Combining atlas and reference tracks for bundle detection.

E.Garyfallidis, M.Brett, I.Nimmo-Smith

3rd December 2009

Abstract

We propose a new method which accurately detects preselected bundles from the training set in the other brains. We generate reference tracks for every bundle in the training set and detect the corresponding tracks in the other datasets using the average mean minimum distance. Then we normalize the LONI atlas onto each of the track dataset brains. For a given ROI value in the normalized atlas we extract all the tracks going through the corresponding ROI in each dataset and then we remove tracks that are not similar enough to the reference tracks. Our results are very promising however we learned about the PBC competition at a late stage and we didn't manage to complete our work towards the unsupervised learning tasks. We have developed some interesting visualization tools, as well as techniques for automatic identification of the entire corpus callosum and its removal from our datasets. This is to overcome the difficulty that the corpus callosum intersects many other white matter areas and complicates any classification procedure. In addition, we we have developed algorithms for down-sampling the tracks in an intelligent fashion so that we do not lose any local or global shape characteristics.

1 Introduction

Figure 1 shows the whole dataset for subject 1, scan 1 and the bundles identified by the expert. The supervised learning challenge is to create an algorithm to label this and two other brains to maximise the agreement with the expert. We identified that there were critical issues to do with the registrations of these brains which we have addressed below. In order to achieve consistency between bundles identified in the different brains we have concentrated on combining prior information from a published atlas together with the selection of a small set of tracks from each labelled bundle to use as reference tracks. Beyond this we have chosen to use simple but effective geometric measures based on the individual tracks.

2 Methods

2.1 Registration

We wanted to use the LONI (ICBM DTI-81) atlas labels to find candidate regions in the individual brains. To make something like an anatomical image from the tracks, we took a binarized track count image - that is an image for which there was a 1 in voxels that had one or more tracks passing it, and zeros elsewhere. We used SPM8 to calculate warping parameters to match the binarized track count image to the ICBM white matter image. To reslice the LONI label template, we inverted the SPM normalization parameters using the SPM deformations

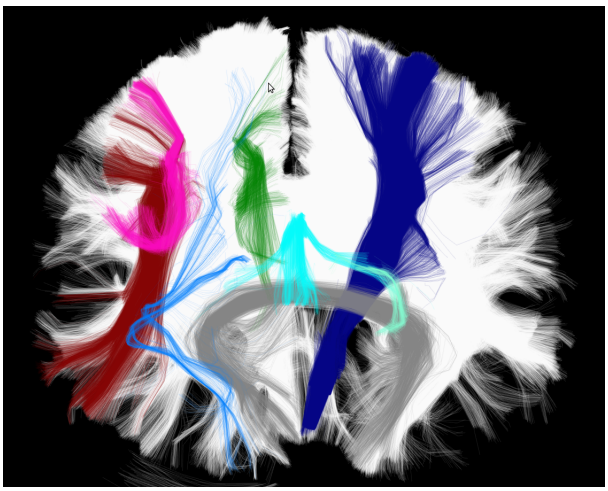

Figure 1: In non-white color are the 8 bundles identified by the PBC competition. The purpose of the supervised learning task was to identify these bundles in other brain datasets.

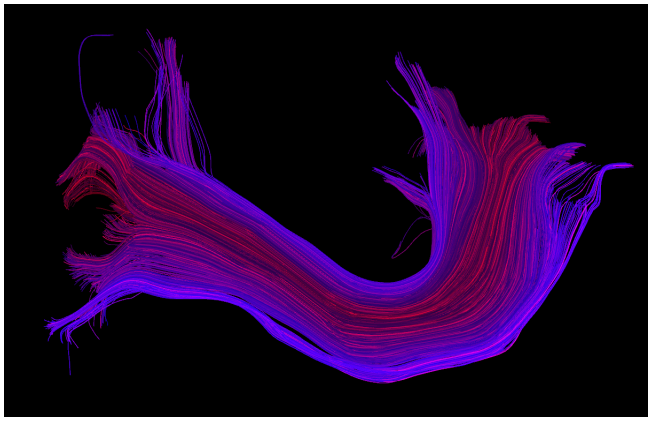

Figure 2: Arcuate fasciculus using the average mean minimum distances.

utilities, and resliced with the inverted parameters and nearest neighbour resampling. Because of the size and orientation differences between the binarized track count images for the same subject, we did this separately for each scan for every subject. This gave us a labeled image from the LONI atlas in the space of each brain image.

2.1.1 Skeleton Track Atlas

While developing techniques for this competition we have created tools to create a skeletonised version of a track dataset in which the skeleton tracks are ones which have multiple neighbours which are very similar. This is detailed below.

