

Knowledge-Based Classification of Neuronal Fibers in Entire Brain

Yan Xia, And U. Turken, Susan L. Whitfield-Gabrieli, and John D. Gabrieli

Psychology Department, Stanford University
yanxia@stanford.edu, andt@stanford.edu, suewhit@stanford.edu

Abstract. This work presents a framework driven by parcellation of brain gray matter in standard normalized space to classify the neuronal fibers obtained from diffusion tensor imaging (DTI) in entire human brain. Classification of fiber bundles into groups is an important step for the interpretation of DTI data in terms of functional correlates of white matter structures. Connections between anatomically delineated brain regions that are considered to form functional units, such as a short-term memory network, are identified by first clustering fibers based on their terminations in anatomically defined zones of gray matter according to Talairach Atlas, and then refining these groups based on geometric similarity criteria. Fiber groups identified this way can then be interpreted in terms of their functional properties using knowledge of functional neuroanatomy of individual brain regions specified in standard anatomical space, as provided by functional neuroimaging and brain lesion studies.

1 Introduction

Delineation of functional brain networks is an important issue for understanding brain function. Most of what is known about functional organization of the brain is based on contributions of gray matter regions distributed across the brain. Relating this knowledge to structural organization of white matter has become possible with recent DTI methods, [1]. Tractography allows characterization of connectivity patterns across cortical and subcortical regions by specifying extent and orientations of the neural fibers in the white matter. With this information, normal brain function, its development, pathologies, and the effects of normal aging can be better understood if the principles of white matter organization can be described [2]; the nature of the interactions between functionally related brain regions but geometrically separated can be investigated. Conversely, functional interpretation of white matter fiber bundles can be made in reference to the gray matter regions that they connect.

An important goal in this area of research is to achieve systematic identification and characterization of axonal fiber tracts. Current attempts for manually and automatic grouping and labeling tend to focus on similarities between geometric properties of fibers. Ding *et al.* [3] combined a corresponding segment ratio with the mean distance over the segments to delineate the similarity between two streamlines generated from a region of interest (ROI), and grouped the geometrically similar streamlines with nearby seeding points into the bundle. Corouge *et al.* [4, 5] defined three distance measures based on the point sets on the pair of streamlines, and classified a set of streamlines

within specified ROI(s) into meaningful bundles. However, variations across subjects in white matter pathway organization can make extracting and labeling the fiber paths by hand time-consuming and error-prone. The task becomes more challenging when classifying the paths over the entire brain instead of a ROI since no human input about the path information will be implied. Zhang *et al.* [6] explored an unsupervised classifying algorithm on a dense set of paths over the entire brain, and used a serial of strategies in setting seeding points, path constraints, culling distance and setting distance metrics to minimize the misclassifications. Brun *et al* [7] compared the fiber path pairwise to create a weighted undirected graph that was partitioned into coherent sets using the normalized cut criterion.

Our approach starts by identification of commonalities in terminations of white fibers in gray matter. For this, we use a standard brain atlas which provides a parcellation scheme for brain gray matter [Talairach and Tournoux, 1986], which is used by most functional neuroimaging studies for assigning functional roles to individual brain regions in a common standard space. The rationale is that fibers that connect neighboring gray matter regions that are considered to form functional units are most likely to be part of the same white matter tract. At this stage of analysis, alternate pathways between the same regions may appear to form a single tract, and fibers whose gray matter terminations are nearby but labeled differently in the atlas might be assigned to different clusters. After this initial classification, subsequent steps eliminate unlikely fibers, determine subgroups of fibers based on pairwise comparisons of geometric properties and blend groups of fibers with similar trajectories. This approach also considerably reduces computational load, since not all possible comparisons between fibers need to be computed. Additional anatomical constraints related to global organization of brain structures, as well as functional neuroimaging results, can easily be incorporated into this procedure for fiber classification.

