2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Group-Wise Cortical Surface Parcellation Based on Inter-Subject Fiber Clustering*

Christopher Vergara[1], Felipe Silva[1], Isaías Huerta[1], Narciso López-López[1], Andrea Vázquez[1], Josselin Houenou[2], Cyril Poupon[2], Jean-François Mangin[2], Cecilia Hernández[1,3] and Pamela Guevara[1]

***Abstract*— We present an automatic algorithm for the group-wise parcellation of the cortical surface. The method is based on the structural connectivity obtained from representative brain fiber clusters, calculated via an inter-subject clustering scheme. Preliminary regions were defined from cluster-cortical mesh intersection points. The final parcellation was obtained using parcel probability maps to model and integrate the connectivity information of all subjects, and graphs to represent the overlap between parcels. Two inter-subject clustering schemes were tested, generating a total of 171 and 109 parcels, respectively. The resulting parcels were quantitatively compared with three state-of-the-art atlases. The best parcellation returned 69 parcels with a Dice similarity coefficient greater than 0.5. To the best of our knowledge, this is the first diffusion-based cortex parcellation method based on whole-brain inter-subject fiber clustering.**

I. INTRODUCTION

The study of the human connectome, the network of connections between different brain regions, has become a key research area over the last decade. This is due to its role in neuroscience of furthering the comprehension of brain function and organization [1]. It also has application to the study of neurological disorders such as Alzheimer's or Parkinson's disease, among others [2]. The development of Diffusion-weighted Magnetic Resonance Imaging (dMRI) techniques has enabled the acquisition of higher quality data, which results in a more accurate description of brain structure at the macroscopic level [3]. Diffusion MRI tractography is able to obtain a representation of the main white matter (WM) tracts in the brain in a non-invasive and in-vivo way. When applied to the whole brain, it results in a dataset containing a large number of 3D curves called fibers [4], which connect different gray matter (GM) regions. Then, the connected areas and their connectivity patterns can be used to develop Connectivity-Based Parcellations (CBP) methods, with the goal of defining GM sub-regions with similar intra-group connectivity profiles, while maximizing the difference between groups [5].

However, high dimensionality, noise, and inter-subject variability of the data make group-wise parcellation a complicated task. Several strategies have been developed to solve this problem, where all of them apply some sort of dimensionality reduction scheme to address the high inter-subject variability. Lefranc et al. [6] implemented a group-wise CBP method based on dMRI, where connectivity profiles were averaged with a Gaussian kernel, after which a watershed algorithm was used to split the cortical surface. The final parcellation was obtained using the k-medoids algorithm on the joint profiles. Silva et al. [7] developed a method for individual cortex parcellation, based on segmented fiber bundles using a superficial white matter (SWM) bundle atlas. The regions were defined based on the connectivity of each bundle, given by the intersecting regions (or parcels) in the cortex. The overlap between parcels was solved using a graph representation. This method has the advantage of having a direct correspondence between subjects, but its individual strategy cannot be easily extended to a groups of subjects due to the high variability of the parcels among subjects. López-López et al. [8] extended this method to a group-wise framework, where a diffusion-based subdivision of anatomical regions is performed. Further improvements include the selection of superficial and deep WM bundles from two additional atlases, filtering of irregular parcels, and the processing of intersected regions across all the subjects based on the probability of connection for each parcel. A disadvantage of our previous methods [7], [8], is that they do not consider inter-hemispheric connections.

Furthermore, the use of an anatomical parcellation and segmented bundles, as in [8] are strong constraints. Hence a group-wise parcellation, based on the main connections for a population, and for the whole brain would be suitable. To achieve this goal, we propose a method that relies on inter-subject clustering for the dimension reduction, without anatomical prior information.

II. MATERIALS AND METHODS

A. Database and tractography datasets

79 healthy subjects from the HARDI ARCHI database [9] were used. Special acquisition sequences were obtained on a 3T MRI scanner with a 12-channel head coil (Siemens, Erlangen). The MRI protocol included the acquisition of T1-weighted images using a MPRAGE sequence (160 slices, matrix=256×240, voxel size=$1\times1\times1.1mm$), a B_0 field map, and a single shell HARDI SS-EPI dataset along 60 optimized DW directions, b=1500 s/mm^2 (70 slices, matrix=128×128, voxel size=$1.71875\times1.71875\times1.7mm$). The database also includes affine transformation matrices to convert data between T1, T2 (diffusion) and Talairach spaces.

