itself.

This chapter is aimed towards providing a paradigm for voxel-wise tensor statistics by determining the underlying statistical distribution of the data and using this distribution for subsequent voxel-wise analysis. To achieve this, a novel method named kernel-based morphometry (KBM) is developed which demonstrating that it can accurately estimate the underlying distribution of the tensor data compared to other existing methods. A standard statistical test is then performed on these projections with appropriate testing for multiple comparisons.

2 Kernel Based Approach to Group-Wise Voxel Based DTI Statistical Analysis

Group-wise voxel-based statistical analysis of DTI data involves spatially normalizing all DT images to a common DTI template using a suitable technique [12, 22, 43, 46] and then using an appropriate voxel-wise statistical test to infer regional differences between groups based upon the tensors at (or around) a voxel. In this kernel based approach, the underlying distribution of the data at each voxel is determined and subsequently statistics are performed.

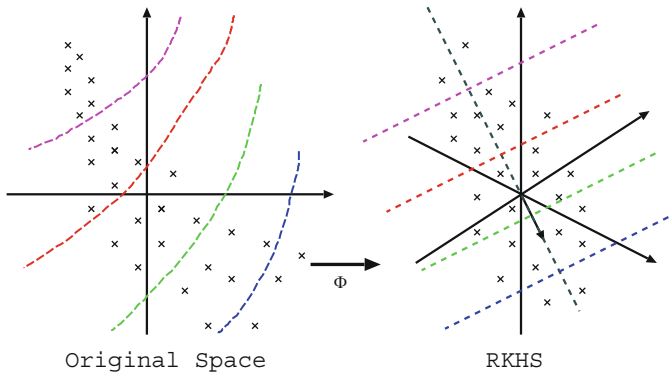


Fig. 1 Kernel-based projections: The mapping ϕ takes points (marked with crosses) from the original non-linear space to the linearized RKHS. Hyperplanes having constant projections onto a vector in the RKHS become curved lines in the original space. Such curved lines can give us important insight into how the corresponding RKHS projection parameterizes the original points

Kernel Principal Component Analysis (kPCA) [33] is a suitable method for learning the underlying data distribution. The common idea in any of the kernel-based techniques is to transform the samples into a higher-dimensional reproducible kernel Hilbert space (RKHS). Samples can be expressed using an appropriate kernel in a higher dimensional space using the well known “kernel trick” [33]. The non-linear hypersurfaces in the original space are mapped into hyperplanes in RKHS. These hyperplanes separate the given samples linearly in RKHS which is equivalent to a non-linear separation in original space. Subsequently, statistical operations can be performed in this “kernelized” space. Thus, in case of DTI, at each voxel, the intensities are kernelized to hyperplanes in the RKHS. Figure 1 illustrates the idea behind obtaining such components. Since these components are linear in the RKHS, linear tests for statistical inference, such as the Hotelling T^2 test, can be reliably applied to these projections in order to identify separation between groups.

Before presenting the kPCA technique in detail, here is a brief note on our mathematical conventions: Vectors are denoted by bold-faced lower case letters, e.g. \mathbf{x} , and matrices by upper-case letters, e.g. S for tensor matrix with U as the eigenvector matrix and D as the diagonal matrix containing the eigenvalues of the tensor. Vector of all 1’s is denoted by \mathbf{e} while the identity matrix is denoted by I . Vector of all 1’s in m -dimensional space is represented by \mathbf{e}_m . Matrix transpose and the matrix inverse are denoted by superscripts T and $^{-1}$ respectively. The sample mean of a set of vectors $\{\mathbf{x}_i, i = 1, \dots, K\}$ is represented as $\bar{\mathbf{x}}$, while the inner product of two vectors $\mathbf{x}_i, \mathbf{x}_j$ is denoted by $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. Group-wise study includes a statistical analysis of the DT images of N subjects, with N_+ subjects in one class (the positive class) and the remaining N_- subjects in a second class (the negative class).

3 Kernel Principal Component Analysis (kPCA)

We now describe the kPCA technique [33] by which one can find a rich linear representation of our voxel-based samples as well as provide an accurate estimate of the probability density underlying these samples. In conventional PCA, principal directions in the vector space of the samples that maximize the variance of the components of the samples along those directions and which also minimize the least-squares representation error for the samples are determined. In kPCA, similar principal eigen-directions in higher-dimensional RKHS are found where can be safely assumed to be normally distributed.

A DT image consists of a 3×3 positive-definite symmetric matrix or tensor S at each voxel in the image. Earlier work presented by Khurd et al. [24] used the diffusion tensor directly as a $6D$ vector that could potentially lead to inaccurate results since the tensors lie on the geodesic sub-manifold which is known to be a cone embedded in \mathfrak{R}^6 . Therefore, the log-Euclidean form of tensor was employed that retained the key attributes of affine-invariant Riemannian metric, and allowed standard Euclidean computations in the space of matrix logarithms, as was described by Arsigny et al. [3]. A tensor S can be represented as $S = UDU^T$ where matrix U is a matrix of its eigenvectors and D is a diagonal matrix that consists of the three eigenvalues. The log-Euclidean form of the tensor is given by Eq. 1.

$$S^{le} = \log(S) = U \log(D) U^T \quad (1)$$

$$\mathbf{x} = (S_{xx}^{le}, S_{yy}^{le}, S_{zz}^{le}, \sqrt{2}S_{xy}^{le}, \sqrt{2}S_{xz}^{le}, \sqrt{2}S_{yz}^{le})^T \quad (2)$$

A similarity invariant log-Euclidean form of the tensor is computed using Eq. 1 [3]. The 6-dimensional vectors are then obtained using Eq. 2 for all our N subjects $\mathbf{x}_1, \dots, \mathbf{x}_N$. Let us denote the nonlinear mapping of this vector \mathbf{x} into the Hilbert space by $\phi(\mathbf{x})$, and let us denote the underlying kernel by $k(.,.)$, where $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j)$. Let $\bar{\phi}$ denote the mean of $\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)$. Since a principal eigenvector \mathbf{v} in the higher-dimensional Hilbert space lies in the span of the vectors $\phi(\mathbf{x}_i) - \bar{\phi}, i = 1, \dots, N$, it can be conveniently represented as $\mathbf{v} = \sum_i \alpha_i (\phi(\mathbf{x}_i) - \bar{\phi})$, where α is an N -dimensional vector. Components of any sample along the eigenvector \mathbf{v} can now be conveniently computed using this new representation in the kernel basis.

