A Method for Clustering White Matter Fiber Tracts

Lauren O'Donnell[1,2,5], Marek Kubicki[3,4], Martha E. Shenton[3,4], Mark H. Dreusicke[3], W. Eric L. Grimson[1], Carl-Fredrik Westin[1,5]

1. MIT Computer Science and AI Lab (CSAIL) , Cambridge, MA
2. Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA
3. Psychiatry Neuroimaging Laboratory, Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
4. Clinical Neuroscience Division, Laboratory of Neuroscience, VA Boston Healthcare System, Brockton, MA
5. Laboratory of Mathematics in Imaging, MRI Division, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

American Journal of Neuroradiology (AJNR)

Volume 27 Issue 5 (In Press)

ABSTRACT

Background/Purpose: Despite its potential for visualizing white matter fiber tracts in vivo, diffusion tensor tractography has found only limited applications in clinical research, where specific anatomical connections between distant regions need to be evaluated. Here, we introduce a robust method for fiber clustering which guides the separation of anatomically distinct fiber tracts and enables further estimation of anatomical connectivity between distant brain regions.

Methods: Line-scan-diffusion-tensor images (LSDTI) were acquired on a 1.5 Tesla magnet. Regions of interest (ROIs) for several anatomically distinct fiber tracts were manually drawn, then white matter tractography was performed using the Runge-Kutta method to interpolate paths (fiber traces) following the major directions of diffusion, where traces were seeded only within the defined ROIs. Next, a fully automatic procedure was applied to fiber traces, grouping them according to a pairwise similarity function which takes into account the shapes of the fibers and their spatial locations.

Results: We demonstrated the ability of the clustering algorithm to separate several fiber tracts which are otherwise difficult to define (left and right fornix, uncinate fasciculus and inferior occipito-frontal fasciculus, and corpus callosum fibers).

Conclusion: This method successfully delineates fiber tracts that can be further analyzed for clinical research purposes. Hypotheses regarding specific fiber connections and their abnormalities in various neuropsychiatric disorders can now be tested.

INTRODUCTION

Diffusion Tensor Imaging (DTI) is one of the first methods that makes it possible to visualize and to quantify the organization and integrity of white matter fiber tracts in the human brain in vivo. This method, since its introduction (1), has become increasingly popular in clinical research. Numerous studies have investigated either global diffusion changes within the white matter in various neuropsychiatric diseases, or particular fiber tracts hypothesized to be affected in conditions such as schizophrenia, Alzheimer's disease, multiple sclerosis, brain tumor, etc. (2, 3). DTI characterizes the behavior of local water diffusion for each image voxel. To quantify this phenomenon, several indices of anisotropy (the extent to which local water diffusion remains restricted by linearly organized axonal fibers) have been introduced. Among the most popular scalar indices measured from DTI are fractional anisotropy (FA) and mean diffusivity (D). These measures, believed to be related to axonal coherence and density in each voxel, are usually obtained on several slices, either within a small, manually defined region of interest, or within large, standardized ROIs. Neither method, however, captures properties of water diffusion for the whole single fiber tract.

Fiber tractography (4) is a promising new method for visualizing bundles of white matter fiber tracts in the brain. The method follows the principal diffusion direction (direction of maximal diffusion) in small steps, producing long fiber tracts that connect anatomically distant brain regions. These tractographic paths are commonly referred to as fibers, though the data resolution is much too low to measure any individual fibers or axons: instead the paths represent large-scale features of the diffusion data. The remaining problem in clinical research, however, is how to quantify the features defined by tractography, and in particular, how to extract tracts of interest. Since the introduction of fiber tractography, several methods have been proposed to

delineate anatomically distinct fiber tracts. So far, the most frequently used method utilizes multiple regions of interest (ROI). This is a guided method that performs fiber tractography starting from seed points within the first predefined ROI, and then preserves only traces that touch the other predefined ROI.

