Segmentation of Short Association Bundles in Massive Tractography Datasets Using a Multi-subject Bundle Atlas

Pamela Guevara[1,2,3], Delphine Duclap[1,2], Cyril Poupon[1,2],
Linda Marrakchi-Kacem[1,2], Josselin Houenou[1,2,4],
Marion Leboyer[4], and Jean-François Mangin[1,2]

[1] Neurospin, CEA, Gif-sur-Yvette, France
[2] Institut Fédératif de Recherche 49, Gif-sur-Yvette, France
[3] University of Concepción, Concepción, Chile
[4] AP-HP, University Paris-East, Department of Psychiatry, INSERM, U955 Unit

Abstract. This paper presents a method for automatic segmentation of some short association fiber bundles from massive dMRI tractography datasets. The method is based on a multi-subject bundle atlas derived from a two-level intra-subject and inter-subject clustering strategy. Each atlas bundle corresponds to one or more inter-subject clusters, presenting similar shapes. An atlas bundle is represented by the multi-subject list of the centroids of all intra-subject clusters in order to get a good sampling of the shape and localization variability. An atlas of 47 bundles is inferred from a first database of 12 brains, and used to segment the same bundles in a second database of 10 brains.

1 Introduction

Diffusion MRI allows noninvasive study of brain white matter (WM) structure through the measurement of the restricted diffusion of water. The fiber orientation can be inferred from this data and fiber bundles can be reconstructed using tractography algorithms [1]. Until now, several WM bundle atlases have been proposed [2,3] for the bundles belonging to deep white matter (DWM). However, short fibers of superficial white matter (SWM) have been barely considered, probably because these are more variable across sujects. Furthermore, the partial volume effect in subcortical regions prevents accurate delineation of small fiber bundles. The continuous improvement of DW-MRI acquisition schemes, diffusion models and tractography algorithms leads to increasingly complex and large tractography datasets, with known DWM tracts composed by various fiber fascicles of different shapes and lengths, and a big amount of short SWM association bundles. This improvement allows deeper analyses of WM bundles, but, at the same time, increases the requirements of tractography datasets analysis and segmentation techniques. The segmentation of human brain WM fiber bundles is therefore a complex and not completely solved problem. In particular, the cartography of fiber bundles of SWM is still an unachieved task. In [4], the authors

performed a group analysis to study SWM using a voxel-based approach relying on linear brain normalization. They could identify only four U-fiber bundles because of the blurring occurring with such a normalization. Most recently, this method was improved using non-linear normalization, and was able to detect 29 short association bundles in 20 subjects [5]. These results are very interesting but as a ROI (region of interest) based approach was used, there is no guarantee that the fibers present the same shape across subjects.

The usual strategies proposed for the segmentation of fiber bundles follow two complementary ideas. The first approach is based on ROIs used to select or exclude tracts [3,5]. The second strategy is based on tract clustering using pairwise similarity measures [6]. This last approach requires less interaction than manual approaches and integrates fiber shape and position information in the analysis, which is not the case of most ROI-based segmentation approaches. It can also embed a priori knowledge represented by a bundle template [7]. However, the clustering-based methods commonly present a limitation on the number of fibers that can be analyzed. In spite of two recent works that describe the analysis of huge datasets (120,000 [8] and 480,000 fibers [9]), the segmentation of huge tractography datasets, presenting more than one million tracts, is still a challenge.

Hence, this paper presents a method for the segmentation of SWM fiber bundles from massive tractography datasets using a priori information embedded in a multi-subject (MS) fiber bundle atlas. The method builds upon a multiresolution intra-subject clustering that can compress millions of tracts into a few thousand consistent bundles, described in [10]. A second level of clustering is performed across subjects in order to infer a list of generic bundles with consistent shape and localization in a normalized space [11]. The most reproducible inter-subject (IS) clusters computed from a database of 12 brains were manually labeled to build the atlas. This MS strategy, embedding the shape and localization variability, has been shown recently to be more efficient than the usual single template approach for brain structure recognition because of weaknesses of the spatial normalization paradigm [12]. New tractography datasets are first compressed with the same intra-subject clustering. The resulting clusters are then labeled using pairwise distances to the centroids representing the MS atlas bundles. To the best of our knowledge, this is the first SWM clustering-based segmentation method.

