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

Enhanced Automatic Segmentation for Superficial White Matter Fiber Bundles for Probabilistic Tractography Datasets

C. Mendoza[1], C. Román[1], A. Vázquez[1], C. Poupon[2], J.-F. Mangin[2], C. Hernández[1,3], and P. Guevara[1]

***Abstract*— This paper presents an enhanced algorithm for automatic segmentation of superficial white matter (SWM) bundles from probabilistic dMRI tractography datasets, based on a multi-subject bundle atlas. Previous segmentation methods use the maximum Euclidean distance between corresponding points of the subject fibers and the atlas centroids. However, this scheme might include noisy fibers. Here, we propose a three step approach to discard noisy fibers improving the identification of fibers. The first step applies a fiber clustering and the segmentation is performed between the centroids of the clusters and the atlas centroids. This step removes outliers and enables a better identification of fibers with similar shapes. The second step applies a fiber filter based on two different fiber similarities. One is the Symmetrized Segment-Path Distance (SSPD) over 2D ISOMAP and the other is an adapted version of SSPD for 3D space. The last step eliminates noisy fibers by removing those that connect regions that are far from the main atlas bundle connections. We perform an experimental evaluation using ten subjects of the Human Connectome (HCP) database. The evaluation only considers the bundles connecting precentral and postcentral gyri, with a total of seven bundles per hemisphere. For comparison, the bundles of the ten subjects were manually segmented. Bundles segmented with our method were evaluated in terms of similarity to manually segmented bundles and the final number of fibers. The results show that our approach obtains bundles with a higher similarity score than the state-of-the-art method and maintains a similar number of fibers.**

***Clinical relevance*— Many brain pathologies or disorders can occur in specific regions of the SWM, automatic segmentation of reliable SWM bundles would help applications to clinical research.**

I. INTRODUCTION

Diffusion-weigthed imaging (dMRI) is a non-invasive technique able to characterize water molecules diffusion, enabling the study of the microstructure of brain white matter (WM) in vivo [1]. Through the use of tractography algorithms, the main WM pathways can be reconstructed in the form of 3D streamlines, also called fibers. Tractography datasets, calculated from methods such as constrained spherical deconvolution, coupled with probabilistic streamlines tracking algorithms, provide a better delineation of white matter tracts than deterministic approaches [2]. However, these datasets present a higher complexity and more false positives. Hence, some analysis algorithms, such as WM bundle segmentation, need to be adapted. This is specially important for short association connections, located in the superficial white matter (SWM), right under the cortex, connecting near or adjacent gyri. These bundles are characterized by their small size, high variability among subjects and a U-shaped form [1].

Automatic segmentation methods extract WM bundles, to perform clinical studies and comparison between populations. The algorithm proposed in [3] enables the bundle extraction on massive tractography datasets based on a multi-subject bundle atlas. Also, recent optimizations [4], [5] have allowed a fast classification of white matter fibers. However, the algorithms need to improve the identification of the U-shaped morphology of SWM bundles and filter out spurious fibers. We propose a processing pipeline, based on different filters that successfully improve the segmentation of short association bundles in probabilistic tractography datasets. Preliminary results are shown for precentral-postcentral connections (PrC-Poc) [6].

II. MATERIALS AND METHODS

A. Database and tractography datasets

We used ten healthy subjects from the Human Connectome Project (HCP) database, containing multi-shell dMRI data acquired on a Siemens Skyra scanner with customized protocol [7]. The dMRI data was collected over three b-values (1000, 2000, 3000 s/mm^2), with an isotropic voxel of 1.25 mm. Probabilistic tractographies of 3M streamlines were calculated using MRtrix software [2], based on spherical deconvolution model, Anatomically-Constrained Tractography and a Spherical-deconvolution Informed Filtering of Tractograms. Finally, the fibers were non-linearly transformed to MNI space and resampled with 21 equidistant points.

B. PrC-Poc bundle atlas

We used a new atlas of PrC-Poc connections [6], based on probabilistic tractography over HCP data. It contains seven bundles per hemisphere, connecting the brain precentral-postcentral regions, as seen in Fig. 1. These bundle models were constructed from 100 subjects of the HCP database, after the application of several steps, including intra- and inter-subject fiber clustering. These fascicles present higher coverage than previous atlases, enabling a more detailed study of PrC-Poc connections.

