ORIGINAL ARTICLE

Tractography Processing with the Sparse Closest Point Transform

Ryan P. Cabeen[1] (ORCID) **· Arthur W. Toga[1] · David H. Laidlaw[2]**

Abstract

We propose a novel approach for processing diffusion MRI tractography datasets using the sparse closest point transform (SCPT). Tractography enables the 3D geometry of white matter pathways to be reconstructed; however, algorithms for processing them are often highly customized, and thus, do not leverage the existing wealth of machine learning (ML) algorithms. We investigated a vector-space tractography representation that aims to bridge this gap by using the SCPT, which consists of two steps: first, extracting sparse and representative landmarks from a tractography dataset, and second transforming curves relative to these landmarks with a closest point transform. We explore its use in three typical tasks: fiber bundle clustering, simplification, and selection across a population. The clustering algorithm groups fibers from single whole-brain datasets using a non-parametric k-means clustering algorithm, with performance compared with three alternative methods and across four datasets. The simplification algorithm removes redundant curves to improve interactive visualization, with performance gauged relative to random subsampling. The selection algorithm extracts bundles across a population using a one-class Gaussian classifier derived from an atlas prototype, with performance gauged by scan-rescan reliability and sensitivity to normal aging, as compared to manual mask-based selection. Our results demonstrate how the SCPT enables the novel application of existing vector-space ML algorithms to create effective and efficient tools for tractography processing. Our experimental data is available online, and our software implementation is available in the Quantitative Imaging Toolkit.

Keywords Diffusion MRI tractography · Clustering · Simplification · Segmentation · Fiber bundles · Sparse closest point transform · Neuroimaging

Introduction

Diffusion MR imaging provides a unique *in-vivo* probe of tissue microstructure through the sensing of water molecule diffusion patterns (Pierpaoli and Basser 1996). This technique is particularly valuable for characterizing the local features of white matter and for reconstructing the large scale structure of fiber bundles through tractography (Basser and Pierpaoli 1996). The size and complexity of tractography datasets can pose a challenge to the practical application of tractography in neuroimaging studies, as delineating fiber bundle pathways from whole brain tractography often involves expert anatomical knowledge and time consuming manual interaction (O'Donnell et al. 2013) (Lenglet et al. 2009). Computational processing of tractography datasets can both reduce the time invested by the human analyst and provide highly reliable measurements in large population studies (O'Donnell and Schultz 2015).

Tractography typically produces geometric models of white matter pathways, or tractograms, which are a collection of space curves that are each represented by a sequence of 3D coordinates sampled along the route taken through the brain. When developing algorithms to process tractograms, two main complications arise. First, each curve may have an arbitrary number and distribution of sampled points along its length, e.g. doubling the sampling density would give approximately the same pathway. Second, the start-to-end ordering is arbitrary and may be reversed without changing the path taken (because the diffusion process underlying the tractogram itself has no preferred forward-backward direction). Researchers have worked to

✉ Ryan P. Cabeen
rcabeen@ini.usc.edu

1 Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA

2 Department of Computer Science, Brown University, Providence, RI, USA

address these issues by using either distance-based or feature-based approaches to tractogram processing, which each have their own strengths and weaknesses.

Distance-based approaches process tractogram curves based on inter-curve proximity. A number of inter-curve distances have been proposed, including Hausdorff measures (O'Donnell et al. 2013), distance-based integrals (Zhang et al. 2008), endpoint distances, and others. Pairwise curve distances can then be used in any of number a clustering algorithms, such as spectral clustering (Brun et al. 2004) (O'Donnell and Westin 2007), hierarchical clustering (Zhang et al. 2008), or Dominant Sets (Dodero et al. 2015). These approaches can tease apart subtle distinctions between fibers (Moberts et al. 2005); however, the computational cost of total pairwise comparison of curves is prohibitively expensive for full datasets. This is usually mitigated by sub-sampling the full dataset to a feasible size, e.g. from a million down to less than 20,000.

In contrast, feature-based approaches map each curve to a summary coordinate representation to be more directly compared. The simplest approach is to measure connectivity between regions of interest (gray-matter areas or manually drawn volumetric masks) to select curves (Zhang et al. 2010) (Yendiki et al. 2011) (Wang et al. 2011). More sophisticated approaches have used implicit volumetric representations of curves, for example with distance fields (Maddah et al. 2008), spatial-angular histograms (Wang et al. 2011), and blurred indicator function (Wassermann et al. 2010); however, these can produce high-dimensional representations that can exhaust system memory when used for whole brain tractograms. Other approaches have aimed to derive arclength parameterizations of fiber curves with a fixed number of points (Clayden et al. 2007) (Corouge et al. 2006) (Gerig et al. 2004) (Garyfallidis et al. 2012). The bundles can be analyzed with respect to these points, but such algorithms require repeated flipping of the start-to-end orientation of each curve to reach some optimal consensus.

In either case, processing tractograms using distance- or a feature-based representations leads to limited options for segmentation and clustering. This motivated us to explore a representation that can fit more easily into the wealth of machine learning (ML) algorithms that already exist, as this could greatly simplify tractography pipelines. To fit this bill, we devised a novel feature-based representation that puts each curve into a vector-space, that is, a fixed-dimension coordinate system, which can plug directly into existing ML algorithms. We designed this approach to be sufficiently low-dimensional that entire tractograms can be stored, while at the same time retaining relevant anatomical information. The general idea of our approach is to use a set of anatomical landmarks to represent each curve, where the curve is summarized by the concatenated coordinates of points on the curve that are closest to each of the landmarks.

For example, if there were 100 landmarks, each curve would be represented by a 300 dimensional vector, consisting of the 100 points on the original curve that are closest to the landmarks. The landmarks are then defined in an atlas and deformed to an individual, but for some tasks, it is sufficient to define them on a subject-specific basis, as described later.

The primary goal of this paper is to investigate the general usefulness of this approach and to determine the extent that existing ML algorithms can be applied using it. We explore this through experiments involving three common tractography processing tasks: fiber bundle clustering, bundle simplification, and population-based bundle selection. The proposed fiber bundle clustering algorithm is sufficiently time efficient for interactive use and includes a mechanism to select the number of fiber bundles from the data. The proposed simplification algorithm can reduce the memory usage and improve performance of 3D rendering, while preserving the geometric structure of the full resolution dataset. The proposed population-based analysis builds a one-class classifier for each bundle, which can be used to select bundles from whole brain tractography without parameter tuning or manual intervention. We also perform evaluation experiments to gauge performance with *in vivo* human brain data and compare with other existing methods.

Methods

We now describe our proposed approach in detail and how it can be used in conjuction with several existing ML algorithms to solve three basic tractography processing tasks. These components are illustrated in Figs. 1 and 2. First, we will describe our method for mapping tractogram curves to a vector space representation using a closest point transform. Next, we will describe how bundle clustering and simplification can then be performed in single subjects with the DP-means algorithm. Finally, we will describe a population-based analysis for segmenting specific fiber bundles across multiple subjects with a one-class classifier and deformable volumetric registration.

