Fast and Fully Automated Clustering of Whole Brain Tractography Results Using Shape-Space Analysis

Greg D Parker[1], David Marshall[2], Paul L Rosin[2], Nicholas Drage[3], Stephen Richmond[3], and Derek K Jones[1]

[1]CUBRIC, School of Psychology, Cardiff University, Cardiff, South Glamorgan, United Kingdom, [2]School of Computer Science, Cardiff University, Cardiff, South Glamorgan, United Kingdom, [3]School of Dentistry, Cardiff University, Cardiff, South Glamorgan, United Kingdom

Purpose Fully automated segmentation of large tractography data sets into distinct anatomical bundles is an important but challenging problem. Recent solutions [1-2] tend to incorporate clustering algorithms that require a huge number of computationally expensive inter-streamline distance comparisons (while ref. [1] reduces computational load, it still requires a set of ~3000 streamline comparisons for embedding). Here we propose a novel method for streamline segmentation that eliminates explicit comparisons by re-casting the problem as a shape recognition task. Principal component analysis (PCA) allows us to learn a 'shape space' that covers all expected streamline shape variations within the human brain. Appropriate partitioning of this shape space reduces streamline clustering to a fast and trivial, but robust, projection operation.

Methods

Data Acquisition DW images were acquired for 16 healthy adult brains (10 datasets for training, 6 for evaluation): 60 directions, b = 1200 s/mm^2, (+6 b=0 s/mm^2), 2.4mm isotropic resolution. Following motion correction, whole brain tractography was performed using the damped Richardson-Lucy algorithm [3] with 2 mm seedpoint resolution, 0.5 mm step size and 0.1/45^o fODF magnitude/angular termination thresholds, respectively, resulting in approximately 55,000 streamlines per dataset.

Shape Space Creation The fractional anisotropy (FA) maps from the training images were first co-registered (affine, without shear) to the FMRIB58 FA template image in MNI space, and the corresponding transformations applied to each native space tractography dataset (to remove head rotations & minimize size differences). All streamlines were then reparameterized to 30 data points (equidistant knots), translated to the origin (no rotation applied) and the *x*, *y* and *z* spatial co-ordinates concatenated to form a 90 element descriptor; we will refer to the descriptor creation process as 'normalisation'. Following PCA of these streamline descriptors, the resultant eigenvectors describe the axes of a streamline 'shape space'. Examination of the corresponding eigenvalues revealed that the first 10 dimensions accounted for 98% of explained variation in streamline shape, so we truncated the space to these first 10 dimensions to both reduce computational load and ameliorate the effects of noise and inter-subject variability (i.e., 10 dimensions is sufficient to capture general shapes but not subject-specific detail).

Space Partitioning There are two practical methods for partitioning the shape space: (1) Learn the partitioning through *k*-means clustering of projected (Eq. 1) data (for which we use ~ 500,000 data points and k = 450 clusters), then label the clusters according to their contents (similar to [1]); and (2) Label all streamlines within the training datasets as either spurious or belonging to one of a number of tracts, normalise and project the streamlines into the shape space and then train a bank of support vector machines (SVM's) to segment each labelled class individually.

Segmentation Segmentation of novel data is comparatively trivial. For method 1, streamlines are normalised and projected into the shape space and assigned a label according to the nearest cluster centroid, grouping similarly labelled streamlines into a bundle For method 2 streamlines are normalised/projected as before then classified using the SVM bank.

Eq.1 $C = E^T * (s - m)$: C - Projected coordinates, E - 1st 10 PCA eigenvectors, s – individual descriptors, m - mean descriptor.

Results Fig. 1 demonstrates shape space trajectories (mean over 10 datasets, +/- 1SD) for the left uncinate and arcuate fasciculi. Note that while these tracts have *similar* shapes (with respect to the gamut of possible shapes within the brain), the shape-space trajectories remain distinct. Fig. 2 demonstrates the result of segmenting a novel whole-brain tracking dataset (n ~ 55,000 streamlines) using the proposed method, (total run time < 5 minutes). The results are qualitatively consistent with those derived through manual segmentation (the corpus callosum, corticospinal tracts, cingulum, fornix, arcuate and uncinate fasciculi are shown). Fig 3 shows fully-automated segmentations of the arcuate fasciculus using the SVM-based classification method in four novel datasets. Again results are consistent with manual segmentation

Discussion/Conclusion We have demonstrated that shape parameterisation provides a descriptive, stable, feature space in which streamlines belonging to a particular white matter tract map to predictable sub-regions in shape-space and that those sub-regions are sufficiently separated to distinguish between similarly shaped though anatomically distinct structures. The advantage of shape-space embedding is that it removes the need to explicitly calculate streamline similarity matrices during segmentation. Rather, this is now implied by proximity in a feature space that can be reached through a simple projection. Thus, while the initial creation and partitioning of the shape space is computationally intensive, this is a one time operation. Subsequent segmentations need only a projection into that space and so are both rapid and low on memory demands.

Fig. 1

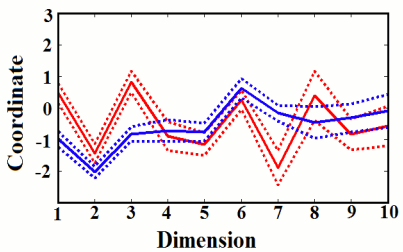

Fig. 2

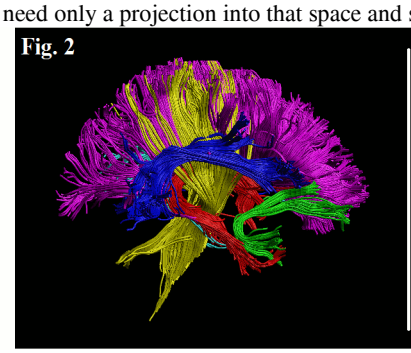

Fig. 3

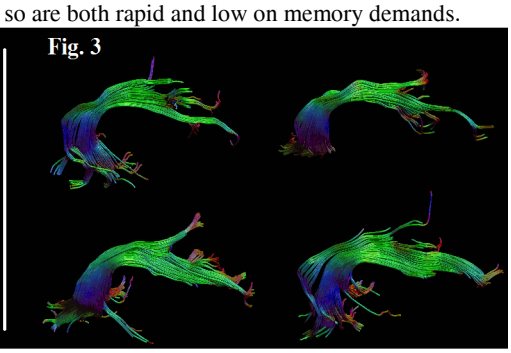

Fig. 1*: Parallel axis plot of mean shape space coordinate trajectory (solid line) +/- 1 standard deviation (broken lines) for the arcuate (Red) and uncinate (blue) fasciculi* ***Fig. 2:*** *Subset of a whole brain segmentation result: corpus callosum (purple), arcuate fasciculus (blue), fornix (red), uncinate fasciculus (green), corticospinal tracts (yellow), cingulum (partially visible, cyan).* ***Fig. 3:*** *The arcuate fasciculus automatically segmented from 4 novel whole brain tracking datasets*

[1] O'Donnell & Westin. 2007. *IEEE TMI* 26**:**1562-1575; [2] Leemans & Jones. 2009, *Proc ISMRM* 856. [3] Dell'Acqua *et al*. 2010. *NeuroImage* 49:1446-1458