The entire kPCA procedure is summarized below [10]:

Another alternative to using LE form of tensors and then computing the kernel $K(x_i, x_j)$ is employing directly a kernel $K(\log(s_i), \log(s_j))$ where s is the original tensor (in a 6 dimensional vector form).

In addition to finding the orthogonal directions of maximal variance in the higher-dimensional RKHS, kPCA also provides an estimate of the probability density underlying the samples. It has been pointed out by Girolami et al. [18] that kPCA with a Gaussian radial basis function kernel amounts to orthogonal series

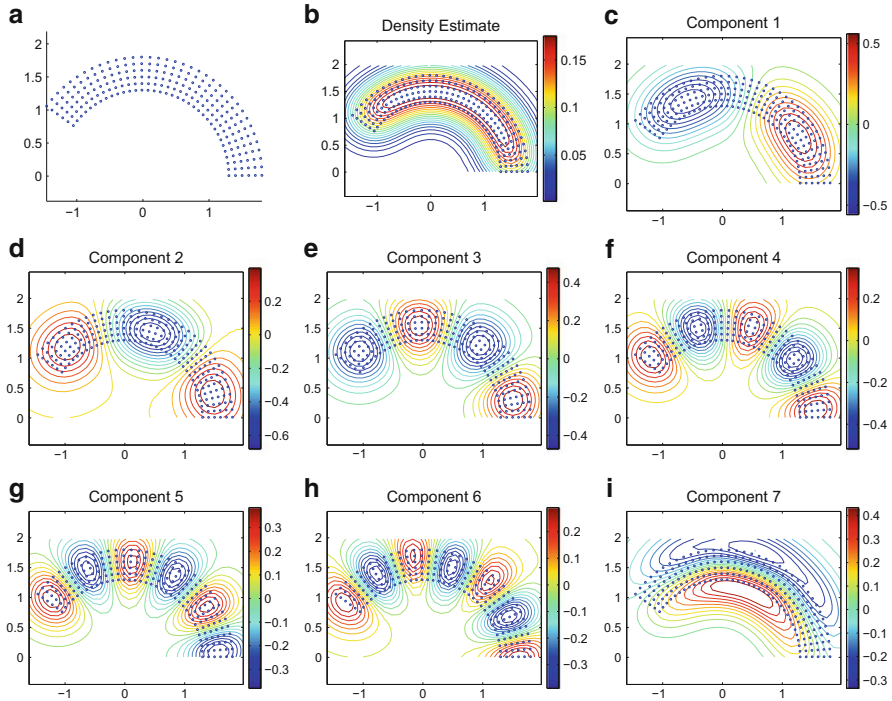


Fig. 2 (a) Synthetic dataset, (b) Contour plot for kernel probability density estimate, (c) Contour plot for 1st kPCA component, (d) Contour plot for 2nd kPCA component, (e) Contour plot for 3rd kPCA component, (f) Contour plot for 4th kPCA component, (g) Contour plot for 5th kPCA component, (h) Contour plot for 6th kPCA component, and (i) Contour plot for 7th kPCA component (please see text for explanation in Sect. 3)

density estimation using Hermite polynomials. Gaussian kernels are frequently employed in alternative kernel-based classifiers such as support vector machines [33]. The advantages of using Gaussian kernel are multifold; it non-linearly maps the samples into RKHS, involves less number of parameters than a polynomial kernel and is known to be robust. The Gaussian σ value was chosen to be based on the average distance between nearest neighbors (NN) x_i and x_j for e.g. $\|x_i - x_j\|$ and our choice was motivated by the desire to obtain meaningful representations for the different kPCA components.

In Sect. 5.1, simulated example is presented (see Fig. 2) where kPCA provides an accurate parametrization of the underlying density of the dataset. We note that the kPCA components constitute a linear representation of the tensors in the RKHS, which considerably simplifies further statistical analysis that will need to be performed on the dataset. An important issue is selecting the number of kPCA

Algorithm 1 kPCA

1. Form the kernel matrix K , where $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $i = 1, \dots, N$, $j = 1, \dots, N$.
2. Center the kernel matrix to obtain $K_c = (I - \frac{1}{N}\mathbf{e}\mathbf{e}^T)K(I - \frac{1}{N}\mathbf{e}\mathbf{e}^T)$.
3. Eigen-decompose K_c to obtain its eigenvectors $\alpha^{(i)}$ and eigenvalues λ_i , $i = 1, \dots, N$ ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$).
4. Normalize the eigenvectors $\alpha^{(i)}$ to have length $\frac{1}{\sqrt{\lambda_i}}$ so that the eigenvectors $\mathbf{v}^{(i)}$ in the RKHS have unit length.
5. The i^{th} kPCA component for training sample \mathbf{x}_k is given by:

$$\langle \phi(\mathbf{x}_k) - \bar{\phi}, \mathbf{v}^{(i)} \rangle = \lambda_i \alpha_k^{(i)}$$