*This work has received funding by the ANID FONDECYT 1190701, ANID PIA/Anillo de Investigación en Ciencia y Tecnología ACT172121, ANID-Basal Project FB0008 (AC3E), and ANID-Basal Project FB0001 (CeBiB).

[1] Faculty of Engineering, Universidad de Concepción, Concepción, Chile, e-mail: pguevara@udec.cl

[2] I2BM, Neurospin, CEA, Gif-sur-Yvette, France

[3] Center for Biotechnology and Bioengineering (CeBiB), Chile

The data were pre-processed using BrainVISA / Connectomist 2.0 software[1]. The main sources of artifacts were corrected and defective slices were discarded. The analytical Q-ball model [10] was computed to obtain ODF fields in each voxel. Finally, a whole-brain deterministic streamline tractography was performed, using a T1-based propagation mask with 1 seed per voxel at T1 resolution, a maximum curvature angle of 30°, and a forward step of 0.2 mm. In average, the resulting datasets contain around 1 million fibers per subject. Cortical meshes were also calculated from T1 images using Freesurfer[2] and BrainVISA, for a total of 81,920 triangles per subject.

B. Fiber clustering

First, an intra-subject fiber clustering scheme is applied to the whole-brain tractography dataset. The objective is to obtain compact clusters composed of fibers with similar trajectories and shapes, representing the main connections of the brain.

Step 1: Pre-processing. First, the tractography is resampled to 21 equidistant points per fiber, which is enough to represent the topology of the curves accurately. Next, all fibers with a length less than 40 mm are discarded, as they are considered too short and noisy.

Step 2: Intra-subject clustering. A fast intra-subject clustering algorithm called FFClust [11] is applied to the fibers of each subject, and clusters with less than 15 fibers are discarded, leading to an average of 9,733 clusters per subject. As the resulting clusters are composed of tightly-packed fibers, the trajectory of each individual group can be represented by its centroid. Then, the centroids of every subject are transformed from T2 to Talairach space, enabling an inter-subject analysis.

Step 3: Inter-subject clustering. With all the centroids gathered in a common space, an inter-subject clustering method is applied to them. The first method is QuickBundles (QB) [12], which is based on a distance threshold defined by the minimum average direct-flip distance (MDF) between fiber centroids. This value is set at 20 mm, in line with previous WM inter-subject studies [13], [14]. The second method corresponds to a density-based clustering scheme [15], which uses an autoencoder with LSTM layers to compress the fibers into a latent space (in this case using a size of 11 floats). After that, a modified version of the HDBSCAN clustering algorithm [16] is applied, using default parameters (minimum samples=1 and minimum cluster size=70). The neural network is trained using the pre-processed tractography data from all 79 subjects, but only the centroids are clustered. Note that neither of these inter-subject clustering schemes uses the number of clusters as an input parameter, but other parameters, such as MDF distance and density, to determine them.

In both cases, with the goal of maintaining high representativeness, only clusters formed by centroids coming from 70 different subjects or more are selected. Then, the fibers associated to the intra-subject centroids that make up the selected clusters are aligned with respect to the inter-subject cluster centroid using the MDF distance. Furthermore, as most inter-subject clusters contain several centroids from the same subject, the fibers corresponding to those centroids are fused, and the centroid of the fused cluster is calculated as the pointwise mean of the fused fibers. This effectively groups the fibers of each subject, which are labeled according to the inter-subject clusters, obtained using the data from the whole population. Hence, the obtained clusters have correspondence between subjects. Finally, the clusters of every subject are transformed from Talairach to T1 space.