In the following analysis, we make a simplifying assumption that a tractogram consisting of 3D curves has already been extracted, for example by streamline integration tractography (Zhang et al. 2003). We will then assume the curves are represented by polylines, i.e. sequences of points and the line segments connecting subsequent points. We will take the dataset to have N curves, where the i-th fiber has N_i points and each curve is denoted by $C_i \subset R^3$ for $i \in [1, N]$. However, we make no assumption about the spatial resolution or uniformity of the points sampled along each curve, or the start-to-end orientation.
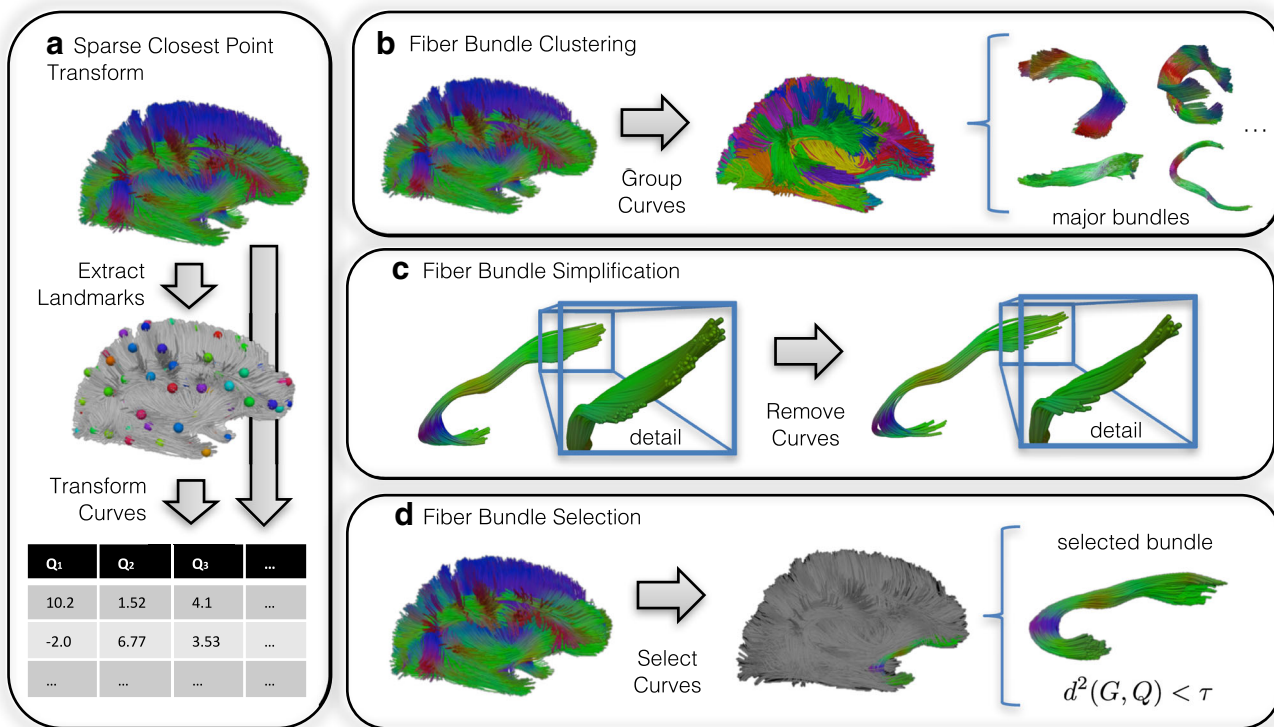
Fig. 1 Illustration of the major components presented in the paper. The left panel **A** shows the landmark extraction and sparse closest point transform steps, which produce a vector for each curve. The right panels show how several ML algorithms that can use this representation in several tractography processing tasks. The top right panel **B** shows an application of this to fiber bundle clustering, where whole brain tractography is decomposed into bundles using a variant of the k-means algorithms. The middle right panel **C** shows an bundle simplification, where redundant curves are removed to improve render speed and reduce memory usage. The bottom right panel **D** shows population-based analysis, in which a one-class-classifier is used to segment the bundle in each case

Tractogram Curve Representation

The proposed vector-space mapping of tractogram curve data uses the sparse closest point transform, which is a two step procedure described as follows.

Landmark Extraction. In the first step, a set of representative landmark positions is extracted algorithmically from the dataset. The process is illustrated in an example dataset in Figs. 1A and 2A. When peforming group analysis, e.g. selection of a specific bundle across a population, these
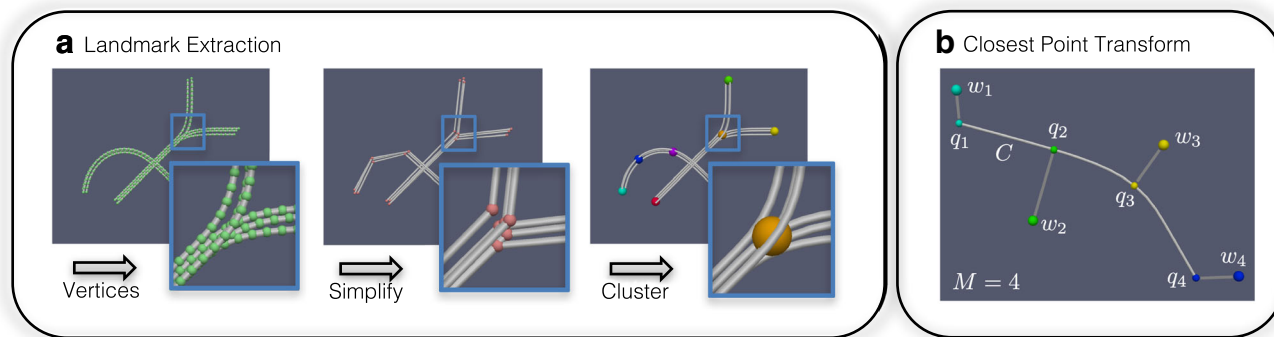

Fig. 2 Detailed illustration of the landmark extraction and closest point transform steps with digital phantom data. Panel A shows landmark extraction, including steps for vertex simplification (using the RDP algorithm) and clustering (using the DP-means algorithm). Panel B illustrates a simple case of the closest point transform. In this particular case, a curve C is transformed using four landmarks w_i, which results in four closest points q_i that are then contatenated to produce a single 12-dimensional vector representing the curve

landmarks should be defined in an atlas brain and then deformed to each subject; otherwise, the landmarks may be generated on subject-specific basis, avoiding the need for registration. Either way, the number of landmarks determines the coverage of the bundle representation, but fewer landmarks can reduce memory and compute time; therefore, we aim to select a small but representative set of landmarks to balance these two factors.

To extract an optimal set of landmarks, we first subsample the curves through random selection to obtain a small number of fibers, e.g. less than 5,000 (note: the full dataset is still retained for later analysis). These subsampled curves are then geometrically simplified using the Ramer-Douglas-Peucker (RDP) algorithm (Saalfeld 1999) to reduce the complexity of the curve vertices. The RDP procedure works by removing redundant points from the polyline while keeping the maximum simplification error below a given threshold, e.g. 2 mm. The vertices of the remaining simplified fibers are then taken together and further reduced by clustering. This clustering can include a specific number of points M by using the k-means algorithm, or alternatively determine the number of points M based on a spatial threshold by using the DP-means algorithm, which is described in detail later in the paper. In practice, we use the DP-means approach with a threshold of $\lambda = 5.0$. This produces a collection of landmarks that tend to be placed at salient features, such as points of high curvature and termination. We will denote the j-th landmark by $w_j \in R^3$ for $j \in [1, M]$.