6. For a general test point \mathbf{x} , the i^{th} kPCA component is:

$$\langle \phi(\mathbf{x}) - \bar{\phi}, \mathbf{v}^{(i)} \rangle = \sum_m \alpha_m^{(i)} k(\mathbf{x}, \mathbf{x}_m) - \frac{1}{N} \sum_{m,n} \alpha_m^{(i)} \alpha_n^{(i)} k(\mathbf{x}, \mathbf{x}_n)$$

components used for subsequent statistical analysis. This number can be chosen by looking at the kPCA eigenvalue spectrum and selecting only those eigenvectors that correspond to large eigenvalues. The notion of “large” eigenvalues is empirically defined using an application-specific threshold in one of two ways. The threshold may either specify the minimum energy $\frac{\sum_{i=1}^L \lambda_i}{\sum_{i=1}^N \lambda_i}$ that should be present in the retained L eigenvalues, or it may specify a minimum value for the ratio of the smallest retained eigenvalue λ_L to the largest retained eigenvalue λ_1 . For good discriminatory performance between the groups, the number of kernel PCA components chosen should not exceed the number of samples in either class. Statistical p-value maps are then computed using Hotelling’s T^2 statistic on the retained kPCA projections. This procedure is repeated at each voxel to obtain the kernelized version of the tensors at that voxel. These are now vectors in a high-dimensional linear space. Thus linear statistical tools for high dimensional data such as Hotelling’s T^2 can then be applied to the retained kPCA components. The resulting p-value map can then be thresholded to obtain regions of interest.

To overcome the multiple comparisons problem associated with voxel-wise analysis, False Discovery Rate (FDR) is implemented. This method controls the expected proportion of falsely rejected hypotheses [9]. The FDR threshold is determined from the observed p-value distribution, and therefore is adaptive to the amount of information in a given dataset [17].

The entire computational procedure for statistical analysis of tensors using kPCA also referred as KBM is summarized below:

This method can also be applied directly to eigenvectors of the tensors for studying the groupwise orientation changes. Performing KBM on a DTI population encompasses changes in scalar maps like FA as well as the orientation.

Algorithm 2 KBM of DTI data:

- **Input:** DTI datasets spatially normalized to a standard template. (N_+ subjects in one group and N_- subjects in the other group).
- **Output:** p-value maps indicating regional differences between the two groups (in general patients and controls).
- *Parametric p-value map:*
 - For each voxel $v = 1, \dots, V$
 - Compute the log-Euclidean form of tensors from N subjects.
 - Apply kPCA (refer to algorithm 1 above).
 - Select the number of kPCA components (using the energy threshold criterion).
 - Compute Hotelling's T^2 statistic $T^2(v)$ on the kPCA components and the parametric p-value $p(v)$.

Regions of significance can be identified by controlling the FDR using a suitable p-value threshold. Genovese et al. [17] have recommended the usage of threshold p-values of < 0.1 .

4 kPCA Based kFDA

It has been showed that kernel Fisher Discriminant Analysis (kFDA) could be an alternative tensor analysis method to kPCA [24]. kFDA focuses on finding non-linear projections of the tensorial data which can optimally discriminate between two groups. It computes a direction in higher order RKHS such that the projection along this direction maximizes a separability measure known as Rayleigh coefficient (or Fisher discriminant ratio). To quantify the group difference a T^2 test was performed on the kFDA components [24].

It is important to note that kFDA solution uses the group labels in obtaining the scalar projections and therefore permutation tests on the T^2 statistic computed from these projections are essential in finding meaningful p-value maps. The permutation tests on a million voxels can be computationally expensive. To circumvent the permutation tests, we provide an alternative analytical kFDA solution, based upon eigen-decomposition, as is shown by Baudat et al. [7]. This analytic solution has been shown to be mathematically equivalent [45] to first performing kernel PCA on the input data, followed by ordinary FDA. Therefore, we shall refer to this alternative solution as kPCA-based kFDA. An advantage of the analytic solution is that one can reduce the number of kernel PCA components used in the subsequent ordinary FDA and obtain superior discriminatory performance. For good discriminatory performance, the number of kernel PCA components used in the subsequent ordinary FDA should not exceed the number of samples in either class. In practice, this number is chosen based upon the kernel PCA eigenvalue spectrum as discussed in Sect. 3. On account of the equivalence between the T^2 statistic and FDA (Appendix), a second advantage is that it is possible to compute p-values using the T^2 statistic on the retained kPCA components in a faster parametric manner, with a small loss in accuracy, in comparison to performing permutation tests on kFDA components. The p-value map computation procedure using kPCA-based kFDA is identical to Algorithm 2 presented in Sect. 3.

5 Application

The kPCA framework is applied on three types of datasets: (1) simulated 2D datasets with the purpose of testing parameters of the kPCA analysis; (2) real datasets in which changes in shape and orientation have been simulated to study the practical applicability of KBM to group-wise population studies. Knowing the ground truth, that is, the magnitude of changes introduced, makes it easier to evaluate the differences captured by kPCA analysis and identify possible false positives; and (3) A group analysis between children with Autism spectrum disorder (ASD) and typically developing (TD) controls. We now describe the details of each of the experiments.

5.1 Kernel Based Analysis of Simulated Datasets

In this experiment the aim is to establish that the kernel-based method is able to identify the changes in shape and orientation in tensors when the changes occur in combination as could be in the case of pathology-induced changes. A 2-dimensional dataset with variation in the radial and angular directions was created that modeled a tensorial dataset with changes in the principal eigenvalue and eigen-direction. The purpose for using only two-dimensions was to make understanding and visualization straightforward (Fig. 2).

