

Overview

Algorithms for clustering tractographies are typically quite slow. Much of the computational burden comes from the need for detailed geometric comparisons between pairs of tracks in large datasets often containing hundreds of thousands of tracks. We have developed an approach that leaves this more detailed comparison to a later stage after an initial first pass through the dataset to create prototype bundles.

The method presented here is the fastest currently known method. Less than 5' are required to produce clusters from a whole brain deterministic tractography datasets of ~250,000 tracks. Our method is inspired by BIRCH (Zhang, 1996).



Algorithm

Split Phase:Select the first track t_1 , and place it in the first cluster $r_1 \leftarrow \{t_1\}$. Then for all remaining tracks $n \leq 2N$ (where N is the number of tracks):

- 1: Goto next track t_n .
- 2: Calculate 3TED between this track and virtual tracks of all current clusters r_m (where $1 \leq m \leq M$ and M is the current number of clusters).
- 3: If the minimum 3TED distance is smaller than threshold, add the track to the cluster with the minimum 3TED, and update by joining t_n to r_m ; otherwise create a new cluster $r_{m+1} \leftarrow \{t_n\}$; $m \leftarrow m+1$.

Merge phase:Create a higher node that aggregates nearby clusters by comparing their virtual 3-tracks. The new cluster is the union of the two previous clusters.

Creating Hierarchies - Fornix

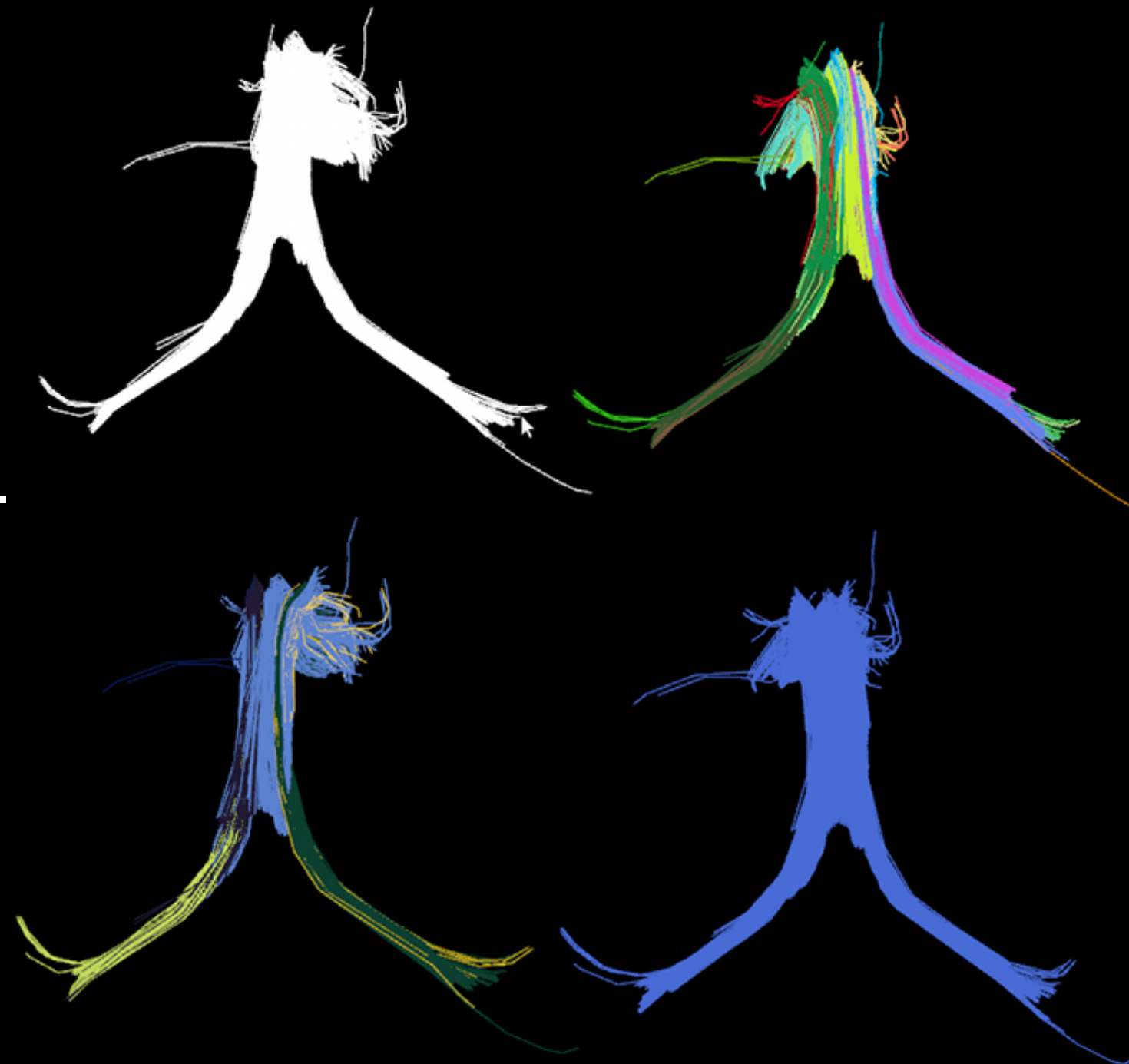
The figure shows how the algorithm clustered the fornix bundle from the Fall 2009 Pittsburgh Brain Competition (PBC, 2009).