2   Material and Method

2.1   Diffusion and Tractography Datasets

The atlas was constructed from 12 subjects of a High Angular Resolution Diffusion Imaging (HARDI) adult database (DB1). This database provides high quality T1-weighted images and diffusion-weighted (DW) data acquired with a GE Healthcare Signa 1.5 T Excite scanner. The diffusion data presents a high angular resolution based on 200 directions and a b-value of $3000 \, \text{s/mm}^2$ (voxel size of $1.875 \times 1.875 \times 2 \, \text{mm}$).

Ten subjects of another adult HARDI database (DB2), were used to test the segmentation method. This database provides high quality T1-weighted images and DW data acquired with a Siemens 3.0 T Tim Trio system. The DW data is based on 41 directions and a b-value of $1000 \, s/mm^2$ (voxel size of $2 \, x \, 2 \, x \, 2 \, mm$).
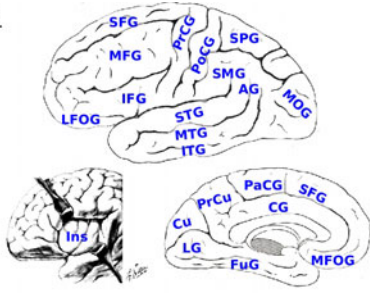
DW data were acquired using a twice refocusing spin echo technique compensating Eddy currents to the first order. Geometrical distortions linked to susceptibility artifacts were corrected using a phase map acquisition. T1 and DW data were automatically realigned using a rigid 3D transform. The diffusion Orientation Distribution Function (ODF) was reconstructed in each voxel. For subjects from DB1, a spherical deconvolution (SD) of the fiber ODF was used. It is a SD transform reconstructed from q-ball imaging with a constrained regularization [13], using a maximum spherical harmonic order $SH_{max} = 8$ and a Laplace-Beltrami regularization factor $\lambda_{LB} = 0.006$. For subjects from DB2, an analytical solution of the q-ball model was determined [14], using a $SH_{max} = 6$ and a $\lambda_{LB} = 0.006$. Whole-brain tractography was performed using an improved tractography propagation mask (using T1 data rather than FA) and a regularized deterministic tractography algorithm. Tractography was initiated from two seeds in each voxel of the mask (with T1 resolution), in both retrograde and anterograde directions, according to the maximal direction of the underlying ODF. Tracking parameters included a maximum curvature angle of 30° and a minimum and maximum fiber length of 20 mm and 250 mm, respectively, leading to a set of about 1.5 millions tracts per subject.

2.2   HARDI Multi-subject Fiber Bundle Atlas

The two-level clustering was performed using the method described in [11] applied on database DB1, with some improvements. First, intra-subject clustering [10] was applied to each dataset. This **intra-subject clustering** reduces the tractography dataset information from more than one million of tracts to a few thousand fiber bundles. The obtained bundles are thin and regular fiber fascicles composed by fibers presenting similar length and shape. In addition, during the analysis most of noise fibers are discarded, leading to a cleaner fiber dataset. Due to its regular shape, each resulting fiber bundle can be represented by a single fiber, called a bundle centroid. This compressed representation of a tractography dataset allows the application of further processing steps that could not be applied to the whole fiber dataset.

The second clustering level aimed at matching the putative bundles produced by the previous level across the population of subjects. In this **inter-subject clustering**, fiber centroids from all the subjects were aligned by an affine transformation to the Talairach space (TS), estimated from the T1-weighted image. Then, the centroids were clustered using pairwise distance measures [15] in order to match bundles with similar shapes and positions in TS. In order to get population representative clusters, only clusters composed by centroids from at least half of the subjects were selected. The final addition of closest centroids described in [11] was not performed with the aim of keeping very tight clusters.

