

# Inference of a HARDI Fiber Bundle Atlas Using a Two-Level Clustering Strategy

Pamela Guevara<sup>1,2\*</sup>, Cyril Poupon<sup>1,2</sup>, Denis Rivière<sup>1,2</sup>, Yann Cointepas<sup>1,2</sup>,  
Linda Marrakchi<sup>1,2</sup>, Maxime Descoteaux<sup>4</sup>, Pierre Fillard<sup>1,2,3</sup>,  
Bertrand Thirion<sup>1,2,3</sup>, and Jean-François Mangin<sup>1,2</sup>

<sup>1</sup> Neurospin, CEA, Gif-sur-Yvette, France

<sup>2</sup> Institut Fédératif de Recherche 49, Gif-sur-Yvette, France

<sup>3</sup> Parietal Team, INRIA Saclay-Ile-de-France, Saclay, France

<sup>4</sup> Université de Sherbrooke, Sherbrooke, Canada

**Abstract.** This paper presents a method inferring a model of the brain white matter organisation from HARDI tractography results computed for a group of subjects. This model is made up of a set of generic fiber bundles that can be detected in most of the population. Our approach is based on a two-level clustering strategy. The first level is a multiresolution intra-subject clustering of the million tracts that are computed for each brain. This analysis reduces the complexity of the data to a few thousands fiber bundles for each subject. The second level is an inter-subject clustering over fiber bundle centroids from all the subjects using a pairwise distance computed after spatial normalization. The resulting model includes the large bundles of anatomical literature and about 20 U-fiber bundles in each hemisphere.

## 1 Introduction

An interesting way to compare the different DW-MRI acquisition schemes, diffusion models and tractography algorithms proposed in the literature lies in the exploitation of the large sets of generated tracts to infer atlases of the fiber bundles. Hence, this paper presents a method taking as input the sets of diffusion-based tracts of a population of subjects and producing as output a list of generic fiber bundles that can be detected in most of the population. The usual strategies proposed for the reconstruction of fiber bundles follow two complementary ideas. The first approach is based on regions of interest (ROI) used to select or exclude tracts [1]. The second strategy is based on tract clustering using pairwise similarity measures [2]. This last approach is potentially less intensive in terms of user interaction and can also embed predefined knowledge represented by a bundle template [3]. Furthermore, when applying the clustering after spatial normalization with a set of tracts stemming from several subjects, this strategy can help to discover new reproducible bundles. The fiber clustering approach has

---

\* This project was supported by grants from Région Ile-de-France (France), French Embassy in Chile, CONICYT Chile and University of Concepción (Chile)

been successfully used to map the well-known fiber bundles of deep white matter (DWM). Until now, short fibers of superficial white matter (SWM) have been barely considered. The cartography of fiber bundles of SWM is a complex and unachieved task for the human brain. In a recent paper, Oishi et al. performed a group analysis to study SWM using a voxel-based approach relying on linear brain normalization [4]. They could identify only four U-fiber bundles because of the blurring occurring with such a normalization. In this paper we show that the fiber clustering approach can overcome this weakness.

We propose a two-level strategy chaining intra and inter-subject fiber clustering. The first level can be viewed as a compression procedure reducing a huge set of fibers to a few thousand bundles. This step is developed following a multi-resolution paradigm. A key point is the use of a voxel-based parcellation of white matter, allowing the analysis of any number of fibers. This parcellation produces small fiber subsets that can be split further using additional clustering performed in the space of fiber extremities. The second level is an inter-subject clustering of the resulting fiber bundles. This group analysis relies on a pairwise distance between bundles computed after affine spatial normalization. A simulation is performed to prove that affine normalization is sufficient to create consistent clusters in the bundle space.