C. Fiber-mesh intersection

In this stage, the intersection between the cortical mesh and the fiber clusters is calculated separately for each subject in T1 space, using a modified version of the algorithm in [7], enabling the computation of inter-hemispheric connections, in addition to the intra-hemispheric ones. First, since checking every triangle for an intersection is inefficient, the 3D space is subdivided into cubic cells of side $\approx 1.5mm$, which is a good trade-off between computational cost and number of intersections found. Then, each fiber is projected two points forwards and one point backward at each endpoint, thus considering a cell and its neighborhood in the calculation. Finally, the ray-triangle intersection at both fiber ends is calculated by the Möller-Trumbore algorithm [17]. Every cluster-cortical mesh intersection is obtained for each of the 79 subjects.

D. Cortical parcellation

To integrate the connectivity information of all the subjects, we use the correspondence that exists between the cortical mesh of different subjects, i.e., an intersection found in an arbitrary triangle T_i represents an intersection in roughly the same anatomical location for every subject. Then, using the index of the triangle and the label of the cluster that intersects it, enables the processing of individual intersections at a population level. Several processing are similar to those proposed in [8], but had to be reimplemented to deal with the whole cortical meshes and inter-hemispheric connections.

Step 1: Preliminary parcels. Since every cluster considered in this step intersects the cortical mesh at both endpoints, each one of them defines two preliminary regions or parcels, one for each end. These parcels are named according to the label of the cluster in question, followed by an **A** and a **B** suffix for each extremity, respectively.

Step 2: Density and probability maps. As whole-brain connections are used, most cortical mesh triangles are intersected by fibers from different clusters, and multiple fibers belonging to the same cluster. Hence, each triangle T_i can represent different parcels with a defined probability for each one. First, for each parcel, a density map is calculated, based on the count of intersections that occur in each triangle.

[1] http://www.brainvisa.info/

[2] https://surfer.nmr.mgh.harvard.edu/

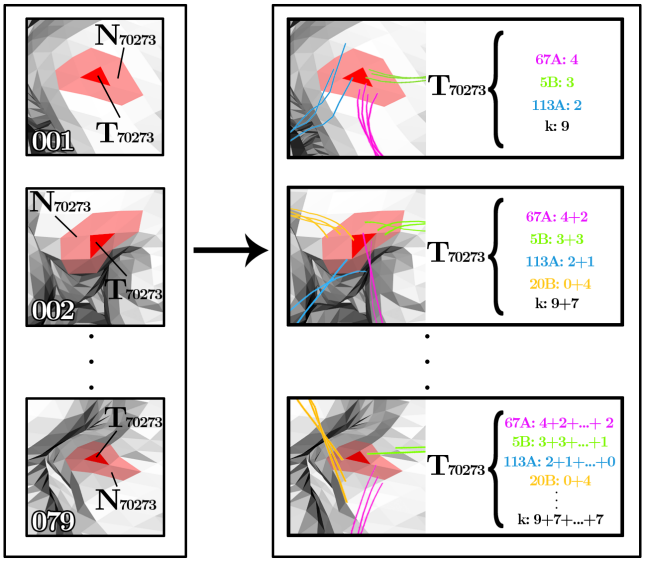

Fig. 1. Creation of density maps based on cluster-mesh intersections across all 79 subjects.

This is performed by the calculation of the probability of connection for the parcels in each triangle. Thus, for each triangle T_i of the cortical mesh:

1) A global counter k_i is created, which registers how many intersections occur in T_i.
2) The neighborhood N_i is determined, corresponding to all the triangles that share either a vertex or a side with T_i.
3) All the clusters that intersect with either T_i or N_i are obtained across all 79 subjects. Then, for each cluster found:
   a) A label with the name of the cluster C_j plus the suffix (**A** or **B**, depending on the extremity) is created. This label, corresponding to a preliminary parcel, is added to a list associated with T_i.
   b) A counter $C_{j,i}$ (plus **A** or **B**) and k_i are increased in an amount equal to the number of times the cluster intersects with T_i or N_i.

