Identifying white-matter fiber bundles in DTI data using an automated proximity-based fiber-clustering method

Song Zhang[1], Stephen Correia[2], and David H. Laidlaw[3]
[1]Department of Computer Science and Engineering, Mississippi State University, Mississippi State, MS, United States
szhang@cse.msstate.edu
[2]Department of Psychiatry and Human Behavior, Brown Medical School, Providence, RI, United States
stephen_correia@brown.edu
[3]Department of Computer Science, Brown University, Providence, RI, United States
dhl@cs.brown.edu

***Abstract*— We present a method for clustering diffusion tensor imaging (DTI) integral curves into anatomically plausible bundles. An expert rater evaluated the anatomical accuracy of the bundles. We also evaluated the method by applying an experimental cross-subject labeling method to the clustering results. Our approach is guided by assumptions about the proximity of fibers comprising discrete white-matter bundles, and proceeds as follows. We first employ a sampling and culling strategy for generating DTI integral curves and then constrain the curves so that they terminate in gray matter. This approach seems likely to retain anatomically plausible fibers. We then employ a clustering method based on a proximity measure calculated between every pair of curves. We interactively selected a proximity threshold to achieve visually optimal clustering in models from four DTI datasets. An expert rater then assigned a confidence rating about bundle presence and accuracy for each of 12 target fiber bundles of varying calibers and type (i.e., commissural, association, projection) in each dataset. The interactive clustering and evaluation information was incorporated to create a fiber-bundle template. We then used the template to cluster and label the fiber bundles automatically in new datasets. According to expert evaluation, the automated proximity-based clustering and labeling algorithm consistently yields anatomically plausible fiber bundles, although fiber bundles with smaller calibers and those that are not highly directionally coherent are identified with lower confidence. This work has the potential to provide an automatic and robust way to find and study neural fiber bundles within DTI.**

***Index Terms*— cluster, clustering, diffusion tensor imaging, DT-MRI, DTI**

I. Introduction

Diffusion-tensor imaging (DTI) is a magnetic resonance imaging (MRI) technique that provides three-dimensional *in vivo* information about the structural integrity of cerebral white matter based on the diffusion of water molecules. DTI tractography is a method for visualizing white-matter fiber structure [3]. Tractography essentially involves producing, within the image volume, a set of integral curves that follow the principal direction of diffusion. The resulting tractography models produce a dense set of "fibers" that bear a close resemblance to known white-matter pathways [17]. These models provide a potentially valuable tool for understanding white-matter anatomy and for testing hypotheses about the cognitive and behavioral correlates of specific white-matter tracts in both healthy and diseased brains. Such tests require the ability to identify and segment white-matter tracts-of-interest (TOI). Once segmented, the structural integrity of a TOI can be quantified by a variety of scalar metrics (e.g., average fractional anisotropy, length) [10]. However, accurately identifying and isolating specific bundles in the visually dense tractography models is a challenge. Manual segmentation of the fiber bundles is often time consuming, typically requires an expert rater with detailed knowledge of white-matter anatomy, and is prone to human error and experimenter bias. A method that automatically clusters and labels the integral curves into anatomically plausible pathways could vastly improve the efficiency of data processing and facilitate hypothesis testing.

Our approach builds on the assumption that a proximity measure that compares fiber trajectories can represent anatomical relationships and that there exists a threshold that segregates a set of trajectories into discrete clusters that are anatomically significant. This assumption has been demonstrated for some fiber structures in Moberts *et al.* [13] and Corouge *et al.* [6]. This paper presents a method that will, in such cases, automatically cluster integral curves into anatomically plausible bundles. Our method extends previous work in this area in that it enables an expert to interactively specify a proximity threshold between tracts-of-interest and ensures that this proximity threshold is satisfied between the resulting clusters. Another contribution is our evaluation of the clustering method by rating the presence of a pre-specified set of bundles on the resulting fiber clusters.

In section II, we review prior work in this area. Section III-A outlines the acquisition and preparation of DTI data. In section III-B, we describe the proximity measures we used for culling integral curves and for clustering. In section III-C, we describe our sampling method and its constraints and the culling methods used for generating a set of DTI integral curves. In section III-D, we describe our clustering algorithm and the proximity threshold used to group these integral curves into anatomically related bundles. In section III-E, we demonstrate an interactive interface by which an expert achieves visually optimal clustering results in four datasets. The rater then assigns confidence ratings for each of 12 target white-matter fiber bundles of varying calibers and type (i.e., commissural, association, projection) in each dataset (section III-F.1). We describe the construction of a fiber bundle template with the expert's clustering and rating results and use this template for fully automated clustering and labeling of a subset of

the 12 selected fiber bundles on two new subjects in section III-F.2. We show the results in section IV.

II. Related Work

A number of methods have been developed for classifying, clustering, and labeling fiber bundles in DTI.

Fiber bundles can be generated by first manually selecting one or more regions-of-interest (ROIs) and then grouping the integral curves that pass through these ROIs (Catani *et al.* [5], Wakana *et al.* [17], and Maddah *et al.* [12]). Maddah *et al.* [12] used hand-selected ROIs in white matter to construct a bundle template to which curves from a new subject can be registered. Our method is similar to Maddah *et al.*'s in that we use an expert rater to identify the clusters and use the result to build a template; however, instead of manually specifying ROIs, our method automatically clusters and labels fiber bundles, which may improve the efficiency of the process.

Recently, automatic DTI fiber clustering methods have been developed on DTI fibers by Ding *et al.* [7], Corouge *et al.* [6], Brun *et al.* [4], O'Donnell *et al.* [14], [15], Maddah *et al.* [11], and Moberts *et al.* [13]. Moberts *et al.* [13] evaluated different fiber clustering methods by manually constructing a set of fiber bundles as "ground truth" and using that to test the results of different similarity measures and clustering algorithms. They concluded that the single-link algorithm used in our earlier technical report [21] gives the best results in the clustering methods they tested. We build on and expand the methods in [21].

O'Donnell *et al.* [14] applied a spectral clustering method on DTI fibers from a population of brains to generate a white-matter atlas automatically. In another work [15], they had the experts label clusters in the atlas and used the labeled atlas to cluster new subjects. Our method is similar to O'Donnell *et al.* in that we also build a template of fiber bundles from automatic clustering results and use the template for clustering new subjects. However, instead of using the expert to label the clustering results, our method incorporates the expert's anatomical knowledge into the clustering by implementing a minimal proximity threshold set by the expert (see also section V-A). We also ask our expert to evaluate the clustering results after incorporating the expert's minimum proximity threshold in the method. Our method also incorporates anatomical assumptions to constrain both the curve generation and clustering algorithms (section III-C). We have presented some of the preliminary results as conference abstracts [19], [22].