## 2 Material and Method

### 2.1 Diffusion and Tractography Datasets

Analysis was performed for twelve subjects of the NMR public database. This database provides high quality T1-weighted images and diffusion data acquired with a GE Healthcare Signa 1.5 Tesla Excite scanner. The diffusion data presents a high angular resolution (HARDI) based on 200 directions and a b-value of  $3000 \text{ s/mm}^2$  (voxel size of  $1.875 \times 1.875 \times 2 \text{ mm}$ ). DW-weighted data were acquired using a twice refocusing spin echo technique compensating Eddy currents to the first order. Geometrical distortions linked to susceptibility artifacts were corrected using a phase map acquisition. T1 and DW-weighted data were automatically realigned using a rigid 3D transform.

The diffusion Orientation Distribution Function was reconstructed in each voxel using a spherical deconvolution of fiber Orientation Distribution Function. It is a spherical deconvolution transform (SDT) reconstructed from q-ball imaging with a constrained regularization [5], using a maximum spherical harmonic (SH) order 8 and a Laplace-Beltrami regularization factor  $\lambda = 0.006$ .

Tracts were reconstructed using a deterministic tractography algorithm [6] provided by BrainVISA public software (<http://brainvisa.info>). Tractography was initiated from two seeds in each voxel of the mask (voxel size of  $0.94 \times 0.94 \times 1.2 \text{ mm}$ ), in both retrograde and anterograde directions, according to the maximal direction of the underlying ODF. Tracking parameters included a maximum curvature angle of  $30^\circ$  and a minimum and maximum fiber length of 20 mm and 200 mm, respectively, leading to a set of about 1.5 millions tracts per subject.

## 2.2 The Two-Level Clustering Method

**Intra-subject clustering.** The intra-subject clustering follows a multiresolution strategy including five steps (see Fig. 1 A):

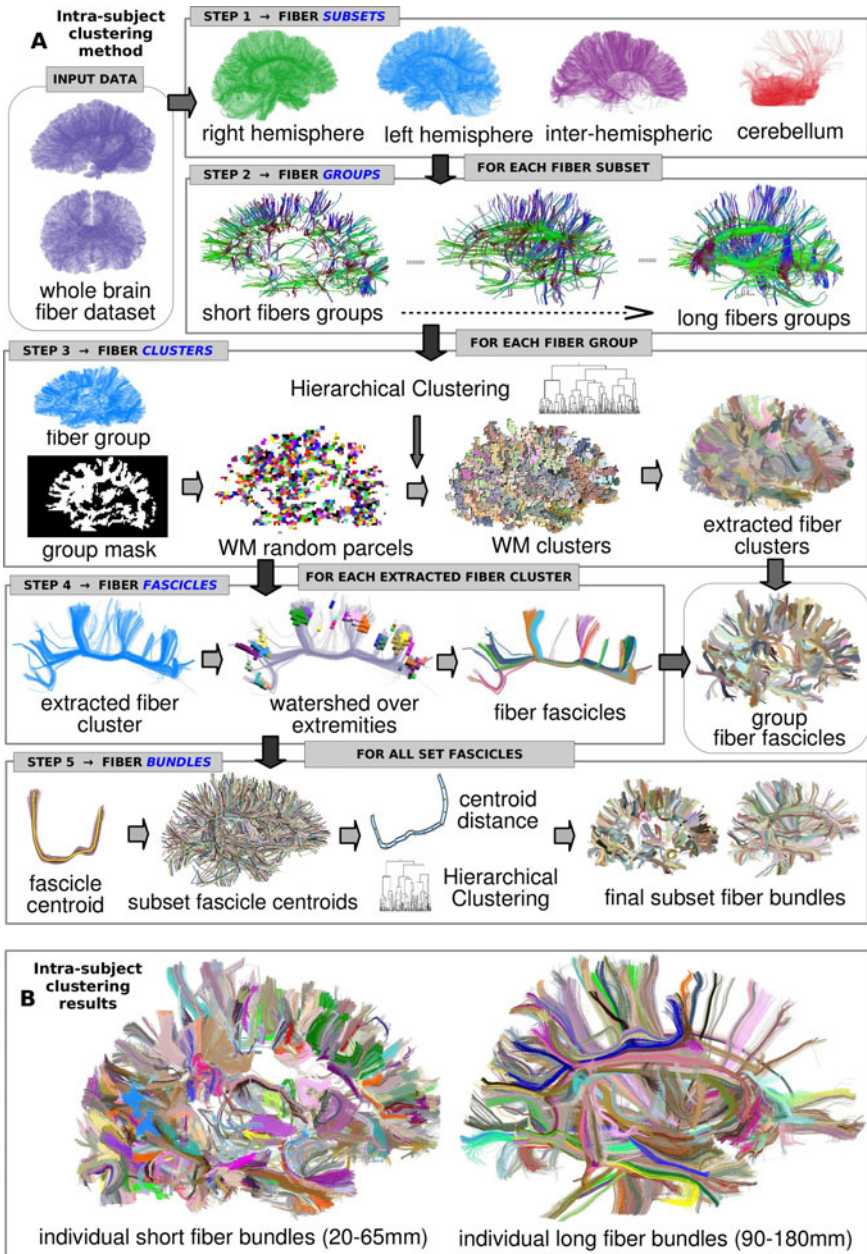
**Step 1: Hierarchical decomposition:** The complete tract set is segmented into four parts, called *fiber subsets*: right hemisphere, left hemisphere, inter-hemispheric and cerebellum tracts. The following steps are applied separately to each subset.