Instead of this guided method, fiber clustering approaches have been proposed, which are fully automatic, unguided, and take advantage of the similarity of the fiber paths. Fiber clustering methods analyze a collection of tractographic paths in 3D, and separate them into bundles, or clusters, that contain paths with similar shape and spatial position. These bundles are expected to contain fiber paths with similar anatomy and function. Several fiber clustering methods have been described in the literature. One method by Brun et al. (5) uses a 9D path shape descriptor, performs a pairwise path shape comparison, and then uses the Normalized Cuts spectral clustering method (6-8) to group the paths into bundles. This work is an extension of pioneering work by Brun et al. which performed soft grouping of fiber paths using spectral embeddings (9). Another method (10, 11) compares path shapes using variants on the Hausdorff distance (the largest distance separating pairs of points on the paths) (12). This latter method then groups paths using a hierarchical clustering algorithm, an iterative approach which groups the most similar paths at each step. Extension of clustering techniques to simultaneously cluster multiple subjects has been explored in (13). Our method groups tract paths based on shape and location using k-way Normalized Cuts spectral clustering. Here, we test the method on several brain fiber bundles, and demonstrate its utility, advantages and disadvantages.

METHODS

Acquisition

Line scan diffusion imaging (14) was used to acquire diffusion tensor data on 1.5 Tesla GE Echospeed system (General Electric Medical Systems, Milwaukee, WI), permitting maximum gradient amplitudes of 40 mT/m. Coronal oblique images, perpendicular to the AC-PC line were acquired with a quadrature head coil. For each line, six images with high (1000 s/mm^2) diffusion-weighting along six non-collinear directions, and two with low (5 s/mm^2) diffusion-weighting were collected. The following scan parameters were used: receiver bandwidth +/- 4kHz; TE (echo time) 64ms; effective TR (repetition time) 2592ms. We acquired a total of 32-36 slices (1.7 x 1.7 x 4 image resolution), covering the entire brain, depending upon brain size. As part of the same protocol, 5 midsagittal oblique LSDI scans (again 1.7 x 1.7 x 4 mm) were also acquired, parallel to the interhemispheric fissure. These images were co-registered to coronal slices, and used for corpus callosum ROI definition.

Subjects were normal controls, recruited from the general community and participating in other studies through the Laboratory of Neuroscience. Inclusion criteria for all subjects were: right-handedness, age between 18 and 55 years, no history of electroconvulsive shock treatment, no history of neurological illness, no alcohol or drug abuse in the last 5 years, no medication with known effects on MR such as steroids, verbal IQ above 75, and an ability and desire to cooperate with the procedures as evidenced by written informed consent. The study was approved by the local IRB committee.

ROI definition

For the purpose of this study, we manually delineated several ROIs. Since the resolution of our DTI acquisition is much higher in-plane (coronal), two of our three ROIs were defined on the coronal plane. The first ROI included the fornix (left and right together). The second ROI

was drawn on the temporal stem, where two fiber tracts, the uncinate fasciculus (UF) and inferior occipito-frontal fasciculus (IOFF), merge. The third ROI for the corpus callosum was drawn on sagittal DTI images acquired at the same time as the coronal scans. Since the sagittal reconstruction of coronal data was characterized by poor resolution, this allowed us to take advantage of the high in-plane resolution of the sagittal slices. All ROIs were drawn on the Fractional Anisotropy (FA) map, since the fiber bundles were easier to define than on the regular structural images. FA is a measure of tensor anisotropy that is high in voxels where a single fiber direction is present. Seeding for the fiber tracking was done for each voxel included in the ROI.

Data analysis

Our method of data analysis involved two steps. First, fiber tractography produced fiber paths, then fiber clustering delineated the paths into separate bundles. Both of these analyses, tractography and fiber clustering, were performed using 3D Slicer software (www.slicer.org).