2.2 Track Metrics

2.2.1 Average Mean Minimum Distance Metric

We explored a number of metrics for the similarity or closeness of two tracks. This is based on the work of [1] as developed by [4]; the method was also used by [3]. If $P = [p_0, p_1, \ldots, p_N]$ and $Q = [q_0, q_1, \ldots, q_M]$ are two tracks then the mean minimum distance of P from Q is defined as

$$\text{m.m.d.}(\text{P}, \text{Q}) = \frac{1}{N} \sum_{i=0}^{N-1} \min_{0 \le j \le M-1} |p_i - q_j|.$$

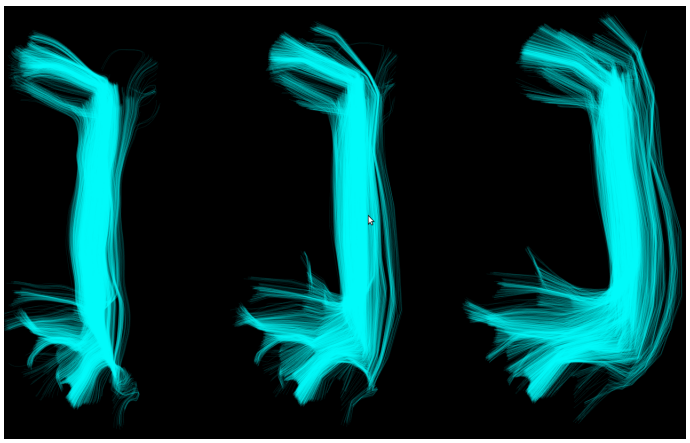

Figure 3: Left: Raw bundle, centre: MDL approximation, Right: Simple down-sampling along the length.

In general

$$\text{m.m.d.(P, Q)} \neq \text{m.m.d.(Q, P)}$$

To provide a symmetric distance measure we define the average mean minimum distance between P and Q as

$$\frac{\text{m.m.d.}(P, Q) + \text{m.m.d.}(Q, P)}{2}$$

. [1] introduced two other symmetrisations of the two m.m.d. measures, either selecting the greater or the lesser of the two. This appears to be a very effective metric for filtering a collection of candidate tracks to identify ones which are reasonably similar to a specified reference track (see Figure 2).

2.3 Down-sampling

Doing calculations with all the miles of identified tracks is computationally very demanding. Therefore, we developed three different types of down-sampling to speed up our calculations. We used the two first types of down-sampling (see Figure 3) to reduce the number of points needed to describe a track and the third type we used to select tracks that had many other tracks that were very similar.

2.3.1 Minimum description length approximation

This is based on the work of [2] who devised an information theoretic approach based on minimum description length (MDL) to generate adaptive sub-samplings of tracks as a precursor to breaking them up into their constituent component line-segments. The criteria for the sub-sampling are based on the degree to which sections of the track can be approximated by the chord joining its end points, with an MDL penalty function that compares the loss of local position and direction information in the approximation with the information required to represent the full data. However, instead of then partitioning the tracks into segments, we used this MDL version as an approximation for the entire track. The sub-sampling ratio was approximately 1:8 and automatically produced tracks with long segments on the straighter elements and a higher density of shorter segments where there was significant curvature. This appears to preserve essential shape characteristics of the tracks with good fidelity (see Figure 3).

2.3.2 Three Point Tracks

We needed a very fast way to remove tracks that are very far away from a reference track. In order to do this we have used a three-point distance as a coarse filter to remove tracks which are distant from a reference track or bundle. This metric is the arithmetic average of the euclidean distances between the end points and the mid-points of the two tracks.

2.3.3 Skeletal Tracks

Here we tried to identify some core tracks that could possibly be good representatives of many of their neighbouring tracks. In order to do this we randomly selected 5000 tracks from the 250000 given tracks (2%) and for these tracks we kept only those who had at least 50 other tracks in the full dataset which had every single point closer than 5mm from the closest point in the reference track. For all three brains the number of skeletal tracks was reduced from 5000 to approximately 1800 by this technique.

2.4 Object Detection

2.4.1 Corpus Callosum

The corpus callosum (CC) is the major white matter fibre structure connecting the two hemispheres. We found that it accounted for more than 40% of the tracks in the competition datasets. Because the CC is such a massive structure with ramifications throughout the brain we saw that if we could remove it, we would be better able to identify other structures near the CC. To do that we tried the following very simple technique. We know that in a normal brain there will be plenty of fibres passing through the mid plane separating the two hemispheres. Finding this mid-plane is very easy with normalized brains in MNI space because this is the sagittal plane passing though the centre of the volume. Then we find all the points of intersection of tracks with this plane and mapped them from the track space to a 2d binary image space so that every pixel in the image plane has 1 if one or more tracks passes through it and 0 if none. In this 2d image the biggest visible object is the corpus callosum (CC) which is very easy to identify and label, thus separating it from the other smaller objects in the plane using standard morphological operations i.e. erosion followed by dilation. After the CC has been detected in the 2d image it is straightforward to find the tracks that pass through the corresponding voxel in MNI space. The corpus callosums detected in this way are shown in Figure 4 for two different brains.