**Step 2: Fiber segmentation based on length:** The subset is split into different *fiber groups*, containing fibers of similar length. While looking insignificant, this second step is of key importance because it partially overcomes a big limitation of the voxel-based approach: the difficulty to separate fiber bundles with different shapes overlapping for a large part of their voxel support.

**Step 3: Voxel-based clustering:** Each fiber group obtained in the preceding step is divided using a connectivity-based parcellation of white matter voxels. A T2 fiber mask is calculated containing voxels crossed by tracts. The mask is randomly parcellated using a geodesic k-means leading to about 12,000 parcels per fiber group. Parcels are clustered using an average-link hierarchical clustering (HC) based on a parcel connectivity measure defined as the number of tracts passing through the pair of parcels. The tree resulting from the hierarchical clustering is analyzed in order to get an adaptive partition where each cluster contains ideally only one putative fiber bundle. The tree analysis discards small isolated clusters and split the large clusters until reaching sizes compatible with the largest actual bundles. Finally, each cluster mask is used to extract corresponding diffusion-based *fiber clusters*.

**Step 4: Extremity-based fiber clusters subdivision:** Fiber clusters are divided into several thin and regular bundles, called *fascicles*, based on the fiber extremities. For each cluster, the list of fiber extremities is converted into a voxel-based density image. This image is segmented by a 3D watershed into maxima-based regions. These regions are used to divide the extracted fiber clusters into several fascicles, each one composed of the fibers whose extremities pass through two particular regions.

