Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2011 April ; 2011: 276–280. doi:10.1109/ISBI.2011.5872405.

ATLAS-BASED FIBER CLUSTERING FOR MULTI-SUBJECT ANALYSIS OF HIGH ANGULAR RESOLUTION DIFFUSION IMAGING TRACTOGRAPHY

Gautam Prasad^{*}, Neda Jahanshad^{*}, Iman Aganj[†], Christophe Lenglet[‡], Guillermo Sapiro[†], Arthur W. Toga^{*}, and Paul M. Thompson^{*}

^{*}Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA, USA

[†]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA

[‡]Department of Radiology - CMRR, University of Minnesota Medical School, Minneapolis, MN, USA

Abstract

High angular resolution diffusion imaging (HARDI) allows *in vivo* analysis of the white matter structure and connectivity. Based on orientation distribution functions (ODFs) that represent the directionality of water diffusion at each point in the brain, tractography methods can recover major axonal pathways. This enables tract-based analysis of fiber integrity and connectivity. For multi-subject comparisons, fibers may be clustered into bundles that are consistently found across subjects. To do this, we scanned 20 young adults with HARDI at 4 T. From the reconstructed ODFs, we performed whole-brain tractography with a novel Hough transform method. We then used measures of agreement between the extracted 3D curves and a co-registered probabilistic DTI atlas to select key pathways. Using median filtering and a shortest path graph search, we derived the maximum density path to compactly represent each tract in the population. With this tract-based method, we performed tract-based analysis of fractional anisotropy, and assessed how the chosen tractography algorithm influenced the results. The resulting method may expedite population-based statistical analysis of HARDI and DTI.

Index Terms

tractography; clustering; Dijkstra's shortest path; multi-subject analysis; fiber bundles

1. INTRODUCTION

Fiber integrity and connectivity in the living brain may be measured using high angular resolution diffusion imaging (HARDI), which reconstructs the local profile of water diffusion at each point in the brain. Following the dominant directions of local water diffusion throughout the white matter, is one way tractography methods can recover the geometry and connectivity of the major white matter fiber pathways.

In clinical research, tractography may be used to reconstruct white matter tracts for surgical planning [1]. Analysis of structural connectivity is also useful for understanding coherent activity in functional networks. There are several ongoing efforts to map the human connectome, and to detect altered patterns of connectivity in disease. Additionally, fiber tracts may be extracted to study hemispheric asymmetries and functional lateralization [2], and to identify genetic effects and sex differences in neural network organization [3]. However, before group differences in specific tracts can be studied, the large collection of 3D curves generated by tractography needs to be organized into bundles that correspond to well-known white matter tracts [4]. Corresponding tracts and bundles also need to be matched across subjects, as a basis for statistical comparisons.

A wealth of methods have been developed to cluster fibers extracted using diffusion-based tractography [5, 6]. Some methods embed the problem in $\mathbb{R}^n$ [7]; they compute geometrical invariants for each curve; these are then stacked into high-dimensional vectors that are clustered. The resulting clusters of points may correspond to known white matter commissures such as the corpus callosum, or major fasciculi. Some methods are inspired by standard unsupervised pattern recognition methods. They divide a set of n observations into g groups, to maximize some metric of similarity among members of a group versus between members of different groups. Algorithms based on K -means, fuzzy clustering, hierarchical or agglomerative methods, and self-organizing maps can all be adapted, in principle, to cluster tracts. One such approach, spectral clustering [8, 7] groups fibers based on pair-wise distance measures among fibers as well as their spatial locations. Probabilistic clustering has also been used, without anatomical constraints, based on polynomial regression mixture models [9]. One problem with clustering methods that do not use prior anatomical information is that in many cases, known fiber bundles may be split into two or combined by the algorithm. A consistent partition across subjects is often difficult to find, and results may not correspond to known tracts in neuroanatomical atlases. Overall, the goal of this work is to (1) generate representative curves for tract based analysis of HARDI; (2) derive them guided by regions of interest in a brain atlas, so that the resulting tracts better reflect known anatomy, and (3) to make the tracts analyzed robust to discrepancies between the atlas and each individual subject, which tend to make atlas-based methods harder to apply to new subjects. To assess the utility of the method, we show an illustrative application assessing hemispheric asymmetry in tract-based FA, and we assess how the results depend on the tractography method that provides a basis for the analyses.