Sparse Closest Point Transform. In the next step, a vector-space representation of the tractogram curve dataset is produced by applying the sparse closest point transform relative to the extracted landmarks. We refer to this approach as "sparse" to distinguish it from related approaches that densely sample closest points on a pixel or volumetric grid (Mauch 2000). The goal of using sparse landmarks is not only to reduce the dimensionality of the resulting representation but also to allow landmarks to be placed at salient points of the dataset to be emphasized, which may not otherwise be well aligned with a dense volumetric grid. The process is illustrated in Figs. 1A and 2B.

The closest point transform represents each tractogram curve by the positions along its length that are nearest to each of M reference landmarks. Given a curve C, the transformed curve is given by $Q = (q_1, ..., q_M) \in R^{3M}$, with the j-th closest point q_j given by:

$$q_j = \arg \min_{p \in C} \| p - w_j \|^2 \tag{1}$$

Note that with the polyline representation, this minimization can be performed not just over vertices but also along the connecting segments. This allows the algorithm to be applied to irregularly sampled curve data, such as those produced by the RDP algorithm. This can potentially reduce time and memory requirements compared to the uniform sampling typically required by other approaches.

Finally, the M closest closest points q_j are concatenated to produce a $3M$ dimensional vector for each curve, which is suitable to be used in vector-space ML algorithms. In the following sections, we describe how such an approach enables the novel application of existing ML algorithms to tractography processing tasks including fiber bundle clustering, simplification, and population-based analysis.

Fiber Bundle Clustering and Simplification

Based on this representation, fiber bundle clustering and simplification can be readily performed with an ML algorithm known as DP-means, which was introduced by Kulis et al. (Kulis and Jordan 2012) as a variant of the k-means algorithm. These steps are illustrated in Fig. 1B and C. As with k-means clustering, the goal here is to group vectors according to spatial proximity; however, the DP-means variant has the benefit of learning the number of clusters from the dataset. In comparison to other clustering algorithms, the k-means and DP-means algorithms are both notable for their efficient performance, which allows them to be applied to large datasets. This is a significant issue for clustering whole brain tractography, as the typical dataset size is far larger than is practical with $O(n^2)$ distance-based clustering algorithms. Past approaches have used subsampling to work around this problem, but the method proposed here is possible without subsampling.

This DP-means clustering algorithm can be defined by a minimization problem similar to the standard k-means algorithm with an additional regularization term. Prior work has shown this extension is theoretically related to Dirichlet Process mixture modeling with Gibbs sampling, through small-variance asymptotic analysis. This provides the algorithm a mechanism to choose the number of clusters in the dataset with the regularization term added to the typical k-means objective. When applied to a tractography dataset with the above fiber curve mapping, this process decomposes the set of curves into K bundles, where the k-th bundle is represented by a prototype $B_k \in R^{3M}$, and the optimal partitioning π is found by solving the following minimization problem:

$$\pi = \arg \min_{\pi} \sum_i^N \| Q_i - B_{\pi(i)} \|^2 + \lambda^2 M K \tag{2}$$

where $\pi(i) \in \{1, ..., K\}$ encodes the label for the i-th curve. This objective can be optimized with an efficient

algorithm that is described in detail in the referenced related work (Kulis and Jordan 2012). Since this method estimates the number of clusters, no user-defined value of K is explicitly specified. Instead, a regularization parameter λ is chosen, which can be interpreted as the largest average distance allowed between each curve Q_i and its associated bundle $B_{\pi(i)}$. When used for extracting bundles from whole brain tractography, a relatively large threshold can be used, e.g. $\lambda = 20$ mm. When used for simplification, a smaller threshold can be used, e.g. $\lambda = 2$ mm, and then each cluster is inspected to select the curve closest to the cluster centroid.

Fiber Bundle Selection

Next, we focus on a way to select a fiber bundle from a novel dataset based on an example bundle in a brain atlas. This is an important task for population-based analysis, where a brain structure is matched across a group. Our approach requires an atlas bundle, which can be generated in any way, e.g. by manual tractography seeding or by clustering whole brain tractography in an population average dataset. The process is illustrated in Fig. 1C. In our tests, we created a tractography bundle atlas using tensor-based deformable registration and selected several well known bundles from among the clusters. The goal is then to match curves from each subject to the bundles in the atlas. We accomplish this using an ML approach known as *one-class* classification, described as follows.

For a particular bundle, we build a statistical model based on the transformed space representation of a collection of fiber curves $\hat{Q} = \{\hat{Q}_1, ..., \hat{Q}_C\}$ by a multivariate Gaussian G with mean μ and covariance Σ. We fit the mean and the covariance using maximum likelihood (ML):

$$\mu_{ML} = \frac{1}{C}\sum_i^C \hat{Q}_i \tag{3}$$

$$\Sigma_{ML} = \frac{1}{C}\sum_i^C (\hat{Q}_i - \mu)(\hat{Q}_i - \mu)^T \tag{4}$$

Maximum likelihood estimation of covariance matrices can suffer from a number of problems, such as ill-posedness and overfitting for small sample sizes, so we regularize the model with a shrinkage estimator:

$$\Sigma_{Shrink} = \omega\,\mathrm{Diag}(\sigma_0^2) + (1 - \omega)\Sigma_{ML} \tag{5}$$

given a shrinkage coefficient ω and prior spatial variance σ_0^2. The shrinkage estimator not only ensures Σ is well-defined but also incorporates prior knowledge about expected misregistration errors, e.g. typical performance of

the atlas registration process as measured by the distance of misalignment. In practice, we used $\omega = 0.3$ and $\sigma_0^2 = 80$ mm.

The fitted model G can then be used to segment a bundle in a particular subject by first deforming the subject trac-togram to the atlas, taking the sparse closest point transform (relative to atlas space landmarks), and then finding the subset of curves that are sufficiently close to the modeled bundle. This last step is accomplished with a one-class classifier (Moya and Hush 1996), and we use a special case with a Gaussian distribution (Tarassenko et al. 1995). This approach is useful for classification tasks when only positive examples are available or when out-of-class examples are particularly complex (Tax 2001). To evaluate the classifier, we define a bundle-to-fiber distance $d(G, Q)$ with the squared Mahalanobis distace (Barnett and Lewis 1994):

$$d^2(G, Q) = (Q - \mu)^T \Sigma^{-1}(Q - \mu) \tag{6}$$

This has the benefit of measuring the statistical distance to known fibers in the bundle, as opposed to a spatial distance, whose optimal threshold would instead depend on the bundle size. By calculating the distance relative to the bundle covariance Σ, the distance can also reflect anisotropic spread along the bundle's length, for example due to fanning. This also provides a test statistic that can be used to decide if a given fiber is included in the bundle, i.e. if $d^2(G, Q) < \tau$ for some threshold critical value τ. Given that the Mahalanobis distance is Chi-squared distributed as $d^2 \sim \chi_{3M-1}^2$, we can look up such a value τ where $p(d^2(G, Q) < \tau) = 0.99$; using this threshold will retain 99% of fibers sampled from the distribution G. As long as the same number of landmarks are used, this threshold requires no tuning, and the same value can be used for all subjects, as well as other bundles.

Experiments and Results

We evaluated the proposed approach using *in vivo* and population-averaged human brain datasets in five experiments that examine the following: first, the inter-curve distance of our proposed representation; second, bundle clustering; third, bundle simplification; fourth, reliability of the automated bundle extraction; and fifth, an application to age-related neurodegeneration. All statistical analysis were implemented using R 3.1.1 (R Core Team 2015) with the ggplot2 package for plotting (Wickham 2009).