Region	Abbrev.	Region	Abbrev.
Superior frontal gyrus	SFG	Superior temporal gyrus	STG
Middle frontal gyrus	MFG	Middle temporal gyrus	MTG
Inferior frontal gyrus	IFG	Inferior temporal gyrus	ITG
Medial fronto-orbital gyrus	MFOG	Cuneus	Cu
Lateral fronto-orbital gyrus	LFOG	Pre-cuneus	PrCu
Precentral gyrus	PrCG	Cingulate gyrus	CG
Postcentral gyrus	PoCG	Paracentral gyrus	PaCG
Supramarginal gyrus	SMG	Fusiform gyrus	FuG
Angular gyrus	AG	Lingual gyrus	LG
Superior parietal gyrus	SPG	Insular	Ins
Middle occipital gyrus	MOG		

Fig. 1.   Cortical surface anatomical regions used to identify the atlas short association bundles. Images where adapted from *http://www.bartleby.com/107/* and *http://www.netterimages.com/.*

The inter-subject clusters belonging to SWM were manually labeled by an expert using gyral parcellation of cortical surfaces, in order to give an anatomical name to each reproducible bundle (see Fig. 1). Each atlas bundle is then represented by the complete set of individual centroids belonging to the underlying intra-subject clusters. A last visual inspection led to discard a few artefactual centroids clearly including spurious parts like loops. The resulting multi-subject representation provides a good sampling of the inter-subject variability of the bundle trajectory after affine normalization. The atlas inference was done for the bundles of the left hemisphere (LH), with a length between 35 and 110 mm. The bundles of the right hemisphere (RH) were obtained using the symmetric of those of the LH with respect to Talairach inter-hemispheric plane. The goal is to get a symmetric atlas for the validation described in this paper. Ongoing work aims at performing the same inference for the RH in order to remove any bias. The current atlas includes a total of 47 SWM bundles; see details in Fig. 2.

2.3   WM Tracts Segmentation

The segmentation of a new tractography dataset begins with a compression into a few thousand bundles equivalent to the compression used during the atlas inference, described in [10]. Then, the resulting bundles are labeled using a supervised classification based on the fiber bundle atlas. The bundle centroids are normalized to the TS using an affine transformation. Then pairwise distances are computed between each centroid of the new subject and all the centroids of the atlas. The distance measure used is the maximum of the Euclidean distances between corresponding points (dM), defined for two fibers A and B, described by N_p points, as