The synthetic dataset consisted of points forming a semi-circular band (see Fig. 2a) and was generated using 36 angles (in the 0–144° range) and 6 radial values (in the range 1.3–1.8) resulting in a total of 216 points. The aim was to check whether kernel-based morphometry paradigm was able to capture both these changes. The kernel-based procedure (Algorithm 1) was applied to this dataset using a Gaussian radial basis function (RBF) with the kernel width σ^2 set to 0.1 ($\sigma = 0.316$). The kernel width parameter was based on the average distance between nearest neighbors \mathbf{x}_i and \mathbf{x}_j , i.e. $\|\mathbf{x}_i - \mathbf{x}_j\|$ and the number of samples.

Figure 2c–i shows the iso-contour plots for 7 principal kPCA components representing the hyperplanes having constant projections onto the corresponding 7 RKHS eigenvectors, as was described earlier using Fig. 1. It was observed that the first 6 components (Fig. 2c–h) represented the angular changes in the data using varying scales, the third kPCA component (Fig. 2e) divided the angular variation in the data into four regions and alternately assumed positive and negative values as we move along the angular direction across these four regions. Only the seventh kPCA component (Fig. 2i) individually captured the radial change in the data and it smoothly increased from negative to positive values in the radial direction.

5.2 Analysis of DTI Datasets with Simulated Changes

The simulation in the previous Sect. 3 was then extended to a more realistic scenario as now the changes were simulated in real datasets. This was performed to determine whether KBM is able to capture combined shape and orientation changes, that a simple voxel-based morphometry of the scalar maps of anisotropy and diffusivity is unable to obtain.

The DTI data consisted of scans of 36 healthy volunteers (17 male and 19 female). These DT images were acquired on a Siemens TrioTM 3.0 Tesla scanner, using a single shot spin-echo, echo planar imaging (EPI), with 12 diffusion directions with a b-value of 800 s/mm² and TR/TE = 6,400/97 ms. Forty axial slices with 128 × 128 matrix, were acquired with a voxel-size of 1.72 × 1.72 × 3.0 mm. The diffusion tensor images were reconstructed from the DWI data using multivariate linear fitting [32]. The FA images that were computed from the tensors, were deformably registered elastically to a chosen healthy subject as template, by hierarchically matching features that provide a rich morphological signature for each voxel [36], The deformation was applied to the tensors while reorienting them using the underlying rotation component of the transformation [44]. We then identified an ROI on the template in the corpus callosum, as shown in Fig. 3a, and introduced spatially smooth random changes in the principal eigenvalue and the azimuthal angle for the principal eigenvector of each tensor into the appropriate ROI for all unwarped subject DT images. The random changes were designed to slightly increase the principal eigenvalue (average 4.6 % change) and the principal azimuthal angle ($\leq 20^\circ$), on average, but were subtle enough so that these changes could not be visibly easily discerned on an FA map or a colormap for the principal direction. These changes emulated changes in FA and orientation in the tracts. The DT images with the introduced random changes were then warped back to the template resulting in 35 DT images belonging to the class with induced pathology. The tensors were then transformed to log-Euclidean space. KBM method was tested using two different cases: (1) using LE tensors (without smoothing) (2) by applying 4 mm FWHM Gaussian blur to the LE tensors. All the tests were performed using 3 kPCA components, 8 kPCA components and 12 kPCA components. Following kPCA, the statistical p-value maps were computed using Hotelling T^2 test. KBM method uses a vector (LE tensor in this case) as an input and hence can be applied for analyzing of one of the eigenvectors. To demonstrate this adaptability, we applied our method to analyze the orientation of the simulated tensors. The principal direction (PD) defined by a 3D vector was chosen for the analysis. The issue of antipodal symmetry of eigenvectors was resolved by making sure that all the vectors lay in the positive z-hemisphere. The kPCA framework described in Sect. 3 was applied to the PDs.

The effect of applying kPCA to PDs, tensors and the effect of change of parameters was quantitatively evaluated based on the percentage overlap of voxels

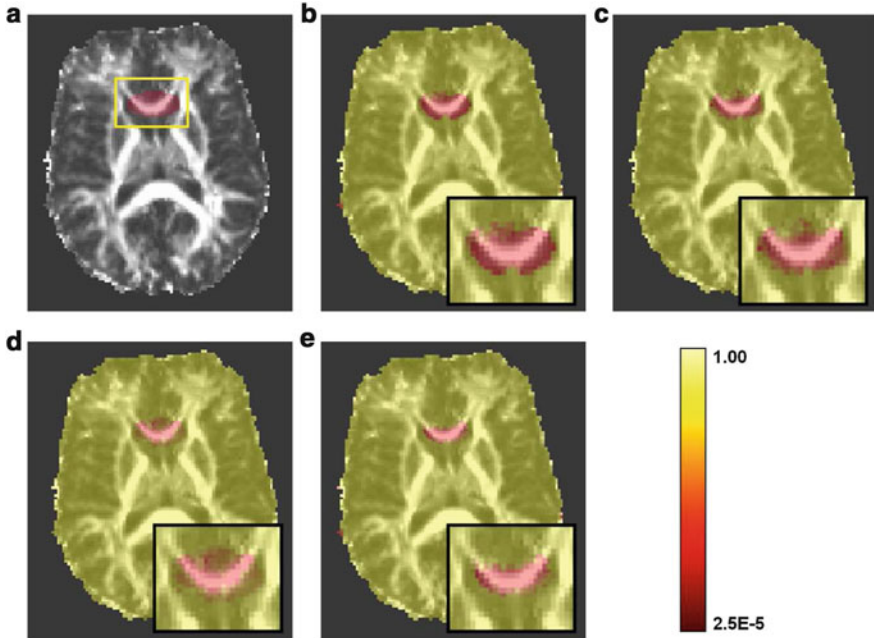


Fig. 3 (a) ROI with changes (highlighted) overlaid on the template FA map. P-value map computed from voxel-wise kPCA using 8 components is overlaid on the template FA when (b) Smooth tensors are used (c) when original tensors are used, (d) voxel-wise t-test on FA and (e) when principal eigen-directions are used. Low p-value regions in (b) better ($\approx 66\%$) match the true ROI in (a) than in (c) ($\approx 59\%$) after thresholding at p-value of 0.1. For FA analysis in (d), the sensitivity is lower than kPCA on tensors. Although most of the ROI is detected, the p-values are higher. For PD analysis in (e) changes in anisotropic areas are better detected than in isotropic regions. The ROI's are zoomed in each case for better visualization