To create consistent fiber maps across subjects, we extracted fibers with a whole-brain tractography method, based on the Hough transform, and developed an atlas-based method to cluster them. First we retain all curves that are consistent, according to a similarity metric, with those in a probabilistic atlas. We then remove spurious curves by applying a median filter to a binary map representing the preliminary fiber clusters. We create a graph from a density image of the clustered fibers in each subject. With the Dijkstra shortest path algorithm [10], we compute a representative path through this graph. This path follows areas with the highest fiber density. These maximum-density shortest paths offer a compact way to compare fiber bundles across a population, and compute multi-subject statistics. We assessed how the results depend on the choice of tractography algorithm, using one

algorithm that used the full HARDI ODF, and a more standard streamline method, that follows the principal eigenvector of the diffusion tensor. We concluded that both methods gave satisfactory results for tract-based statistical analyses, with some characteristic differences in the paths extracted; we comment on these in the Discussion.

2. METHODS

2.1. Image data

We analyzed 105-gradient high-angular resolution diffusion images (HARDI) acquired from 20 healthy young adults [11], on a 4 T Bruker Medspec MRI scanner. 3D volumes consisted of 55 2-mm thick axial slices with a $1.79 \times 1.79 \text{ mm}^2$ in-plane resolution. 94 diffusion-weighted images ($b = 1159 \text{ s/mm}^2$) were acquired with a uniform distribution of gradient directions on the hemisphere. 11 b_0 (non-diffusion encoding) images were also acquired. Images were corrected for motion and eddy current distortions using FSL (www.fmrib.ox.ac.uk/fsl/).

2.2. Tractography methods

We analyzed two different tractography methods to explore the robustness of our clustering technique: (1) a novel Hough transform method to compute optimal paths through a field of constant solid angle orientation distribution functions (CSA-ODFs) derived from the HARDI data [12], and (2) a streamline-based method based on the diffusion tensor [13]. We used *TrackVis* [14] to visualize fibers and clusters from both tractography methods.

The Hough transform method [12] tests 3D fibers that pass through seed points throughout the image space. It assigns a score to each fiber, based on the log-probability of the existence of the fiber, according to a cost function. The cost function measures the generalized fractional anisotropy (GFA) [15] along the fiber path, as well as the probability of a tract propagating in any given direction, derived from the ODF. The algorithm applies a Hough transform based voting process to a vast set of 3D curves in the volume. It assigns a score to each of them, and chooses those with the highest scores as potential tracts. The Hough transform is used to assign the highest scoring curve through a seed point as the potential fiber tract by having the voxels vote for the possible curves. For a single subject, each curve's score is defined as:

$$\int (\log[ODF(\vec{x}(s), \vec{t}(s))GFA(\vec{x}(s))] + \lambda) ds \quad (1)$$

where $x(\vec{s})$ and $t(\vec{s})$ are respectively the location and the unit tangent vector to the curve under consideration, at the arc length s and λ is a positive constant used as a prior on fiber length.

We also used the Diffusion Toolkit [14] to calculate standard streamline based fibers from the diffusion tensor model. Tracking follows the local principal diffusion direction, from a seed point in the image, to form a fiber. The stopping criterion is based on a strong change in fiber direction,

$$R = \sum_i^s \sum_j^s |v_i \cdot v_j| / s(s-1), \quad (2)$$

where s is the number of data points referenced on the fiber, v_i is the direction of maximal diffusion at point i . *TrackVis* was used to visualize fibers and filter them based on regions of interest (ROI).

2.3. Probabilistic atlas clustering

Each subjects T1-weighted image was aligned by 9-parameter transformation to the Colin27 high-resolution brain template [16] after manual skull stripping. The 11 b_0 images from the DWI were averaged and aligned to the corresponding registered T1-weighted image, after they were masked using BET [17]. In this space, FA maps were generated for each subject. We also created a geometrically-centered study-specific mean FA template (or minimal deformation template; MDT) [18].