C. Automatic segmentation of WM bundles

The segmentation algorithm [4], [5] uses the maximum Euclidean distance (d_{ME}) to label subject streamlines to the closest atlas bundle, provided that the distance to this bundle is lower than a predefined threshold. It assumes

[1]C. Mendoza, C. Román, A. Vázquez, P. Guevara and C. Hernández are with Faculty of Engineering, Universidad de Concepción, Chile. cristobal.mendoza@biomedica.udec.cl

[2] C. Poupon and J.-F. Mangin are with I2BM, Neurospin, CEA, Gif-sur-Yvette, France

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

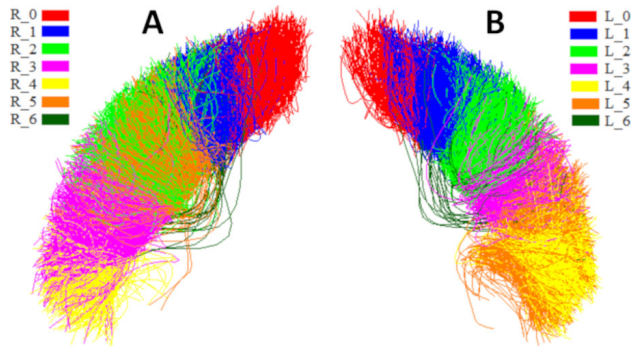

Fig. 1: SWM bundles connecting the precentral and postcentral regions [6]. Bundles are in different colors. A label indicates the hemisphere (R or L), and a correlative number: (A) right hemisphere and (B) left hemisphere bundles.

that fibers are resampled with N equidistant points. The Euclidean distance d_E between N-point subject streamline S and centroid atlas C is computed using Eq. 1, where S_i and C_i are corresponding 3D points.

$$d_E(S_i,C_i) = \|(S_i - C_i)\| \tag{1}$$

Then, d_{ME} is computed as Eq. 2:

$$d_{ME}(S,C)=\min(\max_i(d_E(S_i,C_i)),\max_i(d_E(S_i,C_{N-i}))) \tag{2}$$

A penalization term TN is added to penalize the difference in length between streamlines (Eq. 3):

$$TN = \left(\frac{abs(l_s - l_c)}{max(ls,lc)} + 1\right)^2 - 1 \tag{3}$$

where l_s and l_c are the lengths of subject streamline S and atlas centroid C respectively. Finally, the penalized Euclidean distance d_{NE} is computed using Eq. 4 as:

$$d_{NE}(S,C) = d_{ME}(S,C) + TN \tag{4}$$

D. Enhanced segmentation pipeline for SWM bundles

We propose a new pipeline based on three main steps, that identify fibers with a high probability to belong to the bundle, while filtering out fibers that do not follow the main U-shape morphology of the bundle model.

1) Intra-subject fiber clustering: First, a fiber clustering is applied to the fibers and the segmentation is performed based on the cluster centroids. This step removes outliers and enables a better identification of fibers.
2) Cluster filtering: For each segmented cluster, a filtering based on fiber similarity is applied. Two methods were tested: i) filtering based on Symmetrized Segment-Path Distance (SSPD) over 2D ISOMAP, and ii) filtering based on an adapted version of SSPD for 3D space.
3) Filtering of fibers based on connectivity patterns: Finally, to eliminate noisy fibers, we remove the fibers connecting regions that are far from the main atlas bundle connections.

In the following, we describe the three steps.