The measure used to define the spatial proximity between integral curves is of fundamental importance for clustering. Proximity measures represent the similarity between two data elements and therefore can be used to determine the geometric similarity of the two curves. A number of proximity measures have been used to describe the geometric similarity between two curves. In general, proximity measures can be categorized into two varieties: those that measure the Euclidean distance between two selected points on two curves and those that summarize all points along two curves as the mean Euclidean distance along their arc lengths. Examples of the former type of proximity measure include the closest point measure, the Hausdorff distance [6], and the Fréchet distance [2]. Examples of the latter type include the average point-by-point distance between corresponding segments defined by Ding *et al.* [7], the mean of closest distances defined by Corouge *et al.* [6], and the mean of thresholded closest distances defined by Zhang *et al.* [20]. Brun *et al.* [4] embedded the curves in a feature space and then calculated the distance between the two curves in the feature space. Our clustering method uses a variation of the proximity measure defined in Zhang *et al.* [20], which captures any difference between two curves while considering the spatial proximity of the two curves along their entire arc length. Moberts *et al.* [13] implemented several previously published distance measures in their evaluation of fiber clustering methods and concluded that the mean of closest distances performs better than closest point distance, Hausdorff distance, and end points distance.

III. Method

A. Data acquisition and preparation

Siemens MDDW protocol was used to collect three co-registered sagittal double-spin-echo, echo-planar diffusion-weighted volumes of the entire brain in 5 healthy adults (2 males, 3 females; mean ($\pm$ standard deviation) age = 56.8 ± 10.23). All subjects provided written informed consent to participate in a DTI research project approved by the Institutional Review Board at Butler Hospital in Providence, RI.

The volumes were spatially offset in the slice direction by $0.0mm$, $1.7mm$, and $3.4mm$. Parameters for each acquisition were as follows: $5mm$ thick slices, $0.1mm$ inter-slice spacing, 30 slices per acquisition, $matrix = 128 \times 128$, $FOV = 21.7cm \times 21.7cm$, $TR = 7200$, $TE = 156$, no partial echoes, $NEX = 3$. Diffusion encoding gradients ($b = 0, 1000mm/s^2$) were applied in 12 non-collinear directions. Total time for the three acquisitions was slightly less than 15 minutes. A vacuum pillow was used to limit subject motion. The three acquisitions were interleaved to achieve $1.7mm^3$ resolution images and then up-sampled (equivalent to zero-filling) to $0.85mm^3$ isotropic voxels for analysis.

Diffusion tensors were calculated with a non-linear sequential quadratic programming (SQP) method [1].

B. Proximity measure

The integral curves in the cerebral white matter can be quite long (in this paper curves are defined as a set of 3D points). An ideal proximity measure should consider the matching on the whole arc lengths of the two curves. Measuring proximity by the Euclidean distance between two selected points on two curves ignores all but these two points on those two curves. Therefore, it is not surprising that the mean of closest distances defined by Corouge *et al.* [6] is found to give better clustering results than the closest points measure or the Hausdorff distance [13]. Corouge *et al.* defined the mean of closest distances between curves Q and R as:

$$d_{MC}(Q,R) = \frac{d_m(Q,R) + d_m(R,Q)}{2},$$

where $d_m(Q,R) = \underset{a \in Q}{\text{mean}} \min_{b \in R} \parallel a-b \parallel$, a and b are vertices on Q and R respectively, and $\parallel . \parallel$ is the Euclidean norm. Note that d_m is not symmetric while d_{MC} is. Symmetry is a desirable property for a proximity measure since it removes the inconsistency of a proximity value between two curves when the order of the two curves is switched, and subsequently results in symmetric proximity matrices. We can also define two other symmetric proximity measures based on d_m: the shorter mean of closest distances $d_{SC}(Q,R) = \min(d_m(Q,R), d_m(R,Q))$ and the longer mean of closest distances $d_{LC}(Q,R) = \max(d_m(Q,R), d_m(R,Q))$.

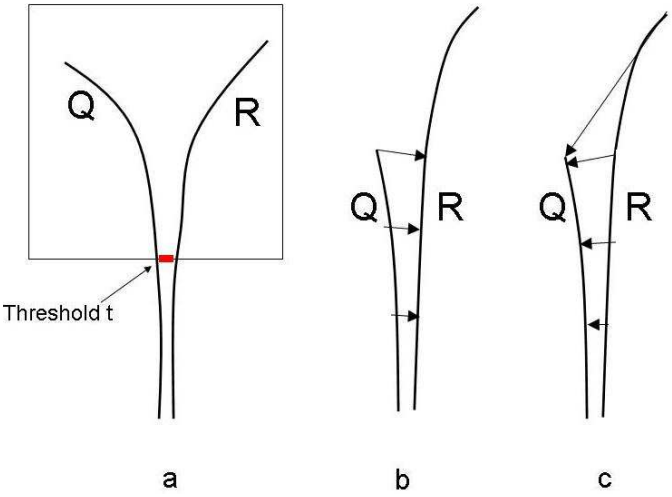


Fig. 1. (a) shows the motivation behind the mean of thresholded closest distances. Q and R are considered different if they branch for a portion of their lengths. Without the threshold, the mean of closest distances between Q and R is low if they stay close for a large part of their lengths. (b) and (c) show the difference between $d_{St}(Q,R,t)$ and $d_{Lt}(Q,R,t)$. The arrows are samples of distances to the other curve from the selected curve (Q in (b) and R in (c)). Note the distance from the top of R to Q in (c) is missing in (b). Also note the different arrow orientations at similar locations between (b) and (c).

When two curves form a boundary between two branching fiber bundles, they might run very closely together for a long course, and then diverge abruptly for a relatively short course. Although two such fibers should be grouped into different clusters, the mean closest distances might be low due to their long overlapping course and they may end up being considered to be part of the same cluster (see Fig. 1(a)). To mitigate this problem, we modify d_m by using a threshold on the minimum contributing distance for the mean of closest distances between curves Q and R:

$$d_t(Q,R,t) = \text{mean}_{a\in Q,(\min_{b\in R}\|a-b\|)>t} \min_{b\in R} \| a-b \|,$$

where t is the minimum threshold so that distances below it are not counted toward the mean. a and b are the vertices on Q and R respectively.

We then define two symmetric proximity measures by replacing d_m with d_t in d_{SC} and d_{LC}: the shorter mean of thresholded closest distances

$$d_{St}(Q,R,t) = \min(d_t(Q,R,t), d_t(R,Q,t)),$$

and the longer mean of thresholded closest distances

$$d_{Lt}(Q,R,t) = \max(d_t(Q,R,t), d_t(R,Q,t)),$$

d_{St} is a discrete approximation of D_t defined in [20]:

$$D_t = \frac{\int_{s_0}^{s_1} \max(\text{dist}(s) - T_t, 0)\text{d}s}{\int_{s_0}^{s_1} \max(\frac{\text{dist}(s)-T_t}{|\text{dist}(s)-T_t|}, 0)\text{d}s},$$

where s parameterizes the arc length of the shorter curve, s_0 and s_1 are the starting and end points of s, and dist(s) is the shortest distance from location s on the shorter curve to the longer curve.

d_{St} and d_{Lt} are the proximity measures implemented in this paper. Fig. 1 (b) and (c) show the difference between d_{St} and d_{Lt}. If a shorter curve Q runs along a longer curve R, $d_{St}(Q,R,t)$ is usually smaller than $d_{Lt}(Q,R,t)$ since the unmatched part of R only counts in d_{Lt}. We used d_{St} in our integral curve culling scheme in section III-C because partial-volume effect or noise might result in broken short curves along a long curve and we wanted to cull out those curves. On the other hand, once we culled those short curves along their longer neighbors, we used d_{Lt} to make sure any difference between two curves, including the unmatched part, was captured when we applied the single linkage clustering algorithm in section III-D.