in the detected ROI (the voxels that showed to be significantly different based on a threshold of p-value 0.1) with the voxels in the original ROI in which the changes were introduced. We compared our method with voxel-wise t-test on FA and ADC as well as the Isomap-based method introduced in [40]. The kPCA on PD's was compared with the Bipolar-Watson method on the PD's introduced by Schwartzman et al. in [34].

Results are shown in Fig. 3b–e. Figure 3b displays the p-value map after performing group analysis on the kernelized data at each voxel. It can be observed that the simulated ROI has a very low p-value range. Similarly, Fig. 3c shows the p-value map after performing kPCA analysis on tensors. Figure 3d, e show the ROI detected by t-test on FA and kPCA on PD's respectively. Table 1 gives all the percentage overlap values from kPCA compared to FA analysis, ADC analysis and Isomaps. The FA and ADC analysis involved a voxel-wise t-test.

Table 1 Percentage overlap of detected ROI (p-value map threshold at a cut-off of 0.1) with the ground-truth ROI in which changes had been introduced (please refer to Sect. 5)

	Analysis method	Original DTI	Smooth DTI
1.	ADC	14.92	21.71
2.	FA	39.14	42.39
3.	ISOMAP – 3 components	36.19	45.50
4.	kPCA on tensors – 3 components	57.39	65.56
5.	kPCA on tensors – 8 components	59.43	66.34
6.	kPCA on tensors- 12 components	64.38	87.42
7.	kPCA on PD – 3 components	42.15	45.76
8.	kPCA on PD – 8 components	42.87	47.55
9.	Bipolar Watson on PD	39.84	42.92

5.3 *kPCA Analysis on Autism Spectrum Disorder*

Finally, the method of KBM was applied to a real population of subjects with ASD pathology and typically developing (TD) controls. In this study 26 TD controls (mean age=10.7) and 44 subjects with ASD (mean age=9.8) were used. The images were taken using Siemens 3T VerioTM scanner using a 32 channel head coil. DTI was performed using a single shot spin-echo, echo-planar sequence with the following parameters: TR/TE=16,900/70 ms, b-value of 1,000 s/mm² and 30 gradient directions. Eighty axial slices of 128 × 128 matrix (FOV 256 mm) were acquired yielding 2 mm isotropic data. The diffusion tensors were estimated using the least squares fitting method and then spatially normalized to a standard template described in Wakana et al. [41]. The deformable registration utilized the full tensor information by integrating intensity and orientation into a hierarchical matching framework [22]. KBM analysis was then carried out on the tensors using an energy threshold of 80% and with a sigma of 4.0. FA and mean diffusivity (MD) maps were computed from the spatially normalized dataset. A comparative voxel-wise FA and MD analysis was performed on the same dataset by employing a standard t-test between the groups. In all the analyses, the p-values were thresholded at a significance level of $p < 0.05$ and the results were overlaid on the template FA image. The resulting images are displayed in Fig. 4 indicating the regions of differences between subjects with ASD and the TD controls. The kernel based method could capture multiple areas of significance that included left superior longitudinal fasciculus (SLF), left inferior longitudinal fasciculus (ILF) and left inferior fronto-occipital fasciculus (IFO) and parts of right and left internal (IC) and external capsule (EC). The conventional FA and MD voxel-wise results are shown in Fig. 4b, c respectively. From Fig. 4b it can be observed that FA captures only the differences in the right external capsule (EC), while MD analysis shows significance only in inferior regions that include the ILF and IFO as seen in Fig. 4c. Multiple comparisons using FDR at 0.1 threshold (on tests for FA, MD and KBM on tensors) could not survive any of the voxels.