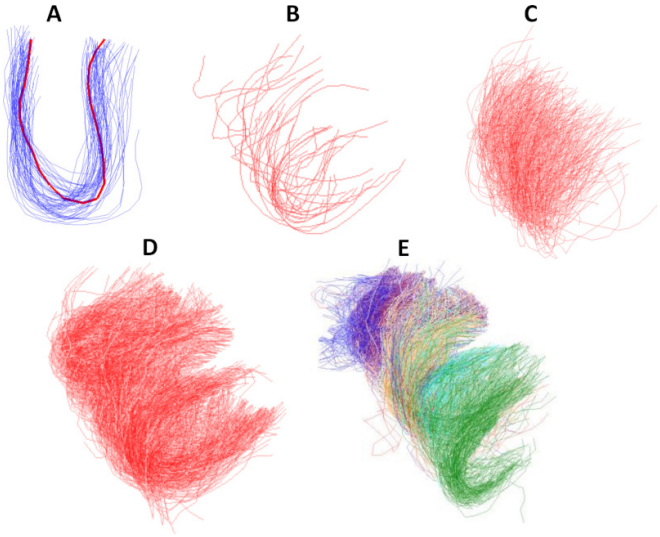

Fig. 2: Segmentation of cluster centroids: (A) an arbitrary cluster (blue) and his centroid (red) that corresponds to the mean streamline of the cluster, (B) centroids segmented for bundle R_0, (C) atlas bundle R_0, (D) bundle R_0 obtained using the original segmentation method and (E) bundle R_0 after replacing each centroid in (B) with its cluster, clusters are shown with different colors. The same threshold was used for both segmentations.

1) Intra-subject fiber clustering: Before applying the segmentation, we use a fiber clustering algorithm (FFClust) to identify compact and homogeneous clusters [8]. Briefly, FFClust proceeds in four steps: (1) builds point clusters, (2) generates preliminary streamline clusters grouping fibers with common point clusters, (3) reassign small preliminary clusters and (4) merges candidate streamline clusters. This method provides outlier removal and a better disentanglement's of the fibers. The algorithm retrieves two datasets: (1) the fiber clusters, and (2) the cluster centroids (Fig. 2-A). The automatic segmentation algorithm is applied to the centroids of a subject, to label them according to the atlas bundles and predefined thresholds. This results in segmented fascicles made of cluster centroids. Segmented fiber bundles are obtained by replacing each centroid by its corresponding cluster fibers (Fig. 2-B,E). This processing obtains bundles with fibers that otherwise would have been discarded due to the predefined threshold, i. e., makes the algorithm less sensitive to the threshold.

2) Cluster filtering: Even though the clustering obtains compact clusters, a filtering is required to discard noisy fibers in each segmented cluster, before fusing them to the final bundles. Two methods were tested: Symmetrized Segment-Path Distance (SSPD) over 2D ISOMAP of fibers, and SSPD distance between fibers for 3D space. The ISOMAP algorithm is a nonlinear dimensionality reduction method, based on the geodesic distance between all pair of points, preserving the intrinsic geometry of the data [9].

i) Cluster filtering based on Symmetrized Segment-Path Distance (SSPD) over 2D ISOMAP: The SSPD [10] is

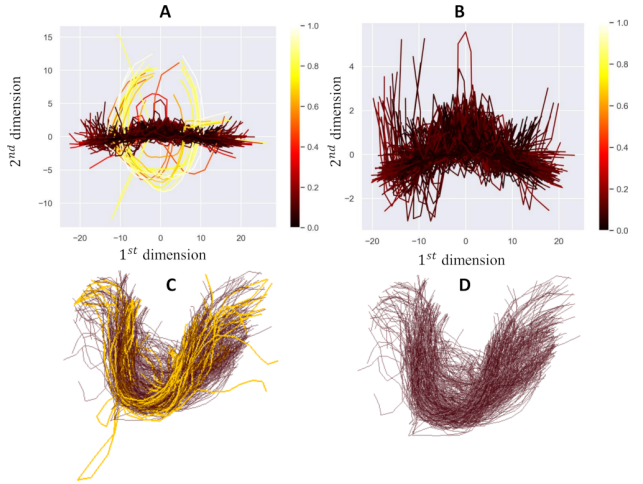

Fig. 3: ISOMAP of clusters processing: (A) ISOMAP of an arbitrary cluster, streamlines with a higher sum of SSPD distances to every other streamline are displayed in yellow, (B) ISOMAP after discarding the 10% of streamlines with the highest trajectory dissimilarity, (C) the same cluster but in 3D space, with the discarded fibers in orange and (D) final cluster after the processing.

a shape based distance useful for regrouping 2D trajectories with similar shape and length and is less affected by noise. This distance is available in the *traj-dist* package (https://github.com/bguillouet/traj-dist).