We can also define the mean of closest thresholded distances by replacing d_m with d_t in d_{MC}. This measure is not used in this paper since it has no apparent advantage over d_{St} and d_{Lt} in our approach to culling curves or capturing differences between remaining curves.

C. Generating DTI integral curves

The integral curves that make up our tractography models were generated by solving the following equation:

$$p(r) = \int_0^r \vec{v}(p(s))ds,$$

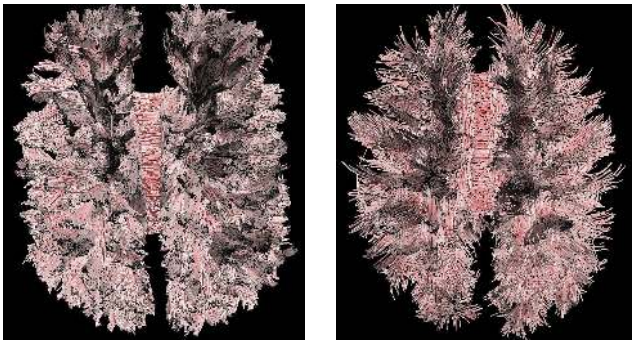
where r is the arc length of the generated streamline, $p(r)$ is the generated streamline and $\vec{v}$ corresponds to the vector field generated from the major eigenvector $\vec{e_1}$ of the diffusion tensor. $p(0)$ is set to the initial point of the integral curve, often called seed point. The integral curve was extended to both $\vec{e_1}$ and $-\vec{e_1}$ directions from the seed point using a second-order Runge-Kutta integration method [16], [3] with a stepsize of 1mm. The integration stopped when the curve went out of the data boundary, or went into a region of low linear anisotropy or low signal-to-noise ratio. We used the linear anisotropy defined by Westin *et al.* [18]:

$$\frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}},$$

where λ_1, λ_2, and λ_3 are the three eigenvalues of the diffusion tensor. A T2-weighted image that was coregistered with DTI data was used to identify low signal-to-noise regions such as air. These regions result in lower intensities in T2-weighted images compared to water, gray matter, and white matter. A threshold on the T2-weighted image was set to be lower than the intensities of water, white matter, and gray matter, but higher than the intensity of air. At any point within the data volume, tricubic B-spline functions were used to interpolate the tensor field [3].

We first generated seed points on a fine regular grid in the data volume and then jittered their locations. This produced a dense set of seed points that ensured there is no place in the image volume that was under-sampled. To reduce artificial clusters among integral curves and maintain some space between these curves, we culled the shorter member of any two pairs of curves that exceeded a pre-specified threshold for proximity d_{St}.

Curves with errant trajectories that seem anatomically implausible might also be generated due to noise, low image resolution, or partial volume effects. To limit the number of spurious curves and maximize the anatomical correctness of our integral curves, we also culled these curves based on a minimum linear anisotropy of the curve and a minimum length of the integral curve since curves running in low linear anisotropy regions and very short curves are often affected by partial-volume effects and noise [20].



a b

Fig. 2. Integral curve models generated from a $128 \times 128 \times 90$ dataset with a voxel size of $1.7mm \times 1.7mm \times 1.7mm$. (a) shows 438,042 integral curves generated from a jittered regular grid of $256 \times 256 \times 180$, with minimum length set to $13mm$ and minimum linear anisotropy set to 0.15. (b) shows the model generated from the same dataset and the same jittered regular grid sampling, and with the added culling and gray matter projection constraints. The culling threshold on d_{St} is set to $1.0mm$ with $t = 0.5mm$. The model in (b) has 6,113 integral curves. Color is mapped to the linear anisotropy value. Redder means higher linear anisotropy.

We also set a constraint that an integral curve should project into the gray matter. This constraint removes incoherent fibers that break in the white matter due to the partial-volume effect. This is desirable since we focus our effort in this paper on clustering coherent fiber structures. We accomplished this constraint by segmenting the brain into white matter, gray matter, and cerebrospinal fluid compartments using the FAST [23] segmentation tool. The segmentation was performed using the non-diffusion-weighted image as well as scalar maps of the trace of the diffusion tensor and fractional anisotropy. We then discarded the integral curves that would not have extended into the gray matter within $3mm$. We extended a curve by following the direction of each curve at its two end points. We estimated the $3mm$ threshold based on the observation of the gaps between the ends of some curves that appear to project into the gray matter and the segmented gray matter. The estimation was also dependent on the minimum linear anisotropy threshold and the parameters on the FAST segmentation algorithm that affect the shapes of the segmented areas. We colored and rendered the resultant curves and superimposed them on background anatomical images to provide context.

Fig. 2 shows the difference between two integral curve models without and with the culling and gray matter projection constraints. Both sets of the integral curves were generated on a $128 \times 128 \times 90$ dataset with a voxel size of $1.7mm \times 1.7mm \times 1.7mm$. Both sets of the integral curves were generated from a jittered regular grid of $256 \times 256 \times 180$, with minimum length set to $13mm$ and minimum linear anisotropy set to 0.15. Without the culling and gray matter projection constraints, Fig. 2(a) shows 438,042 integral curves. With the constraints, Fig. 2(b) shows the generated model of 6,113 integral curves. The culling threshold on d_{St} was set to $1.0mm$. The minimum threshold t was set to $0.5mm$, much smaller than the resolution along each of the data dimensions, to remove the distance contribution from the parts of the two curves that ran very closely and were likely from the same feature in the data. Some of the thresholds for generating the curves such as minimum length, minimum linear anisotropy and minimum threshold t were set empirically. We tried to cull the spurious fibers from noise and other artifacts and retain anatomically meaningful fibers by visually examining the effect of different settings on these thresholds.