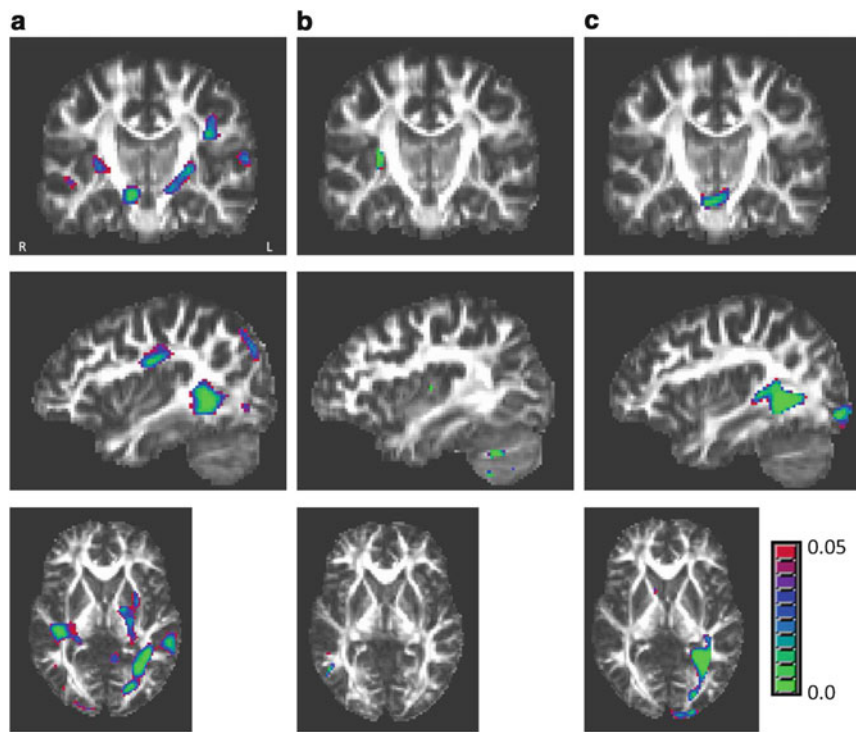


Fig. 4 Displays the differences between ASD and TD groups. (a) Result from k-PCA on tensors. The areas include left SLF (includes arcuate fasciculus and acoustic radiation), ILF, IFO, IC and EC while (b) show the result from voxel-wise FA analysis in which only the right EC is captured and (c) significant areas captured by MD changes that include ILF and IFO

6 Summary

In this chapter we address the problem of tensor-based population statistics of DTI data by employing a kernel based morphometric method that can capture the underlying distribution of the data. Our sequence of experiments shows that the mapping of data to a kernelized higher dimensional space enhances group separation and also models the underlying changes in the data.

To validate the kernel-based procedure we first applied it Sect. 5.1 on simulated data. This aided in determining whether it was able to capture combined changes in shape and orientation and whether this depended on the number of components. The experiment demonstrated that using different number of components achieves different degrees of separability in the data, between the different kinds of changes. Moreover, we found that it is important to utilize an adequate number of features/components for better group separation. Therefore, an energy threshold

criteria was defined as was described in Sect. 3. It was also noted that if a high value of σ was used (e.g. 10 times of NN distance) then the separability of the shape and orientation could not be obtained even when maximum allowable components were used. On the other hand, a low σ value would have lead to overfitting of the data. Thus selecting an optimum σ value was critical.

In the next experiment (Sect. 5.2) pathological changes were simulated in the genu of the corpus callosum and the surrounding CSF (in the form of spatially smooth subtle random changes in the principal eigenvalue and the PD). The main reason behind picking such an area was to evaluate KBM in areas with high anisotropy (genu) as well as with low anisotropy (CSF). These simulations were mainly created for better validation of the KBM method as there are no models or ground truth for the variability that a disease may introduce in the data. From the results shown in Table 1, rows 1 and 2 are the outcome of a conventional t-test on ADC and FA and row 3 presents the results using the Isomap technique from [40]. The t-test on FA detected only regions of higher percentage change in eigenvalues with a significance lower than 0.1. The Isomap technique (row 3) performed better than the first two approaches (indicating the non-linear nature of data variation), but perhaps it suffered as it does not utilize the knowledge of the underlying distribution of data. Knowledge of the statistical distribution led to improved results using the kPCA technique on the LE tensors (Table 1 rows 4–6). It was observed that the resulting overlap from kPCA was better when the LE tensors were smoothed and when larger number of features were used (87 % when 12 components were used on smoothed datasets). The results also improved when richer number of features were used (that is features had more components). In CSF, since the average eigenvalues were slightly increased, we expected to see shape differences which were significantly caught by kPCA method whereas FA picked it up subtly. Row 7 and 8 present results using kPCA on PD's while row 9 shows the results of using Bipolar-Watson model introduced in [34] to determine changes in PD. Although the ROI overlap using kPCA on PD's using 8 components, ($\approx 47\%$) was much lower than the tensor overlap ($\approx 66\%$), it was better than using the Bipolar-Watson method which showed only 42 % overlap after thresholding at a p-value of 0.1. Although the changes were introduced in the principal eigenvalue and eigenvector, it seems that methods that targeted each of these changes individually (FA-based and PD-based) were unable to capture it fully as opposed to when the full tensor information was used for statistical analysis. Thus suggests that changes in shape and orientation are difficult to detect by conventional methods, however the KBM is able to capture mixed changes. Since combination changes are expected in pathology, we have established the importance of our method in performing large population studies in which changes cannot be hypothesized a priori. It may be noted that if it were known a priori that the changes were only in shape or orientation, we could examine just that aspect, but in the absence of such knowledge, it is important to study the full tensor.

Finally, we performed KBM on a population of subjects with ASD (Sect. 5.3), to demonstrate the applicability of our method on clinical datasets. DTI based research in ASD has mainly involved studying WM changes using anisotropy and diffusivity values [39]. Abnormalities have been reported in WM structures like the genu and splenium of the corpus callosum [1], the internal capsule [23] and in the temporo-parietal regions [5, 21].

The results from our analysis displayed in Fig. 4 indicate the regions of differences between subjects with ASD and the TD controls. The resulting differences from the kernel-based method suggest WM abnormalities and hypo-connectivity between brain regions which may be strong contributors to the social deficits that are hallmarks of the ASD phenotype. For example, the changes observed in SLF (which includes the arcuate fasciculus) can be linked to the language impairment often observed in the ASD population [16] while the differences in the internal capsule were comparable to the previous finding by Keller et al. [23]. The p-values computed from the kernelized analysis as well from FA and MD analyses could not survive the FDR correction (at 0.1 threshold), perhaps owing to the heterogeneity in the population, small sample size and/or subtle differences between the groups. However, the aim here was to demonstrate future clinical applicability of the method as it was able to capture more changes than were observed using conventional analysis of DTI.

It was shown in Sect. 5.2, that tensor analysis could capture interplay between combined shape and orientation changes which individual (FA or PD) analysis could not capture. Similarly in the ASD example, the significant regions using kPCA on tensors in Fig. 4a–c included many areas like SLF and IC that are known to be affected in ASD while FA failed to capture the pathological abnormalities in ASD. This demonstrates that KBM of tensors is able to find combined FA and PD changes that other methods are unable to, underlining the importance of full tensor statistics, in comparison to statistics on the scalar maps alone.

