

QuickBundlesX: Sequential clustering of millions of streamlines in multiple levels of detail at record execution time

Figures

Eleftherios Garyfallidis¹, Marc-Alexandre Côté¹, François Rheault¹, and Maxime Descoteaux¹¹Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada

Synopsis

QuickBundlesX shows a remarkable 20+X speedup over its predecessor who was until today the fastest clustering algorithm for streamlines. In addition, it returns a useful tree of clusters at different resolutions which allows to query streamlines and easily process millions of streamlines by comparing only with their neighbours.

Introduction

Today, it has become common for fiber tracking algorithms to generate many millions of streamlines^{7,8} and for this reason we need to improve our algorithms to handle and simplify these large datasets at higher speeds than before. This is why we propose QuickBundlesX (QBX), which is a substantial improvement over its predecessor QuickBundles (QB)^{1,5}. Quickbundles was until today⁴ the fastest algorithm for clustering streamlines using streamline distances. The complexity of QB is $O(kN)$ where k is the number of clusters and N is the number of streamlines. Because k is usually much smaller than N the complexity of the algorithm is near to $O(N)$ with large distance thresholds (e.g. 30mm). However, when we start using small thresholds (e.g. 10mm), the number of clusters increases considerably and the complexity goes up accordingly. At same time k increases when N increases. The more streamlines we add the more clusters will be created.

Methods

QuickBundlesX (QBX) is a new method which allows to cluster many millions of streamlines sequentially² and in multiple resolutions by only passing one time through the whole set of streamlines. QBX does this by building a tree of cluster nodes on the fly, in which each layer uses a different distance threshold. As in QB, the distance metric used to compare the streamlines with the centroids of the clusters is the minimum average direct and flip distance (MDF)¹.

The QBX algorithm works as follows. Input variables are a list of streamlines and a list of predefined thresholds e.g. 30mm, 25mm, 20mm. Each streamline is sequentially inserted in the tree from the root until it reaches the bottom. At each step (i.e. node that is visited), two things can happen: a) a new child node is created or b) the streamline is inserted into an existing child node. In either case, we move to the next layer and the process is repeated until the last layer is reached.

In more detail, a new child node is needed when the closest child node to the streamline being inserted is too far (greater than the threshold of the current layer). Otherwise, we update the centroid of the closest child node and then repeat the process with the child node.

The output of QBX is a tree with the same number of layers as the number of predefined thresholds (see circular tree example in Fig. 2 and a part of the same tree shown hierarchically in Fig. 3).

Results

In this experiment, we used probabilistic tracking using anatomical priors³ and random white matter seeding applied on a subject from the Human Connectome Project⁶ datasets and generated from 1 million up to 5 million whole brain streamlines. Then we compared the execution time between QB and QBX (see Fig. 1). QBX showed a speedup (22X - 25X) beyond the existing state-of-the-art. This means that we can easily cluster 5 million streamlines in less than 5 minutes with a standard laptop (one Intel i7 core, 1.8 GHz). In contrast, QB needs at least 2 hours for the same task. In this experiment, we used the following distance thresholds 30mm, 25mm, 20mm, 15mm which produced a tree with 4 layers giving access to the clusters of those resolutions. For simplicity we show a smaller representation of the tree in Fig. 2 and 3.

Discussion and conclusions

QuickBundlesX shows a 20+X speedup over its predecessor who was until today the fastest clustering algorithm for streamlines⁴. In addition, it returns a useful tree of clusters at different resolutions which allows to query streamlines. This has been a major issue when processing streamlines because it is computationally very expensive to know the exact neighbour of a streamline as you will have to go through all the streamlines and compare their pairwise distances. However, using this tree you can always start from the root, find the centroid that is closest to you from the top layer and continue recursively searching the tree. So, now with the QBX output tree, we have a better structure for searching in the space of streamlines. A current disadvantage of QBX the number of clusters generated at the bottom layer is often larger than the number of clusters generated by QB because some clusters split at higher layers. A simple solution would be to merge clusters at each layer according to the current threshold. Nonetheless, QBX provides the new standard for efficient clustering in the space of streamlines.

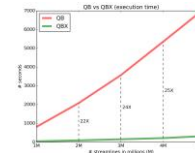
Acknowledgements

We are grateful to the Fonds de recherche du Québec - Nature et technologies (FQRNT) and the Natural Sciences and Engineering Research Council of Canada (NSERC) programs for funding this research.

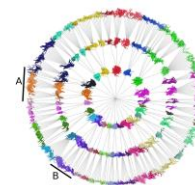
References

- [1] Garyfallidis, E., M. Brett, M. M. Correia, G.B. Williams, I. Nimmo-Smith. QuickBundles, a method for tractography simplification. *Frontiers in Neuroscience*, 6-175, 2012.
- [2] Tou, J.T., and Gonzalez, R.C., Pattern recognition principles. *Pattern Recognition in Physics*, 1, 1974.
- [3] Girard, G., K. Whittingstall, R. Deriche, M. Descoteaux. Towards quantitative connectivity analysis: reducing tractography biases, *Neuroimage*, 98, 266--278, 2014.
- [4] Reichenbach, A., M. Goldau, H. Christian, M. Hlawitschka, V-bundles: Clustering Fiber Trajectories from Diffusion MRI in Linear Time, *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015*, 191--198, 2015.
- [5] Garyfallidis, E., M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, and I. Nimmo-Smith. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, 1-18, 2014.
- [6] Sotiropoulos, S.N. and S. Jbabdi, and others. Advances in diffusion MRI acquisition and processing in the Human Connectome Project, 80, 125--143, *Neuroimage*, 2013.
- [7] Rheault, F, J.C. Houde, and M. Descoteaux. Real time interaction with millions of streamlines. In *proc. of ISMRM*, 23, 6005, 2015.
- [8] Smith, R.E., J-D Tournier, F. Calamante, and A. Connelly. The effects of SIFT on the reproducibility and biological accuracy of the structural connectome, 104, 253--265, 2015.

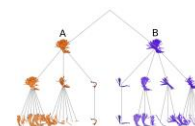
Abstract ID: 4187



Comparison of time spent in seconds between QuickBundles (QB) and our proposed approach QuickBundlesX (QBX). Notice that QBX is at least 20X faster than QB which is currently the fastest method for clustering streamlines.



Full brain clustering obtained with QBX and showed as a circular tree. Only the layers 35mm, 30mm and 20mm are shown for clarity. For a close-up of A and B sub-trees see Fig. 3.



Close-up of the tree shown in Fig. 2. The left CST (orange) and the right CST (purple) at layers 35mm, 30mm and 25mm