The Segment-Path Distance from n_1-points trajectory T^1 to n_2-points trajectory T^2 is defined as:

$$D_{SPD}(T^1,T^2) = \frac{1}{n_1}\sum_{i_1=1}^{n_1} D_{pt}(p_{i_1}^1,T^2) \quad (5)$$

where, p_k^i corresponds to the k^{th} location of T^i, and D_{pt} distance from a 2D point p to a trajectory T is the minimun of the distances between this point and all segments s that compose T.

Then, SSPD is computed as:

$$D_{SSPD}(T^1,T^2) = \frac{D_{SPD}(T^1,T^2) + D_{SPD}(T^2,T^1)}{2} \quad (6)$$

Before we replace each centroid by the corresponding cluster fibers, we compute an ISOMAP of each streamline in a cluster. This non-linear dimensional reduction takes the fibers to a 2D plane. For each cluster a streamline distance matrix is calculated for all the streamlines in the cluster, based on D_{SSPD} distance, applied to the 2D fiber coordinates obtained from the 2D ISOMAP. Fibers with the highest sum of SSPD distances are more dissimilar and are candidates to being discarded. Fig. 3 shows and example of filtering, discarding 10% of fibers with the highest sum of SSPD distances.

The final filtering threshold was selected using an heuristic approach, where a value of 50% of the most dissimilar fibers was found to be adequate for all the bundles. Fig. 4-A,B shows and example of filtering with SSPD distance over 2D fiber ISOMAP.

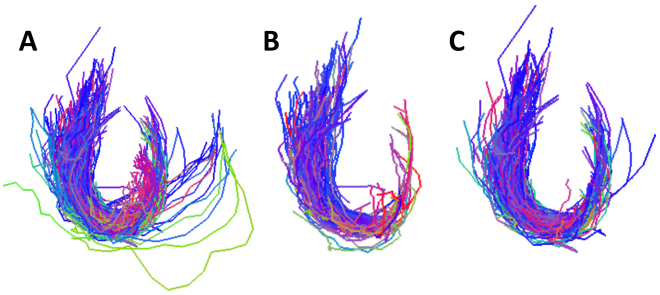

Fig. 4: Filtering of noisy fibers within a cluster: (A) An arbitrary cluster, (B) the cluster after ISOMAP processing (threshold of 50%), (C) the cluster in (A) after filtering based on SSPD3D.

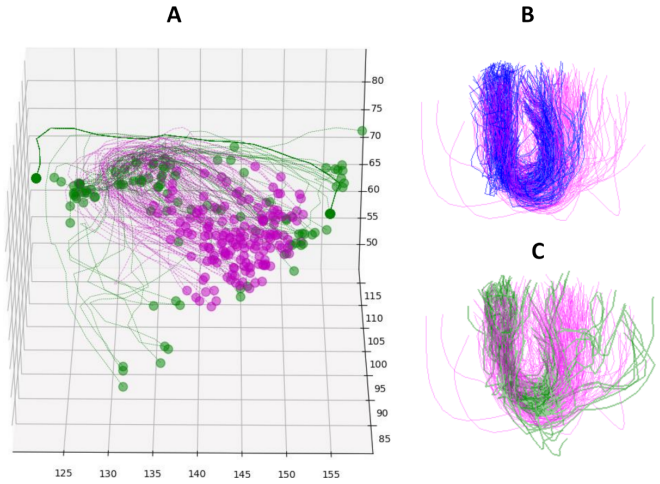

Fig. 5: Filtering based on fiber connectivity patterns: (A) Atlas bundle L_3 (in pink), and streamlines with inconsistent endpoints of a segmented bundle (in green); endpoints are highlighted. Green fibers are likely to be discarded using the d_{END} distance, (B) Segmented bundle (in blue) after discarding the 20% of the streamlines with the highest sum of d_{END} distance to the atlas streamlines, and the atlas bundle (in pink), (C) Discarded fibers with erratic connectivity patterns (in green), and the atlas bundle (in pink).

ii) Cluster filtering based on SSPD distance in 3D space: Furthermore, we modified the D_{SSPD} distance to work in the 3D space. In this case, instead of using the 2D coordinates from the ISOMAP of a cluster, we calculated the distance matrix between all the fibers in the cluster directly using the 3D coordinates, based on the 3D SSPD distance (SSPD3D). After testing, we discarded the 40% of fibers with highest sum of 3D SSPD distances (Fig. 4-C).

