



Published in final edited form as:

Med Image Anal. 2008 April ; 12(2): 191–202. doi:10.1016/j.media.2007.10.003.

A Unified Framework for Clustering and Quantitative Analysis of White Matter Fiber Tracts

Mahnaz Maddah^{a,b,*}, W. Eric L. Grimson^{a,b}, Simon K. Warfield^c, and William M. Wells^{a,b}

^aComputer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar St., Cambridge, USA., Phone: +1(617) 258-8832, Fax: +1(617) 258-6287, Email: mmaddah@mit.edu

^bSurgical Planning Laboratory, Harvard Medical School and Brigham and Women's Hospital, 75 Francis St., Boston, USA

^cComputational Radiology Laboratory, Children's Hospital, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Abstract

We present a novel approach for joint clustering and point-by-point mapping of white matter fiber pathways. Knowledge of the point correspondence along the fiber pathways is not only necessary for accurate clustering of the trajectories into fiber bundles, but also crucial for any tract-oriented quantitative analysis. We employ an expectation-maximization (EM) algorithm to cluster the trajectories in a Gamma mixture model context. The result of clustering is the probabilistic assignment of the fiber trajectories to each cluster, an estimate of the cluster parameters, i.e. spatial mean and variance, and point correspondences. The fiber bundles are modeled by the mean trajectory and its spatial variation. Point-by-point correspondence of the trajectories within a bundle is obtained by constructing a distance map and a label map from each cluster center at every iteration of the EM algorithm. This offers a time-efficient alternative to pairwise curve matching of all trajectories with respect to each cluster center. The proposed method has the potential to benefit from an anatomical atlas of fiber tracts by incorporating it as prior information in the EM algorithm. The algorithm is also capable of handling outliers in a principled way. The presented results confirm the efficiency and effectiveness of the proposed framework for quantitative analysis of diffusion tensor MRI.

Keywords

Clustering; Diffusion tensor MRI; Gamma mixture model; Tract-based quantitative analysis; White matter fiber tracts

Introduction

Diffusion tensor MR imaging (DT-MRI) is considered a powerful tool for identifying pathologies in white matter well before any structural change is visible in other imaging modalities (Werring et al., 1999). It measures the local diffusivity of water molecules within the tissue, and since the water diffusion is restricted in the direction normal to the fibers (Basser et al., 2000), it provides valuable information about the orientation of white matter fiber tracts.

*Corresponding author. *Email address:* mmaddah@mit.edu (Mahnaz Maddah).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Using a tractography scheme (eg. Basser et al. 2000), pathways of the fiber tracts, called trajectories in this work, can be extracted from the DT data and visualized as estimates of fiber bundles.

In recent years, a significant amount of work has been devoted to extracting information from DT images to study brain changes related to development (Huang et al., 2006), aging (Salat et al., 2005), and different pathologies such as multiple sclerosis (Werring et al., 1999), schizophrenia (Park et al., 2004), Parkinson's (Schocke et al., 2004), and Alzheimer's (Bozzali et al., 2002). Developing tools to perform accurate and comprehensive quantitative analysis is thus of great interest (Gerig et al., 2004). However, most clinical studies performed to date are limited to the analysis of local parameters, such as fractional anisotropy, measured in a manually defined region of interest (ROI) and averaged over groups of healthy and patient cases. Such ROI-based methods are not time-efficient as they require user interaction to specify the ROIs, either manually or semi-automatically, and their accuracy is limited by the reliability of specifying the ROIs by an expert. Others have performed a voxel-based analysis of a registered DTI dataset, which requires non-linear warping of the tensor field (Ruiz-Alzola et al., 2000), which in turn needs reorientation of the tensors (Alexander et al., 2001; Park et al., 2003). With a different strategy, Smith et al. (2006) perform FA Analysis by registering the skeleton of FA volumes across the subjects.

An alternative approach is to compute the quantitative parameters of interest along the tracts which makes more sense as the underlying anatomical unit in DT images is a fiber tract, not a voxel (Fillard and Gerig, 2003; Gerig et al., 2004). In fact, neuroscientists are very interested to do quantitative analysis along the tracts (See for example Goldberg-Zimring et al. (2005)). Such tract-based analysis is valuable, especially if performed on a group of trajectories that correspond to an anatomical fiber tract. The principal benefit offered by a tract-based analysis is that when comparing two groups, one would compare the same tracts. So, observed differences will be due to differences in the properties of the specific tracts rather than differences in the overall anatomy/shape of the individual brains. To accomplish this goal, algorithms are required to segment the trajectories into anatomically-known bundles. As suggested in our earlier work (Maddah et al., 2005), the anatomical information can be introduced to the algorithm by incorporating an atlas of fiber tracts. This would also solve the issue of determining the correspondence across subjects which is essential for statistical analysis.

There has been some work that addresses the issue of grouping the trajectories into bundles (Shimony et al., 2002; Brun et al., 2004; O'Donnell and Westin, 2005) but without going further into quantitative analysis. The outcome of these methods is a set of labeled trajectories, each assigned to a cluster. Point-by-point correspondence between the trajectories of each cluster, however, is not determined rigorously. In one of the early works on DT-MRI analysis, Ding et al. (2003) tackle the issue of quantification of tracts by finding the corresponding segments, which they defined as the portion of a trajectory that has point-wise correspondence to a portion of another trajectory. They assumed that the seed points of the two trajectories to be compared correspond to each other, which is not the case unless all trajectories are seeded from a small ROI. The algorithm is thus inadequate for whole brain fiber analysis.