Datasets

Data Acquisition. Imaging data included diffusion-weighed MRI data acquired from healthy volunteers with a GE 1.5T scanner with a voxel size of 2 mm^3 and image size 128x128

and 72 slices. For each volunteer, a total of 71 volumes were acquired, with seven T_2-weighted volumes (b-value = 0 s/mm^2) and 64 diffusion-weighted volumes with distinct gradient encoding directions (b-value = 1000 s/mm^2). 80 volunteers were scanned with ages ranging between 25 and 65 years old and equal numbers in each sex; an additional five volunteers were scanned with three repetitions each to assess scan-rescan reliability (Cabeen et al. 2013).

Data Preprocessing. Diffusion-weighted MR image data was preprocessed using FSL (Jenkinson et al. 2012) as follows. First, the diffusion-weighted MR images were corrected for motion and eddy current artifacts by affine registration to the first T_2-weighted volume using FSL FLIRT with the mutual information cost function. The gradient encoding directions were rotated to account for the alignment (Leemans and Jones 2009), and non-brain tissue was removed using FSL BET. A diffusion tensor atlas was constructed from the 80 subjects. This was done by first fitting single tensor models using FSL DTIFIT and then constructing a population-specific atlas by deformable tensor registration using DTI-TK (Zhang et al. 2007). The 3x5 scan-rescan volumes were registered to the atlas, and the deformation fields were retained for each case.

Tractography. We generated tractography curves through the standard deterministic streamline approach implemented in the Quantitative Imaging Toolkit (QIT) (Cabeen et al. 2018), in which fiber trajectories are considered a 3D space curves whose tangent vector is equated with the fiber orientation of the voxelwise diffusion models (Zhang et al. 2003). This process proceeds by evolving a solution to a differential equation with some initial condition at a given seed position. During tracking, we computed fiber orientations from the principal tensor direction, estimated from the tricubicly interpolated diffusion-weighted image. We used the following tracking parameters: an angle threshold of 45°, step size of 1.0 mm, minimum fractional anisotropy 0.15, and 2 seeds per voxel. Whole brain tractography was performed in both the population atlas and each individual subject.

Experiment: Evaluating Inter-Curve distances

Design. This experiment evaluated how well several inter-curve distances reflect the structure of manually selected fiber bundles, comparing to the proposed fiber curve representation described in "Tractogram Curve Representation". We manually selected eight bundles from the atlas tractography dataset using a multiple region of interest approach with two inclusion and one exclusion mask. This included the anterior thalamic radiation, arcuate fasciculus, cingulum bundle, corticospinal tract, inferior longitudinal fasciculus,

forceps minor, forceps major, and uncinate fasciculus. The Dunn Index (DI) was used to measure the ability of a given inter-curve distance to distinguish between bundles. Given a distance measure $d(x, y)$ and manual labels π, this index measures the ratio of the minimum distance between clusters to the largest distance within clusters:

$$DI = \frac{\min_{\pi(a)!=\pi(b)} d(C_a, C_b)}{\max_{\pi(c)=\pi(d)} d(C_c, C_d)} \tag{7}$$

A higher score implies that clusters are well separated relative to the spread within the clusters, suggesting the given distance measure is useful for distinguishing the given bundles. We applied this test to the Euclidean distance in the feature space described in "Tractogram Curve Representation". For comparison, we also computed the Dunn Index for numerous other inter-curve distances, including the minimum endpoint distance, the Chamfer distance, and the Hausdorff distance. We also symmetrized the Chamfer and Hausdorff measures with the minimum, maxmum, and mean of the left and right oriented distances. We also included two variants on the closest point transform approach in which the landmarks are either chosen at random or placed on a volumetric grid. Each experiment was repeated with 30 resampling iterations to obtain an estimate of the expected Dunn Index and its uncertainty.

Results. We found the proposed vector-space Euclidean distance using the sparse closest point transform to have the highest Dunn Index of 0.95 (Fig. 3). However, performance depended on the number of landmarks included, with a critical number of landmarks being 15, afterwhich performance increased little. We also found runtime costs to increase linearly with the number of landmarks. Among the other distance measures, we found the mean-symmetrized Hausdorff distance performed best with a Dunn Index of 0.8. In comparison to other approaches to landmark selection, random landmarks performed worst, and gridded landmarks performed second best.

Experiment: Evaluating Fiber Bundle Clustering

Design. This experiment examined the clustering algorithm described in "Fiber Bundle Clustering and Simplification" and tested its ability to recover labels from manually delineated bundles. We used several test cases of varying difficulty, that is, varying degrees of bundle separation and overlap. We first manually delineated bundles similarly to the previous experiment, which represent somewhat idealized bundles without the heterogeneity of individual subjects. We also made an additional "harder" set of bundles with overlapping and contiguous pathways, including three segments of the body of the corpus callosum,
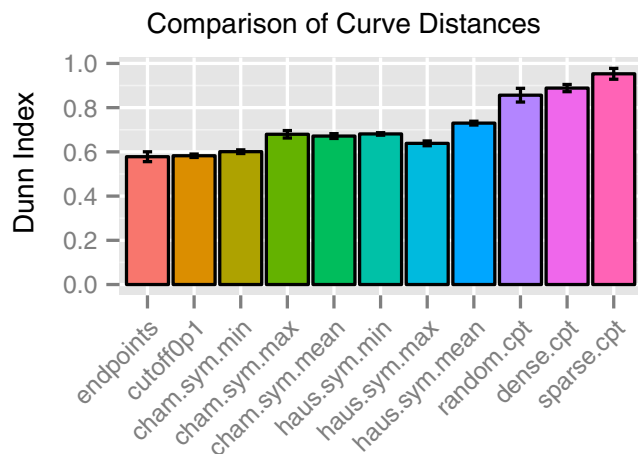
Comparison of Curve Distances

Fig. 3 Results from the first experiment described in "Experiment: Evaluating Inter-Curve distances", showing the Dunn Index for several inter-curve distance measures. A higher score indicates a closer relationship between the distance measure and manual bundle labeling. This shows that the Euclidean distance with the sparse closest point transform performs well relative to other options, although we found performance was dependent on the number of landmarks (not shown). In this dataset, the performance leveled out at 15 landmarks; however, we found that more complex datasets require more landmarks to see similar convergence. We also found runtime costs were also linearly related to the number of landmarks (not shown)

fornix, superior thalamic radiation, and posterior thalamic radiation. We further created a pair of easy and hard bundles for an individual subject, which includes more anatomical detail than the atlas. Together, this resulted in four test cases, which we refer to as *easy atlas*, *hard atlas*, *easy subject*, and *hard subject*.

We tested the proposed clustering algorithm by removing the bundle labels from each case, applying the clustering algorithm, and measuring agreement between the segmentation results and the original bundles labels with the Adjusted Rand Index (ARI) (Hubert and Arabie 1985). The Rand Index measures similarity between clustering labels, as represented by the proportion of label pairs that either agree or disagree among two datasets. The ARI gives a high score to labels that agree and includes an "adjustment" to give an expected score of zero to uniformly random labels. This was repeated with 30 resampling to obtain a the mean ARI and its uncertainty. We also ran this experiment over a range of λ threshold values, and performed similar experiments with spectral (O'Donnell and Westin 2007), hierarchical (Zhang et al. 2008), and quickbundles (Garyfallidis et al. 2012) clustering for comparison. We used the mean-symmetrized Hausdorff distance with spectral and hierarchical clustering, as it was the best performing in the previous experiment.