Finally, after this step, for each bundle, we discard small clusters, that are mainly noise, and replace the remaining centroids with their corresponding clusters. Due to the rigorous criteria employed, we augmented the threshold of all bundles by 2.0 mm.

3) Filtering of fibers based on connectivity patterns: Then we focused on the connectivity patterns of the segmented bundles, instead of the streamlines shape. Endpoints of

two fibers in a bundle must present the same pattern of anatomical connectivity [11]. To compare the endpoints of subject bundle streamline S_i and corresponding atlas bundle centroid C_j, we used the metric proposed in [11]:

$$d_{END}(S_i,C_j) = \frac{1}{2}(min(\left\|s_1^i - c_1^j\right\|_2, \left\|s_1^i - c_{n_j}^j\right\|_2) + min(\left\|s_{n_i}^i - c_1^j\right\|_2, \left\|s_{n_i}^i - c_{n_j}^j\right\|_2)) \quad (7)$$

where s_1^i, $s_{n_i}^i$ are the endpoints of subject streamline S_i and c_1^j, $c_{n_j}^j$ are the endpoints of atlas centroid C_j.

Next, for each bundle, a distance matrix is calculated based on the distance d_{END}, between all the bundle fibers and the centroids of the corresponding atlas bundle. Fig. 5 shows an example for bundle L_3. In this test, we discarded 20% of fibers with the highest sum of d_{END} distance to the bundle atlas. For the final evaluation, by using an heuristic approach, we selected the optimal discarding threshold. It was fixed to a 60% of fibers with the highest sum of distances.

E. Manual segmentation

To evaluate the proposed method and compare it with Vázquez et al. algorithm [5], we performed a semi-automatic segmentation (called manual segmentation) of the 14 bundles in the 10 subjects, based on manually selected parameters for each bundle. We first applied a semi-automatic delineation of the regions connected by each atlas bundle. Then, for each bundle, we selected the fibers connecting the two corresponding atlas bundle regions. Finally, the noisy fibers from each bundle where filtered out using a cluster confidence index [12]. First, for each atlas bundle, we intersected its streamline endpoints with the voxels of an image in MNI space, to create a connection density image for the two connecting regions, one in the precentral gyrus and the other in the postcentral gyrus. Next, for each bundle, we created a mask with a different label for each region, using an adapted threshold for each bundle.

Once, the atlas bundle ROI masks were created, the fiber extremities of the tractography datasets were intersected to obtain the segmented bundles for the 10 subjects. Next, to filter out anatomically implausible fibers, we used *dipy* to calculate the cluster confidence index as a scoring method [12]. This metric compares the pathway of each trajectory in a cluster and scores each fiber by considering its similarity to the other fibers. Streamlines with the lowest scores were discarded. The score was manually selected for each bundle.

F. Bundle similarity measure

We used the manually segmented bundles as our ground-truth to compare the segmentation algorithm proposed by Vázquez et al.[5] (original segmentation) with the enhanced segmentation algorithm. For each bundle, we calculated a distance matrix based on distance d_{ME}, between all the fibers segmented by a method and the fibers manually segmented. Then, the mean d_{ME} distance was calculated for each bundle, and averaged for the ten subjects.

We used predetermined thresholds, and the enhanced segmentation using SSPD distance over the 2D ISOMAP, and SSPD3D distance, both with augmented thresholds (+2.0 mm). We also included a comparison with the original segmentation using thresholds augmented by 2.0 mm.

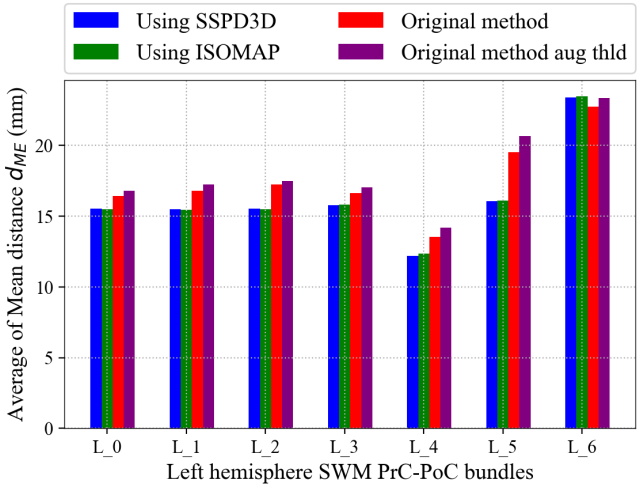

Fig. 6: Average of Mean d_{ME} of PrC-PoC bundles from the left hemisphere for 10 subjects of the HCP database.