2.4.2 Bottleneck Finding - Cut Plane

One of the techniques that we have partially developed as a tool for the unsupervised learning task is based on the use of cutting-planes generated from a reference track. Given a bundle of tracks B, and a reference track $R = [r_0, r_1, \ldots, r_N]$, we construct the family of planes π_j normal to R at each point $r_j, j = 1, \ldots, N-1$, and consider the local geometry of the hit sets, i.e. the intersections $\pi_j \cap B$. The metric we have introduced is the radial divergence metric (RDM), which is given by the radial component towards or away from r_j of the tangent vector of each track $b \in B$ where it meets π_j. Preliminary evidence is that RDM is a useful metric in identifying tracks which maintain a course parallel to a reference track though displaced some distance. This is particularly relevant in trying to identify tracks which may belong to a broad, thin, strap-like bundle. Figure 5 shows features of this implementation where the blue points indicate cutting points with low divergence.

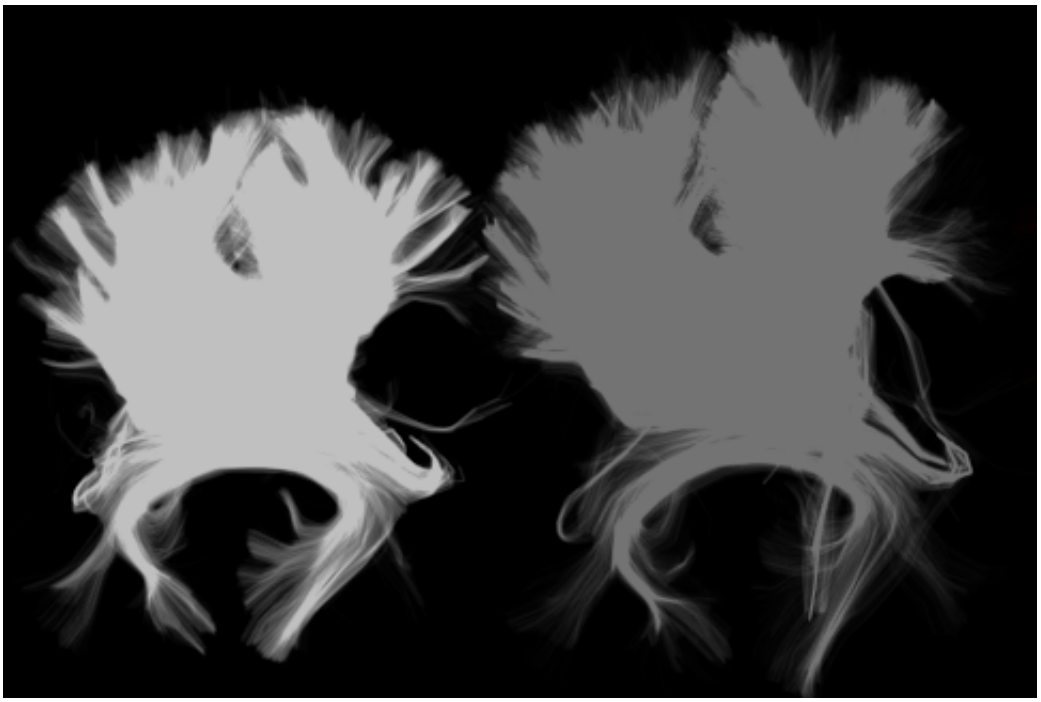

Figure 4: Detecting and separating corpus callosum.