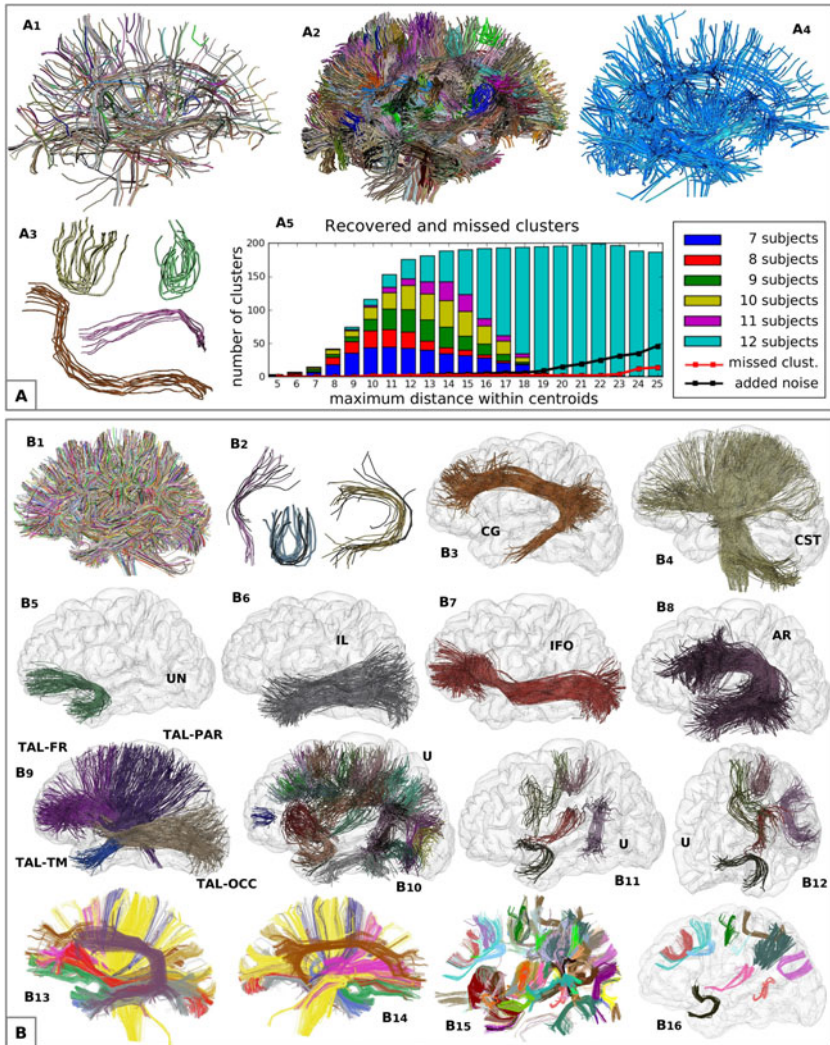
**Step 5: Fiber fascicle centroids clustering:** This step considers all the fascicles from all the fiber length groups of a subset. It consists in a second clustering, aiming to agglomerate fiber fascicles that were over-segmented in the fiber clusters subdivision step (*Step 4*) or in the length-based segmentation step (*Step 2*). For this, a centroid tract, localized in the center of each fascicle, is computed as a representative of the fascicle. It is determined as the tract minimizing a distance to the rest of the fascicle fibers. The distance measure employed is a symmetrized version of the mean closest point distance [7,2,3]. Centroids from all the fascicles of a set are clustered using an average-link hierarchical clustering over a pairwise distance between centroids. The distance is defined as the maximum distance between corresponding points, which is more stringent than the mean closest point distance. This step aims only at regrouping fascicles that have very similar shapes and positions.



**Fig. 1. A: General scheme of the intra-subject clustering: Step 1:** Hierarchical decomposition, **Step 2:** Fiber segmentation based on length, **Step 3:** Voxel-based clustering, **Step 4:** Extremity-based fiber clusters subdivision, **Step 5:** Fiber fascicles centroids clustering. **B: Example results for the intra-subject clustering.** Short and long bundles obtained for the right hemisphere fiber set of one subject.

**Inter-subject clustering.** The second clustering level aims at matching the putative bundles produced by the previous level across the population of subjects. This step is very similar to the clustering performed in section 2.2 (*Step 5*) but this time the calculation considers the bundles obtained from all the subjects in the same fiber subset. A centroid is first calculated for each bundle using the mean of the two mean closest point distances [7,2,3]. Once all the centroids are computed, they are transformed to the Talairach space (TS) using an affine transformation estimated from the T1-weighted image. Then, a bundle centroid affinity graph is computed using the maximum distance between corresponding points, normalized by the bundle length. This restraining measure puts a focus on matching bundles with similar shapes and positions in TS. The affinity graph is used to compute an average-link HC. The resulting tree is analyzed in order to extract only very tight clusters, where the distance between all the fibers within a cluster is inferior to a maximum distance ( $tM_{dcp}$ ). The resulting clusters are discarded if they do not contain more than half of the subjects. A final procedure aims at relaxing the constraints in order to recover some instances of the generic bundles that were missed during the stringent clustering analysis. The goal is to be less demanding on the match between centroids, which is specially important for the subjects that present a deficient normalization in Talairach space. For each non attributed centroid, we compute the distance to each of the centroids of the tight clusters. When the distance to the nearest neighbor is below a threshold, the non attributed centroid is added to the final generic bundle representation. Most of the added centroids belong to long fiber bundles.