$$d_M(A,B) = \min\left(\max_i \| \mathbf{a}_i - \mathbf{b}_i \|, \max_i \| \mathbf{a}_i - \mathbf{b}_{N_p-i} \|\right), \tag{1}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the position of the points of fibers A and B respectively, for $i = 0..N_p-1$. This distance is a good representation of the similarity between two
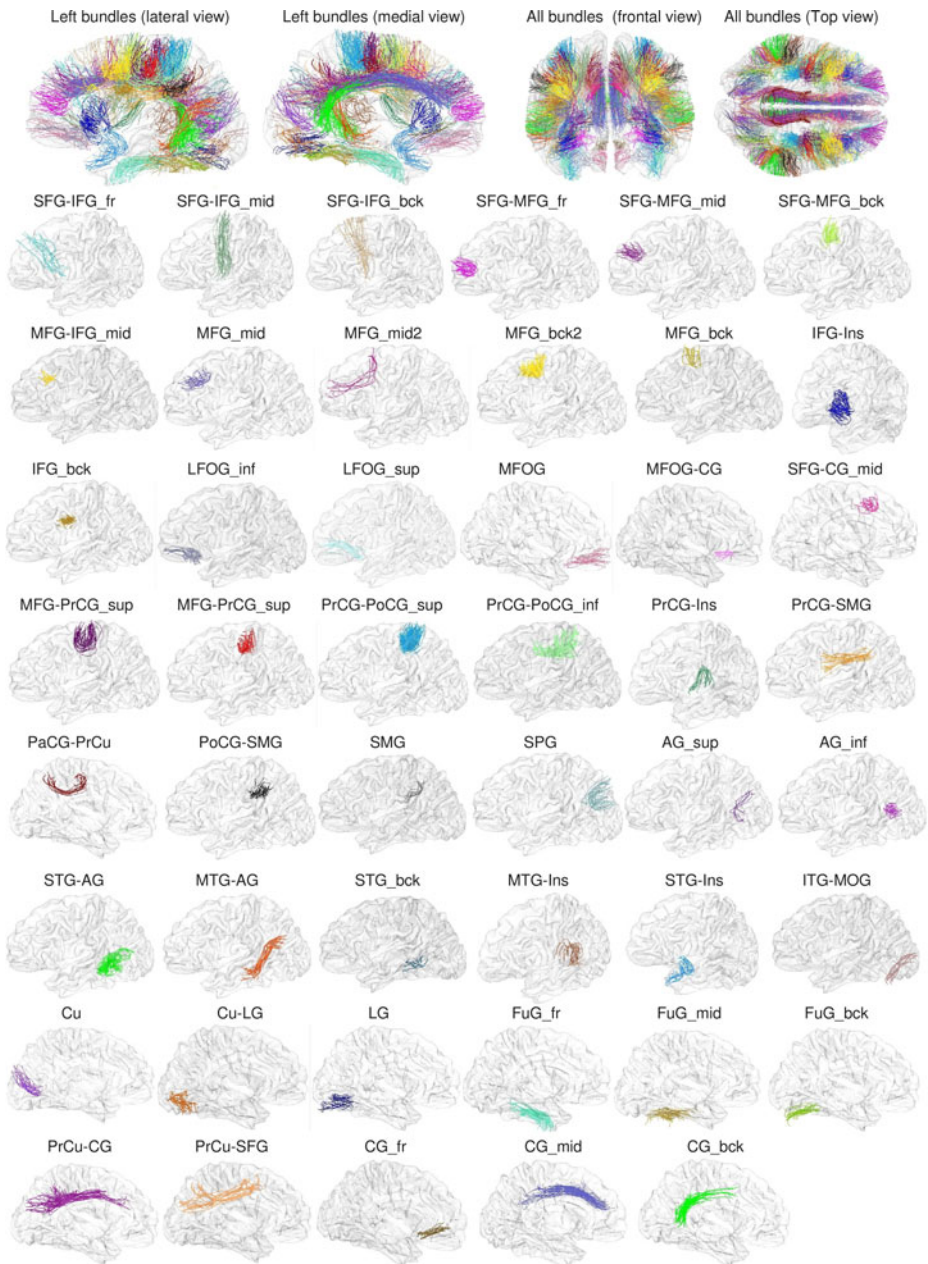
Fig. 2. Short association bundle atlas (47 bundles per hemisphere). The first row shows all the atlas bundles. The remaining rows show each bundle in a separated figure. Bundle names were assigned in function of the regions that the bundles connect, following the names illustrated in Fig. 