A Pointwise Correspondence Based DT-MRI Fiber Similarity Measure

Sema Berkiten and Burak Acar

Abstract— **Diffusion Tensor Magnetic Resonance Imaging (DTI) fiber tractography is a way to reconstruct fiber tracts underlying data according to local anisotropic diffusion characteristics. Reliability of fiber tracts as a result of tractography decreases due to noise in the data, error accumulation during integration and stochastic nature of the underlying data. We proposed a new similarity measure based on point-wise correspondence between tracts. Laplacian Eigenmaps are used to embed the fiber tracts into $\Re^3$ based on the new similarity measure. We compared our method with a previously proposed method, on real and phantom data, that uses a 9D feature space to measure fiber similarity and showed that the new similarity measure results in a low dimensional manifold representing the fiber bundles. We presented preliminary results demonstrating that the fibers that fall far from this manifold correspond to outliers.**

I. INTRODUCTION

The human brain mapping refers to understanding the functional and the physiological structure of human brain. Diffusion Tensor Magnetic Resonance Imaging (DTI) is the unique technique for in-vivo visualization of the brain mapping. The DTI data has diffusion information (with a zero mean 3D Gaussian distribution) of a finite volume. DTI tensor field, which are 3×3 symmetric, positive semi-definite second-order tensors has direct relation with the covariance matrix of this distribution. Diffusion has an anisotropic nature and the principal eigenvectors of tensors are parallel to the underlying fiber orientation [1], [2]. In the basis of this knowledge, DTI tractography is done by integration of principal eigenvectors starting from some specific seed points [3]. Reliability of tractography decreases due to noise, high curvatures, and kissing/crossing tracts which can lead to misinterpreted fiber tracts [4]. In order to differentiate erroneous/unreliable fiber tracts and to visualize underlying anatomical meaning, several clustering techniques were proposed. Clustering algorithms are composed of the main parts: *i)* fiber tract similarity measure, *ii)* clustering the fiber tracts according to that similarity measure.

As similarity measure, Brun et al. proposed to represent the fibers in $\Re^9$ with feature vectors composed of the first and the second order moments of the x-, y-, z- coordinates along a given fiber tract[6]. The fiber similarity is defined as the closeness of Euclidean distance in $\Re^9$. For another work, they proposed a coarse method which is the Euclidean

Authors are with Department of Electrical and Electronics Engineering, Boğaziçi University, İstanbul, Turkey. Corresponding author: burak.acar@ieee.org, http://www.vavlab.ee.boun.edu.tr

distance between two fibers' endpoints in $\Re^3$ [7]. Ding et al. introduced the corresponding segment method based on measuring point-wise *corresponding* portions of two fibers [8]. A corresponding segment ratio is calculated as the ratio of the length of the corresponding segment over the total length of two fibers decreased by the length of the corresponding segment. Two fibers are said to be similar if their corresponding segment ratio is high and the mean point-wise Euclidean distance between them small. O'Donnell et al. proposed to use the average Hausdorff distance between two fiber tracts as the similarity measure [10], [9] while Zhang et al. proposed to use the average of distances above a threshold between fiber points [12], [11].

For solution of clustering problem, there are two commonly used methods: hierarchical clustering and spectral clustering. Brun et al. proposed the Normalized Cut (NCut) algorithm [6], [13]. As in all spectral clustering methods, the NCut algorithm depends on an $N \times N$ symmetric affinity matrix, $\mathbf{W}$, with w_{ij} being the similarity between i^{th} and j^{th} fiber tracts. In terms of graph theory, $\mathbf{W}$ is the edge weights of a graph with N vertices corresponding to N fiber tracts. Cost minimization for graph-cut is solved with the NCut algorithm as a generalized eigenvalue problem. It divides whole data into two groups. Bi -partitioning is applied by thresholding the entries of the N-D eigenvectors starting from the optimal solution, and continues on entries of disjoint parts. Brun et al. colored fiber tracts in a continuous manner, soft-coloring, with using three optimal solutions (forming $\Re^3$ space which is used as RGB color space) instead of performing an explicit clustering. Another spectral clustering technique, Laplacian Eigenmaps (LE), was proposed by Belkin et al. [5]. This method is based on the same generalized eigenvalue problem in NCut algorithm. LE finds the optimal solution to embed an input space (space of fiber tracts) into $\Re^n$, $1 \leq n \leq N$. Each resulting eigenvector is a mapping function from $\Re^N$ to $\Re^1$. The optimal solution is the eigenvector corresponding to second smallest eigenvalue as in NCut algorithm. As commonly, the three most optimal solution ($n = 3$) is used to map each fiber into $\Re^3$ and then each fiber is colored with three components as R-, G-, B- values. Alternatively, Zhang et al. proposed to use an agglomerative hierarchical clustering in which fiber tracts are merged until a specified number of clusters are formed [12]. Moreover, K-most Similar algorithm is used by Ding et al. as clustering method[8].