Results. We found that the best performing method depended on the dataset (Fig. 4). For the *easy atlas* dataset, all methods performed nearly perfectly (ARI > 0.98). For the *easy subject* dataset, spectral clustering performed worst

(ARI = 0.87), with the rest performing very well (ARI > 0.95). For the *hard atlas* dataset, hierarchical performed worst (ARI = 0.65), quickbundles was best (ARI = 0.82), and the proposed method (ARI = 0.77) was comparable to spectral (ARI = 0.77). For the *hard subject* dataset, hierarchical performed worst (ARI = 0.32), spectral was third from best (ARI = 0.54), and quickbundles (ARI = 0.60) performed slightly worse than the proposed method (ARI = 0.69).

Experiment: Evaluating Bundle Simplification

Design. The goal of this experiment was to evaluate the performance of the simplification algorithm on several fiber bundles in comparison to random resampling. The experiment was conducted with four bundle datasets selected from a single subject: forceps minor, forceps major, arcuate fasciculus, and uncinate fasciculus. The proposed simplification algorithm was applied with λ ranging from zero to five, and each condition was repeated 10 times to estimate stability. To provide a baseline comparison, we also tested a simple subsampling procedure was also tested: for given threshold from 0 to 100% a subset was randomly chosen and removed from the bundle. To evaluate performance, 2 mm^3 volumetric masks were created from the simplified bundles in each condition, and the Dice coefficient (Dice 1945) was measured between each of these masks and that of the full resolution bundle. Performance was plotted in relation to the percentage of curves retained from simplification.

Results. We found the proposed simplification algorithm to perform better than random subsampling (Fig. 5) across all tested bundles. The performance difference of the methods was greatest when 10 to 25% of the fibers were retained, in which case, the difference in Dice coefficient was greater than 0.05, representing about 15% improvement in performance relative to the low end dice score. Looking at aggregated runtime across datasets and parameter settings, subsampling was found to take 16.92 $\pm$ 0.63 ms, and the proposed method took 1179.87 $\pm$ 25.32 ms, which is practical for interactive usage.

Experiment: Evaluating Reliability

Design. This experiment applied the population-based analysis described in "Fiber Bundle Selection" to scan-rescan data and tested its ability to produce similar fiber bundle metrics across multiple *in vivo* scans of the same individual. We used 50 landmarks created from the full atlas tractography dataset to construct one-class classifiers for each bundle. For each of the 3x5 scan-rescan subjects, we followed the process described in
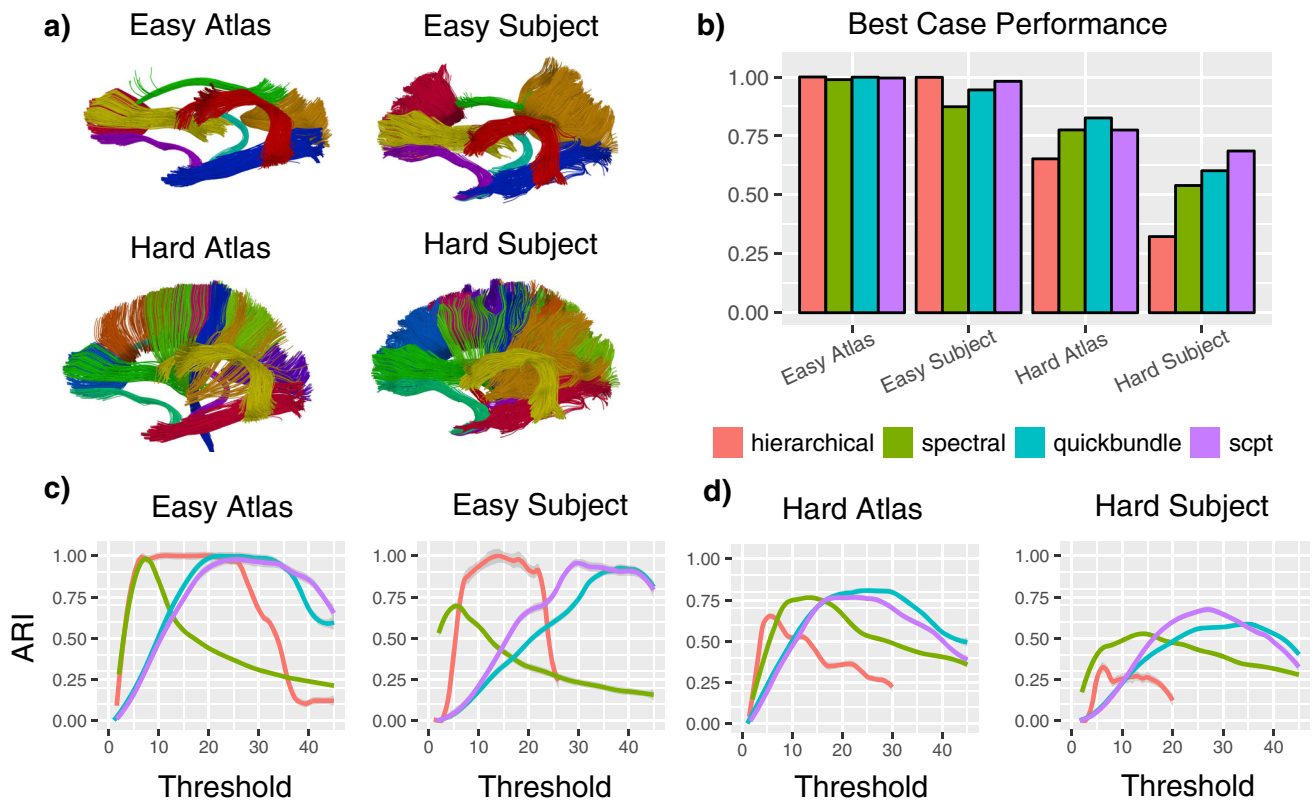
Fig. 4 Results from the second experiment described in "Experiment: Evaluating Fiber Bundle Clustering". The proposed clustering algorithm was applied to four datasets, which include atlas and subject data with a small (eight) larg number (fifteen) of bundles, as shown in the top left. We compared this approach to quickbundles clustering and spectral and hierarchical clustering with the mean symmetrized Hausdorff distance. The bottom four plots show performance as a function of clustering threshold, and the best case performance is shown in the bar chart at the top right. It should be noted that the behavior of

the threshold depends on the algorithm, so the peaks should not be expected to align, but the height of the peaks can be compared. The results show the performance varied across datasets. All methods performed well in the *easy atlas* and *easy subject* datasets. The *hard atlas* dataset showed quickbundles performed best, the proposed method showed performance that was comparable to spectral, and hierarchical performed the worst. The *hard subject* dataset showed the proposed method to outperform all others