1. In some cases, an additional spatial specification was used: fr (frontal), mid (middle), bck (back), sup (superior), inf (inferior).
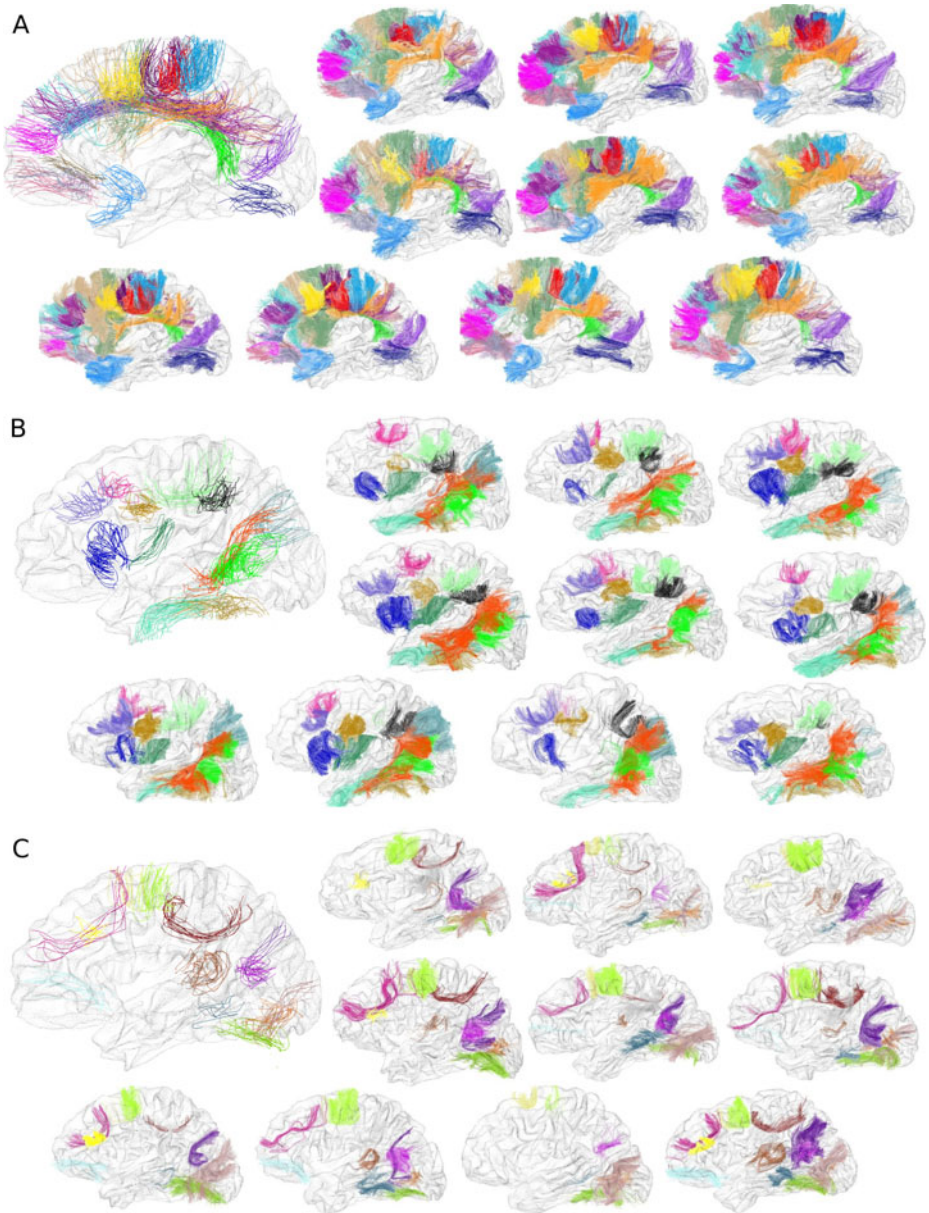
Fig. 3. Automatic fiber bundle segmentation results. Only left hemisphere bundles are shown. Colors are the same as for the bundle atlas (Fig. 2). The bundles were divided into three groups, in function of their reproducibility. Atlas bundles are shown in the upper-left corner of each image. **A**: Fiber bundles found in all the subjects (21 bundles). **B**: Fiber bundles found in 9 of the 10 subjects (12 bundles). **C**: Fiber bundles found in 5 to 8 subjects (14 bundles).