Batchelor et al. (2006) proposed different tools to quantify the shape of fiber tracts. They noted the problem of point correspondence but made the assumption that it is approximately achieved by proper choice of the seed point and regularly sampling the arc-length. With point correspondence roughly known, they applied the Procrustes algorithm to register the trajectories. There is also a series of work by Gerig *et al.* (see for example Gerig et al. 2004), which described methods and applications for tract-oriented quantitative analysis. They dealt with the issue of correspondence by letting the user define a common origin for the set of

trajectories in each cluster, based on geometric criteria or based on anatomical landmarks. In their latest work (Corouge et al., 2006), they also proposed the Procrustes algorithm for the registration of the trajectories to compute the average tensor. Although these methods provide some valuable information about the fibers, their applicability is limited by their need to manual intervention for setting common start points for all the trajectories in a cluster. Also, they assume the trajectories in a cluster have the same length, which is a reasonable assumption only if a small ROI is considered as the tractography seed points and they end roughly in a common area. Otherwise a thorough preprocessing is required. In our earlier work (Maddah et al., 2006), we used a string matching algorithm to align all extracted trajectories with each cluster center at each iteration of our expectation-maximization (EM) clustering. The accuracy of this approach was limited by the simple curve matching algorithm used. More sophisticated three-dimensional (3-D) curve matching methods could be performed but at the expense of increased computational effort (Batchelor et al., 2006).

This paper presents a clustering method that aims to facilitate quantitative analysis as the next step in the study of DT images in a single subject or over a population. A statistical model of the fiber bundles is calculated as the mean and standard deviation of a parametric representation of the trajectories. Using this model representation, expectation-maximization (EM) is performed to cluster the trajectories in a mixture model framework. We obtain correspondence between points on trajectories within a bundle by building distance maps from each cluster center. To include anatomical information in the method, we take user-drawn curves or manually selected trajectories, each representing a cluster, as the initial cluster centers, and the algorithm probabilistically finds all trajectories in the dataset that are similar in shape and location to these curves. Note that in this work we only deal with clusters defined as a group of trajectories that are similar in shape and nearby in space. Also, the proposed method can potentially benefit from an atlas for the initialization step and as the prior map in the EM algorithm, which also makes the correspondence between different subjects known. All these tasks are done in a unified framework and the results are soft assignment of trajectories to labels and the point-by-point correspondence to each cluster center. The latter is essential for tract-based analysis, on which some examples are provided in Section 6. The approach proposed in this work for establishing point correspondences is computationally efficient and thus makes the framework capable of population studies in which the number of trajectories is quite large.

2 Similarity measure and point correspondence

Determining the point correspondence between a pair of trajectories is not a trivial task (Ding et al., 2003; Brun et al., 2004) but if achieved, not only makes the computation of their similarity, needed for clustering, straightforward, but also makes it possible to measure the quantitative parameters within a bundle. Although many authors acknowledge that point-by-point correspondence of the trajectories should be defined by a curve matching algorithm for accurate clustering and quantitative analysis (Gerig et al., 2004; Batchelor et al., 2006; Ding et al., 2003), to our knowledge this problem has not been solved to date. The difficulty lies in the fact that the number of trajectories is usually very large, especially when tractography is performed on the whole brain. This makes it computationally very inefficient, if not impossible, to perform a rigorous curve matching algorithm on every pair of trajectories. Leemans et al. (2006) find the closest subcurves in the curvature-torsion space and vary the length of the subcurves to deal with the curve matching problem. They further use a scale-space method to construct the space curves at different levels of detail. The spatial information is, however, not utilized in their approach. Furthermore, the time-efficiency of their method for whole brain clustering is questionable as the complexity of their approach grows with the square of the number of subcurves.

To get around the difficulty of performing a sophisticated curve matching algorithm, many authors attempt to define simple, yet reasonable similarity measures. For instance, Brun et al. (2004) defined a 9-D shape descriptor for each trajectory, including the mean and square root of the covariance matrix of its points and then computed the pairwise Euclidean distance of the descriptors. Corouge et al. (2004) implemented several distance measures such as closest point distance, mean distance of closest points, and the Hausdorff distance. Each distance metric has certain advantages and shortcomings (see Corouge et al. 2004). Here, we propose a novel approach for measuring the similarity of 3-D curves in a large dataset that includes the whole information of the curve for more accurate clustering and further quantitative analysis.

2.1 Basic Definitions

We treat each trajectory as a sampled three-dimensional (3-D) curve, i.e., an ordered set of points, $\mathbf{r}_i = \{\mathbf{r}_{ij}\}$. As will be shown in Section 3, the set of trajectories is clustered to a number of subsets by assigning a membership probability, p_{ik} , to each trajectory, \mathbf{r}_i , to denote its

membership in the k th cluster ($\forall i, \sum_{k=1}^K p_{ik} = 1$). The number of clusters, K , is defined by the number of curves (trajectories) that the user specifies as the initial cluster centers. For each cluster, a three-dimensional curve, $\boldsymbol{\mu}_k = \{\boldsymbol{\mu}_{kj}\}$, is defined as the cluster center where each point, $\boldsymbol{\mu}_{kj}$, is obtained as the average of all of its corresponding points on the trajectories. In an abstract form

$$\boldsymbol{\mu}_k = \sum_i p_{ik} \mathbf{r}_i^{(k)}, \quad (1)$$

where $\mathbf{r}_i^{(k)}$ is the trajectory \mathbf{r}_i , re-parametrized to have point correspondence to cluster k , and the summation is performed over all trajectories. We also define $\boldsymbol{\Lambda}_k = \{\boldsymbol{\Lambda}_{kj}\}$ where $\boldsymbol{\Lambda}_{kj}$ denotes a 3×3 covariance matrix attributed to the point $\boldsymbol{\mu}_{kj}$.