In summary, DTI analysis has the advantage of being more sensitive than the standard scalar or PD analyses, especially when the changes appear in combination (that is shape and orientation as could be the case in real data). However more work needs to be done for interpretation of results which will vary based on the dataset on which KBM has been applied. However, this work establishes the need for full tensor statistics in group-wise population studies. As the effect of pathology is not known, tensor analysis can be thought of as an unbiased method rather than using scalar indices computed from combination of tensor eigenvalues and eigenvectors. The wide range of experiments demonstrate applicability of kernel-based tensor morphometry for population statistics and provides a novel method of statistical analysis, based on capturing the underlying distribution of the data, that is specific to the disease that introduced the change.

Acknowledgements This work was supported by the NIH grants RO1-MH079938 and R01-MH092862.

Appendix

Proof of the equivalence between Hotelling's T^2 test and FDA [24].

Let $\mathbf{x}_{1i}, i = 1, \dots, N_1$ be the p -dimensional data vectors belonging to class 1 and let $\mathbf{x}_{2i}, i = 1, \dots, N_2$ be the p -dimensional data vectors belonging to class 2. Let $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ denote the means for classes 1 and 2. Let $S_x = \left(\frac{1}{N_1+N_2-2}\right) \left(\sum_{i=1}^{N_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)^T + \sum_{i=1}^{N_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)^T\right)$.

Hotelling's T^2 test: Then the T^2 test statistic is:

$$T_x^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (3)$$

Assuming normal distributions for $\mathbf{x}_{1i}, i = 1, \dots, N_1$ and $\mathbf{x}_{2i}, i = 1, \dots, N_2$ implies that $F_x = \frac{N_1+N_2-p-1}{(N_1+N_2-2)p} T_x^2$ has the cdf $F(p, N_1 + N_2 - p - 1)$. Therefore, the parametric form of Hotelling's T^2 test is sometimes referred to as the F-test.

FDA: The FDA optimal linear discriminant direction is $w = S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ and the corresponding scalar mapping is $y = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} x$.

Therefore, $\bar{y}_1 - \bar{y}_2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$,

$$\begin{aligned} S_y &= \left(\frac{1}{N_1 + N_2 - 2}\right) \left(\sum_{i=1}^{N_1} (y_{1i} - \bar{y}_1)(y_{1i} - \bar{y}_1)^T + \sum_{i=1}^{N_2} (y_{2i} - \bar{y}_2)(y_{2i} - \bar{y}_2)^T\right) \\ &= \left(\frac{1}{N_1 + N_2 - 2}\right) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} \left(\sum_{i=1}^{N_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)^T \right. \end{aligned} \quad (4)$$

$$\left. + \sum_{i=1}^{N_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)^T\right) S_x^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

$$= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (5)$$

and

$$\begin{aligned} T_y^2 &= \frac{N_1 N_2}{N_1 + N_2} (\bar{y}_1 - \bar{y}_2)^T S_y^{-1} (\bar{y}_1 - \bar{y}_2) \\ &= \frac{N_1 N_2}{N_1 + N_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \\ &= T_x^2 \end{aligned} \quad (6)$$

Thus, the T^2 statistic computed on vectorial input samples, i.e. T_x^2 , is mathematically equivalent to a one-dimensional T^2 statistic T_y^2 computed on the scalar samples obtained by performing Fisher discriminant analysis on the input vectorial samples.

References

1. Alexander, A.L., Lee, J.E., Lazar, M., Boudos, R., DuBray, M.B., Oakes, T.R., Miller, J.N., Lu, J., Jeong, E.K., McMahon, W.M., Bigler, E.D., Lainhart, J.E.: Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage* **34**, 61–73 (2007)
2. Ardekani, B.A., Tabesh, A., Sevy, S., Robinson, D.G., Bilder, R.M., Szeszko, P.R.: Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. *Hum. Brain. Mapp.* **32**, 1–9 (2011)
3. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.* **56**, 411–421 (2006)
4. Ashburner, J., Friston, K.J.: Voxel-based morphometry—the methods. *Neuroimage* **11**, 805–821 (2000)
5. Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., Reiss, A.L.: White matter structure in autism: preliminary evidence from diffusion tensor imaging. *Biol. Psychiatry* **55**, 323–326 (2004)
6. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *J. Magn. Reson. B* **111**, 209–219 (1996)
7. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. *Neural Comput.* **12**, 2385–2404 (2000)
8. Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* **34**, 144–155 (2007)
9. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., Golani, I.: Controlling the false discovery rate in behavior genetics research. *Behav. Brain. Res.* **125**, 279–284 (2001)
10. Burges, C.: Data mining and knowledge discovery handbook. Kluwer, New York (2005)
11. Calamante, F., Masterton, R.A.J., Tournier, J.D., Smith, R.E., Willats, L., Raffelt, D., Connelly, A.: Track-weighted functional connectivity (TW-FC): a tool for characterizing the structural-functional connections in the brain. *Neuroimage* **70**, 199–210 (2013)
12. Cao, Y., Miller, M., Mori, S., Winslow, R.: Diffeomorphic matching of diffusion tensor images. In: Proceedings of CVPR-MMBIA, New York (2006)
13. Field, A.S., Alexander, A.L.: Diffusion tensor imaging in cerebral tumor diagnosis and therapy. *Top. Magn. Reson. Imaging* **15**, 315–324 (2004)
14. Fletcher, T., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.* **87**(2), 250–262 (2007)
15. Fletcher, T., Lu, C., Pizer, S., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging* **23**, 995–1005 (2004)
16. Fletcher, P.T., Whitaker, R.T., Tao, R., DuBray, M.B., Froehlich, A., Ravichandran, C., Alexander, A.L., Bigler, E.D., Lange, N., Lainhart, J.E.: Microstructural connectivity of the arcuate fasciculus in adolescents with high-functioning autism. *Neuroimage* **51**, 1117–1125 (2010)
17. Genovese, C.R., Lazar, N.A., Nichols, T.: Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002)
18. Girolami, M.: Orthogonal series density estimation and the kernel eigenvalue problem. *Neural Comput.* **14**, 669–688 (2002)
19. Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O.: Mapping the structural core of human cerebral cortex. *PLoS Biol* **6**, e159 (2008)
20. Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V.J., Meuli, R., Thiran, J.P., Grant, P.E.: White matter maturation reshapes structural connectivity in the late developing human brain. *Proc. Natl. Acad. Sci. U S A* **107**, 19067–19072 (2010)
21. Ingalhalikar, M., Parker, D., Bloy, L., Roberts, T.P.L., Verma, R.: Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of ASD. *Neuroimage* **57**, 918–927 (2011)