Another reason why it is desirable to use the spatial organization of gray matter as a reference point for identifying white matter pathways is that this might make easier to combine tractography data across subjects, and to make comparisons across group. Once pathways within an individual brain have been identified, there is also the problem of deciding which pathways across individuals correspond to the same anatomical tract so that quantitative comparisons can be made. Variability of spatial extents of pathways across individual brains is currently not well known. Correspondences in the relative positions of gray matter structures have been better characterized, and addressed by spatial normalization methods. More concretely, a tract connecting two cortical regions, such as dorsal frontal and posterior parietal cortices, might show considerable variability in its trajectory, so that spatial overlap of its extent across individuals might be poor. Choosing tracts based on the regions that they connect rather than the regions that they cover in space might be a better basis for comparing the properties of tracts across individuals.

2 Materials and Methods

2.1 Tracking Streamlines

The extraction of fiber paths for the entire brain is performed with the tractography tool described in [8]. The tracking algorithm used is based on the STT algorithm. In brief, tensor components are calculated from a voxel size of $1.85 \times 1.85 \times 1.85 \text{ mm}^3$

diffusion weighted images acquired along 13 gradient directions. Streamlines following the principal diffusion directions are produced from seed points along a regular 1.85 mm grid, constrained only to voxels with fractional anisotropy (FA) value exceeding 0.15, and following a local continuity constraints. The integration stops at points where FA value falls below 0.15, which indicate low diffusion directionality.

Figure 1 shows whole-brain streamline sets extracted by this method from a test dataset, and transformed to a common stereotactic space using the Montreal Neurological Institute (MNI) T2 template with SPM2 software. There are more than 46,700 paths which are longer than 20mm. Since the tracking results are not constrained by geometric criteria, a significant portion these 40K paths can be due to effect of noise and include multiple outliers.

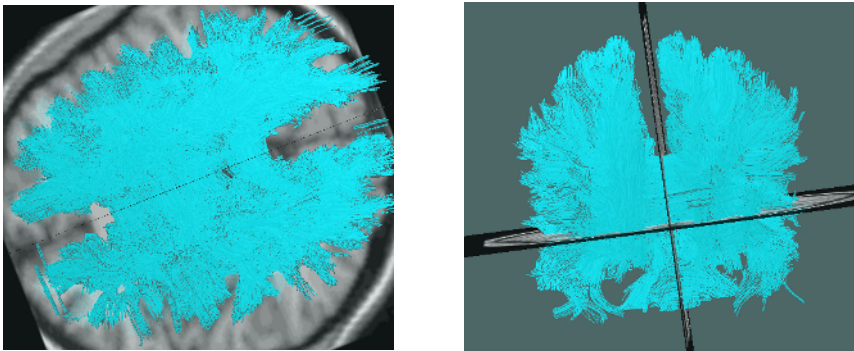


Fig. 1. Whole-brain streamline sets extracted by the method from a test dataset

2.2 Atlas-Based Labeling and Classification of the Fiber Bundles Based on Initial Labels

As a first step, Talairach brain atlas is used to provide labels for each streamline or path and for each fiber bundle separately on the basis of coordinates of two end-points. The labeling of each streamline is done as following:

- If both of two end-points of the streamline fall within regions labeled as gray matter in the atlas space, the streamline is labeled with the names of the regions that it connects (eg., “BA 7_Thalamus”).
- If only one end-point of the streamline falls into labeled gray matter, the second termination is not identified in the atlas as gray matter, the streamline is marked as the name of the gray matter and unidentified (eg., “BA7_ unidentified”).
- If both end-points of the streamline are within regions not identified in the atlas as gray matter, it is marked as “unidentified_unidentified”.