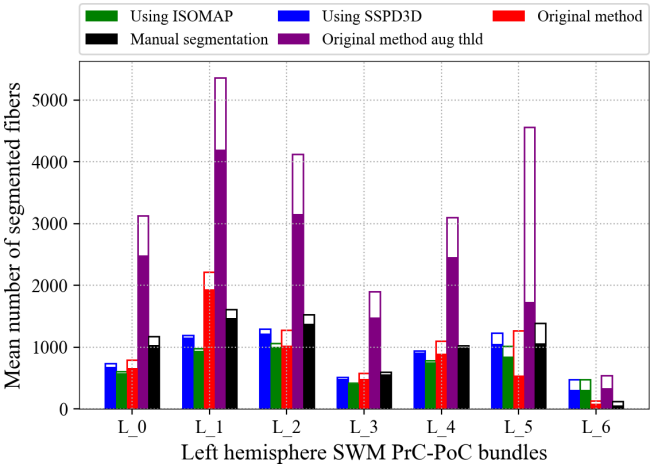

Fig. 7: Mean number of segmented fibers (mNF) for each bundle of the left hemisphere. In solid color are displayed the mNF with $d_{ME} < 15mm$ to the centroid of the bundle, while in white are displayed the mNF with $d_{ME} \geq 15mm$ to the centroid of the bundle,.

III. RESULTS

Fig. 6 shows the average mean d_{ME} distance between bundles obtained by the evaluated methods and the manual segmentation, for the seven PrC-PoC bundles of the left hemisphere. All the mean distances d_{ME} were found to be bigger for the original segmentation method except for one bundle L_6, which got a lower mean of d_{ME} distance for the original method. This bundle has an irregular shape. We can observe that, in general, the proposed methods (SSPD3D and SSPD+ISOMAP), obtain slightly smaller distances. On the other hand, the original method with the threshold augmented in 2.0 mm gets higher distances.

Fig 7 shows the mean number of fibers of the segmented bundles for all the methods. Note that the number of fibers increases dramatically for the original method with the augmented threshold. The enhanced method using either SSPD3D or SSPD+ISOMAP processing within clusters, detects a non negligible number of reliable fibers, filtering the ones with the most dissimilar connectivity patterns and trajectory path. Our method, composed of three steps gets a number of segmented fibers similar to the manually segmented fibers for almost all the bundles, showing that is a better approach for selecting SWM fibers, than the solely use of a predefined threshold, as seen in Fig. 7.

Furthermore, augmentation of thresholds using the original segmentation method adds mostly spurious fibers, by contrast, our method using either SSPD+ISOMAP or SSPD3D enables the augmentation of thresholds while keeping a low mean d_{ME} and the main U-shaped fibers, as seen in Fig. 8.

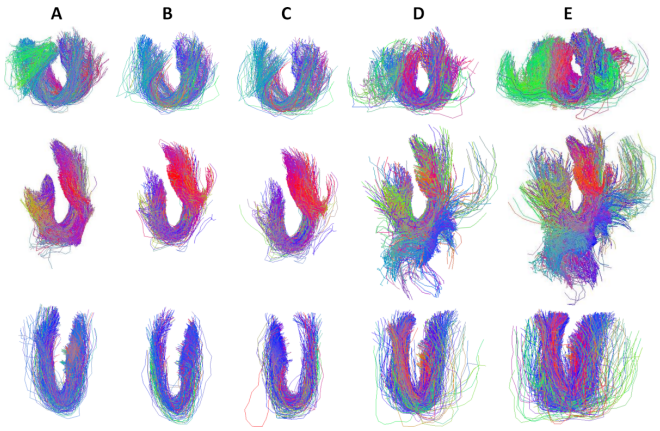

Fig. 8: Column (A): manual segmented bundles, column (B): bundles obtained with SSPD3D processing, column (C): bundles obtained with SSPD+ISOMAP processing, column (D): original segmentation with predefined threshold and column (E): original segmentation with augmented threshold. First row: bundle L_0, second row: bundle L_5 and third row: bundle R_1.