In this work, our contributions are in fiber tract similarity measurement and detection of erroneous fiber tracts. We proposed a new technique to calculate similarity between fiber tracts which is based on point-wise correspondence.
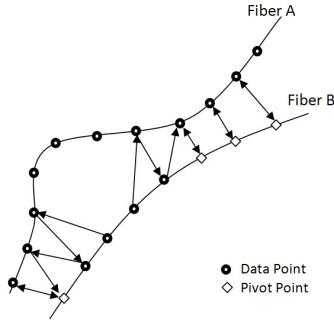
Fig. 1. The Pivot Points (PP) for a fiber tract is defined to be the points that are themselves closest to their nearest neighbours on the other tract. The arrows in the figure mark the nearest neighbours. The white marked points are the PPs where as the dark ones are normal data points. Note how the PPs take care of the high frequency deviations along the course of tracts, ie the short range deviation between the tracts decreased the number of PPs in this example.

Then we map fiber tracts to $\Re^3$ with using LE [5]. Outlier tract detection is achieved by a simple algorithm based on thresholding the neighbouring mean distance to k-nearest fiber tracts in $\Re^3$. Experiments are conducted on phantom and real DTI data.

II. METHOD

A. Fiber Tract Similarity Based on Pivot-Point

Similarity measure based on Pivot-Point (PP) has been inspired a method proposed by Ding et al. [8]. The basic idea underlying similarity measure depends on the assumption of similar fiber tracts following similar pathways throughout their courses. So pivot points have one-to-one mapping between each other. Although corresponding segment method is derived from similar idea to PP method, it does not take into account short deviations (high frequency noise) along tracts. First step in PP method is detection of them, of course. A PP can be defined as a point on the shorter tract, whose closest point on the longer tract is closest to itself along the shorter tract. PP pairs are shown in Figure 1. For this example shorter fiber is Fiber B, so the PPs are defined for Fiber B only. Each PP on Fiber B has a pairing point on Fiber A. Let $\rightarrow$ denote the *closest point* in terms of Euclidean distance in $\Re^3$. Let A and B denote the sets of equally spaced points along fibers A and B respectively. Then,

$$p \in B \text{ is a PP} \iff (p \rightarrow r) \wedge (r \rightarrow p) , \, p \in B, r \in A \quad (1)$$

After the PPs are determined, the similarity between Fibers A and B (w_{AB}) is defined in terms of the Pivot Ratio (PR_{AB}) and the mean Euclidean distance between pivot point pairs d_{AB} defined as,

$$PR_{AB} = \frac{N_{PP}}{N_A + N_B - N_{PP}} \quad (2)$$

$$d_{AB} = \|\mathbf{r_A} - \mathbf{r_B}\|_2 \quad (3)$$

$$w_{AB} = PR_{AB} \times e^{-\frac{d_{AB}}{K}} \quad (4)$$

where N_{PP} is number of PPs, N_A and N_B are total numbers of points on Fibers A and B respectively, $\mathbf{r}$ is position vector in $\Re^3$ of tract points, and K is a scaling factor that acts as a similarity threshold on d_{AB} only. When the two fibers are equal in length and totally parallel, PR_{AB} will have the maximum value of 1. Zero valued Pivot Ratio means that there is no PP detection, in other words, two are totally different from each other. d_{AB} is the mean Euclidean distance between PP pairs and measures the spatial proximity of fiber tracts. The purpose of such a metric is to differentiate two totally parallel fibers which are far away from two totally parallel fibers which are close to each other. w_{AB} is the final similarity measure, by definition, $0 \leq w_{AB} \leq 1$.