Streamlines that can actually be interpreted as corresponding to axonal fibers should be connecting functional brain regions, and organized in groups (fiber bundles) that correspond to white matter tracts. After the first label assignment, the streamlines are classified into the groups of fiber bundles based on correspondence of atlas labels for their terminations. Figure 2 shows the fiber bundles that connect left and right

hemisphere and classified into more than 150 bundles in the step. Since the robustness of fiber tracking remains limited at junctions and in noisy regions, the extracted fiber set contains outlier paths, paths that are part of other anatomical tracts, and spurious paths linking between anatomically unrelated tracts. These paths should be removed or classified into the correct bundles during the initial classifying stage:

- All streamlines linking between identified anatomic or functional gray matter structures are grouped into the different fiber bundles according to their names. For instance, all streamlines labeled as “BA7_Thalamus” are incorporated into a single group.
- The streamlines named as ‘unidentified _unidentified’ are assigned to one of the set of the fiber bundles defined for the whole brain based on the similarity measures defined in the next section. Figure 3 shows that red color fiber bundle is the source bundle that is within left hemisphere regions not identified in atlas as gray matter, and incorporated into the white color fiber bundle based on the similarity.
- The streamlines labeled as “gray matter _unidentified” are incorporated into one of the set fiber bundles which share the identified gray matter label for this streamline as a termination based on the similarity.
- Streamlines that are shorter than 30 mm or that cannot be assigned to any fiber group based on the similarity measure are removed.

2.3 Classifying Fiber Bundles

The way the fibers bundles are clustered in 3D space, the degree to which fibers in the same bundle run parallel, and their distinct physical and geometric properties should match the properties that brain white matter tracts have in general to be biologically plausible. Low spatial resolution of DT images, limited robustness of fiber tracking methods at junctions and in noisy regions, and interindividual variability of white matter anatomy introduce errors into the fiber tracking procedure. Therefore, the fiber bundles generated by the steps above may contain biologically implausible outlier paths and paths forming spurious bridges between fiber bundles that correspond to actual pathways. The next classification stage is introduced to address these problems. The goals of classifying are removing the outliers, partitioning the fibers in the bundle into natural clusters, and agglomerating the clusters into correct bundles according to the position and shape similarity of fibers. Two measures of similarity between pairs of single fibers are defined below.

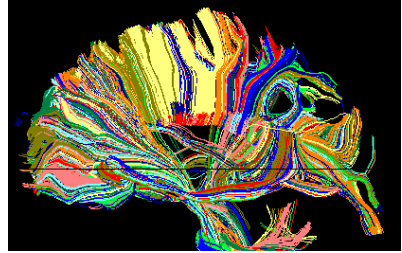


Fig. 2. Fibers connecting the left and right hemisphere are classified into the bundles viewed with different colors

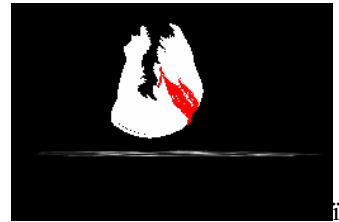


Fig. 3. Red-color fiber bundle is incorporated into the white-color fiber bundle

Similarity Measure. Two fibers are considered similar when they have comparable length, similar shape, and are separated by a short distance. Two pairwise distances between fibers F_i and F_j are used for the similarity measure:

1. Mean distance d_M :

$$d_M(F_i, F_j) = \min(d_m(F_i, F_j), d_m(F_j, F_i))$$

with $d_m(F_i, F_j) = \text{mean}_{p_l \in F_i} \min_{p_k \in F_j} \|p_k - p_l\|^c$ $\|\dots\|$ being point-by-point Euclidean distance▷

2. Hausdorff distance d_H :

$$d_H(F_i, F_j) = \max(d_h(F_i, F_j), d_h(F_j, F_i))$$

with $d_h(F_i, F_j) = \max_{p_l \in F_i} \min_{p_k \in F_j} \|p_k - p_l\|▷$

The mean distance d_M is a good indicator of shape similarity and spatial closeness of a pair of fibers. When two fibers that have comparable length are similar in shape and close in location, d_M is small; when the distance between two fibers is large, or their shapes are different, d_M is large. On the contrary, the Hausdorff distance d_H being a worst-case distance, it is a useful metric to reject outliers and prevent classifying paths with high dissimilarity.

Classifying Fiber Bundles. Classifying the fibers in the bundle involves finding subsets of fibers that correspond to natural anatomical groups by identifying fibers that are clustered closely together in 3D space, running parallel to each other, and have distinct physical and geometric properties. If a bundle formed in Section 2.2 contains more than one such group, it is split into natural sub-clusters. After splitting of bundles, similarities between neighboring sub-clusters are assessed in order to identify those that are likely to belong together in the same tract, so that they are regrouped into a new bundle, and to reject sub-clusters containing small numbers of fibers and showing low similarity with other groups as outliers that are likely to be due to artifacts.