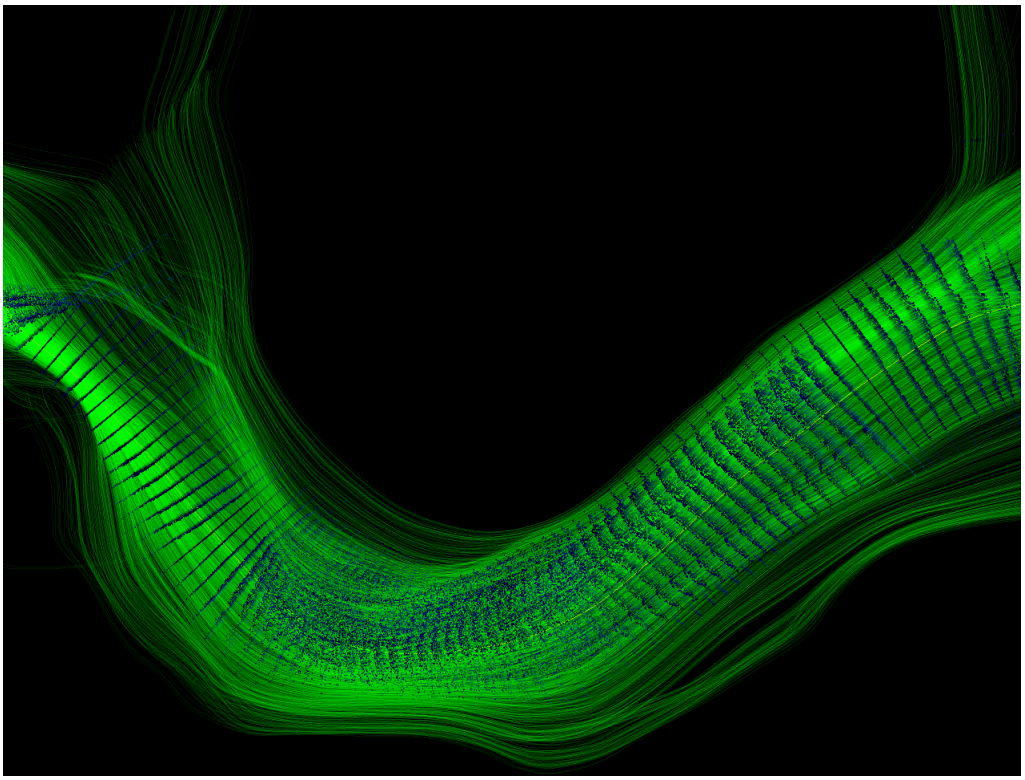

Figure 5: Cut plane - hidden dragon

2.5 Algorithmic Description

We have developed many more tools than those we actually incorporated in our submission - for example the corpus callosum finding algorithm and cut plane. Here we give a short algorithmic description restricted to those used to make the results we submitted for the supervised learning challenge.

1. Normalize ICBM atlas with all brains, as described in the registration section.

2. Generate most similar tracks (REFERENCE TRACKS) using minimum average distances and manually pick some extra fibres so you can have a better shape description of the bundle.

3. For an ROI value in the atlas corresponding to each dataset, generate the tracks (VALUE TRACKS) which pass through the region having that value.

4. Remove very far tracks from the REFERENCE TRACKS using the 3-point method.

5. Compare the REFERENCE TRACKS with the VALUE TRACKS. If they do not have any tracks in common then THE VALUE TRACKS are used in place of the REFERENCE TRACKS.

6. Finally, use the minimum average minimum distances to reduce the number of VALUE TRACKS which are far from every corresponding REFERENCE TRACK.

7. Compare with the training set.

8. If the intersection of the training set with the current bundle set is not maximised go to 6 else STOP.

The results we obtained after running this algorithm were very accurate as can be seen in Figure 6 below.

3 Software

We developed all our software in Python for fast code development, and Cython for fast execution whenever it was necessary. Our core-detection functions and IO readers were build in the DIPY module which we plan to embed in the Neuroimaging in Python (NIPY) suite. We developed our visualization methods in a package we called FOS which is our python-vtk implementation for visualizing tractographic and brain imaging datasets. All pictures shown in this paper were obtained from FOS. FOS has also the property that it is able to visualize many track datasets simultaneously. For example you could visualize in the 3d space simultaneously all 5 brains of the datasets or even more. It also allows you to pick and select specific fibres with the mouse.

4 Conclusion

With a simple geometric approach to this problem, incorporating prior information, we have managed to get very promising results for all 5 datasets. Unfortunately, we didn't have time to test these for the unsupervised learning Challenges 2a/b/c, however this competition has inspired us to continue working on this problem full hearted and batteries included!

Figure 6: Detecting all 8 bundles

References

[1] I. Corouge, S. Gouttard, and G. Gerig. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *International Symposium on Biomedical Imaging*, pages 344–347. Citeseer, 2004.

[2] J.G. Lee, J. Han, and K.Y. Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, page 604. ACM, 2007.

[3] L. O'Donnell and C. F. Westin. High-dimensional white matter atlas generation and group analysis. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 9(Pt 2):243–51, 2006.

[4] S. Zhang, S. Correia, and D.H. Laidlaw. Identifying White-Matter Fiber Bundles in DTI Data Using an Automated Proximity-Based Fiber Clustering Method. *IEEE transactions on visualization and computer graphics*, 14(5):1044, 2008.