"Fiber Bundle Selection" to segment fiber bundles with a threshold probability of $\tau = 0.99$. For each bundle, we computed five fiber bundle metrics: mean bundle length, and bundle-averaged fractional anisotropy, mean diffusivity,

radial diffusivity, and axial diffusivity (Correia et al. 2008). For comparison, we also performed bundle selection with a manual region-of-interest based approach described in the previous experiment.
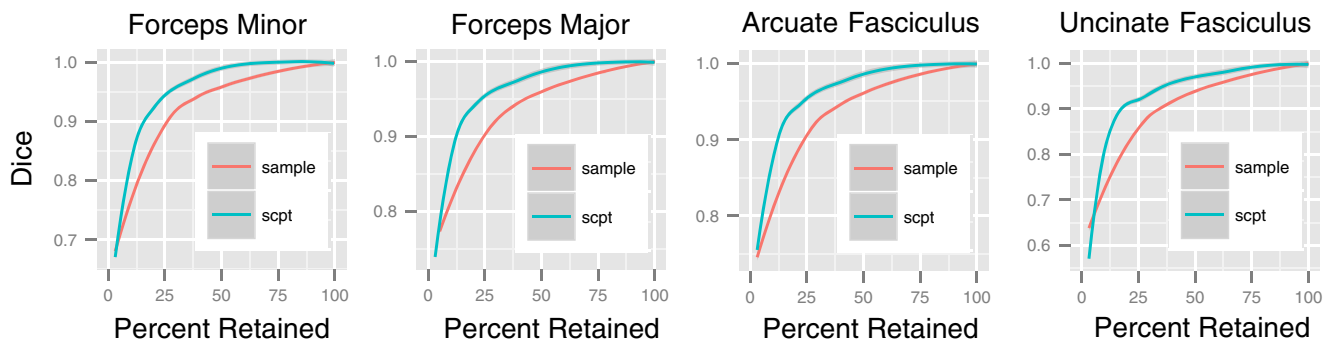

Fig. 5 Results from the third experiment described in "Experiment: Evaluating Bundle Simplification". This shows performance of the simplification algorithm in several bundles. Each plot shows the simplification error measured with the Dice coefficient. The proposed

method was compared to random sampling, and both methods are plotted by the percentage of curves retained from simplification. The results show the proposed method produces more accurate simplified bundles in nearly all cases
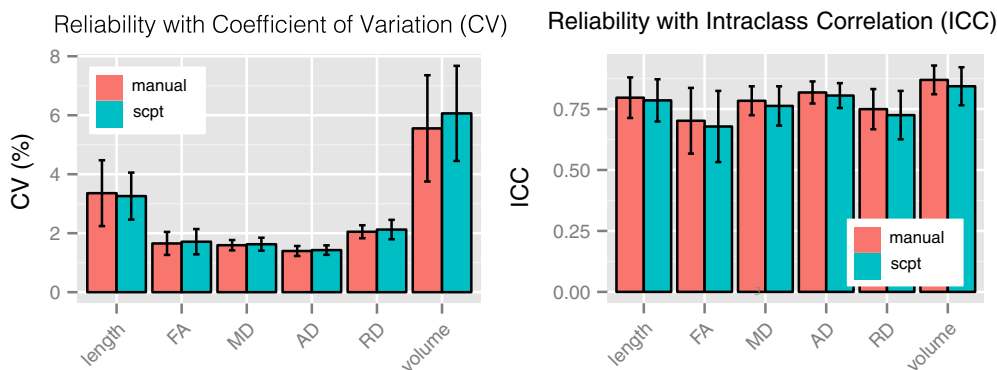
Reliability with Coefficient of Variation (CV)

Reliability with Intraclass Correlation (ICC)

Fig. 6 Results from the fourth experiment described in "Experiment: Evaluating Reliability" testing the scan-rescan reliability of the proposed population-based analysis. Performance was compared to manual region-based selection measured according to six fiber bundle metrics: fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), axial diffusivity, bundle length, and bundle volume.

The experiment included five subjects and three repeated scans, and the results showed comparable reliability between the manual and proposed automated approaches. The left and right plots report the coefficient of variation (lower is better) and intraclass correlation (higher is better) with 95% confidence intervals, respectively

We measured scan-rescan reliability of the fiber bundle metrics with the coefficient of variation (CV) (Hendricks and Robey 2008) and the intraclass correlation coefficient (ICC) (Bland and Altman 1990). The CV was measured for each subject from mean μ_s standard deviation σ_s, $CV = \sigma_s/\mu_s$. A lower CV score indicates higher reliability with units that are normalized to lie roughly between zero and one. The ICC was measured using the between-subjects variance σ_b^2 and within-subjects variance σ_w^2, with $ICC = \sigma_b^2/(\sigma_b^2 + \sigma_w^2)$ A larger ICC indicates there is more variance between than within subjects. This takes a maximum value of one and values above 0.75 indicate high reliability. Our implementation used the R 'ICC' package (Wolak et al. 2012).

Results . Our findings are summarized in Figs. 6 and 7. We generally found the bundles segmented with the proposed method to agree qualitatively with those found with the manual region-of-interest approach. However, quantitatively, the bundle metrics were not significantly different between the proposed and manual region-of-interest conditions. The bundle-averaged diffusion indicies were highly reliable with a CV of 2% and an ICC of above 0.7. The bundle-average length had a slightly higher CV of 5% and comparable ICC.

Experiment: Evaluating Sensitivity to Sging

Design. This experiment applies the proposed population-based analysis to the study of normal age-related neurodegeneration. The experiment was designed to examine the age associations of fiber bundle metrics, specifically, to compare our approach to a manual region of interest approach. The methods were applied to the 80 subject normal population to obtain five fiber bundle metrics of

the forceps minor, which traverses the anterior portion of the corpus callosum and has well-documented changes with age (Cabeen et al. 2017). This experiment compares linear regression models of subject age based on fiber bundles metrics, with intracranial volume included as a covariate. The performance in each condition was assessed using R^2 to indicate the total variance accounted for by
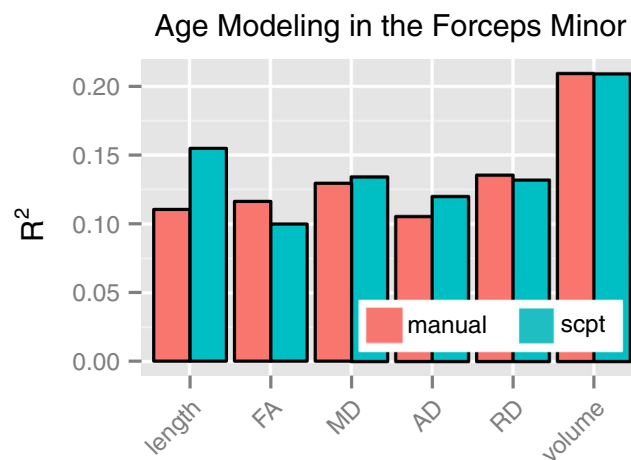
Age Modeling in the Forceps Minor

Fig. 7 Results from the fifth experiment described in "Experiment: Evaluating Sensitivity to Sging" testing the sensitivity to normal aging in a population of 80 subjects. The analysis focused on the forceps minor, which traverses the anterior portion of the corpus callosum and has well-documented changes with age (Cabeen et al. 2017). The proposed population-based analysis was compared to manual region-based selection according to six fiber bundle metrics: fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), axial diffusivity, bundle length, and bundle volume. For each bundle measure and method, multiple linear regression models were fit to model age with respect to intracranial volume, sex and each tractography-based metric. The plot shows the resulting R^2 of each model (bigger is better), showing comparable performance in bundle volume, MD, and RD, improved performance in bundle length, and AD, and slightly worse performance in FA

the model, and normalized coefficient β to indicate the effect size.