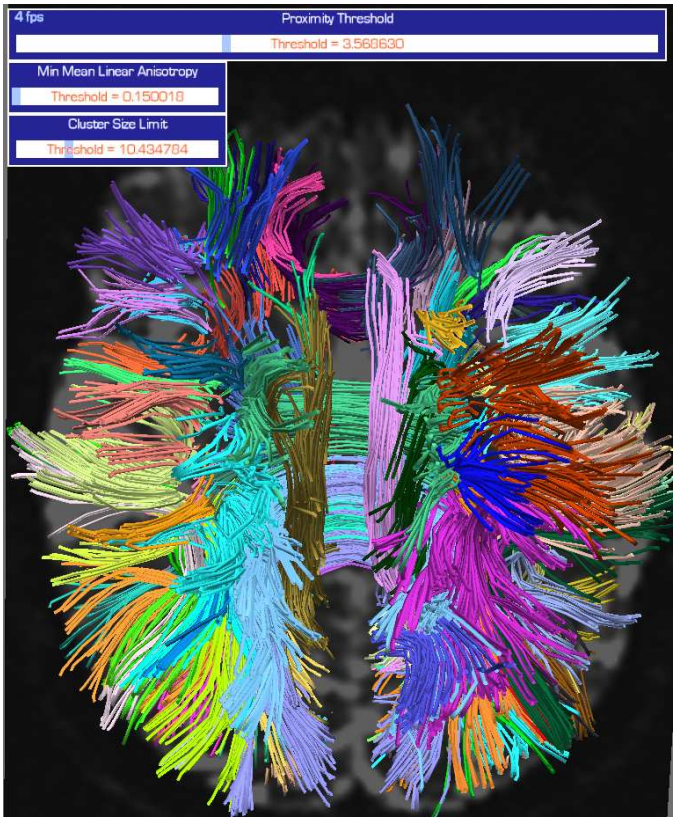


Fig. 3. The clustering result of the model in 2(b) using the single linkage algorithm and d_{Lt} as the proximity measure with a threshold of $3.5mm$. The top left corner of the interactive interface contains three sliders for adjusting the proximity threshold, minimum mean fractional anisotropy values, and minimum number of curves in a cluster.

D. Single linkage clustering algorithm

We used the agglomerative hierarchical clustering method [8] and defined the distance between two clusters as the minimum proximity value (i.e., nearest) between any two curves from two clusters (e.g., one curve from cluster X and its closest neighbor in cluster Y). This is called a single-linkage algorithm because any two clusters with a single link (pairs of similar curves) between them are clustered together. If we define the distance between two clusters as the maximum proximity value (i.e., farthest) between any two curves for two clusters, we have a complete linkage algorithm. The rationale for choosing a single linkage algorithm over a complete linkage algorithm is that some fiber bundles, such as some parts of the corpus callosum and corona radiata, may spread out in a sheet-like structure, although we still consider them to be coherent fiber bundles. The complete linkage algorithm would likely break them into smaller pieces while the single linkage algorithm would group them as whole bundles.

We applied the single linkage algorithm to our curves using the proximity measure d_{Lt}. The algorithm followed the three general broad steps to generate any number of clusters from a set of curves:

1) Obtain a set of n singleton clusters.
2) Merge the two nearest clusters.

3) Repeat step two while the distance between the two merged clusters in the last step is less than a pre-specified proximity threshold.

We chose d_{Lt}, the longer mean of thresholded closest distances, as the proximity measure between two integral curves. We chose this measure to help ensure that any local difference in shape between two curves was captured. In section III-C, we showed how we used d_{St} to cull the short curves that are likely to result from the partial volume effect. Here, we assumed all those short, spurious curves had already been removed from the tractography model and therefore chose d_{Lt} for clustering.

We modified the original algorithm in [8], which pre-specified the number of clusters. The number of clusters generated is inversely related to the proximity threshold. That is, the greater the distance selected for identifying a single cluster, the fewer the number of clusters. When the nearest-neighbor threshold is set to a high value, there will be fewer but larger clusters; when the threshold is set to a low value, the number of clusters increases.

Fig. 3 shows a clustering result of the model in Fig. 2(b) using the single linkage algorithm and the d_{Lt} proximity measure with the proximity threshold set to $3.5mm$.

E. Interactive exploration of clustering models

We built an interactive interface for exploring the clustering results and achieving an optimal representation of white-matter fiber bundles.

The interface is shown in Fig. 3. The model in the picture was generated with methods described in section III-C. Single linkage clustering was pre-computed on the model with a range of proximity thresholds set from $0mm$ to $10mm$ with $0.1mm$ intervals. In the top left corner, three sliders were used to change the values of the proximity threshold, minimum mean linear anisotropy along an integral curve, and the cluster size threshold. The interactive system runs about 10 frames per second with a $30MB$ VRML model on a desktop PC with $2Ghz$ AMD CPU and $1GB$ memory.

F. Evaluation

To evaluate our interactive clustering method, experts rated their level of confidence in identifying a pre-specified set of fiber bundles in the clustering results in section III-F.1. We then designed a preliminary fiber bundle labeling system based on our clustering method in section III-F.2 to evaluate the effectiveness of our clustering method in matching fiber bundles across subjects.

1) Expert evaluation: To evaluate our interactive clustering method, an expert rater interactively selected a proximity threshold to achieve visually optimal clustering in models from four datasets: S1, S2, S3, and S4. Datasets S2 and S4 are the same individual scanned on different occasions. For each model, the expert rater interactively selected a proximity threshold that produced a set of color-labeled bundles that provided the "best" clustering result, i.e. a solution that appeared to be the most accurate representation of known white-matter anatomy when considered globally in whole-brain models. After selecting an optimal threshold, the rater used a four-point scale to assign confidence ratings for identifying each of 12 target white-matter tracts (bilaterally when applicable). To avoid circularity, the target white matter tracts were pre-specified; that is, they were selected as rating targets prior to viewing the models. We chose these particular tracts because they vary in terms of their calibers and type (i.e., commissural, association, projection) and as such were thought to provide a fairly robust test of our method's ability to cluster pathways that are generally readily identified and are reasonably consistent in their location and shape in most individuals. Points on the scale were 0=unlikely located or not located, 1=possibly located, 2=likely located, and 3=very likely located. The confidence ratings are listed in Table 1 in the Results section.

2) Cross-subject fiber bundle labeling: We constructed a white-matter bundle template based on the interactive results presented in the previous section, and then attempted to match integral curve bundles from two new subjects automatically to this template. First, we registered the non-diffusion-weighted image from all six datasets (i.e., the four datasets used in the previous section and two new datasets) to one particular dataset chosen at random using FLIRT [9]. The registration was constrained to translation, rotation, and scaling operations only. We then used the transformation matrices to register all the cluster models. Note that the clustering is done on separate subjects before the registration. Compared to an alternative that registers all the integral curves in the same space and then clusters these curves, our method of clustering in individual subjects preserves the gaps between the fiber bundles that otherwise may be filled by curves from other subjects due to the variation in anatomy among different subjects. For each integral curve cluster X, the centroids of the starting points $C_s(X)$, middle points $C_m(X)$, and end points $C_e(X)$ were calculated. Integral curve clusters from the two subjects were then aligned and compared according to The sum of the distances between these centroids:

$$d_M(X,Y) = \| C_s(X) - C_s(Y) \| + \| C_m(X) - C_m(Y) \| + \| C_e(X) - C_e(Y) \|,$$

where X and Y are clusters of integral curves from two subjects.