22. Ingalhalikar, M., Yang, J., Davatzikos, C., Verma, R.: DTI-DROID: Diffusion tensor imaging-deformable registration using orientation and intensity descriptors. *Int. J. Imaging Syst. Technol.* **20**(2), 99–107 (2010)
23. Keller, T.A., Kana, R.K., Just, M.A.: A developmental study of the structural integrity of white matter in autism. *Neuroreport* **18**, 23–27 (2007)
24. Khurd, P., Verma, R., Davatzikos, C.: Kernel-based manifold learning for statistical analysis of diffusion tensor images. *Inf. Process. Med. Imaging* **20**, 581–593 (2007)
25. Kubicki, M., Westin, C.F., McCarley, R.W., Shenton, M.E.: The application of DTI to investigate white matter abnormalities in schizophrenia. *Ann. N Y Acad. Sci.* **1064**, 134–148 (2005)
26. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. *J. Math. Imaging Vis.* **25**(3), 423–444 (2006)
27. Li, Y., Zhu, H., Y. C., Ibrahim, J., An, H., Lin, W., Hall, C., Shen, D.: Racti: regression analyses of diffusion tensor images. In: *SPIE Conference Proceedings*, Dresden (2009)
28. Lim, K.O., Hedehus, M., Moseley, M., de Crespigny, A., Sullivan, E.V., Pfefferbaum, A.: Compromised white matter tract integrity in schizophrenia inferred from diffusion tensor imaging. *Arch. Gen. Psychiatry* **56**, 367–374 (1999)
29. Nimsky, C., Ganslandt, O., Hastreiter, P., Wang, R., Benner, T., Sorensen, A.G., Fahlbusch, R.: Preoperative and intraoperative diffusion tensor imaging-based fiber tracking in glioma surgery. *Neurosurgery* **56**, 130–7 (2005). Discussion 138
30. Park, H.J., Westin, C.F., Kubicki, M., Maier, S.E., Niznikiewicz, M., Baer, A., Frumin, M., Kikinis, R., Jolesz, F.A., McCarley, R.W., Shenton, M.E.: White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. *Neuroimage* **23**, 213–223 (2004)
31. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66**(1), 41–66 (2006)
32. Pierpaoli, C., Jezzard, P., Basser, P., Barnett, A.: Diffusion tensor MR imaging of the human brain. *Radiology* **201**, 637 (1996)
33. Scholkopf, B., Smola, A.: *Learning with Kernels*. MIT, Cambridge (2002)
34. Schwartzman, A., Dougherty, R.F., Taylor, J.E.: Cross-subject comparison of principal diffusion direction maps. *Magn. Reson. Med.* **53**, 1423–1431 (2005)
35. Schwartzman, A., Dougherty, R.F., Taylor, J.E.: Group comparison of eigenvalues and eigenvectors of diffusion tensors. *J. Am. Stat. Assoc.* **105**(490), 588–599 (2010)
36. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* **21**, 1421–1439 (2002)
37. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487–1505 (2006)
38. van den Heuvel, M.P., Mandl, R.C.W., Kahn, R.S., Hulshoff Pol, H.E.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* **30**, 3127–3141 (2009)
39. Verhoeven, J.S., Cock, P.D., Lagae, L., Sunaert, S.: Neuroimaging autism. *Neuroradiology* **52**, 3–14 (2010)
40. Verma, R., Khurd, P., Davatzikos, C.: On analyzing diffusion tensor images by identifying manifold structure using isomaps. *IEEE Trans. Med. Imaging* **26**, 772–778 (2007)
41. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C.M., Mori, S.: Fiber tract-based atlas of human white matter anatomy. *Radiology* **230**, 77–87 (2004). <http://radiology.rsna.org/cgi/reprint/230/1/77.pdf>
42. Wu, Y.C., Field, A.S., Chung, M.K., Badie, B., Alexander, A.L.: Quantitative analysis of diffusion tensor orientation: theoretical framework. *Magn. Reson. Med.* **52**, 1146–1155 (2004)
43. Xu, D., Hao, X., Bansal, R., Plessen, K.J., Peterson, B.S.: Seamless warping of diffusion tensor fields. *IEEE Trans. Med. Imaging* **27**, 285–299 (2008)

44. Xu, D., Mori, S., Shen, D., van Zijl, P.C.M., Davatzikos, C.: Spatial normalization of diffusion tensor fields. *Magn. Reson. Med.* **50**, 175–182 (2003)
45. Yang, J., Frangi, A.F., Yang, J.Y., Zhang, D., Jin, Z.: KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 230–244 (2005)
46. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Med. Image Anal.* **10**, 764–785 (2006)