This process is illustrated in Fig. 1, using triangle T_{70273} as an example. Then, in order to eliminate noisy regions, parcels that are considered too small are discarded. Since the cortical area is distributed reasonably evenly between each triangle, and considering that a reasonable parcellation has at most 500 parcels per hemisphere, a parcel is considered noisy and is discarded when its size is less or equal to half of the area covered by a single parcel in the extreme case described, i.e., less than 81920/(2·500) $\approx$ 82 triangles. Finally, the probability of a parcel in a triangle T_i is defined as the counter associated to T_i divided by the global counter k_i of T_i. Thus, for each triangle, every parcel has a probability value between 0, when the parcel does not appear in the triangle, and 1, when the parcel is the only one present in the triangle.

Step 3: Parcel overlap. In most cases, multiple probabilistic parcels overlap. Then, if their intersection is significant, it is useful to fuse and assign them a common label. However, as these parcels generally have a non-uniform probability distribution, the overlap may occur in zones of different densities. To address this, density centers are calculated for each parcel based on a probability threshold, heuristically set at 0.20. In general, the overlap between two parcels can be divided into three cases, as shown in Fig. 2, where the density center boundary is marked as a dashed line. In the first case, the parcels do not intersect, and so the overlap is null. In the second case, the parcels intersect; however, the intersection occurs at zones of low density. Finally, the third scenario occurs when the parcels intersect at points belonging to the density center of each parcel. The latter case is considered potentially significant, were two bundles are partially connecting the same region, and the level of significance is calculated using Equation 1:

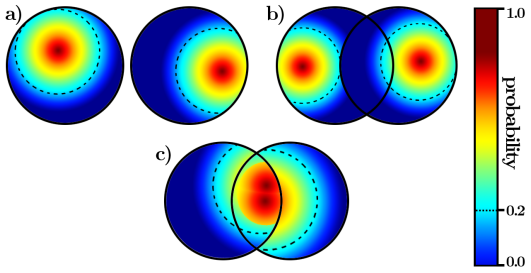

Fig. 2. Different types of intersection between parcels: **(a)** No overlap, **(b)** Non-significant overlap, **(c)** Potentially significant overlap.

$$\Phi(A,B) = \frac{|A \cap B|}{min(|A|,|B|)} \tag{1}$$

To fuse overlapped parcels, an undirected graph is created, where each vertex corresponds to one parcel, and an edge is created between vertices if and only if $\Phi(A,B) \geq 0.1$. Then, the maximal cliques of the graph are calculated and sorted by the number of vertices. Next, the parcels belonging to the same maximal clique are fused in descending order, arbitrarily assigning to them the label of the first parcel of the group. Finally, probability maps are recalculated, and hard parcels are obtained by assigning to each triangle the parcel with the highest probability.

Step 4: Post-processing. In order to obtain homogeneous parcels, a series of morphological operations are applied. First, the parcels are modeled as an undirected graph $G = (V, E)$, where the vertices are triangle vertices, and the edges are triangle sides. Next, the connected components are calculated, selecting only the biggest one as representative of the parcel and discarding the rest. Then, a morphological opening (erosion + dilation) is performed, both to eliminate triangles that are barely connected to the main parcel and to fill empty spaces within the parcel. Note that several processing are similar to those proposed in [8], however, these had to be reimplemented to deal with the whole cortical meshes and inter-hemispheric connections.

