

Cleaning up the mess: tractography outlier removal using hierarchical QuickBundles clustering

Marc-Alexandre Côté¹, Eleftherios Garyfallidis¹, Hugo Larochelle¹, and Maxime Descoteaux¹
¹Université de Sherbrooke, Sherbrooke, Québec, Canada

Target Audience This work is addressed to all researchers who are interested in reproducible bundles generated from dMRI tractography.

Introduction Tractography algorithms reconstruct streamlines that represent white matter fibre bundles for brain connectivity studies. They have helped researchers to produce large structural connectivity datasets (>1M streamlines). However, they generate a large number of invalid connections¹ considered as anatomical outliers which affect connectivity. Having a closer look reveals that several invalid bundles (IB) are rather small in their number of streamlines. This leads to the assumption that tracking algorithms generate these invalid connections in a non-reproducible manner and should be marked as tractography outliers. The common approach to remove outliers remains manual segmentation but clustering tools could be useful as they provide a way to divide (simplify) automatically a full tractography into bundles sharing common features (shape, position, length, etc). This work proposes an outlier removal algorithm that uses a probabilistic tree built from multiple applications of QuickBundles² (QB).

Method Detecting outliers is performed by finding small clusters produced by a clustering algorithm. Most algorithms require that the user knows the true number of clusters beforehand (e.g. K-Means) because otherwise the computational complexity is too high to be applied on millions of streamlines (e.g. hierarchical clustering). QuickBundles performs tractography simplification in $O(nm)$ (n : #streamlines, m : #clusters found) and does not require knowing the number of clusters *a priori*. Instead, a threshold parameter t influences the resulting number of clusters ($t \rightarrow 0$ mm: one cluster per streamline; $t \rightarrow \infty$: one cluster for all streamlines). We rely on QB's speed to recursively cluster streamlines in a hierarchy (Fig. A). We start by applying QB with a high distance threshold and recursively apply QB on each sub-cluster with a lowered threshold. This process goes on until each streamline is assigned to its own cluster. By observing the clustering path of a streamline in the tree, one can identify where its shape starts being different from the others. We assume outliers will find themselves alone earlier in the tree (short clustering path) in contrast to reproducible streamlines since they require a smaller threshold to be separated. Only looking at the length of clustering paths is not robust enough to detect groups of outliers. For example, two outliers only separable by a really small threshold will both have a long clustering path. Instead, we compute the probability that streamline s is an outlier given its clustering path $C = \{c_1, c_2, \dots, c_n\}$ as follows $P(s=\text{outlier} \mid C) = P(|c_n| \mid c_{n-1}) \dots P(|c_2| \mid c_1)$ with $P(|c_{i+1}| \mid c_i) \sim G(\hat{p})$ where G is a geometric distribution parameterized by $\hat{p} = |c_i| / (\sum_{c \in c_i} |c|)$. Geometric distribution assigns a high probability to low values (i.e. small clusters) in contrast to high values (i.e. big clusters). In other words, a cluster is likely to contain outliers if it is much smaller than its siblings. Although QB is sensitive to streamlines ordering², the whole process is repeated using different orderings until a 95% confidence interval around $P(s=\text{outlier} \mid C)$ is below 0.05 for each streamline. Once done, a cut-off can be used on $P(s=\text{outlier} \mid C)$ to keep streamlines less likely to be outliers.

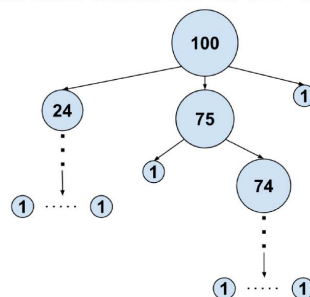
Results Qualitative results (Fig. B) were obtained using diffusion weighted data acquired on a 1.5 T Siemens Magnetom scanner (TE=98ms, TR=12s, 128x128 image matrix) consisting of 64 diffusion encoding gradients with a b-value of 1000 s/mm². Whole brain deterministic tractography was used to generate 500K streamlines using MRtrix with default parameters. TractQuerier³ was used to extract anatomical bundles before applying the pruning algorithm (cut-off between 30-40%) on each of them (fig. B). Quantitative results (Fig. C) were obtained using Phantomas⁴ synthetic data. Then deterministic and probabilistic streamlines were generated using Particle Filtering Tractography⁵ that ensures streamlines connect gray matter. Finally, using Tractometer evaluation¹, we show (Fig. C) that outlier rejection can drastically reduce the number of invalid connections and invalid bundles by 60-70%. It is also important to see that valid bundles that have a small size may have been found accidentally and thus, will be pruned by our outlier removal tool. As shown in Fig. C, it is the case when applied on streamlines generated with deterministic tractography, one valid bundle (VB) is lost after pruning. However, valid connections to connections ratio (VCCR) is still increased considerably.

Discussion/Conclusion We proposed a probabilistic outliers removal based on a hierarchical utilization of QuickBundles. This is different from a previous technique which uses Gamma mixtures⁶ to characterize clusters and a similarity threshold to prune outliers. However, that technique requires knowing the number of clusters. Another method prunes streamlines based on a track density map⁷ whereas we only look at the cluster size. We have shown that our method offers a way to reduce the number of invalid bundles while keeping reproducible valid bundles. This offers a promising tool to safely explore complex networks of brain white matter connectivity.

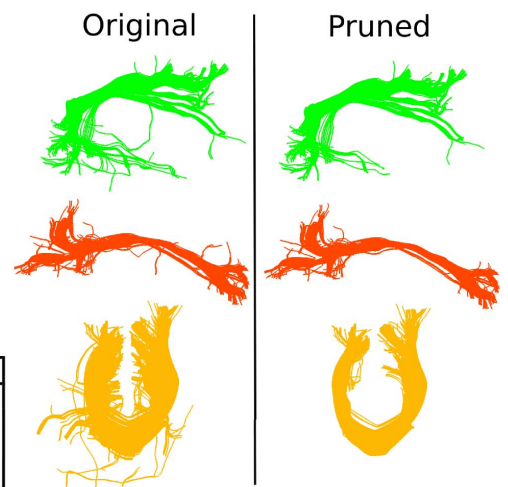
References

- [1] Côté, M.-A., Medical Image Analysis (2013), 17(7), 844-857
- [2] Garyfallidis, E., Frontiers Neuroscience (2012) 6: 175
- [3] Wassermann, D., MICCAI(2013)
- [4] Caruyer, E., ISMRM (2014)
- [5] Girard, G., Neuroimage (2014) 98 (2014) 266–278
- [6] Maddah, M., Inf. Process Med Imaging (2007) 20:372-83
- [7] Guevara, P., Neuroimage (2012)

A. QuickBundles tree example



B. Visual impact of pruning



C. Reduction of invalid bundles

	Deterministic		Probabilistic	
	Initial	Pruned	Initial	Pruned
Count	9568	8431	45500	8123
VCCR	72.2	75.8	50.48	66.64
VB	27	26	27	27
IB	71	28	250	58