2.2 Similarity Measurement

Our space includes a set of 3-D curves and a number of cluster centers. From each center, $\boldsymbol{\mu}_k$, we construct an Euclidean distance map defined at each point \mathbf{x} as:

$$\mathcal{D}_k(\mathbf{x}) = \min_j d(\mathbf{x}, \boldsymbol{\mu}_{kj}) \quad (2)$$

and a label map, \mathcal{L}_k :

$$\mathcal{L}_k(\mathbf{x}) = \arg \min_j d(\mathbf{x}, \boldsymbol{\mu}_{kj}) \quad (3)$$

where $d(\mathbf{x}, \boldsymbol{\mu}_{kj})$ is the Euclidean distance from the point \mathbf{x} in the space to the j th point on the k th center. Each element of \mathcal{L}_k will thus contain the linear index of the nearest point of the center, $\boldsymbol{\mu}_k$. Fig. 1 shows the distance map and label map constructed from a sample 2-D curve. The label map partitions the space into Voronoi cells that each correspond to a point of the center. Now, for every curve, $\mathbf{r}_i = \{\mathbf{r}_{ij}\}$ in the space, the distance to the center $\boldsymbol{\mu}_k$ can be measured simply as:

$$d_E(\mathbf{r}_i, \boldsymbol{\mu}_k) = \sum_j \mathcal{D}_k(\mathbf{r}_{ij}), \quad (4)$$

and by projecting it onto the label map, its point correspondence to the center is readily achieved.

2.3 Adjustments

To model the spread of the trajectories in each cluster, a covariance matrix can be associated with each point on the cluster center. We denote the 3×3 covariance matrix attributed to the

point μ_{kj} on the k th cluster with Λ_{kj} . When computing the similarity of each trajectory to the center, the variability of the covariance matrix along the center should be taken into account, so that the trajectory points associated with a portion of the cluster center that has larger covariance are penalized less. This is done by computing the Mahalanobis distance:

$$d_M(\mathbf{r}_i, \mu_k) = \sqrt{\sum_{j=1}^{L_i} (\mathbf{r}_{ij}^{(k)} - \mu_{kj})^T \Lambda_{kj}^{-1} (\mathbf{r}_{ij}^{(k)} - \mu_{kj})} \quad (5)$$

where L_i is the length of the i th trajectory.

We also note that a spatial distance by itself does not encode enough information for measuring the pair-wise similarity. One obvious issue is the variable lengths of the trajectories. When a trajectory is shorter than the cluster center (Fig. 2.a), a penalty is added for each missing point. We set the penalty equal to the average distance of matched points of the trajectory from the cluster center. Another issue is whether the trajectory has one-to-one point correspondence to the cluster center, which is the case when they are similar in shape. Thus, any repeated or missing match represents shape dissimilarity. Another penalty term is therefore added if there are multiple points on the trajectory that correspond to a single point on the cluster center (Fig. 2.b). However, if multiple correspondence is merely because the trajectory is longer than the cluster center no penalty is added, since the trajectory is already penalized (Fig. 2.c). We add the penalty term, d_{penalty} as described above to the Mahalanobis distance and normalize to the length of the fiber trajectory, L_i . We denote the adjusted distance by $d_a(\mathbf{r}_i, \mu_k)$:

$$d_a(\mathbf{r}_i, \mu_k) = \frac{d_M(\mathbf{r}_i, \mu_k) + d_{\text{penalty}}(\mathbf{r}_i, \mu_k)}{L_i} \quad (6)$$

where d_{penalty} is equal to the number of missing points times the average distance of the matched points of the trajectory from the cluster center, d_{av} . An efficient way of implementing this in the algorithm is to set

$$d_{\text{penalty}}(\mathbf{r}_i, \mu_k) = ((L_k - L_i) + \text{Number of repeated matches}) \times d_{av} \quad (7)$$

where L_k and L_i are the length of the cluster center and the trajectory, respectively.

3 Probability Model

In the previous section, we mapped the variable-length 3-D curves to a distance matrix, $\mathbf{d}_a(\mathbf{r}_i, \mathbf{e}_k) = \{d_{ik}\}_{N \times K}$ where N is the number of curves and K is the number of clusters. The goal of this section is to estimate the membership likelihood of each curve to each cluster based on the values of the d_{ik} 's. Note that calculating the probability density of the points on the curves is not straightforward as they are not statistically independent. Therefore, we use the distance metric defined between the trajectories and cluster centers, d_{ik} 's, to build our probability model.

Density Estimation

A common choice for the density functions of the data points is the Gaussian distribution. However, a Gaussian distribution does not accurately represent the nature of the distance of the 3-D trajectories from the cluster centers. In the simplest form, the number of possible trajectories with a given distance from the center grows with the distance, while the probability that they belong to that cluster decays exponentially. Among the well-known distributions, the Gamma distribution captures this combined behavior very well, the monomial term increases with the distance while the exponential term decays. Given that the d_{ik} 's are non-negative, we assume that distances for each cluster follow a Gamma distribution with shape and inverse scale parameters α_k and β_k , respectively:

$$\text{Gamma}(d|\alpha_k, \beta_k) = d^{\alpha_k-1} \frac{\beta_k^{\alpha_k} e^{-\beta_k d}}{\Gamma(\alpha_k)} \text{ for } d \in [0, \infty) \quad (8)$$

where $\Gamma(\cdot)$ is the gamma function. The shape of the Gamma distribution is controlled by parameter α_k which is indicative of distance variation. For $\alpha_k = 1$ the distribution becomes an exponential distribution. When $\alpha_k > 1$, the distribution is bell-shaped and in the case of $\alpha_k < 1$, the distribution is highly skewed. Parameter β_k is the inverse scale parameter, and controls the decay rate.

Mixture Model

In mixture-model clustering, the data set is modeled by a finite number of density functions, where each cluster is represented by a parametric distribution:

$$p(x|\Theta) = \sum_k \omega_k p_k(x|\theta_k), \quad (9)$$

where x is a feature vector (distances in this case), ω_k 's are mixing weights, and Θ is the collection of parameters which will be defined later in this section, and p_k is the density function of cluster k parameterized by θ_k .