Results. The results support previous findings of age-related changes in prefrontal white matter (Fig. 6). Both manual region selection and the proposed method showed sensitivity of diffusion tensor measures to age, but several measures showed different performance characteristics. Three bundle metrics showed comparable performance: bundle volume ($R^2 = 0.21$, $\beta = -4.35$), RD ($R^2 = 0.13$, $\beta = 2.46$), and MD ($R^2 = 0.13$, $\beta = 2.315$). FA performed slightly better using manual selection ($R^2 = 0.12$, $\beta = -1.98$) than the proposed method ($R^2 = 0.10$, $\beta = -1.43$). AD performed slightly better with the proposed method ($R^2 = 0.12$, $\beta = 2.07$) than manual selection ($R^2 = 0.11$, $\beta = 1.63$). Bundle length performed significantly better using the proposed method ($R^2 = 0.16$, $\beta = -3.28$) than manual selection ($R^2 = 0.11$, $\beta = -2.39$).

Discussion

Our results generally indicate that the sparse closest point transform enables the novel applications of existing ML algorithms for tractography processing tasks, specifically for clustering large datasets, simplifying bundle geometry, and segmenting similar bundles across a population. The key component is that the resulting representation reduces the curves to vectors that retain geometric information and can be directly used with many ML approaches. Furthermore, it does this without introducing a prohibitively high dimensionality in the transformed space, due to the sparseness of the landmarks used in the transform. This contrasts with the standard dense volumetric grid representation of distance and closest point transforms, which require a regular sampling as dense as the smallest discriminative feature, e.g. the voxel size. Higher dimensionality can reduce the performance of learning algorithms, i.e. the peaking phenomenon (Sima and Dougherty 2008), which is another benefit of using this lower dimensional sparse representation.

Our first experiment suggests that careful selection of landmarks through geometry processing can improve performance compared to random or coarse regular sampling. This is likely because it can place landmarks near groups of endpoints and areas of high curvature, which are also likely to be the most discriminative features when comparing curves. The second experiment also indicated strengths of the proposed clustering method relative to the hierarchical, spectral, and quickbundles clustering algorithms. Because the proposed method avoids computing the full pairwise similarity matrix, it can be applied to large whole brain datasets without the typical subsampling used in the pairwise approaches, unlike distance-based clustering. The tests showed some advantages over quickbundles, which is possibly due to the distribution of samples along the curve. In quickbundles, these are equally spaced along the curve, while the proposed method tends to place samples near discriminative features. Another strength of the proposed clustering algorithm is the DP-means algorithm, which provides the efficient runtime and ability to select the number of clusters from the data while not requiring flipping of curves during optimization. Our third experiment shows how the same clustering algorithm is useful for simplifying fiber bundles, which can speed up 3D renderings and reduce disk space required for storing tractograms.

Our fourth and fifth experiments showed that the proposed population-based analysis is reliable across scans and also useful for mapping age-related change in white matter. This transform-based approach succeeds because it retains pose and shape information of the original curves, while being compatible with a simple segmentation algorithm using a Gaussian model. Furthermore, the proposed method compared favorably to manual bundle selection, which can be costly in time and require expert knowledge. It also outperformed the manual selection approach in the normal aging modeling based on fiber bundle length. This suggests that manual region-based selection perhaps excludes short fibers in the bundle that do not reach the regions, but are nevertheless related to age-related decline. Another interesting result was the good performance of the simple one-class classifier. The reliability and sensitivity were comparable to manual segmentation, making it preferable when analyzing datasets large enough to preclude manual intervention. An interesting open question for future work is whether there is a benefit to using more complex one-class classifiers, e.g. using support vectors (Tax 2001) or neural networks (Bishop 1994).

In general, this approach also has potential for broader application, beyond tractography processing. For example, curve clustering can be a useful tool for analyzing trajectory datasets, e.g. to better understand traffic flow or the behavior of particles in physical simulations. The bundle segmentation approach could also be useful in these other areas for categorizing incoming data or detecting anomalies. However, there are several limitations of the proposed method to note. First, the representation obtained from the sparse closest point transform is necessarily lossy, and there is a tradeoff between fidelity and runtime based on the number of landmarks. Our design attempted to address this by determining landmarks in a data-driven approach. The experiments here examined single fiber reconstructions, but multi-fiber tractography should be explored as well (Cabeen et al. 2016), as it is important to understanding

the performance of tractography processing algorithms in brain areas with crossing fibers. It also does not account for any time-domain information, but this could possibly be treated as a fourth dimension with appropriate scaling to be contatenated with the spatial coordinates.

In conclusion, we proposed a novel tractography curve representation using the sparse closest transform, and our experiments show how this enables the novel application of a variety of existing vector-space ML algorithms to tractography processing tasks. Looking forward, this approach is potentially useful for neuroimaging studies, such as anatomical localization in surgical planning and population studies that aim to quantify anatomical variation in white fiber bundles in relation to health and disease.

Information Sharing Statement

We have made our data and software implementation available online [1][2][3]. This work is included in the Quantitative Imaging Toolkit (QIT) [4] in the following modules:

CurvesLandmarks: A module for extracting landmarks, as described in "Tractogram Curve Representation" and Figs. 1A and 2A.

CurvesClosestPointTransform: A module for computing the sparse closest point transform, as described in "Tractogram Curve Representation" and Figs. 1A and 2B.

CurvesClusterSCPT: A module for curve clustering using the DP-means algorithm, as described in "Fiber Bundle Clustering and Simplification" and Fig. 1B and C.

CurvesSegmentBundleFitSCPT: A module for estimating one-class classifier model parameters given an example bundle, as described in "Fiber Bundle Selection" and Fig. 1D.

CurvesSegmentBundlesApplySCPT: A module for segmenting a bundle given a one-class classifier model, as described in "Fiber Bundle Selection" and Figure 1D.

CurvesOutlierSCPT: A module for detecting and removing outlier curves using the Mahalanobis distance, in a variation of the method described in "Fiber Bundle Selection".

---

[1] DOI:https://doi.org/10.35092/yhjc.12441953

[2] http://cs.brown.edu/research/mri/mri_repository.html

[3] https://resource.loni.usc.edu/resources/downloads

[4] https://cabeen.io/qitwiki

References

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* Vol. 3. New York: Wiley New York.

Basser, P.J., & Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance. Series B*, *111*(3), 209–219.

Bishop, C.M. (1994). Novelty detection and neural network validation. In *Vision, Image and Signal Processing, IEEE Proceedings*, (Vol. 141, IET pp. 217–222).

Bland, J.M., & Altman, D.G. (1990). A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Computers in biology and medicine*, *20*(5), 337–40.

Brun, A., Knutsson, H., Park, H.J., Shenton, M.E., Westin, C.-F. (2004). Clustering fiber traces using normalized cuts. *MICCAI*, *2004*(3216), 368–375. https://doi.org/10.1007/b100265.Clustering.

Cabeen, R.P., Bastin, M.E., Laidlaw, D.H. (2013). A Diffusion MRI Resource of 80 Age-varied Subjects with Neuropsychological and Demographic Measures. In *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*.

Cabeen, R.P., Bastin, M.E., Laidlaw, D.H. (2016). Kernel regression estimation of fiber orientation mixtures in diffusion MRI. *NeuroImage*, *127*, 158–172.