B. Embedding in $\Re^3$ through Laplacian Eigenmaps

One of the well known spectral clustering techniques is Laplacian Eigenmaps proposed by Belkin et al.[5]. A graph with nodes (fiber tracts) and edges (assigned with weight between two fibers according to their similarity) is constructed from the data. Then it seeks the optimal solution for a minimization problem of

$$\mathbf{y} = argmin_{\mathbf{y}} \sum_{ij} (y_i - y_j)^2 w_{ij} \quad (5)$$

where y_i is the value of embedding function for the ith node, w_{ij} is the similarity between i^{th} and j^{th} nodes by definition. The problem is formulated as a generalized eigenvalue problem as $(\mathbf{D} - \mathbf{W})\mathbf{y} = \lambda \mathbf{D}\mathbf{y}$ where $\mathbf{W}$ is the affinity matrix composed of weights, $\mathbf{D}$ is a diagonal matrix of row sums of $\mathbf{W}$ [5]. The trivial solution with $\lambda = 0$ is ignored. The eigenvector corresponding to the second smallest eigenvalue is the best solution. It is common to take N (=3) eigenvectors corresponding to N smallest eigenvalues (starting from the second smallest one) as the embedding functions. Let $\mathbf{y}_2$, $\mathbf{y}_3$ and $\mathbf{y}_4$ be the 3 eigenvectors used for embedding, then $\mathbf{e}_i = [\mathbf{y}_2(i), \mathbf{y}_3(i), \mathbf{y}_4(i)]$ represents the ith fiber tract in $\Re^3$. Thus this new embedding space in $\Re^3$ can be used to color each fiber tract as commonly used.

C. Outlier Fiber Tract Detection

Outlier fiber tracts are defined to be the ones that are far away from the embedded manifold, that represents the fiber bundle(s). Thus a simple outlier fiber tract detector can be based on the distance to the learned manifold. To demonstrate the feasibility of this approach, we have devised a simple outlier detector by thresholding on the mean distance of each fiber tract to its K (=12) nearest neighbours as follows,

$$\mu_i = \frac{1}{K} \sum_{j \in \mathcal{N}_i} |\mathbf{e}_i - \mathbf{e}_j| \quad (6)$$

$$\mu_i > \theta \quad \Rightarrow \quad i^{th} \text{ tract is an outlier} \quad (7)$$

where $\mathcal{N}_i$ is the K nearest neighbours of the ith tract, $\mathbf{e}_i$ is the ith tract as represented in the 3D embedding space and θ is an empirically determined threshold set to be 0.3×10^{-3}. Figure 3 shows the number of detected outliers as a function of the threshold θ for real patient DTI data.
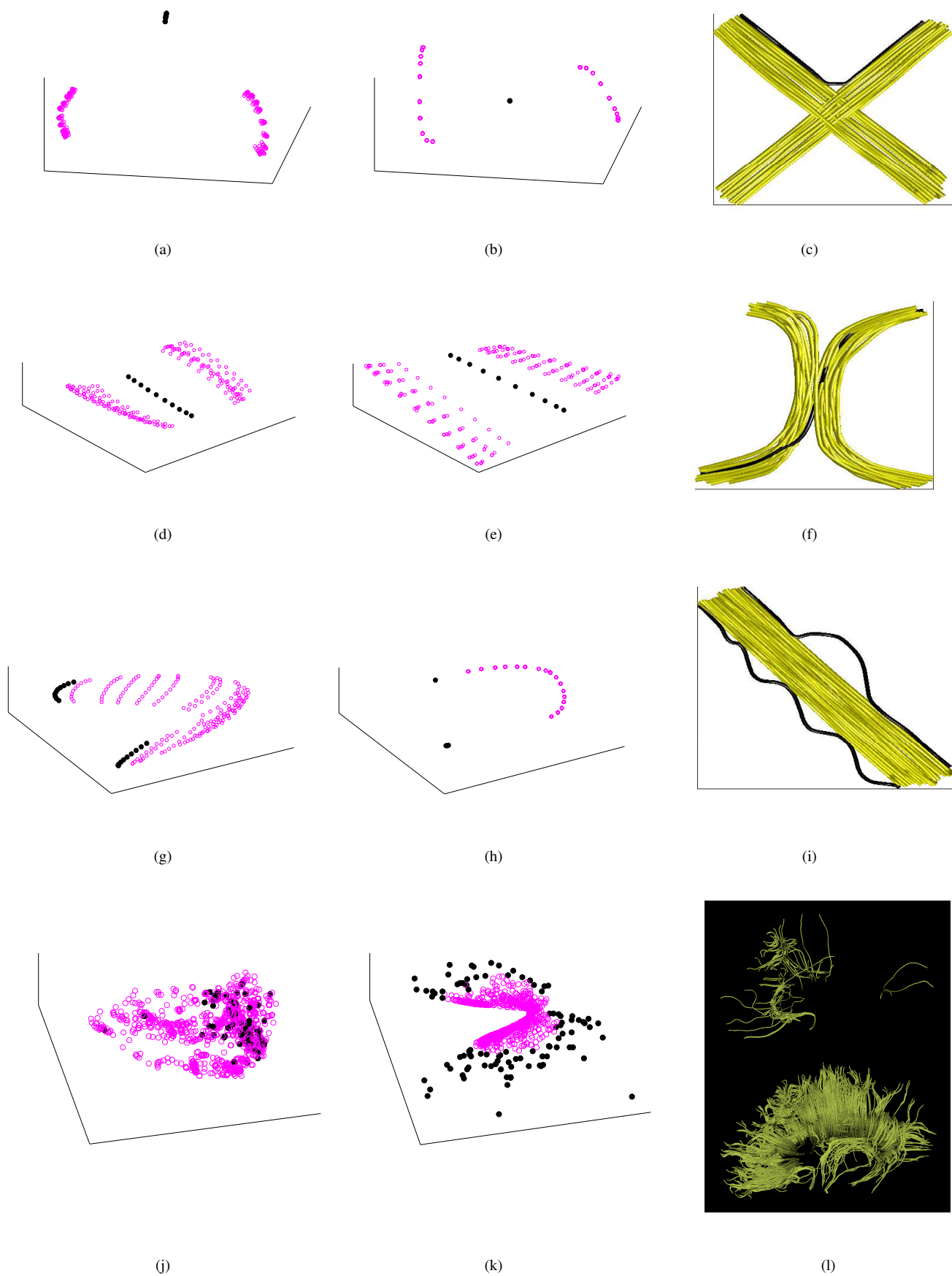
Fig. 2. For all figures, the viewing angle is optimized manually so as to see the structure as much as possible. First three rows correspond to phantom data experiments with crossing, kissing and linear tracts, each with outlier tracts drawn in solid. The fourth row is results of the real patient data experiments. The first column is the the scatter plots of tracts represented in 3D embedding space for 9D feature space based similarity measure, the second column is the same for the new Pivot-Point based similarity measure. The third column shows the fiber tracts where the outlier tracts are drawn in solid for phantom data. (l) shows the identified outlier tracts in real patient DT-MRI data (top) and the non-outlier tracts (bottom). The outlier tracts are marked with ● in all scatter plots, where as the others are marked with ○ (magenta).