As described in the previous section, each trajectory is mapped to a vector of distances, $\mathbf{d}_i = [d_{i1} \dots d_{iK}]$. We define the mixture model in the space of the distance metric, \mathbf{d} , as follows:

$$p(\mathbf{d}_i|\Theta) = \sum_k \omega_k p_k(\mathbf{d}_i|\theta_k), \quad (10)$$

with the density functions given by:

$$p_k(\mathbf{d}_i|\theta_k) = \text{Gamma}(d_{ik}; \theta_k) \prod_{j \neq k} U(d_{ij}; 0, d_0), \quad (11)$$

where, θ_k is the parameter set of the Gamma distribution, $\{\alpha_k, \beta_k\}$, for each cluster and $U(x; 0; d_0)$ is the uniform distribution function over $[0; d_0]$ with d_0 a large enough constant. In the current implementation, ω_k 's are treated as unknown parameters, however they could be assumed as fixed values, taken from an anatomical prior. So, here, the goal is to infer $\Theta = \{\theta_k, \omega_k\}$ from a set of data points, \mathbf{d}_i 's, assumed to be samples of the mixture distribution:

$$\widehat{\Theta} = \arg \max_{\Theta} \{\log p(\mathbf{d}|\Theta) + \log p(\Theta)\} \quad (12)$$

which gives the *maximum a posteriori* (MAP) estimation of parameters, or the *maximum likelihood* (ML) estimation if the $\log p(\Theta)$ term is omitted. Since neither of these estimations can be found analytically, the frequently used approach is to incorporate the *Expectation-Maximization* (EM) algorithm, to find the local maximum of the likelihood function. In this paper, we only address the ML estimation of the unknown parameters along with the probabilistic assignments for each trajectory. However, if a probabilistic distribution of the parameters, i.e. $p(\Theta)$, is supplied by an atlas, we can do the MAP estimation of parameters. Supposing that the atlas is composed of labeled sets of trajectories for a number of subjects, such a prior distribution can be readily calculated.

Since, for each subject the membership of its trajectories to clusters is known, Θ can be calculated for each case. Then the probability distribution of Θ can be estimated over all subjects.

4 Expectation-Maximization Clustering

Expectation Maximization (EM), introduced by Dempster et al. (1977) is frequently used for fitting mixture models. To apply the EM approach in mixture model clustering, an indicator variable is usually defined to represent the hidden variable which is the cluster memberships. We denote this variable with Z , where $z_i = k$ represents the membership of the data point i to the cluster k and $p(z_i = k) = \omega_k$'s are the unknown mixing parameters. We denote the complete data likelihood by $p(\mathbf{d}, \mathbf{z}|\Theta)$. The goal of the EM algorithm is to iteratively find the ML estimates of the parameter Θ by maximizing the expectation of log likelihood of the complete data at each iteration:

$$\begin{aligned}
 \hat{\Theta} &= \arg \max_{\Theta} E_{\mathbf{z}|\mathbf{d}, \Theta_{\text{old}}} \left[\log \prod_i p(\mathbf{d}_i, \mathbf{z}|\Theta) \right] \\
 &= \arg \max_{\Theta} \sum_i E_{\mathbf{z}|\mathbf{d}, \Theta_{\text{old}}} [\log p(\mathbf{d}_i, \mathbf{z}|\Theta)] \\
 &= \arg \max_{\Theta} \sum_i \sum_k p(z_i=k|\mathbf{d}_i, \Theta_{\text{old}}) \log [p(\mathbf{d}_i|z_i=k, \Theta)p(z_i=k|\Theta)] \\
 &= \arg \max_{\Theta} \sum_i \sum_k p(z_i=k|\mathbf{d}_i, \Theta_{\text{old}}) \\
 &\quad \left(\log \text{Gamma}(\mathbf{d}_{ik}; \theta_k) + \log \left[\prod_{j \neq k} U(d_{ij}; 0, d_0) + \log p(z_i=k) \right] \right)
 \end{aligned} \tag{13}$$

Note that the last two terms in the parenthesis do not depend on Θ . The above expression is solved in two alternating steps:

4.1 Membership assignments (E-Step)

Assuming that the parameters of the clusters are known, using Bayes' rule, the probability that the trajectory \mathbf{r}_i belongs to cluster k is

$$p_{ik} = \Pr(z_i=k|\mathbf{d}_i, \Theta_{\text{old}}) = \frac{\Pr(z_i=k)p(\mathbf{d}_i|z_i=k, \theta_k)}{\sum_k \Pr(z_i=k)p(\mathbf{d}_i|z_i=k, \theta_k)} \tag{14}$$

where $p(\mathbf{d}_i|z_i = k, \theta_k) = \text{Gamma}(d_{ik}|\alpha_k, \beta_k) \prod_{j \neq k} U(d_{ij}; 0; d_0)$.

4.2 Updating model parameters (M-Step)

In this step, the new parameters are estimated by solving (13) by setting the derivatives with respect to each parameter equal to zero. Solving with respect to β_k , results in the following equation:

$$\beta_k = \alpha_k \sum_i p_{ik} / \sum_i p_{ik} d_{ik} \tag{15}$$

whereas for α_k , the following equation is obtained, for which no closed form solution exists:

$$\log(\alpha_k) - \psi(\alpha_k) = \log \left(\frac{\sum_i p_{ik} d_{ik}}{\sum_i p_{ik}} \right) - \frac{p_{ik} \log(d_{ik})}{\sum_i p_{ik}} \tag{16}$$

where $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$ is the digamma function. Using the approximation $\log(\alpha) - \psi(\alpha) \approx 1/\alpha(1/2 + 1/(12\alpha + 2))$ (Abramowitz and Stegun, 1964), an approximate value for α_k is calculated as:

$$\alpha_k \approx \frac{3 - x + \sqrt{(x-3)^2 + 24x}}{12x} \tag{17}$$

where

$$x = \log \left(\frac{\sum_i p_{ik} d_{ik}}{\sum_i p_{ik}} \right) - \frac{\sum_i p_{ik} \log(d_{ik})}{\sum_i p_{ik}}. \quad (18)$$

The updated mixing weights are also calculated as

$$\mu_k = \frac{1}{N} \sum_{i=1}^N p_{ik}. \quad (19)$$

where N is the number of trajectories.