To cluster fibers from the tractography algorithm we used the JHU white matter tractography atlas [19], which contains 17 prominent white matter tracts based on a set of fibers traced in 4 healthy subjects. The FA image from the JHU atlas was affinely transformed to the same MDT using FMRIB's Linear Image Registration Tool (FLIRT) [20]. The transformation was also applied to each of the 3D probabilistic tract images.

Figure 1 shows the streamline tractography results with *Track-Vis*. The fibers shown are those that intersect any of six selected regions from the fiber tract atlas: the *forceps major* (shown in green); *forceps minor* (orange); left superior longitudinal fasciculus (purple); right superior longitudinal fasciculus (red); left inferior fronto-occipital fasciculus (burgundy); and right inferior fronto-occipital fasciculus (blue). Fibers were retained if they passed through any of the six regions. The fibers are colored by the direction of their middle segment.

This figure gives insight into how we can use the atlas to select consistent sets of fibers in the brain. *TrackVis* allows set-based operations (logical AND/OR/NOT) with ROIs to select fibers, but generally requires manual effort to interact with the fiber sets on an individual fiber level.

To allow atlas-based fiber labeling independent of *TrackVis*, we fitted a 3D cubic spline curve to each fiber, and regularly sampled it. We found the set of voxels that it intersects with in the encompassing image space, then we computed their overlap with the white matter tract of interest in the atlas. The number of voxels in the intersection measures how well a tract fits into an atlas region. We then selected all tracts that intersected the atlas. This automatically selects fibers from a population of images, or from each individual subject. Figure 2(a) shows the part of the fiber tract atlas that represents the *forceps major* region of the corpus callosum, which includes the splenium. In this region, fibers arch backwards into the occipital lobes at the back of the brain. Figure 2(b) shows all the tracts that intersect it. The tracts shown here were computed using the Hough method. The fibers are colored by the number of voxels they intersect that are within the atlas region.

As shown in Figure 2(b), choosing all possible fibers that intersect a given white matter atlas region does not necessarily yield a good representation of that particular fiber bundle. Spurious curves may be included. Because of this, a threshold based on the voxel intersection count was used to further constrain which fibers were assigned to that fiber bundle (Figure 2(c)). *TrackVis* enables the selection of fibers that intersect an ROI. In the case of the JHU atlas, *TrackVis* allows the user to reduce the size of the ROI. This helps constrain the number of fibers selected. We employed a threshold that differed from *TrackVis* in that it enabled us to select fibers based on how much fiber passed through the atlas. We chose the threshold to be the thickness of the ROI, although varying this did not affect the results drastically.

2.4. Median filtering of tract density

Once a set of tracts has been found in one subject using the JHU atlas, we create a binary volume image where a voxel is 1-valued if a fiber intersects it and 0 otherwise. We computed all the voxels that the selected fibers pass through, to create one binary volume image. We then apply median filtering [21] to this volume image. Median filtering replaces each voxels value by the median of the voxel values in its neighborhood (which for our binary data, is like a voting process). In our experiments, we used a $5 \times 5 \times 5$ box as the neighborhood centered on each voxel. The median filter changes any 1-valued voxels far from the main cluster to 0, to suppress noise and erroneous fibers. We use the median-filtered binary image to further screen the fibers in the cluster by calculating the intersection of the binary image voxels with fibers in the cluster. We remove a fiber if all the voxels it travels through are not included in the filtered binary image. This removes fibers far from the main cluster, and creates a dense set of tracts with a coherent shape.

The resulting set of fibers after this median filtering step is the final set included in the fiber bundle and serve as the basis for the white matter tract. Representative results of median filtering are shown in Figure 3(a).