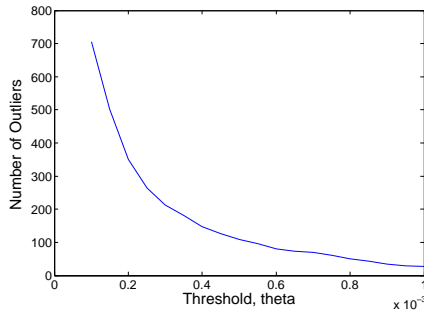Fig. 3. θ (x-axis) vs number of outliers (y-axis) for real patient DTI data.

III. EXPERIMENTS

We applied our method to phantom data composed of 3D tracts generated using MATLAB, that represents the crossing, kissing, and parallel fiber tracts. The tracts are manually drawn in 2D and replicated in 3D with different shifts along the z-axis. Zero mean random Gaussian noise with 0.1 standard deviation is also added along the z-axis. There are 200 tracts in each phantom data. We also used real patient DT-MRI data acquired with a 3 Tesla scanner, using SENSE MR sequence with 16 diffusion weighting gradients at $248 \times 150 \times 24$ mm FOV and $124 \times 124 \times 60$ voxel resolution. Figures 2.a, 2.d, 2.g show the phantom fiber tracts in the new 3D embedding space with the similarity measure based on 9D feature vectors [6]. Figures 2.b, 2.e, 2.h show the 3D representation of the phantom fiber tracts in the embedding space based on the new Pivot-Point based similarity measure. The phantom fiber tracts are shown in Figures 2.c, 2.f, 2.i. The phantom outlier tracts are marked with solid lines and their corresponding 3D representations in the embedding space are marked with •'s, demonstrating that the outliers are easily identifiable by both the previously proposed method [6] and the Pivot-Point method.

Figures 2.j and 2.k show the fiber tracts from the real patient DT-MRI data as represented in the embedding space corresponding to fiber tract similarity measured in the 9D feature space and by the Pivot-Point method, respectively. No clear structure can be observed in Figure 2.j, which makes outlier detection impossible, where as a manifold is clearly identifiable in Figure 2.k. The outlier tracts are marked based on this representation and as described in Section II-C. The same tracts are also marked in Figure 2.j. Both the outlier and non-outlier tracts are shown in Figure 2.l.

The unoptimized C++ implementation of the method takes approximately 20 minutes on a 2.66 GHz desktop computer with 3 GB RAM.