When convergence of the above two steps is reached, values of p_{ik} 's are used to re-calculate the cluster centers by the following abstract formulas:

$$\mu_k = \frac{\sum_i p_{ik} \mathbf{r}_i^{(k)}}{\sum_i p_{ik}} \quad (20)$$

$$\Lambda_k = \frac{\sum_i p_{ik} (\mathbf{r}_i^{(k)} - \mu_k)(\mathbf{r}_i^{(k)} - \mu_k)^T}{\sum_i p_{ik}} \quad (21)$$

Note that in computing covariance matrices, when trajectories have more than one point associated with a particular point on the cluster center, the closest one is considered. Also, note that ideally points that correspond to a particular point on the cluster center lie on the plane orthogonal to the curve at this point. This would be consistent with the fact the natural variation space of the points on the curve is only the orthogonal plane at each point. One could force the tangential component of the $\mathbf{r}_i^{(k)} - \mu_k$ to zero by subtracting it, i.e.

$$\Delta \mathbf{r}_i^{(k)} = \frac{\sum_i p_{ik} \Delta \mathbf{r}_i^{(k)} \Delta \mathbf{r}_i^{(k)T}}{\sum_i p_{ik}}, \quad (22)$$

with

$$\Delta \mathbf{r}_i^{(k)} = (\mathbf{r}_i^{(k)} - \mu_k) \left(1 - \frac{(\mathbf{r}_i^{(k)} - \mu_k) \cdot \mu_k}{\|(\mathbf{r}_i^{(k)} - \mu_k) \cdot \mu_k\|} \right). \quad (23)$$

5 Method

Experimental Data

High resolution line scan DT-MR images were acquired from healthy volunteers on a 3T General Electric Signa scanner (Milwaukee, WI) with repetition time/echo time of 2500/70 ms, and b-factors of 5 and 750 s/mm² for baseline image and six gradient directions, respectively. The spatial resolution of the images was 1.054×1.054×2 mm.

Tractography and Curve Parameterization

Trajectories were reconstructed from 3-D diffusion tensor data with the streamline tractography (Basser et al., 2000). In this method, the pathways of fibers are estimated by sequentially following the direction of the eigenvector associated with the largest eigenvalue, starting from the selected seed points. The tractography algorithm stops when it reaches a point with

fractional anisotropy (FA) less than 0.1 or when a change in the direction greater than 45° occurs.

As a three-dimensional curve, we represent each trajectory with an equally-spaced sequence of control points from its quintic B-spline representation. This provides a smooth representation of the 3-D curve to minimize the effect of local noise on the quantitative analysis, while maintaining enough information to calculate the curvature and torsion along the trajectory. In our implementation, the distance between the successive samples is 5 mm, and hence the number of samples are different for each trajectory.

EM Initialization

When starting the iterative EM algorithm with the E-Step, all the parameters, Θ , must be initialized. We set the shape parameter of the Gamma distribution equal to one to have an exponential distribution which favors those trajectories most similar to the initial center. The algorithm is not very sensitive to the initial value of the inverse scale parameter, β , initialized to ten in our implementation. However, one should be careful not to set it too small. Otherwise many trajectories are labeled as unclustered in the initial step. On the other hand, if β is set too large, the likelihood would be very small even for trajectories very similar to the cluster center. In our implementation, we have a minimum threshold on the membership likelihood (as discussed in next subsection) to ensure the robustness of the algorithm. The cluster centers can be initialized manually by selecting a set of trajectories with different shapes from the data. Dependency of the algorithm on the initial centers will be discussed in Section 6. The associated covariance matrix of each center initially does not play a significant role so we set it to the unity matrix. This reduces the computation of Mahalanobis distance to Euclidean distance through the similarity measurement, which is an acceptable approximation for the first iteration. As an alternative, the initial cluster centers can be supplied by an atlas of fiber tracts, if available, such that the mean trajectories of atlas clusters are employed after registration of the case to the atlas or vice versa.

Evolving Centers

The approach described in Section 3 by itself does not allow the length of the initialized centers to evolve during the EM iterations. In fact, the number of control points on the cluster center remains constant but the distance between successive points changes. It is desired, however, that the method be less dependent on the initial centers and flexible enough so that the centers of clusters are able to grow or shrink in length. To fulfill this goal we perform a uniform re-sampling of the updated centers after the M-stage at each iteration. In doing so, first the spline representation of centers is reconstructed from their updated samples, and then the arc length of the reconstructed center is uniformly sampled. This also ensures that cluster centers remain smooth.

Handling Outliers

In mixture-model clustering, it is assumed that each data point is modeled by the mixture of a finite number of component densities. However, in our case, there might be trajectories resulting from the tractography which do not resemble any of the user- or atlas-initialized cluster centers or are not valid at all due to the inaccuracy of the tractography stage. These outliers will lead to increased variance of the density functions if not properly handled. This would result in excessive spread of the bundles or even instability of the algorithm. An outlier is identified by imposing a threshold on the membership likelihoods. If the membership likelihood of a given trajectory in *all* clusters is less than the specified threshold, that trajectory will be removed from further data processing. In fact, with this threshold the heterogeneity within each cluster is controlled. The larger the threshold is, the more compact are the resulting bundles, and consequently the greater is the number of unclustered trajectories. Handling

outliers is not straightforward in previously proposed clustering schemes (Brun et al., 2004; O'Donnell and Westin, 2005; Maddah et al., 2006). Unlike those methods, we allow the distribution of each cluster to have a different set of parameters (α_k 's, β_k 's), all inferred from the data by using the EM algorithm, and hence the user needs to set only the mentioned threshold to effectively remove the outliers.

Computation Complexity

The computation time of the algorithm is approximately equal to: number of EM iterations $\times K \times (N \times T_t + T_{dm} + T_c)$, where T_t is the processing each trajectory requires to calculate its distance from a cluster center and the associated membership likelihood, T_{dm} is the computation time of constructing a distance map from a given cluster center, and T_c is the processing time to update the cluster parameters, which grows linearly with N . Note that the computational complexity in our algorithm increases linearly with the number of trajectories, whereas performing a common curve matching techniques is of the order of N^2 as it would require pairwise calculations. In terms of memory consumption, although one distance map is computed for each cluster in our approach, only one distance map is stored in the memory at a time.