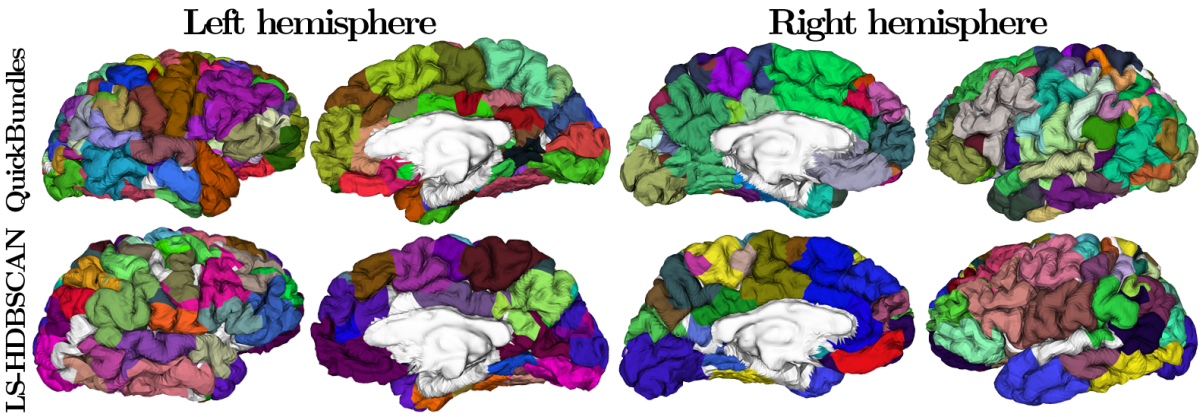

Fig. 3. Resulting parcellations for both implemented versions (QB [12] and LS-HDBSCAN [15]).

III. RESULTS

Table I shows a brief quantitative description of the parcellations obtained with both methods. LS-HDBSCAN clustering results in a relatively small number of parcels, while the QB method is more in line with state-of-the-art methods. However, both methods show a high relative standard deviation, mainly due to the presence of a few very large parcels. Furthermore, while neither method manages to fully label every mesh triangle, QB results in a higher cortical area coverage. This is supported with the parcels shown in Fig. 3, where QB shows less and smaller holes. Also, in Fig. 4 it can be seen that LS-HDBSCAN results in fewer and more compact clusters. In the future, other clustering algorithms and probablistic tractography could be employed, what could lead to better results.

Fig. 5 shows the parcellation obtained for QuickBundles, in three different subjects of the ARCHI database. Note that the size and shape of the parcels change for each subject depending on their anatomy. This information is embedded in the meshes calculated by Freesurfer. Also, fiber clusters are used separately for each subject in T1 space, but thanks to the inter-subject analysis, it is possible to obtain a correspondence between them.

In order to quantitatively evaluate the obtained parcellations, a direct comparison was made with three state-of-the-art methods: *Lefranc* parcellation (dMRI, 237 parcels), *López-López* parcellation (dMRI, 207 parcels), and the *Brainnetome* atlas (multimodal, 246 parcels) [18]. The smaller number of parcels obtained by the proposed method makes difficult the comparison with the other parcellations, however, some similarities were found. The similarity between each obtained parcel and each atlas parcel was evaluated using the Dice coefficient over the corresponding triangle sets, as shown in Table II. Both methods show a higher similarity with the parcellations based on the same modality as this work. This kind of comparison enables the identification of similar parcels found in previous work but is not a form of validation, as there is no ground truth. Moreover, the number and size of the parcels have a high impact in determining the similarity between parcels.

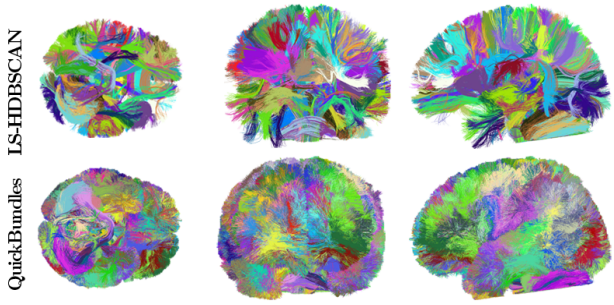

Fig. 4. Inter-subject clustering results from the chosen algorithms (QB [12] and LS-HDBSCAN [15]), for an arbitrary subject.