To be matched as a pair, two clusters X and Y from two subjects $S1$ and $S2$ were compared to all the clusters from the other subject, and they needed to be mutually closest to each other. That is, $d_M(X,Y)$ should be the smallest $d_M(X,Z)$, for all Z in subject $S2$, and $d_M(X,Y)$ should also be the smallest $d_M(Z,Y)$, for all Z in subject $S1$.

From our expert evaluation of the interactive clustering results (section III-F.1), we know that large and distinct fiber bundles are usually well identified in all four subjects, and we believe these four instances should match well against each other. Accordingly, we selected six fiber bundle types that matched in pairs across subjects to construct a color-labeled fiber bundle template. This group of fiber bundle types includes the left and right cingulum bundles, the left and right uncinates, the forceps minor, and the forceps major. Fig. 6 shows four of the six fiber bundle types in the template. Since each fiber bundle type had one instance for each of the four subjects, there were a total of 24 fiber bundles in the template.

With the fiber bundle template, we can automatically cluster and label target fiber bundles from new subjects. The matching score between a fiber bundle X in a new subject and one fiber bundle type $\mathbb{Y}$ was set to

$$\mathbb{M}(X,\mathbb{Y}) = \begin{cases} \min_{i \in \mathbb{S}} \frac{d_M(X,\mathbb{Y}_i)}{|X|} & \text{if there exists a matched pair between } X \text{ and an instance of } \mathbb{Y}, \\ \text{infinity} & \text{otherwise,} \end{cases}$$

where $|X|$ is the number of curves in X, $\mathbb{S}$ is the set of subjects

in the template whose instances of $\mathbb{Y}$ match X as pairs, and $\mathbb{Y}_i$ is the instance of $\mathbb{Y}$ in subject i. $\mathbb{M}$ is designed to favor big clusters in X that are close to $\mathbb{Y}$ because small clusters in the same region are more likely to come from partial-volume effect.

The algorithm searched for a proximity threshold on the new subject that minimizes the sum of the six minimum matching scores between the new subject clusters and the six types of fiber bundles in the template.

IV. Results

To demonstrate our interactive interface described in section III-E, Fig. 4 shows snapshots of clustering results of the model shown in Fig. 3 with different proximity thresholds and cluster size thresholds.

This interface is convenient for qualitatively evaluating new clustering methods or proximity measures. This capability is shown again in Fig. 5, where d_{St} instead of d_{Lt} is chosen as the proximity measure in the single linkage algorithm. The four pictures show the resulting clusters with proximity threshold set to $0.5mm$, $1.4mm$, $1.5mm$, and $1.9mm$. It is clear that using d_{St} as the proximity measure resulted in an abrupt change in cluster structures around the proximity threshold of $1.5mm$. This is because d_{St} deems short and long curves as being similar provided that the short curve lies close to a portion of the long curve, usually resulting in big clusters with curves of different lengths.

The expert confidence ratings assigned to the various bundles after automatic clustering as described in section III-F.1 are shown in Table 1. Each row represents one of the target white-matter tracts. Each rating pair within a row shows the ratings for that feature in both hemispheres of the brain. A comparison across the subjects shows that long, thick, coherent white-matter tracts, or those whose geometry is distinct from their neighbors (e.g. the uncinate fasciculus, cingulum bundles, and forceps major) are identified with high confidence in the fiber bundle models. In contrast, shorter, thinner white-matter tracts like the anterior commissure are almost completely missing from these optimally thresholded models. The difficulty in identifying these smaller structures in the bundle models may be due to the combined effects of their small calibers, and the limited resolution of the images. Other tracts were identified with intermediate confidence. For example, some tracts like the forceps minor are rated with reasonably good confidence in most subjects but are missing or rated somewhat lower in one subject, suggesting a possible error in the model due to image artifacts. Some tracts, such as the superior longitudinal fasciculus, inferior longitudinal fasciculus, arcuate fasciculus, and inferior cerebellar peduncle, received inconsistent ratings. It is important to note that the rater incidentally noticed that some models contained some clusters that appeared spurious or at least anatomically questionable.

Despite these limitations, the results show that our method has promise as a means for interactive clustering of DTI integral curves into anatomically plausible clusters, at least for some of the more prominent white-matter structures. Improvement in image quality, in motion reduction, and in the algorithm itself may provide better results in the future.

Using the matching and labeling algorithm from section III-F.2, the matching scores for two new subjects S5 and S6 are shown in Fig. 7. The search for a proximity threshold that minimized the sum of the matching scores between the new subject fiber bundles and the fiber bundle types in the template is easy since the range of clustering proximity thresholds for a good match is rather narrow. Note, for example, that the whole curve in Fig. 7 spans $1.0mm$. Besides automatically searching for the proximity threshold, we also ask our expert to select a proximity threshold that is visually optimal using the method in section III-F.1. For $S5$, the proximity threshold is $3.6mm$ for the optimal match and $3.9mm$ for the expert-selected threshold. For $S6$, the proximity threshold is $3.2mm$ for the optimal match, while the expert-selected threshold is $3.5mm$. The small $0.3mm$ difference in both cases suggests that the optimal match emulates the expert's choices in these cases adequately.

WMF	S1	S2	S3	S4
fmajor	3	3	3	3
uf	3,3	3,3	3,3	3,3
cb	3,3	3,3	3,3	3,3
cst	3,3	3,3	3,3	3,3
fminor	3	3	2	3
slf	2,3	2,3	2,3	1,1
ilf	2,1	2,3	1,2	2,2
af	2,3	0,3	1,3	1,1
ac	1	0	2	0
scp	0,0	1,1	2,2	1,1
mcp	2,2	3,3	3,3	3,3
icp	1,1	0,0	0,1	1,1

Table 1: Confidence ratings for the 12 white-matter features in the clustering models of four datasets, S1-S4. 3 indicates very likely located, 2 indicates likely located, 1 indicates possibly located, and 0 indicates unlocated. A rating pair shows the ratings for the same tract in each cerebral hemisphere: the first number indicates the tract in the left hemisphere, and the second number indicates the tract in the right hemisphere. WMF – white-matter fiber tract; fmajor – forceps major; uf – uncinate fasciculus; cb – cingulum bundle; cst – cortico–spinal tract; fminor – forceps minor; slf – superior longitudinal fasciculus; ilf – inferior longitudinal fasciculus; af – arcuate fasciculus; ac – anterior commissure; scp – superior cerebellar peduncle; mcp – middle cerebellar peduncle; icp – inferior cerebellar peduncle

We then matched the automatically thresholded clustering result to the fiber bundle template. Fig. 8 shows the matched bundles for S5 and S6. The expert rated these bundles as very likely or likely to be labeled correctly.

Apart from all the spurious or missing curves in the cluster model due to noise, partial-volume effects, etc., the accuracy of the labeling results is also likely to be affected by the anatomical variation of a fiber bundle type across the subjects and the registration errors. Including a large number of subjects in the template should mitigate the impact of anatomical variation. Our labeling results are also biased toward large fiber bundles due to the normalization factor in $\mathbb{M}$.