Splitting approach begins with the calculation of pairwise d_M of the fibers in the same bundle. It is not a time-demanding task since the number of fibers in a bundle is not high generally. Supposed calculating the pairwise d_M between the fiber F_i and the fiber F_j in the bundle C , $j \neq i$. If $d_M(F_i, F_j) < t$ (a threshold to be chosen), F_i and F_j are in the same sub-cluster, and vice versa. After calculating a table of pairwise d_M , the approach propagates from neighboring fiber to neighboring fiber. The threshold t depends on the position and shape of the fiber bundle. The larger t value is, the lower the number of sub-classes will be, and vice versa.

Finding the neighbor candidates of the fiber bundle, and limiting the similarity measurement between sub-cluster fibers and the candidate-bundle fibers are very

important to searching approach. The pairwise d_M comparison of all fiber paths is a time-consuming task because the number of fiber paths in entire brain is huge. So the fiber bundle properties of anatomic position, direction, shape, length, end-points, the mass center, and connectivity are introduced for the finding of neighbor bundle candidates in order to limit the number of similarity measurements.

The following steps are involved to reject a sub-cluster or group it into the neighbor fiber bundle or keep it as a natural fiber bundle:

- 1) For a fiber F_i in the sub-cluster, the fiber candidates are found in the candidate bundles. These candidates and fiber F_i have similar in shape and direction, and the mass centers of them are the neighbors.
- 2) Similarity (mean distance d_M) between F_i and each of the fiber candidates are calculated, and a similarity threshold value is set.
- 3) One fiber in candidates is located, whose similarity is greatest (d_M is smallest).
- 4) If the similarity in step 3 is accepted, the sub-cluster is merged into the bundle with the found fiber in step 3 and the processing is stopped. If the similarity is unaccepted, another fiber F_j in the sub-cluster will be chosen and steps 1-4 are repeated.
- 5) When all fibers F in the sub-cluster are checked and the sub-cluster cannot be merged into other bundle, if the average fiber length of the sub-cluster is less than 30 mm and the number of fibers is less than 3, the sub-cluster is rejected forcibly. Otherwise, the sub-cluster is kept and waited for checking manually by an expert.

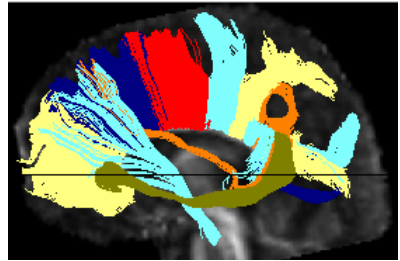


Fig. 4. The processed results of the fiber bundles that connect left and right hemisphere

Repeating the above steps for all sub-clusters and fiber bundles, the algorithm partitions the set of fibers into different fiber bundles. Figure 4 shows the processed results of the fiber bundles that connect left and right hemisphere, there are 14 fiber bundles left.

3 Results

The neuronal fiber classification and quantification DTI method was implemented in Matlab, and tested on the fiber paths from real DTI datasets. Figure 1 shows the streamlines with the length greater than 20 mm before classifying and the path set without the constraint. There are about 46,700 paths; some are short paths stopping within the ambiguous white matter regions and being the part of other paths, which could potentially create artificial links between fiber bundles. Figure 5A shows the

bundles formed after the first step, based on common gray matter projections determined by atlas labels. The final results, derived by redistribution and pruning of fibers based on similarity measurements are shown in Fig 5B.