IV. CONCLUSIONS

We have presented a novel method for the segmentation of reliable SWM bundles. The processing pipeline accomplishes a better disentangling of the SWM connections than its predecessor using either cluster's ISOMAP processing with SSPD or SSPD3D distance metric, and the exploit of endpoints between a segmented bundle and its corresponding atlas bundle. Furthermore, we have shown the impact of trajectory clustering metric in the segmentation of reliable SWM bundles. The intra-subject clustering is good enough to form clusters of fibers with overall similar characteristics, however a higher level of granularity within a cluster is required to form reliable SWM bundles, in particular for probabilistic tractography. Results are very promising even though very preliminary and the method was only tested on PrC-PoC connections. Other limitations are the low number of subjects and the heuristic selection of parameters. Future work will extend the analysis to the whole brain and a larger population, with a deeper analysis of parameter tuning and sensitivity. Also, an adapted selection of the segmentation threshold and the amount of fibers to discard using both, d_{SSPD} and d_{END} distances, can be implemented, to define optimal parameters for each bundle.

ACKNOWLEDGMENT

HCP Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. This work has received funding by the ANID FONDECYT 1190701, ANID-Basal Project FB0008 (AC3E), and ANID-Basal Project FB0001 (CeBiB).

REFERENCES

[1] M. Guevara, P. Guevara, C. Román, and J.-F. Mangin, "Superficial white matter: A review on the dMRI analysis methods and applications," *NeuroImage*, vol. 212, p. 116673, may 2020.

[2] J.-D. Tournier, R. Smith, D. Raffelt, R. Tabbara, T. Dhollander, M. Pietsch, D. Christiaens, B. Jeurissen, C.-H. Yeh, and A. Connelly, "MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation," *NeuroImage*, vol. 202, p. 116137, nov 2019.

[3] P. Guevara, D. Duclap, C. Poupon, L. Marrakchi-Kacem, P. Fillard, D. L. Bihan, M. Leboyer, J. Houenou, and J.-F. Mangin, "Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas," *NeuroImage*, vol. 61, no. 4, pp. 1083–1099, jul 2012.

[4] N. Labra, P. Guevara, D. Duclap, J. Houenou, C. Poupon, J.-F. Mangin, and M. Figueroa, "Fast automatic segmentation of white matter streamlines based on a multi-subject bundle atlas," *Neuroinformatics*, vol. 15, no. 1, pp. 71–86, oct 2016.

[5] A. Vazquez, N. Lopez-Lopez, N. Labra, M. Figueroa, C. Poupon, J.-F. Mangin, C. Hernandez, and P. Guevara, "Parallel optimization of fiber bundle segmentation for massive tractography datasets," in *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019.

[6] C. Román, N. López-López, C. Poupon, J.-F. Mangin, C. Hernandez, and P. Guevara, "Study of precentral-postcentral connections on HCP data using probabilistic tractography and fiber clustering," *IEEE 18th International Symposium on Biomedical Imaging (ISBI 2021)*, 2021.

[7] "HCP Young Adult, 1200 subjects data release." [Online]. Available: https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release

[8] 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, p. 117070, oct 2020.

[9] J. B. Tenenbaum, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, dec 2000.

[10] P. C. Besse, B. Guillouet, J.-M. Loubes, and F. Royer, "Review and perspective for distance-based clustering of vehicle trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3306–3317, nov 2016.

[11] G. Bertò, D. Bullock, P. Astolfi, S. Hayashi, L. Zigiotto, L. Annicchiarico, F. Corsini, A. D. Benedictis, S. Sarubbo, F. Pestilli, P. Avesani, and E. Olivetti, "Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation," *NeuroImage*, vol. 224, p. 117402, jan 2021.

[12] K. M. Jordan, B. Amirbekian, A. Keshavan, and R. G. Henry, "Cluster confidence index: A streamline-wise pathway reproducibility metric for diffusion-weighted MRI tractography," *Journal of Neuroimaging*, vol. 28, no. 1, pp. 64–69, sep 2017.