6 Results and Discussion

Fig. 3 shows the results of clustering roughly 3000 trajectories from corpus-callosum, middle cerebellar peduncle, corticobulbar, and corticospinal tracts into 25 bundles. As the initial centers, 25 trajectories from the data were selected manually, each representing an expected cluster. With a threshold equal to 0.2 of the peak of each membership likelihood, 45 trajectories remained unlabeled. The membership probability of the trajectories to each cluster is obtained using the EM algorithm as described in Section 4. The trajectories in Fig. 3 are colored based on their maximum membership probabilities. To demonstrate how probabilistic labels can be visualized, we further divided one of the clusters in Fig. 3 into two bundles, by selecting two trajectories as the initial centers, as shown in Fig. 4. The trajectories are colored with a linear combination of blue and red based on their degree of membership to the two sub-clusters. When a threshold is applied to the posterior probabilities (or when the algorithm is run with a higher threshold for minimum likelihood function) more distinct bundles are achieved. (Fig. 4(c)).

Fig. 5 illustrates the evolution of the Gamma distribution for the above two sub-clusters. Convergence is achieved just after five iterations of the EM algorithm. The Gamma distribution was initialized with $\alpha = 1$, to value those trajectories that have no distance to the initial cluster center. However, as the algorithm proceeds, the Gamma distribution evolves from a very broad distribution to a narrow distribution with small but non-zero mode. This behavior is in fact expected: the number of possible trajectories with a given distance from the center grows linearly with the distance, however the probability that they belong to that cluster decays exponentially. The combined trend is well modeled with a gamma distribution. This shape evolution emphasizes the fact that Gaussian mixtures with a single Gaussian per bundle are not able to correctly model the distribution.

As a byproduct of the proposed clustering method, a spatial model of the fiber bundles represented by the mean trajectory and its spatial variation is also obtained. This is shown in Fig. 6 in which the abstract models of five fiber bundles are visualized by their spatial mean and isosurfaces corresponding to the mean plus three standard deviations (3σ) of the 3-D coordinates along the cluster center. Such an abstract spatial model for fiber bundles could be used for neurosurgery applications. It enables one to easily visualize the extent of the fiber tracts adjacent to the brain lesions to minimize the damage to the bundles when removing the lesion. Conventional DTMRI-based approaches usually visualize a map of the fractional anisotropy (Laundre et al., 2005) or bundle segmentations based on the FA map. However,

such approaches are not accurate at fiber crossings where FA is low. An alternative approach is to visualize the fiber trajectories (Jellison et al., 2004), but rendering such a large dataset is not efficient in real-time applications.

Having the mean trajectory for each bundle of fiber tract allows us to study the shape of the bundle. As an example, the curvature versus the normalized arc length of the cluster centers are plotted in Fig. 7 for each cluster shown in Fig. 6. For a 3-D curve, \mathbf{r} , parametrized as a function of the arc length, the curvature and torsion are defined as:

$$k = \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^3/2} \quad (24)$$

and

$$\tau = \frac{(\mathbf{r}' \times \mathbf{r}'') \cdot \mathbf{r}'''}{\|\mathbf{r}' \times \mathbf{r}''\|^2}, \quad (25)$$

respectively, where ' denotes differentiation with respect to the arc length. The B-spline representation of the fiber trajectories allows us to evaluate the derivatives analytically and avoid further numerical errors.

To investigate the sensitivity of the clustering to the initial centers, we randomly selected different sets of trajectories from each cluster shown in Fig. 6. At each run with one of the sets as the initial centers, the final centers obtained by the clustering algorithm are almost identical as shown in Fig. 8. This demonstrates the robustness of the algorithm with respect to the variations in the initial centers within each cluster.

To demonstrate the suitability of our framework for tract-oriented quantitative analysis, we compute the mean and variation of the anisotropy parameters along the trajectories for clusters shown in Fig. 6. If any of anisotropy measures, such as fractional anisotropy (FA) and each of the three eigenvalues, or any of the local shape descriptors, such as curvature and torsion, are available along the trajectories, calculation of the mean and standard deviation of that parameter is quite straightforward. In fact, with the point correspondence obtained using the distance map described in Section 2, no further alignment of the quantitative parameters is necessary. Thus, similar to the computation of spatial mean and covariance of the clusters, performed in the M-stage of the EM algorithm, the mean and standard deviation of the parameters are obtained considering the membership probabilities of trajectories. Fig. 9 shows the mean and standard deviation of FA values along the normalized arc length of the cluster centers for each of the bundle in Fig. 6. To highlight the accuracy of our approach in aligning the fiber trajectories, the FA values along the aligned versions of all trajectories are also plotted (blue curves). These plots reveal the pattern of the FA values in the examined bundles.

Being able to perform such an analysis is valuable since it enables us to study the local changes of quantitative parameters along the fibers. This is especially interesting for study of temporal changes of fiber tracts during brain development and also opens new possibilities to compare normal and pathological subjects. The unavailability of efficient tools for tract-oriented quantitative analysis has limited most clinical studies to date to evaluation of scalar averages of parameters over an ROI. Local variations that provide significant information about brain development (Berman et al., 2005) and pathologies (Goldberg-Zimring et al., 2005) are thus lost in the course of study.