2.5. Representing the cluster by the maximum density path

To compactly represent the fiber bundle in each subject, we create a path through the bundle to follow the highest density of fibers. We first create a fiber density volume image by counting, at each voxel, how many fibers pass through it. We then create a graph (a set of nodes and undirected edges connecting them) from this density image. Nodes in the graph represent the voxel locations that have a non-zero density value, and the edges connect each voxel to its surrounding 26 neighboring voxels. The value of the edge connecting nodes i and j is dictated by:

$$\frac{1}{d_i + d_j} \quad (3)$$

where d_i is the density at the voxel corresponding to node i . This makes the cost of traveling from node i to node j inversely proportional to their density.

We then select a start and end node in the graph by first assigning a start and end voxel location in the JHU atlas. The closest non-zero voxel locations in the density image (by

Euclidean distance) to the start and end voxels in the atlas are used as the voxels that correspond to the start and end nodes in the graph. Dijkstra's algorithm [10] is then run on the graph to find the shortest path connecting the start and end nodes. Dijkstra's algorithm is a graph search algorithm that efficiently finds the shortest path from a start node to every other node in the graph. In this case, the shortest path will include edges connecting nodes with high density values. As a result, it follows the path with the highest density of fibers in the image.

This resulting path is an ordered set of voxels. The centers of the voxels in the path are then used as the coordinates that trace the path and a 3D cubic spline is fitted to them. The resulting path is used to compactly represent the fiber bundle. The maximum density path of the fiber bundle in 3(a) is shown in 3(b). By computing the shortest path in each subject, we can represent the fiber bundles in all subjects in the entire dataset. The FA along this fiber path may then be used to compare fiber bundles across subjects or groups [22].

2.6. Paired *t*-test of FA

Next we illustrate how to use the method to obtain tract-based statistics [23]. Once the maximum density curves are computed, we performed a paired-sample *t*-test comparing the FA along the maximum density curves from the Hough transform method to those from the streamline method. This was done pointwise, by uniformly sampling 100 points across each curve and finding the FA value at each point from the corresponding FA image, via cubic interpolation. We then found the mean FA along each curve and used that as an attribute for statistical analysis along that curve.

Another paired-sample *t*-test was used to assess the degree of hemispheric (left/right) symmetry of the tracts using each method. We split the curves in two by separating them at the midsagittal plane. Then we found the mean FA along each of these curves and compared them. This test was applied to the group of maximum density curves for each tractography method.

3. RESULTS

To test the proposed approach, we ran the methods on 20 subjects images. Fibers were found using the Hough transform (*green*) and streamline (*blue*) methods. Figure 4 shows 40 different maximum density paths, one for each of the two tractography methods, per subject. Paths are then readily compared across subjects and methods.

The streamline method generates many more fibers than the Hough transform method. This is partly because the cost function does not rate fibers based on their overall geometry, as the Hough method does. Its density image also covers a larger volume, and all maximum density paths originate at the user-specified start and end points. Fibers were manually selected using *TrackVis* by using the NOT logical function to remove fibers outside the JHU atlas region.

The paired-sample *t*-test comparing the FA along the maximum density curves from the Hough transform method compared to those from the streamline method gave a *p*-value of

8.8×10^{-8} . The paired-sample t -test assessing the hemispheric asymmetry of the maximum density curves from the Hough transform method gave a p -value of 0.56. The test for the symmetry of the curves from the streamline method gave a p -value of 9.4×10^{-3} . The difference may be because there are significantly more curves from the streamline method so that the maximum density curve has a somewhat different trajectory. The fibers generated from each method may also differ in shape. In future, it may be fairer to compare results using the same number of fibers for both methods, although there would have to be a principled decision about which fibers to keep. Also, the Hough method is somewhat more computationally intensive (given its more complex mathematical formulation), making it easier to sample a very high number of curves with the streamline method.

We then found the FA along each of these curves; results are plotted in Figure 5. Shown in green are the curves from the fibers generated using ODF/Hough tractography. In blue, we show the curves generated from the streamline single-tensor method. In general, the FA along the curves from the Hough method was greater, though both approaches had the same general pattern (higher at midline).