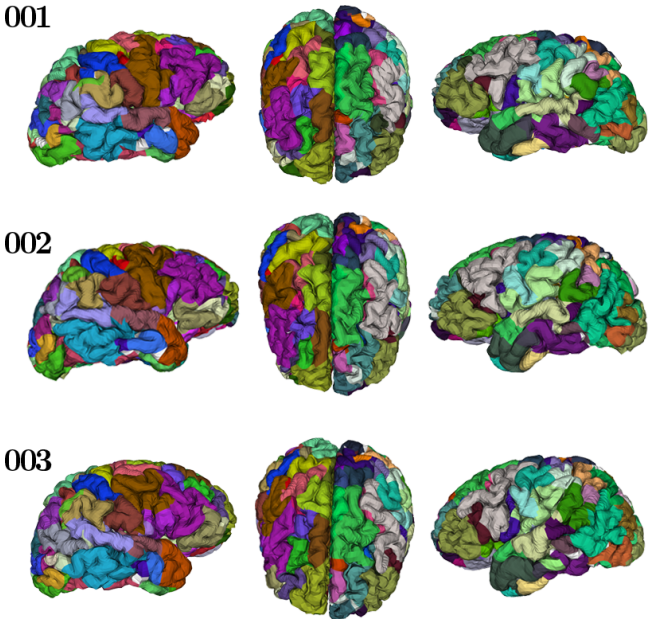

Fig. 5. Final parcellation for QuickBundles, displayed for the first three subjects of the ARCHI database. Colors where randomly selected.

IV. DISCUSSION AND CONCLUSIONS

An automatic group-wise cortical parcellation algorithm was developed, based on the joint structural connectivity of the fiber clusters of a given population. Results are very pre-

TABLE I

COMPARISON OF PARCELLATION PERFORMANCE. AVERAGE SIZE AND STANDARD DEVIATION ARE MEASURED IN NUMBER OF TRIANGLES.

	QuickBundles		**LS-HDBSCAN**	
Parcels per hemisphere	Left	Right	Left	Right
	86	85	54	55
Total parcels	171		109	
Average size	933.5		1231	
Standard deviation	985.1		1054.3	

TABLE II

COMPARISON BETWEEN BOTH PARCELLATIONS AND THE STATE-OF-THE-ART ATLASES, GROUPED BY DICE COEFFICIENT D.

Atlas	**$0.5 \leq D < 0.6$**	**$0.6 \leq D < 0.7$**	**$0.7 \leq D < 0.8$**
QuickBundles			
Brainnetome	31 parcels	13 parcels	3 parcels
Lefranc	32 parcels	21 parcels	11 parcels
López	38 parcels	26 parcels	5 parcels
LS-HDBSCAN			
Brainnetome	20 parcels	3 parcels	1 parcel
Lefranc	16 parcels	12 parcels	4 parcels
López	17 parcels	8 parcels	3 parcels

liminary but promising, where the best parcellation returned 69 parcels with Dice ≥ 0.5, when compared with state-of-the art parcellations. To the best of our knowledge, this is the first approach based on whole-brain inter-subject fiber clustering. This is challenging for this type of analysis, mainly due to inter-individual variability. Previous works in general include anatomical information at some stage of the processing, such as anatomical cortex parcellations. In this work, we propose to use only the connectivity given by tractography. Inter-subject clustering allows us to find reproducible connections across subjects. The performance of the QuickBundles approach is superior, resulting in a higher number of fiber clusters and parcels. Although in the future it is necessary to carry out in-depth evaluations regarding the parameters used, we highlight as the main contribution of this work a first implementation of this type, susceptible to be improved. Hence, future work will be focused on the improvement of the method, based on a detailed analysis of the results for each stage, and the inclusion of an automatic selection of parameters using score metrics based on cortical coverage and trajectory compactness. It is also possible to consider a mixed strategy, using both known labeled fascicles, such as in [8], and inter-subject clusters, as proposed in this paper. In this way, the method would indirectly include anatomical information for the known fascicles, but not in the form of a cortical parcellation, and would not discard information from lesser known fascicles, especially short association bundles. Concerning validation, it is important to create simulated data for validation of the algorithms, since there is no ground truth. Furthermore, in the longer term, the inclusion of data from different modalities, such as fMRI will be evaluated.

COMPLIANCE WITH ETHICAL STANDARDS

The studies involving human participants were reviewed and approved by Comité de Protection des Personnes Ile-de-France VII CPP100002/CPP100022, France. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

[1] O Sporns, G Tononi, and R Kötter, "The human connectome: A structural description of the human brain," *PLOS Computational Biology*, vol. 1, no. 4, September 2005.

[2] M Pievani, N Filippini, M P van den Heuvel, S F Cappa, and G B Frisoni, "Brain connectivity in neurodegenerative diseases–from phenotype to proteinopathy," *Nature Reviews Neurology*, vol. 10, pp. 620, 2014.