Cabeen, R.P., Bastin, M.E., Laidlaw, D.H. (2017). A Comparative evaluation of voxel-based spatial mapping in diffusion tensor imaging. *NeuroImage*, *146*, 100–112.

Cabeen, R.P., Laidlaw, D.H., Toga, A.W. (2018). Quantitative Imaging Toolkit: Software for Interactive 3D Visualization, Processing, and Analysis of Neuroimaging Datasets. In *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*.

Clayden, J.D., Storkey, A.J., Bastin, M.E. (2007). A Probabilistic Model-Based Approach to Consistent White Matter Tract Segmentation. *IEEE Transaction on Medical Imaging*, *26*(11), 1555–1561. https://doi.org/10.1109/TMI.2007.905826.

Corouge, I., Fletcher, P.T., Joshi, S.C., Gouttard, S., Gerig, G. (2006). Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Medical Image Analysis*, *10*(5), 786–798. https://doi.org/10.1016/j.media.2006.07.003.

Correia, S., Lee, S.Y., Voorn, T., Tate, D.F., Paul, R.H., Zhang, S., Salloway, S.P., Malloy, P.F., Laidlaw, D.H. (2008). Quantitative tractography metrics of white matter integrity in diffusion-tensor MRI. *NeuroImage*, *42*(2), 568–581.

Dice, L.R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*(3), 297–302.

Dodero, L., Vascon, S., Murino, V., Bifone, A., Gozzi, A., Sona, D. (2015). Automated multi-subject fiber clustering of mouse brain using dominant sets. *Frontiers in Neuroinformatics*, *8*, 1–12. https://doi.org/10.3389/fninf.2014.00087.

Garyfallidis, E., Brett, M., Correia, M.M., Williams, G.B., Nimmo-Smith, I. (2012). Quickbundles, a method for tractography simplification. *Frontiers in neuroscience*, *6*, 175.

Gerig, G., Gouttard, S., Corouge, I. (2004). Analysis of brain white matter via fiber tract modeling. In *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, (Vol. 2 pp. 4421–4424).

Hendricks, W.A., & Robey, K.W. (2008). The Sampling Distribution of the Coefficient of Variation. *Annals of Mathematical Statistics*, *7*(3), 129–132. https://doi.org/10.1214/193940307000000455.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*(1), 193–218.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M. (2012). FSL. *NeuroImage*, *62*, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015.

Kulis, B., & Jordan, M.I. (2012). Revisiting k-means: New algorithms via bayesian nonparametrics. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (pp. 513–520).

Leemans, A., & Jones, D.K. (2009). The B-matrix must be rotated when correcting for subject motion in DTI data. *Magnetic resonance in medicine*, *61*(6), 1336–49. https://doi.org/10.1002/mrm.21890.

Lenglet, C., Campbell, J.S.W., Descoteaux, M., Haro, G., Savadjiev, P., Wassermann, D., Anwander, Deriche, R., Pike, G.B., Sapiro, G., Siddiqi, K., Thompson, P.M. (2009). Mathematical methods for diffusion MRI processing. *NeuroImage*, *45*(1 Suppl), S111–22. https://doi.org/10.1016/j.neuroimage.2008.10.054.

Maddah, M., Grimson, W.E.L., Warfield, S.K., Wells, W.M. (2008). A unified framework for clustering and quantitative analysis of white matter fiber tracts. *Medical Image Analysis*, *12*(2), 191–202.

Mauch, S. (2000). A fast algorithm for computing the closest point and distance transform. Go online to http://www.acm.caltech.edu/seanm/software/cpt/cpt.pdf.

Moberts, B., Vilanova, A., van Wijk, J.J. (2005). Evaluation of fiber clustering methods for diffusion tensor imaging. In *VIS 05. IEEE Visualization, 2005.*, IEEE (pp. 65–72).

Moya, M.M., & Hush, D.R. (1996). Network constraints and multi-objective optimization for one-class classification. *Neural Networks*, *9*(3), 463–474.

O'Donnell, L.J., & Westin, C.F. (2007). Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas. *IEEE Transactions on Medical Imaging*, *26*(11), 1562–1575.

O'Donnell, L.J., Golby, A.J., Westin, C.-F. (2013). Fiber clustering versus the parcellation-based connectome. *NeuroImage*, *80*, 283–9. https://doi.org/10.1016/j.neuroimage.2013.04.066.

O'Donnell, L.J., & Schultz, T. (2015). Statistical and machine learning methods for neuroimaging: examples, challenges, and extensions to diffusion imaging data. In *Visualization and Processing of Higher Order Descriptors for Multi-Valued Data* (pp. 299–319): Springer.

Pierpaoli, C., & Basser, P.J. (1996). Toward a Quantitative Assessment of Diffusion Anisotropy. *Magnetic resonance in Medicine*, *36*(6), 893–906.

R Core Team (2015). *R: A language and environment for statistical computing,* R Foundation for Statistical Computing. Vienna, Austria.

Saalfeld, A. (1999). Topologically consistent line simplification with the Douglas-Peucker algorithm. *Cartography and Geographic Information Science*, *26*(1), 7–18.

Sima, C., & Dougherty, E.R. (2008). The peaking phenomenon in the presence of feature-selection. *Pattern Recognition Letters*, *29*(11), 1667–1674.

Tarassenko, L., Hayton, P., Cerneaz, N., Brady, M. (1995). Novelty detection for the identification of masses in mammograms. In *Artificial Neural Networks, 1995., Fourth International Conference on*, IET (pp. 442–447).

Tax, D.M. (2001). One-class classification. TU Delft, Delft University of Technology.

Wang, Q., Yap, P.-T., Wu, G., Shen, D. (2011). Fiber modeling and clustering based on neuroanatomical features. *MICCAI*, *14*(Pt 2), 17–24.

Wang, X., Grimson, W.E.L., Westin, C.F. (2011). Tractography segmentation using a hierarchical Dirichlet processes mixture model. *NeuroImage*, *54*(1), 290–302.

Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., Deriche, R. (2010). Unsupervised white matter fiber clustering and tract probability map generation: Applications of a gaussian process framework for white matter fibers. *NeuroImage*, *51*(1), 228–241.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer.

Wolak, M.E., Fairbairn, D.J., Paulsen, Y.R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, *3*(Boake 1989), 129–137. https://doi.org/10.1111/j.2041-210X.2011.00125.x.

Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., Jbabdi, S., Gollub, R., Fischl, B. (2011). Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Frontiers in neuroinformatics*, *5*(October), 23. https://doi.org/10.3389/fninf.2011.00023.

Zhang, H., Yushkevich, P., Rueckert, D., Gee, J.C. (2007). Unbiased white matter atlas construction using diffusion tensor images. *MICCAI*, *10*(Pt 2), 211–8.

Zhang, S., Correia, S., Laidlaw, D.H. (2008). 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), 1044–1053. https://doi.org/10.1109/TVCG.2008.52.

Zhang, S., Demiralp, C., Laidlaw, D.H. (2003). Visualizing diffusion tensor mr images using streamtubes and streamsurfaces. *Visualization and Computer Graphics, IEEE Transactions on*, *9*(4), 454–462.

Zhang, Y., Zhang, J., Oishi, K., Faria, A.V., Jiang, H. (2010). Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage*, *52*(4), 1289–1301. https://doi.org/10.1016/j.neuroimage.2010.05.049.Atlas-Guided.