4. DISCUSSION

In this work, we combined whole-brain HARDI tractography with a standard white matter tract atlas. We were able to automatically group traced fibers in the brain irrespectively of the chosen tractography method. By reconstructing maximum density paths for each individual, we can compactly represent known anatomical fiber tracts in new datasets. This allows us to obtain group statistics for cross-subject comparisons of scalar attributes along these paths (e.g., such as FA, GFA, or mean diffusivity). This method lends itself to population studies as it recovers homologous anatomical tracts across subjects without requiring manual labeling of tracts. It also avoids computationally intensive high-dimensional clustering methods to match corresponding clusters across subjects, which is anyhow not always successful. The threshold and median filter are additional processing steps that do not require manual modification, once established for a specific dataset.

Maximum density paths may be used to compute population statistics for clinical or genetic analyses of connectivity. Features such as curvature along the paths may also be examined. The paths may be registered together across large numbers of subjects to investigate statistical factors associated with differences in tract geometry and fiber characteristics across populations.

The low p -value for the test comparing mean FA along the paths between the two approaches could show that although the overall shape of the maximum density curves is similar, as shown in Figure 5 the distribution of FA along the curves is significantly different. This could demonstrate that each method clusters effectively with sufficient internal coherence to capture the differences in the two tractography methods. In addition, the Hough method tended to pick tracts with consistently higher FA. This may be useful in statistical studies, where high FA regions typically offer greater SNR, making it easier to detect effects of specific clinical or genetic factors on fiber microstructure.

Acknowledgments

This work was supported by NIH grant R01 HD050735 and the National Health and Medical Research Council, Australia, grant NHMRC 496682. Additional support was provided by grants R01 EB008281, P41 RR013642, P41 RR008079, P30 NS057091, R01 EB008432, the University of Minnesota Institute for Translational Neuroscience and NLM T15 LM07356.

REFERENCES

1. Nimsky C, Ganslandt O, Hastreiter P, Wang R, Benner T, Sorensen AG, Fahlbusch R. Preoperative and intraoperative diffusion tensor imaging-based fiber tracking in glioma surgery. *Neurosurgery*. 2005; 56(1):130–138. [PubMed: 15617595]
2. Powell HW, Parker GJM, Alexander DC, Symms MR, Boulby PA, Wheeler-Kingshott CAM, Barker GJ, Noppeney U, Koepp MJ, Duncan JS. Hemispheric asymmetries in language-related pathways: a combined functional MRI and tractography study. *NeuroImage*. 2006; 32(1):388–399. [PubMed: 16632380]
3. Jahanshad, N.; Aganj, I.; Lenglet, C.; Joshi, A.; Jin, Y.; Barysheva, M.; McMahon, KL.; De Zubicaray, GI.; Martin, NG.; Wright, MJ.; Toga, AW.; Sapiro, G.; Thompson, PM. International Symposium on Biomedical Imaging, ISBI 2011. IEEE; 2011. Sex differences in the human connectome: 4-Tesla high angular resolution diffusion imaging (HARDI) tractography in 234 young adult twins. in press
4. Mettler FA. Connections of the Cerebral Cortex. *Archives of Neurology*. 1964; 10(6):637.
5. Visser E, Nijhuis EHJ, Buitelaar JK, Zwiers MP. Partition-based mass clustering of tractography streamlines. *NeuroImage*. 2011; 54(1):303–312. [PubMed: 20673849]
6. Wang X, Grimson WEL, Westin CF. Tractography segmentation using a hierarchical Dirichlet processes mixture model. *NeuroImage*. 2011; 54(1):290–302. [PubMed: 20678578]
7. O'Donnell LJ, Westin CF. Automatic tractography segmentation using a high-dimensional white matter atlas. *Medical Imaging, IEEE Transactions on*. 2007; 26(11):1562–1575.
8. O'Donnell LJ, Kubicki M, Shenton ME, Dreusicke MH, Grimson WEL, Westin CF. A method for clustering white matter fiber tracts. *American Journal of Neuroradiology*. 2006; 27(5):1032–1036. [PubMed: 16687538]
9. Gaffney SJ, Smyth P. Joint probabilistic curve clustering and alignment. *Advances in neural information processing systems*. 2005; 17:473–480.
10. Dijkstra EW. A note on two problems in connexion with graphs. *Numerische mathematik*. 1959; 1(1):269–271.
11. De Zubicaray GI, Chiang MC, McMahon KL, Shattuck DW, Toga AW, Martin NG, Wright MJ, Thompson PM. Meeting the challenges of neuroimaging genetics. *Brain Imaging and Behavior*. 2008; 2(4):258–263. [PubMed: 20016769]
12. Aganj I, Lenglet C, Jahanshad N, Yacoub E, Harel N, Thompson PM, Sapiro G. A Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography. *Medical Image Analysis*. 2011 in press.
13. Mori S, Crain BJ, Chacko VP, Van Zijl PCM. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*. 1999; 45(2):265–269. [PubMed: 9989633]
14. Wang R, Benner T, Sorensen AG, Wedeen VJ. Diffusion Toolkit: A Software Package for Diffusion Imaging Data Processing and Tractography. *Proc. Intl. Soc. Mag. Reson. Med*. 2007; 15:3720.
15. Tuch DS. Q-ball imaging. *Magnetic Resonance in Medicine*. 2004; 52(6):1358–1372. [PubMed: 15562495]
16. Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*. 1998; 22(2): 324–333. [PubMed: 9530404]
17. Smith SM. Fast robust automated brain extraction. *Human Brain Mapping*. 2002; 17(3):143–155. [PubMed: 12391568]