7 Conclusion

We demonstrated the joint clustering and point-by-point mapping of white matter fiber trajectories using an EM algorithm in a Gamma mixture model context. It is not straightforward

to do a comprehensive comparison of different clustering methods. However, compared to other methods used for clustering of the fiber trajectories for the whole brain, our method benefits from several features: Group assignment is probabilistic and the Gamma distribution enabled us to effectively model the normalized distance of the trajectories from each cluster center. Point correspondence is determined in the course of clustering and this information is used for accurate calculation of the similarity measure. Point correspondence of trajectories was obtained by constructing a distance map from each cluster center at every EM iteration. This provides a time-efficient alternative to pairwise curve matching of all trajectories with respect to each cluster center. Probabilistic assignment of the trajectories to clusters was controlled with a minimum threshold on the membership likelihoods, to offer flexibility in trading off between the robustness of the clusters and the number of outliers. In other words, handling of outliers is done in a principled way, only one parameter (likelihood threshold) should be supplied by the user. The similarity measure uses the information from the whole trajectory, sampled uniformly along its length. Spatial information is used explicitly while shape similarity is included implicitly through penalties in the similarity measure. In its current implementation, the algorithm relies on user interaction, providing the initial centers, as the anatomical prior. An anatomical atlas of fiber tracts can be also incorporated in the algorithm as the prior information in a MAP framework. Point correspondence calculated in our algorithm is an essential requirement of the tract-orientated quantitative analysis which is overlooked in previous works.

Acknowledgment

The authors would like to thank C.-F. Westin, S. Haker, Y. Ivanov and A. Khakifirooz. Tensor estimation was performed with the unu tool, part of the Teem toolkit, available at <http://teem.sourceforge.net>. This work is supported in part by NIH grants P41 RR013218, U54 EB005149, U41 RR019703, P30 HD018655, R01 RR021885, R03 CA126466, and R21 MH067054 and by NSF ITR 0426558 from CIMIT, and grant RG 3478A2/2 from the NMSS.

References

- Abramowitz, M.; Stegun, IA. Handbook of Mathematical Functions. Dover Publications; 1964.
- Alexander DC, Pierpaoli C, Basser PJ, Gee JC. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans. Med. Imag* 2001;20:1131–1139.
- Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med* 2000;44:625–632. [PubMed: 11025519]
- Batchelor PG, Calamante F, Tournier JD, Atkinson D, Hill DL, Connelly A. Quantification of the shape of fiber tracts. *Magn. Reson. Med* 2006;55:894–903. [PubMed: 16526017]
- Berman JI, Mukherjee P, Partridge SC, Miller SP, Ferriero DM, Barkovich AJ, Vigneron DB, Henry RG. Quantitative diffusion tensor MRI fiber tractography of sensorimotor white matter development in premature infants. *NeuroImage* 2005;27:862–871. [PubMed: 15978841]
- Bozzali M, Falini A, Franceschi M, Cercignani M, Zuffi M, Scotti G, Comi G, Filippi M. White matter damage in Alzheimer's disease assessed in vivo using diffusion tensor magnetic resonance imaging. *J. Neurol. Neurosurg. Psychiatry* 2002;72:742–746. [PubMed: 12023417]
- Brun, A.; Knutsson, H.; Park, HJ.; Shenton, ME.; Westin, C-F. Seventh International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'04). Rennes - Saint Malo, France: Lecture Notes in Computer Science; 2004 Sep. Clustering fiber tracts using normalized cuts; p. 368-375.
- Corouge I, Gouttard S, Gerig G. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. *IEEE Int. Symp. Biomed. Imag* 2004:344–347.
- Corouge I, Fletcher PT, Joshi S, Gouttard S, Gerig G. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Med. Image Anal* 2006;10:786–798. [PubMed: 16926104]
- Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* 1977;39:138.

- Ding Z, Gore JC, Anderson AW. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Mag. Reson. Med* 2003;49:716–721.
- Fillard P, Gerig G. Analysis tool for diffusion tensor MRI. *MICCAI* 2003:967–968.
- Gerig G, Gouttard S, Corouge I. Analysis of brain white matter via fiber tract modeling. *Proc. IEEE Int. Conf. EMBS* 2004:4421–4424.
- Goldberg-Zimring D, Mewes AUJ, Maddah M, Warfield SK. Diffusion tensor magnetic resonance imaging in multiple sclerosis. *J. Neuroimaging* 2005;15:68S–81S. [PubMed: 16385020]
- Huang H, Zhang J, Wakana S, Zhang W, Ren T, Richards LJ, Yarowsky P, Donohue P, Graham E, v. Zijl PCM, Mori S. White and gray matter development in human fetal, newborn and pediatric brains. *Neuroimage* 2006;33:27–38. [PubMed: 16905335]
- Jellison BJ, Field AS, Medow J, Lazar M, Salamat MS, Alexander AL. Diffusion tensor imaging of cerebral white matter: A pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *American Journal of Neuroradiology* 2004;25:356–369. [PubMed: 15037456]
- Laundre BJ, Jellison BJ, Badie B, Alexander AL, Field AS. Diffusion tensor imaging of the corticospinal tract before and after mass resection as correlated with clinical motor findings: Preliminary data. *American Journal of Neuroradiology* 2005;26:791–796. [PubMed: 15814922]
- Leemans A, Sijbers J, De Backer S, Vandervliet E, Parizel P. Multiscale white matter fiber tract coregistration: A new feature-based approach to align diffusion tensor data. *Mag. Res. Med* 2006;55:1414–1423.
- Maddah M, Mewes AUJ, Haker S, Grimson WEL, Warfield SK. Automated atlas-based clustering of white matter fiber tracts from DTMRI. *MICCAI* 2005:188–195. [PubMed: 16685845]
- Maddah M, Grimson WEL, Warfield SK. Statistical modeling and EM clustering of white matter fiber tracts. *IEEE Int. Symp. Biomedical Imaging* 2006:53–56.
- O'Donnell, L.; Westin, C-F. Eighth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'05). Palm Springs, CA, USA: Lecture Notes in Computer Science 3749; 2005 Oct. White matter tract clustering and correspondence in populations; p. 140-147.
- Park HJ, Kubicki M, Shenton ME, Guimond A, McCarley RW, Maier SE, Kikinis R, Jolesz FA, Westin CF. Spatial normalization of diffusion tensor MRI using multiple channels. *NeuroImage* 2003;20:1995–2009. [PubMed: 14683705]
- Park HJ, Westin C-F, Kubicki M, Maier SE, Niznikiewicz M, Baer A, Frumin M, Kikinis R, Jolesz FA, McCarley RW, Shenton ME. White matter hemisphere asymmetries in healthy subjects and in schizophrenia: A diffusion tensor MRI study. *NeuroImage* 2004;24:213–223. [PubMed: 15325368]
- Ruiz-Alzola, J.; Westin, C-F.; Warfield, SK.; Nabavi, A.; Kikinis, R. Third International Conference on Medical Image Computing and Computer-Assisted Intervention. Pittsburgh: 2000 Oct 11–14. Non-rigid registration of 3d scalar vector and tensor medical data; p. 541-550.
- Salat DH, Tuch DS, Greve DN, van der Kouwe AJW, Hevelone ND, Zaleta AK, Rosen BR, Fischl B, Corkin S, Rosas HD, Dale AM. Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiology of Aging* 2005;26:1215–1227. [PubMed: 15917106]
- Schocke MF, Seppi K, Esterhammer R, Kremser C, Mair KJ, Czermak BV, Jaschke W, Poewe W, Wenning GK. Trace of diffusion tensor differentiates the parkinson variant of multiple system atrophy and Parkinson's disease. *NeuroImage* 2004;21:14431451.
- Shimony JS, Snyder AZ, Lori N, Conturo TE. Automated fuzzy clustering of neuronal pathways in diffusion tensor tracking. *Proc. Int. Soc. Mag. Reson. Med* 2002:10.
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* 2006;31:1487–1505. [PubMed: 16624579]
- Werring DJ, Clark CA, Barker GJ, Thompson AJ, Miller DH. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology* 1999;52:1626–1632. [PubMed: 10331689]