**Inter-subject clustering validation:** In order to study the behavior of the inter-subject clustering over a population of subjects aligned with affine registration, we created a simulated dataset of fiber bundles centroids. First, one subject of the NMR database was selected to generate a set of 200 simulated bundle centroids. These bundle centroids were fibers selected from the right hemisphere of this subject with a minimum pairwise distance across the set. The distance used was the maximum distance between corresponding points. The minimum distance was set to 12 mm (see Fig. 2 A<sub>1</sub>). The obtained bundle centroids set was transformed to the space of each one of the eleven remaining subjects of the database, using a non-rigid transform, calculated between T1 images using MedINRIA (<http://www-sop.inria.fr/asclepios/software/MedINRIA>). Hence, we obtained a set of ground truth clusters, each one containing a centroid in each subject (see Fig. 2 A<sub>2</sub>). In addition, 500 fibers from each subject were selected to simulate noise. These fibers were pairwise separated by a minimum distance equal to 11 mm (see. Fig. 2 A<sub>4</sub>). For each subject, we got a fiber dataset of 700 fibers (200 centroids and 500 added noise fibers), leading to a total number of 8400 fibers for the twelve subjects. We applied the inter-subject clustering to the fibers dataset, with the maximum distance within clusters ( $tM_{dcp}$ ) varying from 5 to 25 mm. Resulting clusters were analyzed and compared with the ground truth. First, only clusters containing centroids from a minimum of seven different subjects were selected. Then, a cluster was counted as *recovered* only if all its centroids belonged to the same simulated cluster, otherwise, it was counted



**Fig. 2. A: Inter-subject clustering validation:** **A1:** the original 200 fibers selected as simulated centroids. **A2:** simulated bundles for the 12 subjects in talairach space. **A3:** a selection of bundles from A2. **A4:** noise fibers set (500 fibers) of one subject. **A5:** inter-subject clustering simulation results presenting recovered clusters (color bars), missed clusters (red line) and clusters with added noise (black line). **B: Inter-subject clustering results for twelve subjects:** **B1:** centroids of one subject. **B2:** examples of tight clusters and the added centroids by the nearest neighbor procedure (in black). **B3-B9:** long generic bundles centroids (manually labelled): cingulum (CG), corticospinal tract (CST), uncinate (UN), inf. longitudinal (IL), inf. fronto-occipital (IFO), arcuate (AR), and thalamic radiations (TAL-(FR,PAR,TM,OC)). **B10-B12:** selections of generic short bundles centroids. **B13-B16:** generic bundles for one subject. B15 shows all the U-fibers while B16 contains a selection of U-fibers.

as a *missed* cluster. Fig. 2 A<sub>5</sub> presents the simulation results as a function of the distance  $tM_{dcp}$ . From the analysis, we note, as expected, that the number of recovered clusters (color bars) increases with  $tM_{dcp}$ , as the number of subjects in the clusters (indicated by different colors, from 7 to 12). For distances  $tM_{dcp}$  superior to 11 mm, a large number of clusters was recovered, but for distances between 11 and 15 mm, most of the clusters miss some centroids. This behavior is accepted by the method, which adds a cluster to the model as soon as it includes more than half of the subjects. The red line indicates the number of missed clusters, which is very low. These are most of the time fused with other clusters. The black line shows the number of recovered clusters that contain also added noise fibers. Finally, a large number of clusters made up of only noise fibers was found but discarded by the method because none of these clusters had fibers from more than six different subjects.