Tractography

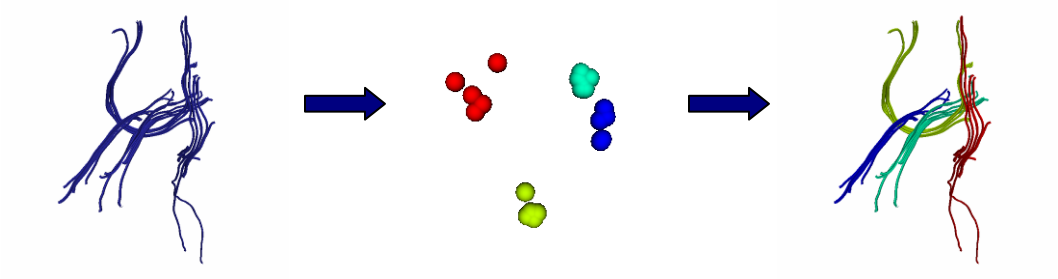
Diffusion tensor tractography was performed on coronal DTI images employing an eigenvector (principal diffusion direction) tracking algorithm based on the second order Runge-Kutta method. First, points were seeded within the predefined ROIs, then tractography was performed by repeatedly following in small steps along the interpolated direction of maximum diffusion. Tracking stopped where FA dropped below 0.1 or where fiber direction rapidly changed (due to noise that had caused a higher curvature than expected anatomically).

Fiber Clustering

Spectral clustering (6-8) is a method which groups data based on its similarity to other data. First, each item to be clustered is compared to every other item to calculate a similarity value. If there are N items, N-squared similarity values will be calculated, and these values are stored in a matrix. For the fiber clustering, we calculated fiber path similarity based on shape and position. To measure similarity between two paths, we used the average distance between pairs of nearest points on the paths (we then averaged the value computed from path A to path B, with the value from path B to path A, since the pairs of nearest points could be different). This is a modification of the Hausdorff distance (12) which was suggested in (10, 11). We also had the option of using the distance between pairs of endpoints (9), or the mean and covariance distance as defined in (5). Next we produced a similarity value from this distance by inverting the values (converting low distances to high similarities, and vice versa) using a Gaussian function (see for example (5)).

After calculating similarity values, the most important shape similarity information was automatically extracted and used to group the tractographic paths into bundles. The term "spectral" in spectral clustering refers to the use of the eigenvectors of the similarity matrix. An eigenvector is a vector, that when multiplied by the matrix, still points in the same direction (the eigenvector is only scaled by the matrix, not rotated). The scale factor is called an eigenvalue. In general, a square matrix such as our N by N similarity matrix can be approximated using its top eigenvectors and eigenvalues. The higher the eigenvalue, the more information the corresponding eigenvector contains about the matrix. In our clustering application we used the top eigenvectors of the fiber similarity matrix to calculate the most important shape similarity information for each fiber path while removing noise. For each fiber path, this information could be visualized as a point, to show the separation of clusters according to similarity, as shown in

Figure 1. The specific type of spectral clustering that we employed, k-way Normalized Cuts (6-8), has been shown to produce clusters that have high within-cluster similarity and low between-cluster similarity.