fibers, as it takes into account the fiber positions and shapes. It is more restrictive than distances based on the closest points [15,6]. For the calculation, the atlas fibers and the individual centroids are resampled using 21 equally distributed points. The whole set of pairwise distances is obtained in a few minutes.

Each individual centroid is labeled by the closest atlas bundle, provided that the distance to this bundle, namely the smallest pairwise distance to the centroids representing this bundle, is lower than a threshold. This threshold was empirically adapted to each atlas bundle (between 8–14 mm) taking into account the bundle mean fiber length and the proximity to other atlas bundles, leading to higher thresholds for long and isolated bundles. A leave-one-out strategy for the determination of the thresholds could be implemented in the future.

3   Results

A general problem for evaluating WM bundle segmentation is the lack of gold standard. This is even more complex for SWM, which cartography is still largely unknown and to the best of our knowledge, no atlas describing the shape of these bundles has been proposed. We evaluate our approach using a second database (DB2). The results for the ten subjects are presented in Fig. 3. All the bundles were found in at least half of the subjects, which is consistent with our atlas construction requirements. Twenty-one bundles were found in all the subjects. Twelve bundles were found in nine subjects and fourteen bundles were found in between five to eight subjects. The segmentations were validated by the expert who defined the atlas. To get an insight of the quality of the results, the bundles were visually compared with those obtained using larger distance thresholds. It was found that the chosen thresholds were close to optimal for most of the bundles. Long and isolated bundles were in general well segmented, when these existed, but some classifications errors were found in short bundles localized very close to other atlas bundles.

4   Discussion and Conclusion

The proposed method shows that it is possible to segment the most reproducible SWM bundles using a clustering-based approach in a population of subjects. The use of a multi-subject representation of bundles and shape information could lead to cleaner bundles than when using a ROI-based strategy, which may improve the sensitivity of morphometric studies. Furthermore, this new atlas and the possibility to manipulate massive tractography datasets allow finer decompositions of the bundles, for instance, we proposed two subdivisions of the bundle connecting the pre- and post-central gyri. Our atlas is bound to be refined with more of such subdivisions in the near future.

However, the proposed method is far from be perfect. This is due in part to the high inter-subject variability of short association SWM bundles and the current limitations of dMRI techniques. Our results depend strongly on the quality of the

tractography results: bundles that are not tracked in individuals can not be segmented, a problem that particularly affects SWM due to the partial volume effect.

But an important improvement will be obtained by the use of non-linear normalization [16]. First, the atlas construction will be performed using this kind of normalization, leading to a better multi-subject representation of the variability of the atlas bundles. Furthermore, the recognition of the bundles should be also improved if non-linear normalization is used between the subjects and the atlas, reducing the classification errors produced in bundles presenting very similar shapes and close positions.

References

1. Mori, S., van Zijl, P.C.M.: Fiber tracking: principles and strategies - a technical review. NMR Biomed. 15(7-8), 468–480 (2002)
2. Mori, S., Wakana, S., van Zijl, P.C.M., Nagae-Poetscher, L.M.: MRI Atlas of Human White Matter, 1st edn. Elsevier Science (August 2005)
3. Catani, M., Thiebaut de Schotten, M.: A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 44(8), 1105–1132 (2008)
4. Oishi, K., Zilles, K., Amunts, K., et al.: Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. Neuroimage 43(3), 447–457 (2008)
5. Zhang, Y., Zhang, J., Oishi, K., et al.: Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. NeuroImage 52(4), 1289–1301 (2010)
6. O'Donnell, L.J., Kubicki, M., Shenton, M.E., et al.: A method for clustering white matter fiber tracts. AJNR 27(5), 1032–1036 (2006)
7. 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)
8. Wang, X., Grimson, W.E.L., Westin, C.F.: Tractography segmentation using a hierarchical dirichlet processes mixture model. Neuroimage 54(1), 290–302 (2011)
9. Visser, E., Nijhuis, E.H.J., Buitelaar, J.K., Zwiers, M.P.: Partition-based mass clustering of tractography streamlines. Neuroimage 54(1), 303–312 (2011)
10. Guevara, P., Poupon, C., Rivière, D., et al.: Robust clustering of massive tractography datasets. NeuroImage 54(3), 1975–1993 (2011)
11. Guevara, P., Poupon, C., Rivière, D., et al.: Inference of a HARDI Fiber Bundle Atlas Using a Two-Level Clustering Strategy. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 550–557. Springer, Heidelberg (2010)
12. Lyu, I., Seong, J.K., Shin, S.Y., et al.: Spectral-based automatic labeling and refining of human cortical sulcal curves using expert-provided examples. Neuroimage 52(1), 142–157 (2010)
13. 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)
14. Descoteaux, M.E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical q-ball imaging. Magn. Reson. Med. 58, 497–510 (2007)
15. Corouge, I., Gouttard, S., Gerig, G.: Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In: ISBI 2004 (2004)
16. Auzias, G., Colliot, O., Glaunes, et al: Diffeomorphic brain registration under exhaustive sulcal constraints. IEEE Transactions on Medical Imaging 99 (2011)