The initial fornix consists of 1076 tracks (shown in white).

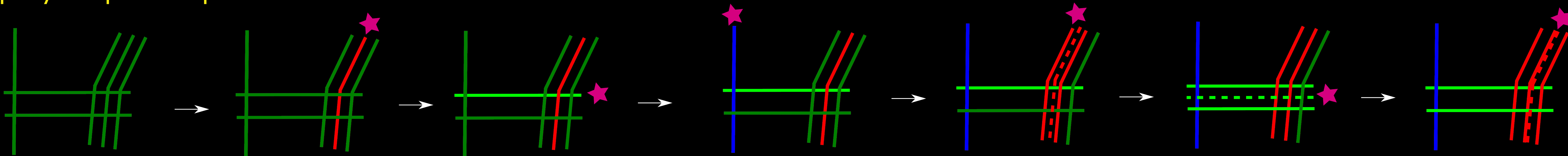
With a 5mm threshold our method generates 22 clusters (top right).

With a 10mm threshold we get only 7 clusters (bottom left).

And with a 20mm the whole fornix is determined by one cluster only (bottom right).

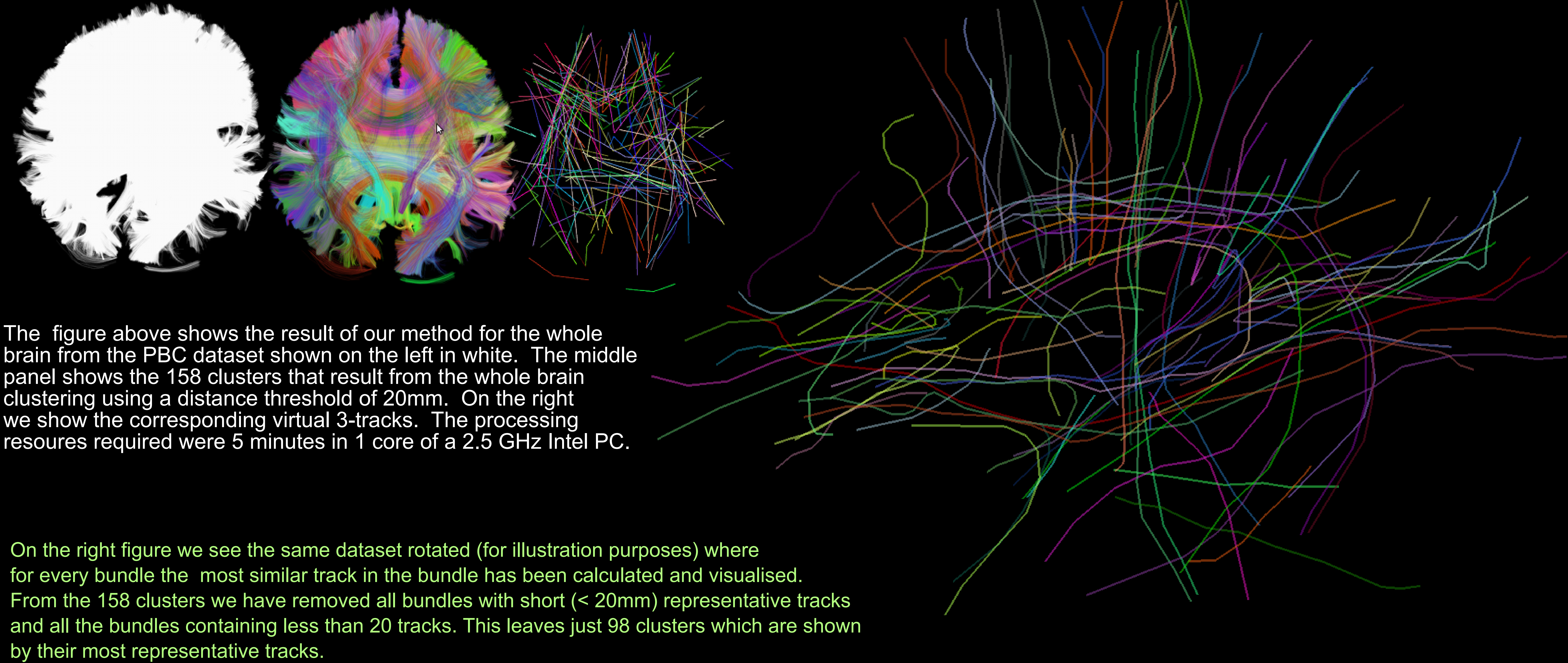


Step by step example



If we wanted to cluster the 6 green tracks on the left in different bundles this is what our method would do step by step. The star indicates the cluster that is being created or modified. The broken lines indicate the virtual tracks when not coincident with a track in the cluster.

Simplifying Tractography Representation in an Accessible Structure is our Goal



The figure above shows the result of our method for the whole brain from the PBC dataset shown on the left in white. The middle panel shows the 158 clusters that result from the whole brain clustering using a distance threshold of 20mm. On the right we show the corresponding virtual 3-tracks. The processing resoures required were 5 minutes in 1 core of a 2.5 GHz Intel PC.

On the right figure we see the same dataset rotated (for illustration purposes) where for every bundle the most similar track in the bundle has been calculated and visualised. From the 158 clusters we have removed all bundles with short ($< 20\text{mm}$) representative tracks and all the bundles containing less than 20 tracks. This leaves just 98 clusters which are shown by their most representative tracks.

Conclusion

Our method reduces the search space between tracks in large streamline datasets from tractography using very low computation and memory time in an online fashion. We named this method LSC (local skeleton clustering). The algorithm returns a graph structure where every node representing a cluster contains (1) the indices of the tracks belonging to that bundle, (2) a virtual (non existing track) which represents the shape of the bundle, and (3) the number of tracks in the bundle.

The results here use only three points (the start, middle and end point). This is not intrinsic to our technique; We can use more points to approximate the tracks and different distance measures (Corouge 2004; Jianu, 2009), to detect similarity. However, we were surprised to see that even three points give a sufficient representation of the real complexity of the tractography. These methods are promising but need further neuroanatomical validation.

References

Zhang, T. (1996), 'Birch: An Efficient Data Clustering Method for Very Large Databases', SIGMOD RECORD, vol. 25, no. 2, pp. 103-114. Corouge, I. (2004), 'Towards a shape model of white matter fiber bundles using diffusion tensor MRI', ISBI, pp. 334-347. Jianu, R. (2009), 'Exploring 3D DTI Fiber Tracts with Linked 2D Representations', IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 1449-1456. PBC (2009), <http://pbc.lrdc.pitt.edu>

Software

nipy.org/dipy