IV. DISCUSSION AND CONCLUSION

We proposed a new similarity metric which takes into consideration not only the mean distance between tracts and their orientation with respect to each other but also the high frequency variations along the tracts, through building a pointwise correspondence using the novel idea of pivot-points. The Laplacian Eigenmaps are used to map the tracts into a minimum dimensional embedding space. We have observed a clearly identifiable manifold structure in this embedding space in both phantom and real patient data experiments. We compared our similarity measure with a 9D

feature space based measure [6] and showed that although both measures perform equally well in phantom data, the new measure results in a manifold in real patient data as well, unlike the 9D feature space based one.

We have proposed to use the distance to the manifold as an outlier detector and implemented this idea by thresholding the distance between K nearest neighbours in the 3D embedding space. Thus, we could identify the outlier tracts in real patient data and furthermore the projection of outlier tracts onto the manifold can be used as an outlier correction scheme. A better projection scheme would require the manifold to be learned with higher precision.

Consequently, our preliminary results suggest that the pointwise correspondence based fiber tract similarity measure is capable of forming a continuum of fiber tracts (within a bundle) in an implicit embedding space which can be exploited through the use of Laplacian Eigenmaps. The resulting low dimensional manifold can be learned and potentially be used for both outlier fiber tract detection and correction. Further research is required to learn and represent the aforementioned curved manifold.

REFERENCES

[1] M. E. Moseley, J. Kucharczyk, H. S. Asgari et al., "Anisotropy In Diffusion-Weighted MRI," Magnetic Resonance in Medicine, vol. 19, no. 2, pp. 321-326, Jun, 1991.

[2] P. J. Basser, J. Mattiello, and D. Lebihan, "Estimation Of The Effective Self-Diffusion Tensor From The NMR Spin-Echo," Journal of Magnetic Resonance Series B, vol. 103, no. 3, pp. 247-254, Mar, 1994.

[3] P. J. Basser, S. Pajevic, C. Pierpaoli et al., "In vivo fiber tractography using DT-MRI data," Magnetic Resonance in Medicine, vol. 44, no. 4, pp. 625-632, Oct, 2000.

[4] C. Poupon, J. F. Mangin, C. A. Clark et al., "Towards inference of human brain connectivity from MR diffusion tensor data," Medical Image Analysis, vol. 5, no. 1, pp. 1-15, Mar, 2001.

[5] Belkin M., Niyogi P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in Neural Information Processing Systems 14, Vols 1 and 2, Advances in Neural Information Processing Systems T. G. Dietterich, S. Becker and Z. Ghahramani, eds., pp. 585-591, Cambridge: M I T Press, 2002.

[6] Brun A., Knutsson H., Park H. J.: Clustering fiber traces using normalized cuts. Medical Image Computing and Computer-Assisted Intervention - Miccai 2004, Pt 1, Proceedings, Lecture Notes in Computer Science C. Barillot, D. R. Haynor and P. Hellier, eds., pp. 368-375, Berlin: Springer-Verlag Berlin, 2004.

[7] Brun A., Park H. J., Knutsson H.:Coloring of DT-MRI fiber traces using laplacian eigenmaps. Computer Aided Systems Theory - Eurocast 2003, Lecture Notes in Computer Science R. MorenoDiaz and F. Pichler, eds., pp. 518-529, Berlin: Springer-Verlag Berlin, 2003.

[8] Ding Z. H., Gore J. C., Anderson A. W.: Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. Magnetic Resonance in Medicine, vol. 49, no. 4, pp. 716-721, Apr, 2003.

[9] O'Donnell L. J., Kubicki M., Shenton M. E.: A method for clustering white matter fiber tracts. American Journal of Neuroradiology, vol. 27, no. 5, pp. 1032-1036, May, 2006.

[10] O'Donnell L., Westin C. F.: White matter tract clustering and correspondence in populations. Medical Image Computing and Computer-Assisted Intervention - Miccai 2005, Pt 1, Lecture Notes in Computer Science J. S. Duncan and G. Gerig, eds., pp. 140-147, Berlin: Springer-Verlag Berlin, 2005.

[11] Zhang S., Laidlaw D. H.: DTI Fiber Clustering and Cross-subject Cluster Analysis. International Society for Magnetic Resonance in Medicine Computer, 2005.

[12] Zhang S., Laidlaw D. H.: Hierarchical Clustering of Streamtubes. Brown University, Providence, RI, United States, 2002.

[13] J. Shi, J. Malik, Normalized Cuts and Image Segmentation IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, August 2000.