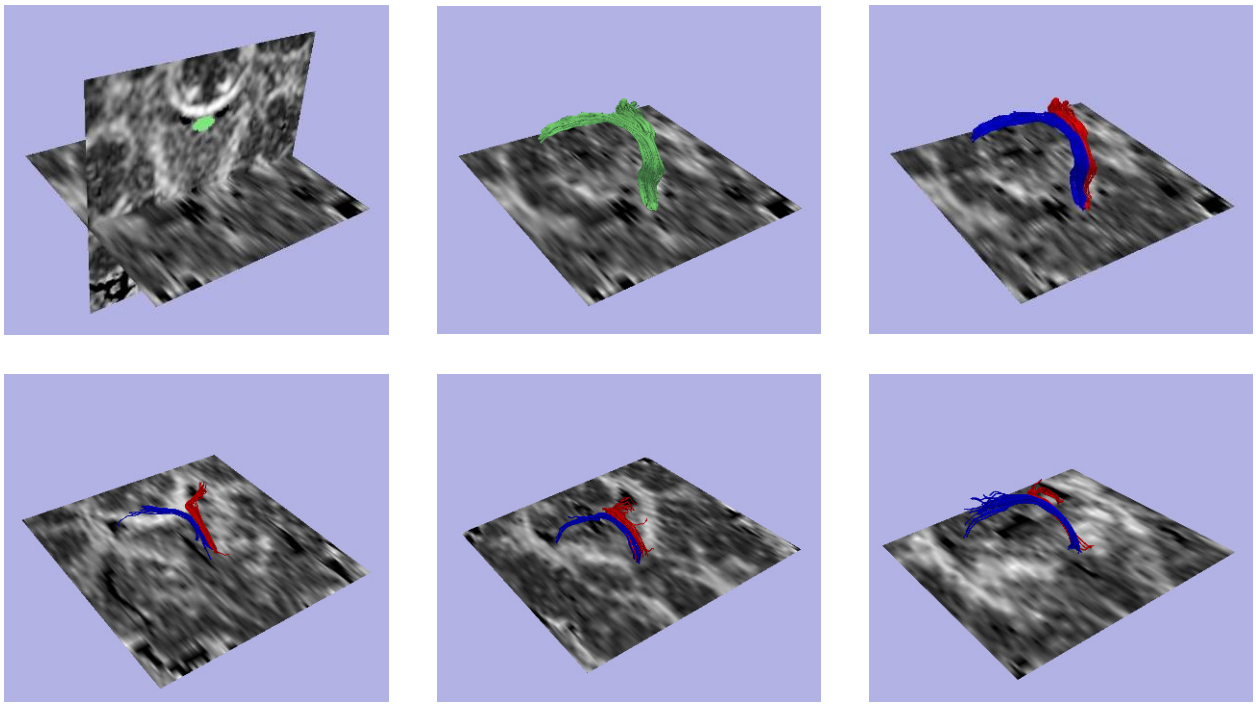


Shape Comparison Spectral Embedding Clusters

RESULTS

Fornix

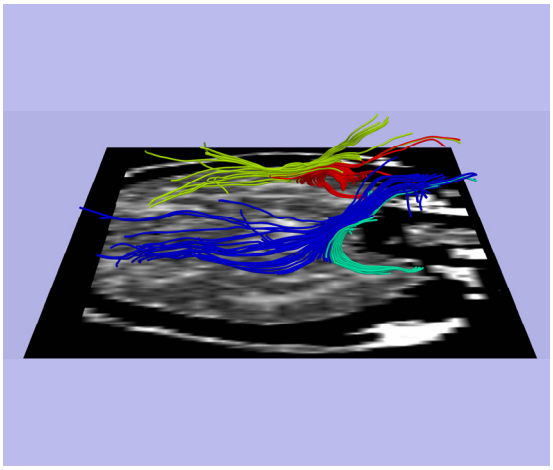
These fibers originate from the crux of the hippocampi, and before they merge, they are relatively thick and short, being theoretically easy to define. Their trajectory, however, is almost parallel to our coronal plane, which makes it difficult to avoid partial volume effects. Starting from the body of the fornix, the two structures are located so close together, that it is extremely difficult to manually divide them into left and right. Moreover, because of the aforementioned partial volume effect due to the curvature of the fornix, an ROI can be drawn reliably on only one coronal slice perpendicular to the body of the fornix. Thus the guided multiple ROI method would have been unreliable in defining left and right fornices. Using the clustering method, we were able to reliably separate left and right fornices (see Figure 2c left (red) and right (blue) fornix.

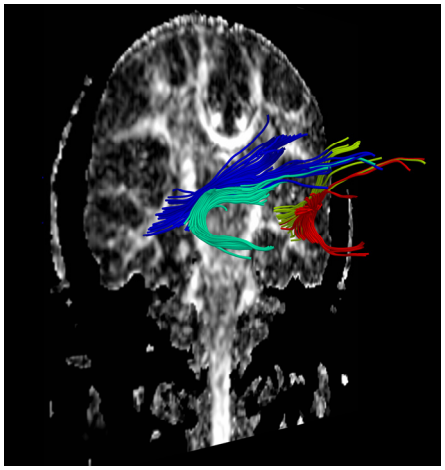


Temporal stem

The temporal stem forms the bridge between frontal and temporal lobes, containing fiber tracts connecting the frontal lobes with more posterior brain regions such as the temporal, parietal and occipital lobes. Thus in a small cross-sectional area the temporal stem includes fibers of different functions. Previous ROI methods have attempted to measure anisotropy within the temporal stem, but were not able to distinguish separate fiber tracts included in this region (2). The clustering method clearly differentiates the uncinate fasciculus (fiber tract connecting orbito-frontal and anterior temporal regions) from the inferior occipito-frontal fasciculus (tract connecting frontal, posterior temporal and occipital lobes), two fiber tracts traveling in close vicinity through the temporal stem. The advantage of clustering over the multiple ROI fiber tracking method for this region lies in the fact that the UF includes many short fibers that arc around the lateral fissure. Consequently, the multiple ROI method would cut off most of them (and would be dependent on the placement of the second ROI), whereas

clustering was able to preserve all the fiber paths of similar function.