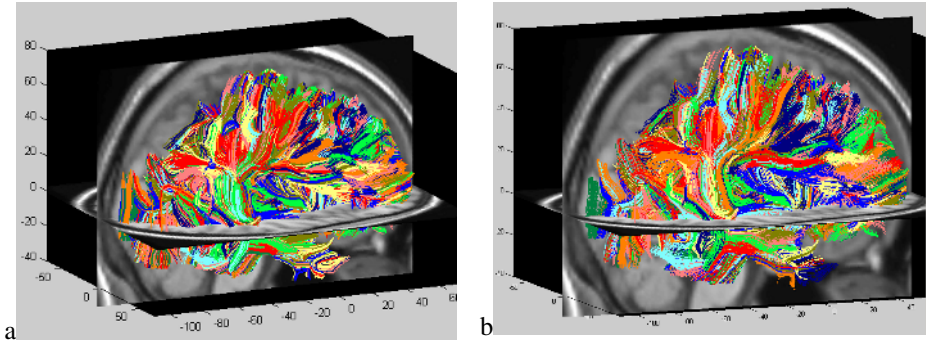


Fig. 5. (a) shows the bundles formed after the first step, there are more than 3000 bundles. (b) The final results, derived by redistribution and pruning of fibers based on similarity measurements, there are 264 bundles left.

4 Discussion and Conclusion

The focus of this work is on classification fiber paths derived from DTI data into bundles associated with distinct anatomical and functional structures. The fiber classification procedure is centered around knowledge of anatomical and functional subdivisions of gray matter obtained from the Talairach atlas. Fibers are first grouped based on common projections to parts of gray matter designated as distinct regions in the atlas. Subsequent stages refine the initial distribution of fiber groups, so that the final set of bundles contain fibers that share similar physical and geometrical properties, and connect gray anatomically and functionally related regions of gray matter. This approach reduces the computational load of whole brain fiber analysis, and automatically produces groups of fibers that can be interpreted in an intuitive manner.

The algorithm extracts whole-brain streamlines from naive dataset, transforms the streamlines to a common stereotactic space using the MNI T2 template with SPM2 software, and names the fiber and fiber bundle with the region names of two end-points of the fiber and fiber bundle. The transform accuracy and interindividual variability do influence the name of one fiber and may put the fiber in an incorrect bundle at initial stage, the algorithm re-classifies the fibers into natural anatomical bundles by identifying fibers that are clustered closely together in 3D space, running parallel to each other, and have distinct physical and geometric properties. If the difference between the individual brain and Atlas is large, some fiber bundles connecting small cortical areas may have an incorrect label names.

Low spatial resolution of DT images and limited robustness of fiber tracking methods at junctions and in noisy regions may issue tracts that are locally consistent but incorrectly connected. The algorithm cannot handle the case currently, and the

fiber bundles have to remove manually. Our future work will focus to validate the fiber tracts and to explore the relationship between the physical and geometric properties of the fiber tracts and the brain diseases and brain development.

References

1. Mori, S., Crain, B. J., Chacko, V.P., and van Zijl, P. C. M. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* Vol 45, pp.265-269, 1999.
2. O'Sullivan, M., Jones, D.K., Summers, P.E., Morris, R.G., Williams, S.C.R., and Markus, A.S. 2001. Evidence for cortical disconnection as a mechanism of age-related cognitive decline. *Neurology* 57:632-638.
3. 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.
4. 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.
5. Guido Gerig, Sylvain Gouttard, and Isabelle Corouge. Analysis of brain white matter via fiber tract modeling. *EMBS* 2004.
6. Song Zhang, and David H. Laidlaw. DTI fiber clustering in the whole brain.
7. Anders Brun, Hans Knutsson, Hse-Jeong Park, Martha E. Shenton, and Carl-Fredrik Westin. Clustering fiber traces using normalized cuts. *Miccai* 2004, pp. 368-375.
8. David Akers, Anthony Sherbondy, Rachel Mackenzie, Robert Dougherty, and Brian Wandell. Exploration of the brain's white matter pathways with dynamic queries. In *Proceedings IEEE Visualization*, pp.77-384, 2004.
9. Dongrong Xu, Susumu Mori, Meiyappan Solaiyappan, Peter C.M. van Zijl, and Christos Davatzikos. A framework for callosal fiber distribution analysis. *NeuroImage* vol. 71, pp. 31-143, 2002.