V. Discussion

We discuss our choice of the clustering methods in section V-A and compare it to optimization-based methods. The expert's evaluation of the fiber bundles is based on the fiber bundles selected a priori by the same expert. We discuss the potential circularity in the fiber bundle selection and evaluation in section V-B. The fiber bundles in the human brain take various sizes and shapes. We discuss the fact that big, coherent, distinct fiber bundles are better identified in the expert evaluation in section V-C.

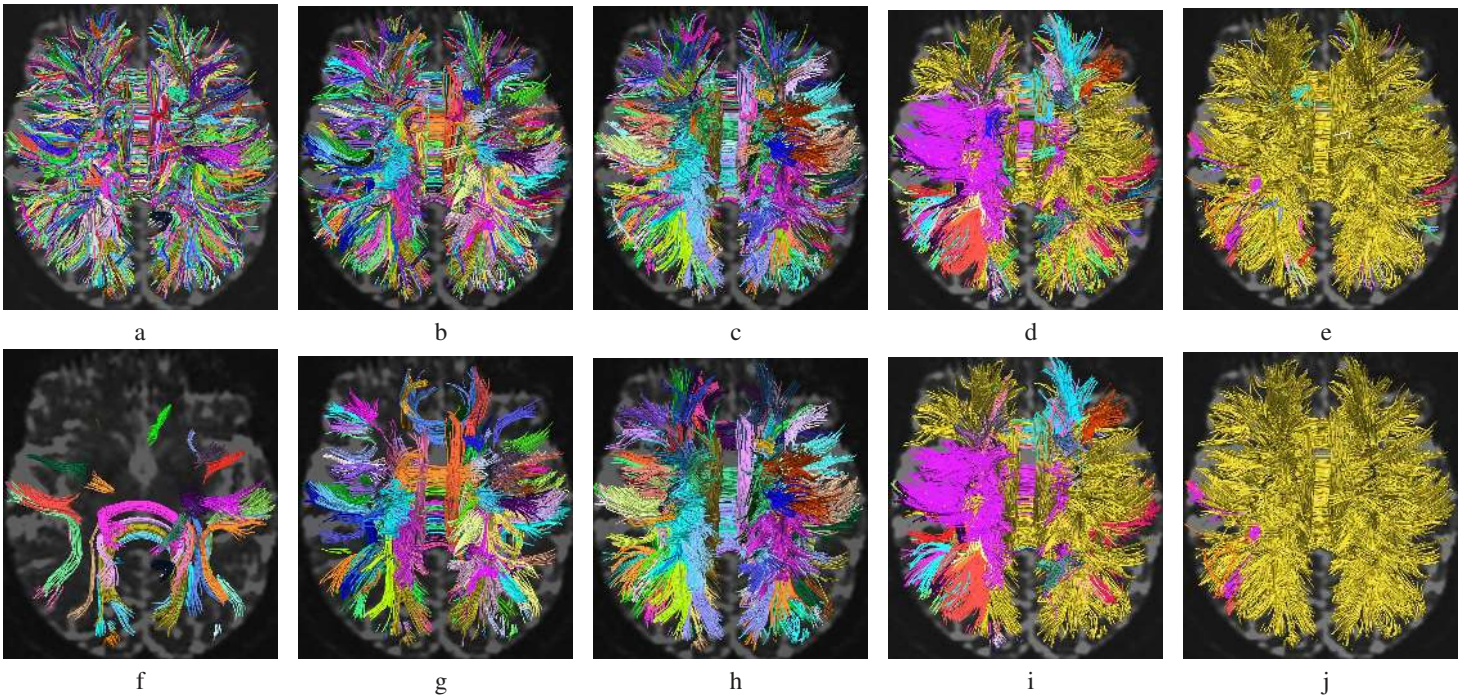


Fig. 4. Snapshots of exploring the clustering models shown in Fig. 3. (a), (b), (c), (d), and (e) show the clustering results with proximity threshold set to $1.5mm$, $2.5mm$, $3.5mm$, $4.5mm$, and $5.5mm$ respectively. (f), (g), (h), (i), and (j) show the same model/threshold with minimum cluster size threshold set to 10.

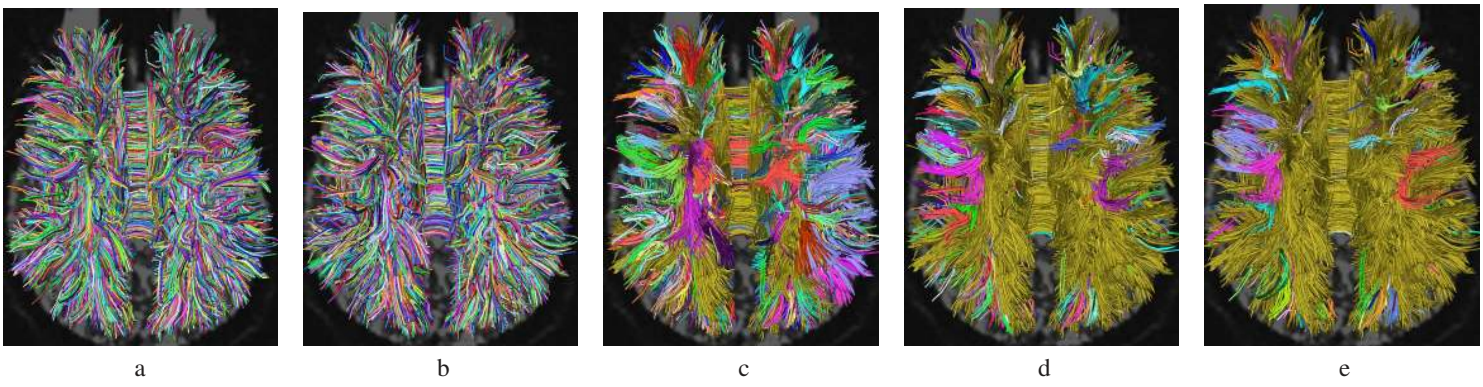


Fig. 5. The interactive interface can be used to evaluate a new clustering algorithm or proximity measure quickly. (a), (b), (c), (d), and (e) show the same model used in Fig. 4 with the same single linkage algorithm, but different proximity measure (d_{St} instead of d_{Lt}). (a), (b), (c), (d), and (e) show the clustering results with the proximity threshold set to $0.5mm$, $1.4mm$, $1.5mm$, $2.1mm$, and $2.5mm$ respectively. The abrupt change of cluster structures around proximity threshold $1.5mm$ suggests that d_{St} is more prone to this kind of jump than d_{Lt}.