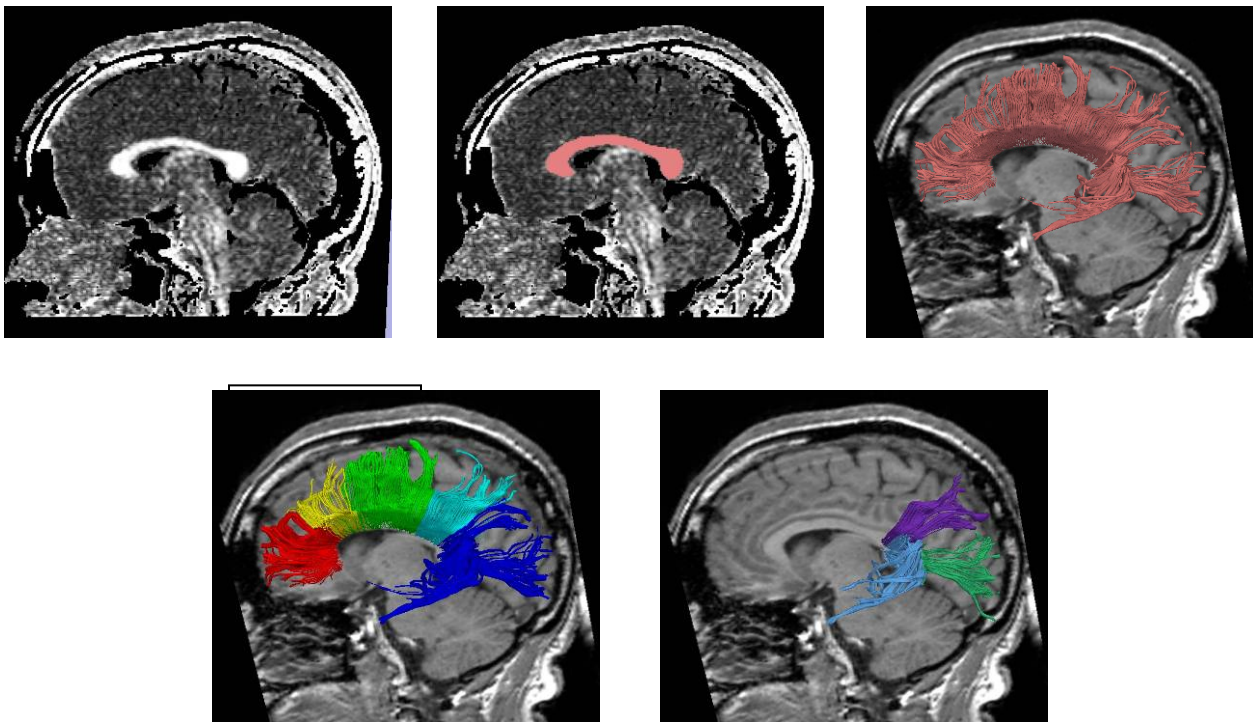


Corpus Callosum

The corpus callosum (CC) is the largest fiber tract of the human brain. Its divisions, however, are not well defined in terms of anatomical connectivity. We used our method to investigate anatomical divisions of CC. Fiber tracking was initiated from seed points placed within the midsagittal cross section of the entire CC.