18. Jahanshad N, Lee AD, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Toga AW, Thompson PM. Genetic influences on brain asymmetry: A DTI study of 374 twins and siblings. *NeuroImage*. 2010
19. Wakana S, Jiang H, Nagae-Poetscher LM, van Zijl P, Mori S. Fiber Tract-based Atlas of Human White Matter Anatomy. *Radiology*. 2004; 230(1):77–87. [PubMed: 14645885]
20. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*. 2001; 5(2):143–156. [PubMed: 11516708]
21. Tyan S. Median filtering: Deterministic properties. *Two-Dimensional Digital Signal Processing II*. 1981; 43:197–217.
22. Fillard P, Gilmore J, Piven J, Lin W, Gerig G. Quantitative analysis of white matter fiber properties along geodesic paths. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*. 2003:16–23.
23. Corouge I, Fletcher PT, Joshi S, Gilmore JH, Gerig G. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2005*. 2005:131–139.

Fig. 1. Selected subsets of curves (extracted using *TrackVis*), based on their intersection with six different white matter regions from the JHU atlas. Atlas regions are shown in solid colors. These regions were used as a basis (ROIs) for curve selection and further analysis.

(a)

(b)

(c)

Fig. 2.

(a) *Forceps major* region of the corpus callosum from the JHU atlas. (b) Fibers from the Hough transform method that intersect the corpus callosum in the JHU atlas. (c) Fibers that remain after a threshold is applied to the number of points that must intersect the atlas. Fibers are colored by the number of voxels they intersect within the atlas.

(a)

(b)

Fig. 3.

(a) Fibers included in the bundle after median filtering of the binary image. (b) Fiber path through the bundle, following the areas of highest density.

Fig. 4. Maximum density curves from the Hough transform method (*green*) and the streamline method (*blue*). There are 40 curves in total: each method was run with the same 20 HARDI datasets.

Fig. 5.

FA along the maximum density curves generated using fibers generated from the Hough transform and streamline method, for each of the 20 HARDI datasets. Interestingly, the Hough method picks curves with a higher FA, and there is some evidence that differences in FA are easier to detect when FA is higher, perhaps due to the higher SNR. For this reason, the Hough-based tracts may be more useful for studying factors that influence tract FA. The FA is slightly reduced away from the midline of the brain, partly due to a less degree of directional coherence among fibers within a voxel. This occurs because the fibers are all essential orthogonal to the midsagittal plane at midline, but mix with other tracts and fan out more as they deviate further from midline.