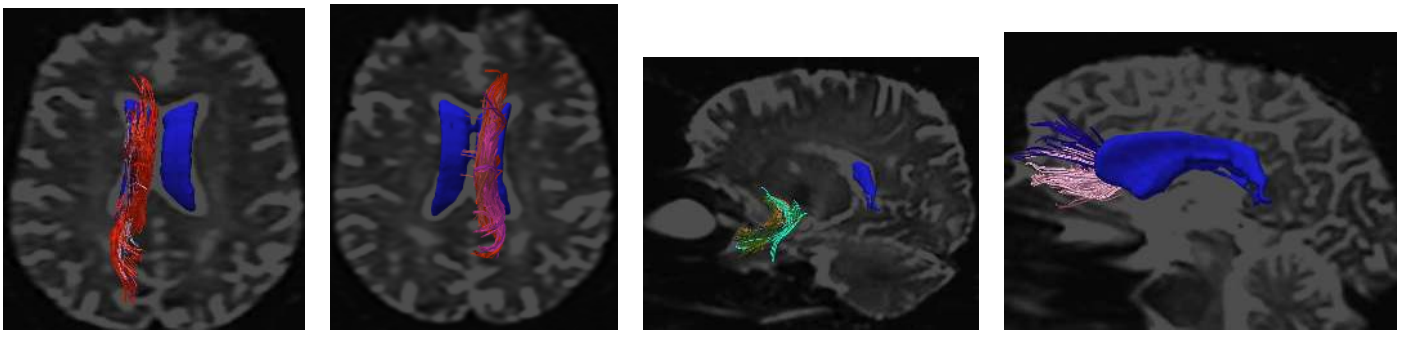
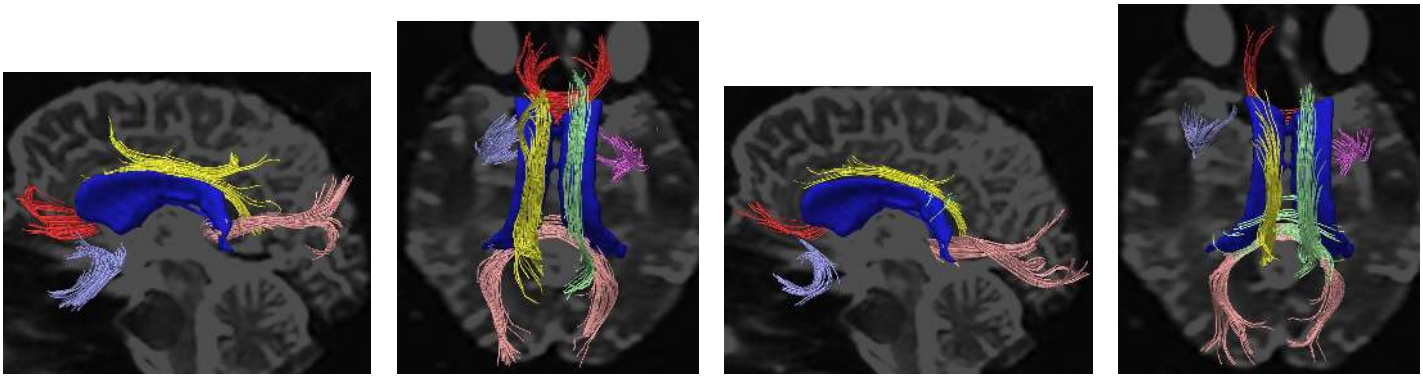
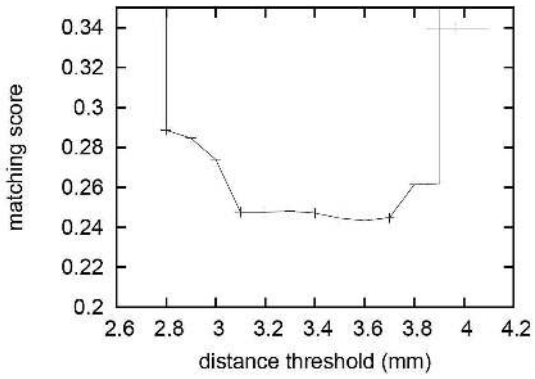


Fig. 6. Four of the fiber bundles in the template: (a) left cingulum bundle, (b) right cingulum bundle, (c) left uncinate, and (d) forceps minor. The blue surface represents the ventricles.

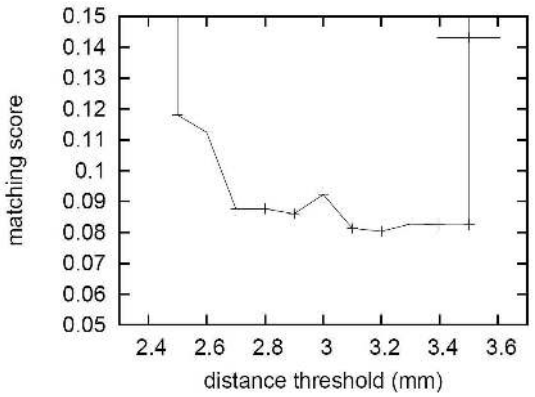


(a) S5 sagittal (b) S5 coronal (c) S6 sagittal (d) S6 coronal

Fig. 8. The matched bundles for subjects S5 and S6. Color is fixed on each label. Blue surface represents the ventricles. There is correspondence between the same bundles across the two subjects.



(a) S5



(b) S6

Fig. 7. The sum of matching scores from the clustering result of a new subject to all the fiber bundles in the template. (a) shows the result of S5. Proximity threshold is $3.6mm$ for the optimal match, while the expert-selected threshold is $3.9mm$. (b) shows the result of S6. Proximity threshold is $3.2mm$ for the optimal match, while the expert-selected threshold is $3.5mm$.

A. Choice of clustering methods

The rationale for using a single linkage algorithm in section III-D to achieve optimal clustering is rooted in our anatomical assumption that in cerebral white matter there exists an implicit physical proximity value between neural fibers within a cluster and between clusters. We assume that a scientist with expert knowledge of cerebral white matter anatomy can make this implicit value explicit by interactively selecting a proximity value that produces a set of clusters that best represents this anatomy. Successful modeling of this implicit physical proximity value should provide a means for parsing individual curves into bundles that accurately reflect the true state of the underlying white matter. Compared to optimization-based clustering methods such as K-means and spectral clustering in Brun *et al.* [4], O'Donnell *et al.* [14], [15], and Maddah *et al.* [11], a single linkage algorithm ensures that one distance threshold is always satisfied between a fiber A and its cluster X; namely there is always another fiber in X whose distance to A is smaller than the threshold. Optimization-based clustering often involves trade-offs among clusters so that while the goal function is optimized, the proximity between a specific pair of clusters may not satisfy a particular threshold. One advantage of an optimization-based clustering method is that it may be more flexible in adjusting to different bundle sizes or different lengths of the fibers. On the other hand, the single linkage algorithm might be useful if a physically-based proximity threshold between the tracts-of-interest can be identified by the experts. Finding a physically-based proximity threshold among fiber bundles is out of the scope of this paper and is a topic of future work. In reality, noise, motion, and partial volume effect also limit our ability to meet our anatomical assumptions. Therefore, our method incorporates an interactive selection of the proximity threshold to maximize the accuracy of the clusters in the face of these artifacts.