When performing clustering of the corpus callosum, the desired results are a division of the corpus callosum into anatomical regions according to the fiber projections, for example orbitofrontal, dorsolateral frontal, parietal, occipital, and temporal regions. However, due to the overall similarity of fiber path shapes (see Figure 4c), clustering of the whole corpus callosum at once does not give us satisfactory results. This is due to the fact that the clustering is based only on shape and position, and thus cannot capture divisions between these anatomical regions unless the tract shapes vary. In the genu and the body of the corpus callosum, the tract shape is so similar from one path to its neighbors that the algorithm cannot differentiate based on shape, and is forced to divide the corpus into approximately even-sized pieces (see Figure 4d). Consequently we concentrated on performing clustering in the splenium of the corpus callosum,

where a shape difference can be noted between fibers projecting to the temporal and occipital lobes (see Figure 4e)



DISCUSSION

We demonstrate that the method of fiber clustering can differentiate fiber bundles that would be otherwise hard to define and separate from the neighboring structures. The extraction of fiber bundles is becoming one of the most important tasks in clinical neuroscience research, where specific hypotheses about connectivity between distant brain areas are being tested. The fiber bundles are more powerful for quantitative analysis than ROIs defined on individual slices, as they allow definition of three-dimensional regions of interest, and they permit quantification of tract shape properties as well as measurement of scalar indices from the tensor field. Imaging studies in neuropsychiatric disorders, such as schizophrenia, bipolar disorder, schizotypal

personality disorder etc, could benefit from the increased sensitivity and specificity of whole fiber tract diffusion anisotropy estimation, since the 2D ROI based studies carried out so far show inconsistent results (2, 15). Automatic definition of fiber bundles may also aid in surgical planning by providing improved 3D visualization of fiber tracts that may closely border a tumor.

In comparison with the commonly used multiple ROI tract selection method, our method is more automatic and depends only on the shape and location of the fiber tracts, rather than on user placement of selection ROIs. Although an advantage of the multiple ROI method is that it displays only the fiber tract of interest to the user, its major limitation is that because of low DTI resolution, it is usually difficult to define more than one ROI for a single bundle. Also, since definition of the ROIs of fiber tracts displaced or interrupted by a lesion can be much more difficult, clustering based on shape similarities can be used in these cases. In addition, the multiple ROI method excludes shorter fibers that, though they run in the same direction, do not touch two ROIs.

The methods described in this paper avoid these problems; however, there are several limitations worth noting. Even though no a priori anatomical knowledge about the entire trajectory of the fiber tract is required, we still have to define, a priori, the number of clusters that we expect to find. It would be of interest to determine this number automatically, however the automatic determination of the number of clusters inherent in data is still a research problem in the field of spectral clustering. Various methods have been proposed, including an iterative method, which measures cluster validity (16) and a method that rotates the affinity matrix (17). We hasten to point out, however, that the approach we have taken is straightforward and simple, as it allows the user to interactively control the number of clusters expected/required, which is based on anatomical a priori knowledge and/or upon a priori hypotheses. As no methods

currently make possible the automatic identification of white matter tracts in the brain, deferring to expert knowledge to determine the number of clusters is the best standard we have at this time.

Second, the example of the corpus callosum demonstrated that the method is not able to subdivide structures with very consistent shapes, though it is effective in separating shapes that differ. This suggests that more anatomical information is necessary to subdivide the anterior corpus callosum. This information could be provided via shape relations to other neighboring tracts or through additional user input about the anatomy.

Finally, the quality of the clustering results depends on the quality of the input paths, which in turn is determined by the data quality and the tractography algorithm. Improvements in data quality such as a higher number of diffusion directions could produce an improvement in the clustering output. The performance of the fiber tractography depends on many factors, including data resolution, noise, image distortions, and partial volume effects caused by multiple tracts in a voxel. At today's image resolutions, this method does not detect water behavior within the individual axons, but instead describes the estimated local diffusion properties in the tensor field, and thus tractography should be regarded as a visualization of features in this field. It is, nonetheless, remarkable, that several studies describe the similarity of fiber tracts obtained with DTI tractography to anatomically defined white matter fiber bundles (18, 19).

CONCLUSION

We have tested the method on several brain fiber bundles, including the fornix, the temporal stem, and the corpus callosum, and we have demonstrated its utilities, advantages and shortcomings. Our method can find applications in surgical planning, clinical neuropsychiatry, and anatomy. Higher spatial resolution of diffusion images, as well as combination of clustering

and multiple ROI methods, should further increase the accuracy and specificity of the proposed method.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the HST Neuroimaging Training Grant (LO), NIH NINDS 1-R01-NS051826-01 (LO), the National Alliance for Research on Schizophrenia and Depression (MK), the Wodecroft Foundation (MK), the National Institute of Health (R03 MH 068464-02 to MK, K02 MH 01110 and R01 MH 50747 to MES), the Department of Veteran Affairs REAP Award (MES), and NIHP41-RR13218 (CFW). This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.