### 3 Results

Individual fiber bundles were extracted for twelve subjects. The results obtained for one subject (right hemisphere) are presented in Fig. 1 B. Inter-subject clustering of the left hemisphere is presented in Fig. 2 B. Ten DWM generic fiber bundles were identified in all the subjects. Some of them are split into several generic fascicles: Inferior fronto-occipital fasciculus, Cortico-spinal tract, Arquate fasciculus, Uncinate fasciculus, Inferior Longitudinal fasciculus, Cingulum, and Frontal, Temporal, Parietal and Occipital Thalamic Radiations. Twenty generic U-Fiber bundles (SWM) occurring in at least seven subjects were detected in the left hemisphere model.

### 4 Discussion and Conclusion

As for any fiber tracts analysis method, our results depend strongly on the quality of the tractography results. Our method can not detect bundles that are not tracked in individuals. Also, spurious bundles can not be differentiated from real bundles if they are reproducible across subjects. Besides, anomalous final bundles can be found due to errors in the propagation mask. Since this mask defines where fibers are tracked, bundles can be erroneously cutted or fused. Nevertheless, independently of the tracking results, our method is a powerfull tool to extract the main bundles that constitute the dataset.

Our method is able to analyze huge fiber datasets and infer a model of the generic bundles present in a population. The first level, composed by an intra-subject clustering, can be seen as a compression of information and a filtering, where bundles representing the individual whole white matter structure are identified. The second level, an inter-subject clustering, deals with a reasonable number of bundle centroids from a population of subjects and is capable to extract generic bundles present in most of the subjects. Long known bundles were identified, but the result of major significance is the capability to identify generic short association bundles. Hence, our approach will scale up easily to the 1 mm

spatial resolution that can now be achieved with highly parallel imaging or very high fields. This spatial resolution is bound to highlight a myriad of U-fiber bundles and better delineate other bigger bundles crossing.

The general idea of hierarchical decomposition underlying our method will be pushed further with the use of a segmentation of the deep grey matter structures to improve the clustering of their connections with the cortex. The corresponding tracts will be filtered out and clustered independently.

We have shown that the affine registration to standard space is sufficient to align reasonably the deep tracts across all the subjects. Each U-fiber bundle inferred in this paper did require a reasonable alignment of the bundles of only half of the subjects, which happens in the most stable brain regions. However, increasing the number of generic U-fiber bundles, will require an improvement of the spatial normalisation used to compare bundles across subject. Further work will lead us to improve iteratively the spatial normalization using the inferred bundles as constraints in order to better align other bundles. Moreover, we will compare our results with the strategy computing bundles after computing an average atlas of diffusion data based on high quality diffeomorphic spatial normalization. Deciding which strategy is the best is one of the goals of our research.

## References

1. Catani, M., Thiebaut de Schotten, M.: A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* 44(8), 1105–1132 (2008)
2. O'Donnell, L.J., Kubicki, M., Shenton, M.E., Dreusicke, M.H., Grimson, W.E.L., Westin, C.F.: A method for clustering white matter fiber tracts. *AJNR Am. J. Neuroradiol.* 27(5), 1032–1036 (2006)
3. O'Donnell, L., Westin, C.F.: Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Transactions on Medical Imaging* 26(11), 1562–1575 (2007)
4. Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A.W., Pike, G.B., Rosa-Neto, P., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P.C.M., Mazziotta, J., Mori, S.: Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. *Neuroimage* (July 2008)
5. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* 35(4), 1459–1472 (2007)
6. Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Rivière, D., Constantinesco, A., Le Bihan, D., Mangin, J.F.: Fiber tracking in q-ball fields using regularized particle trajectories. *Inf. Process Med. Imaging* 19, 52–63 (2005)
7. Corouge, I., Gouttard, S., Gerig, G.: Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In: *ISBI 2004* (2004)