B. Target bundle selection

One potential concern of our method is that the interactive clustering performed by the expert rater cannot be fully separated from the 12 target bundles identified for confidence ratings. This is because some of the 12 bundles are large and generally easily identified and as such would be nearly impossible to ignore visually when trying to achieve an optimal global white-matter clustering solution. However, this potential confusion might be unavoidable, even if different experts did the thresholding and rating, because it seems inherently related to visual pattern detection; that is, some anatomical structures are likely to serve implicitly as visual anchors in determining an initial estimate of the optimal proximity threshold, followed by fine-tuning of the threshold to obtain more and more accurate results. A potential cost for achieving finer clustering is that clusters of large known and easily identified tracts break down into too many components. The fact that the first four tracts in Table 1 are all well identified

suggests they might have been used by the expert as visual anchors for the initial threshold approximation.

C. Sizes and shapes of fiber bundles

In this paper we employ a single proximity threshold on the clustering algorithm. In our evaluation, the bundles we chose were rather distinct in shape and location and, apart from the cerebellar peduncles, do not lie in close proximity to one another. These big and coherent fiber bundles are likely the most consistent ones across subjects and thus may serve as tracts-of-interest in cross-subject studies more reliably. A more challenging test might have been to determine if the method could accurately segment more proximal fiber bundles, for example, the inferior longitudinal fasciculus vs. the fronto-occiptal fasciculus. We plan to capture a variety of the fiber bundles better by developing a clustering method that detects the local size and shape of the fiber bundles from the data and adapts the proximity threshold accordingly, possibly with prior knowledge from the experts.

VI. Conclusion

We present a method for clustering DTI integral curves into relevant bundles and then automatically labeling them. Our method enables an expert to interactively specify a proximity threshold between tracts-of-interest and ensures that this proximity threshold is satisfied between the automatically generated clusters. By doing that, the method applies the expert's anatomical knowledge during the clustering process instead of on the clustering results – a key factor that distinguishes our method from prior automated clustering methods. In this sense, our automatic clustering method may be more strongly "anatomically motivated" than the previous ones and therefore has the potential to produce clustering solutions that have high anatomical accuracy while limiting the amount of human effort. An expert rater evaluated the anatomical accuracy for these bundles. The results of this proof-of-concept study suggest that anatomically motivated methods for integral curve generation and clustering can reliably identify large and coherent fiber bundles that bear close qualitative correspondence with known white-matter anatomy across subjects. The results for these large and coherent tracts are good both when the algorithm is guided with interactive input from an expert rater and when it is implemented automatically. This work has the potential to provide a relatively fast, automatic, and accurate way to identify and study neural fiber bundles in large numbers of DTI datasets.

VII. Acknowledgments

The authors thank Dr. David Tate of Brown Medical School for his help in evaluating the clustering results. The authors also thank Professor John F. Hughes and Çağatay Demiralp of Brown Computer Science Department for the helpful discussions. This work was partially supported by the U.S. National Science Foundation (CCR-0086065), the U.S. National Institutes of Health (NIH EB004155, NIA PAR-03-056), the Alzheimer's Association (NIRG-03-6195), the Brain Science Program at Brown University, and the Research Initiation Program at Mississippi State University.

References

[1] Eric T. Ahrens, David H. Laidlaw, Carol Readhead, Celia F. Brosnan, and Scott E. Fraser. MR microscopy of transgenic mice that spontaneously aquire experimental allergic encephalomyelitis. *Magnetic Resonance in Medicine*, 40(1):119–132, July 1998.

[2] H. Alt and M. Godau. Computing the frechet distance between two polygonal curves. *International Journal of Computational Geometry and Applications*, 5:75–91, 1995.

[3] Peter J. Basser, Sinisa Pajevic, Carlo Pierpaoli, Jeffrey Duda, and Akram Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632, 2000.

[4] A. Brun, H. Knutsson, H.-J Park, M.E. Shenton, and C.-F. Westin. Clustering fiber traces using normalized cuts. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 368–375, 2004.

[5] Marco Catani, Robert J. Howard, Sinisa Pajevic, and Derek K. Jones. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage*, 17:77–94, 2002.

[6] Isabelle Corouge, Sylvain Gouttard, and Guido Gerig. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *International Symposium on Biomedical Imaging*, pages 344–347, 2004.

[7] Zhaohua Ding, John C. Gore, and Adam W. Anderson. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49:716–721, 2003.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[9] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5:143–156, 2001.

[10] Stephanie Lee, Stephen Correia, David Tate, Robert Paul, Song Zhang, Steven Salloway, Paul Malloy, and David H. Laidlaw. Quantitative tract-of-interest metrics for white-matter integrity based on diffusion tensor MRI data. In *Proceedings International Society for Magnetic Resonance in Medicine (ISMRM)*, page 283, 2006.

[11] Mahnaz Maddah, William M. Wells III, Simon K. Warfield, C.-F. Westin, and W. Eric L. Grimson. Probabilistic clustering and quantitative analysis of white matter fiber tracts. In *Information Processing in Medical Imaging: 20th International Conference*, pages 372–383, 2007.

[12] Mahnaz Maddah, Andrea U. J. Mewes, Steven Haker, W. Eric L. Grimson, and Simon K. Warfield. Automated atlas-based clustering of white matter fiber tracts from DTMRI. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 188–195, 2005.

[13] Bart Moberts, Anna Vilanova, and Jarke J. van Wijk. Evaluation of fiber clustering methods for diffusion tensor imaging. In *IEEE Visualization*, pages 65–72, 2005.

[14] Lauren O'Donnell and Carl-Fredrik Westin. White matter tract clustering and correspondence in populations. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 140–147, 2005.

[15] Lauren O'Donnell and Carl-Fredrik Westin. High-dimensional white matter atlas generation and group analysis. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 243–251, 2006.

[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, 1992.

[17] Setsu Wakana, Hangyi Jiang, Lidia M. Nagae-Poetscher, Peter C. M. van Zijl, and Susumu Mori. Fiber tract-based atlas of human white matter anatomy. *Radiology*, 230:77–87, 2004.

[18] C.-F. Westin, S.E. Maier, H. Mamata, A. Nabavi, F.A. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6:93–108, 2002.

[19] Song Zhang, Stephen Correia, David F. Tate, and David H. Laidlaw. Correlating DTI fiber clusters with white-matter anatomy. In *Proceedings International Society for Magnetic Resonance in Medicine (ISMRM)*, 2006.

[20] Song Zhang, Cagatay Demiralp, and David H. Laidlaw. Visualizing diffusion tensor MR images using streamtubes and streamsurfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(4):454–462, October 2003.

[21] Song Zhang and David Laidlaw. Hierarchical clustering of streamtubes. Technical report, Brown University, Department of Computer Science, 2002.

[22] Song Zhang and David H. Laidlaw. DTI fiber clustering and cross-subject cluster analysis. In *Proceedings International Society for Magnetic Resonance in Medicine (ISMRM)*, Miami, FL, May 2005.

[23] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.