REFERENCES

1. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J* 1994;66:259-267

2. Kubicki M, Westin CF, Maier SE, et al. Diffusion tensor imaging and its application to neuropsychiatric disorders. *Harv Rev Psychiatry* 2002;10:324-336

3. Lim KO, Helpern JA. Neuropsychiatric applications of DTI - a review. *NMR Biomed* 2002;15:587-593

4. Basser PJ, Pajevic S. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* 2000;44:625-632

5. Brun A, Knutsson H, Park HJ, Shenton ME, Westin CF. Clustering fiber traces using normalized cuts. *MICCAI*; 2004;368-375

6. Fowlkes C, Belongie S, Chung F, Malik J. Spectral grouping using the Nystrom method. *PAMI* 2004;26:214-225

7. Ng A, Jordan M, Weiss Y. *On spectral clustering: Analysis and an algorithm*. 2001

8. Shi J, Malik J. Normalized cuts and image segmentation. *PAMI* 2000;22:888-905

9. Brun A, Park HJ, Knutsson H, Westin CF. Coloring of DTMRI fiber traces using Laplacian eigenmaps. *EUROCAST*; 2003;564-572

10. Corouge I, Gouttard S, Gerig G. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. *ISBI*; 2004;344-347

11. Gerig G, Gouttard S, Corouge I. Analysis of brain white matter via fiber tract modeling. 2004;426

12. Huttenlocher D, Klanderman D, Rucklige A. Comparing images using the Hausdorff

distance. *IEEE Trans Pattern Anal Mach Intell* 1993;15:850-863

13. O'Donnell L, Westin C-F. White Matter Tract Clustering and Correspondence in Populations. *MICCAI*; 2005;140-147

14. Gudbjartsson H, Maier SE, Mulkern RV, Morocz IA, Patz S, Jolesz FA. Line scan diffusion imaging. *Magn Reson Med* 1996;36:509-519

15. Kubicki M, Park H, Westin CF, et al. DTI and MTR abnormalities in schizophrenia: Analysis of white matter integrity. *Neuroimage* 2005;26:1109-1118

16. Porikli F. Ambiguity Detection by Fusion and Conformity: A Spectral Clustering Approach. *IEEE International Conference Integration of Knowledge Intensive Multi-Agent Systems (KIMAS)*; 2005;366-372

17. Zelnik-Manor L, Perona P. Self-Tuning Spectral Clustering. *NIPS*; 2005;1601-1608

18. Conturo TE, Lori NF, Cull TS, et al. Tracking neuronal fiber pathways in the living human brain. *Proc Natl Acad Sci U S A* 1999;96:10422-10427

19. Stieltjes B, Kaufmann WE, van Zijl PC, et al. Diffusion tensor imaging and axonal tracking in the human brainstem. *Neuroimage* 2001;14:723-735

FIGURE LEGENDS

1. The most important shape information is automatically extracted and used for clustering. The leftmost image shows the input fiber tracts. The middle image shows the clustering step. Each point in this image represents the similarity relationships of one fiber (these points come from the highest eigenvectors of the similarity matrix, in a process called spectral embedding). Finally, in the rightmost image, the tracts are colored by cluster.

2. Fiber tract clustering in the fornix. A fiber path was seeded in each voxel of the single ROI which can be seen in image (a). These fibers are shown in image (b). Next the clustering method was applied to separate the fiber paths into two clusters, the left (red) and right (blue) fornices, as shown in image (c). Images (d) through (f) show similar clustering results for additional subjects.

3. Fiber paths were seeded in each voxel of initial seed ROIs located in the temporal stem. The clustering was then performed to create two groups of paths, corresponding to the IOFF and the uncinate fasciculus.

4. Fiber tract clustering in the corpus callosum. Using a fractional anisotropy map (a), a ROI is drawn on the corpus callosum (b). Since fiber paths have similar shapes (c), clustering produces an inadequate result (d). If the user isolates a region of the corpus callosum containing fiber paths of different shapes, clustering is successful (e).