[3] M Descoteaux, *High Angular Resolution Diffusion MRI: from Local Estimation to Segmentation and Tractography*, Ph.d. dissertation, Université Nice Sophia Antipolis, 2008.

[4] P J Basser, S Pajevic, C Pierpaoli, J Duda, and A Aldroubi, "In vivo fiber tractography using DT-MRI data," *Magnetic Resonance in Medicine*, vol. 44, no. 4, pp. 625–632, 2000.

[5] S B Eickhoff, B Thirion, G Varoquaux, and D Bzdok, "Connectivity-based parcellation: Critique and implications," *Human Brain Mapping*, vol. 36, no. 12, pp. 4771–4792, 2015.

[6] S Lefranc, P Roca, M Perrot, C Poupon, D Le Bihan, J-F Mangin, and D Rivière, "Groupwise connectivity-based parcellation of the whole human cortical surface using watershed-driven dimension reduction," *Medical Image Analysis*, vol. 30, pp. 11 – 29, 2016.

[7] F Silva, M Guevara, C Poupon, J Mangin, C Hernández, and P Guevara, "Cortical surface parcellation based on graph representation of short fiber bundle connections," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1479–1482.

[8] N López-López, A Vázquez, J Houenou, C Poupon, J-F Mangin, S Ladra, and P Guevara, "From coarse to fine-grained parcellation of the cortical surface using a fiber-bundle atlas," *Frontiers in Neuroinformatics*, vol. 14, pp. 32, 2020.

[9] B Schmitt, A Lebois, D Duclap, G Guevara, F Poupon, D Rivière, C Cointepas, D LeBihan, J-F Mangin, and C Poupon, "CONNECT/ARCHI: an open database to infer atlases of the human brain connectivity," in *29th Annual Scientific Meeting of the European Society for magnetic Resonance in medicine and Biology,*, Lisboa, Portugal, 2012.

[10] M Descoteaux, E Angelino, S Fitzgibbons, and R Deriche, "Regularized, fast, and robust analytical q-ball imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 3, pp. 497–510, 2007.

[11] A Vázquez, N López-López, A Sánchez, J Houenou, C Poupon, J-F Mangin, C Hernández, and P Guevara, "FFClust: Fast fiber clustering for large tractography datasets for a detailed study of brain connectivity," *Neuroimage*, vol. 220, pp. 117070, 2020.

[12] E Garyfallidis, M Brett, M M Correia, G B Williams, and I Nimmo-Smith, "QuickBundles, a method for tractography simplification," *Frontiers in Neuroscience*, vol. 6, 2012.

[13] M Guevara, C Román, J Houenou, D Duclap, C Poupon, J-F Mangin, and P Guevara, "Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography," *NeuroImage*, vol. 147, pp. 703 – 725, 2017.

[14] C Román, M Guevara, R Valenzuela, M Figueroa, J Houenou, D Duclap, C Poupon, J-F Mangin, and P Guevara, "Clustering of whole-brain white matter short association bundles using HARDI data," *Frontiers in Neuroinformatics*, vol. 11, pp. 73, 2017.

[15] I Huerta, A Vázquez, N López-López, J Houenou, C Poupon, J-F Mangin, P Guevara, and C Hernández, "Inter-subject clustering of brain fibers from whole-brain tractography," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 1687–1691.

[16] L McInnes, J Healy, and S Astels, "HDBSCAN: Hierarchical density based clustering," *The Journal of Open Source Software*, vol. 2, no. 11, mar 2017.

[17] T Möller and B Trumbore, "Fast, minimum storage ray-triangle intersection," *Journal of Graphics Tools*, vol. 2, no. 1, pp. 21–28, 1997.

[18] L Fan, H Li, J Zhuo, Y Zhang, J Wang, L Chen, Z Yang, C Chu, S Xie, A R Laird, P T Fox, S B Eickhoff, C Yu, and T Jiang, "The human brainnetome atlas: A new brain atlas based on connectional architecture," *Cerebral Cortex*, vol. 26, no. 8, pp. 3508–3526, 2016.