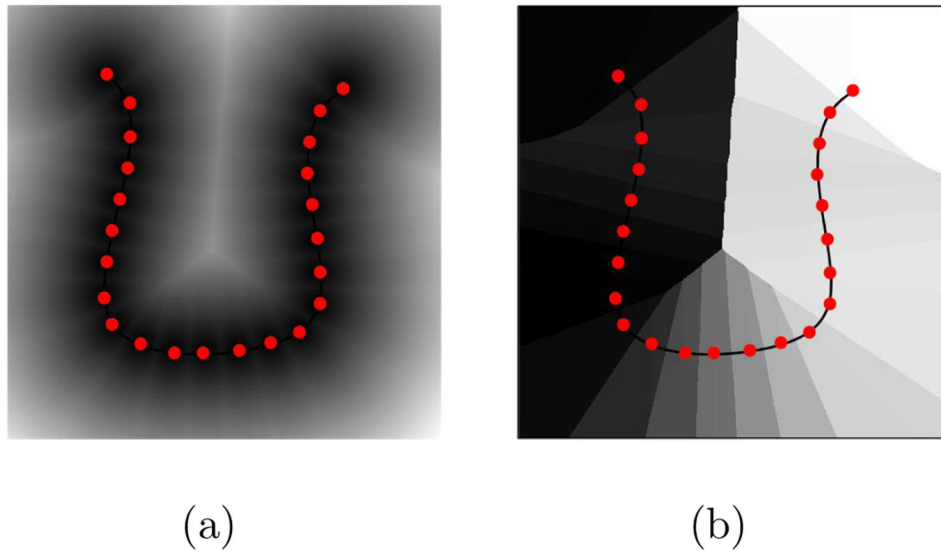


Fig. 1. (a) Distance map from sample points on a cluster center and (b) the point correspondence label map with the center overlaid. Each region in the label map, displayed by a different color, consists of all of the points in the space that have the minimum distance to a specific point on the cluster center. Therefore, projecting any curve onto this label map determines the point correspondence of each of its samples to the center based on which region that sample is located.

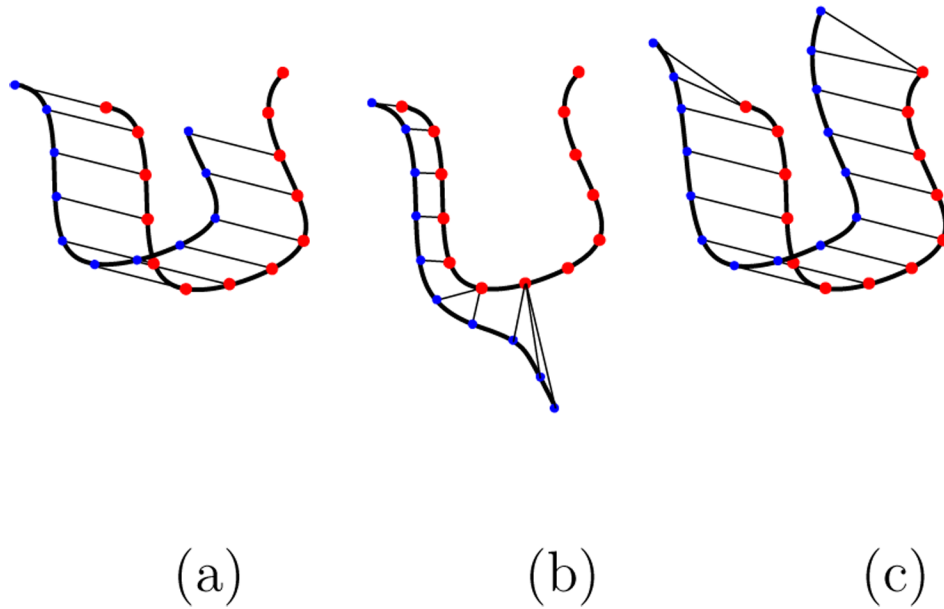


Fig. 2. Adjustments made to the total distance between the trajectory (blue) and the cluster center (red) calculated by Equation 5. The penalty is equal to the number of missing point times the average distance of the matched points, d_{av} . This is implemented more efficiently as $d_{\text{penalty}} = ((L_k - L_i) + \text{Number of repeated points}) \times d_{av}$, where L_k and L_i are the length of the cluster center and the trajectory, respectively. (a) $d_{\text{penalty}} = 2d_{av}$, (b) $d_{\text{penalty}} = 6d_{av}$, (c) $d_{\text{penalty}} = 0$.

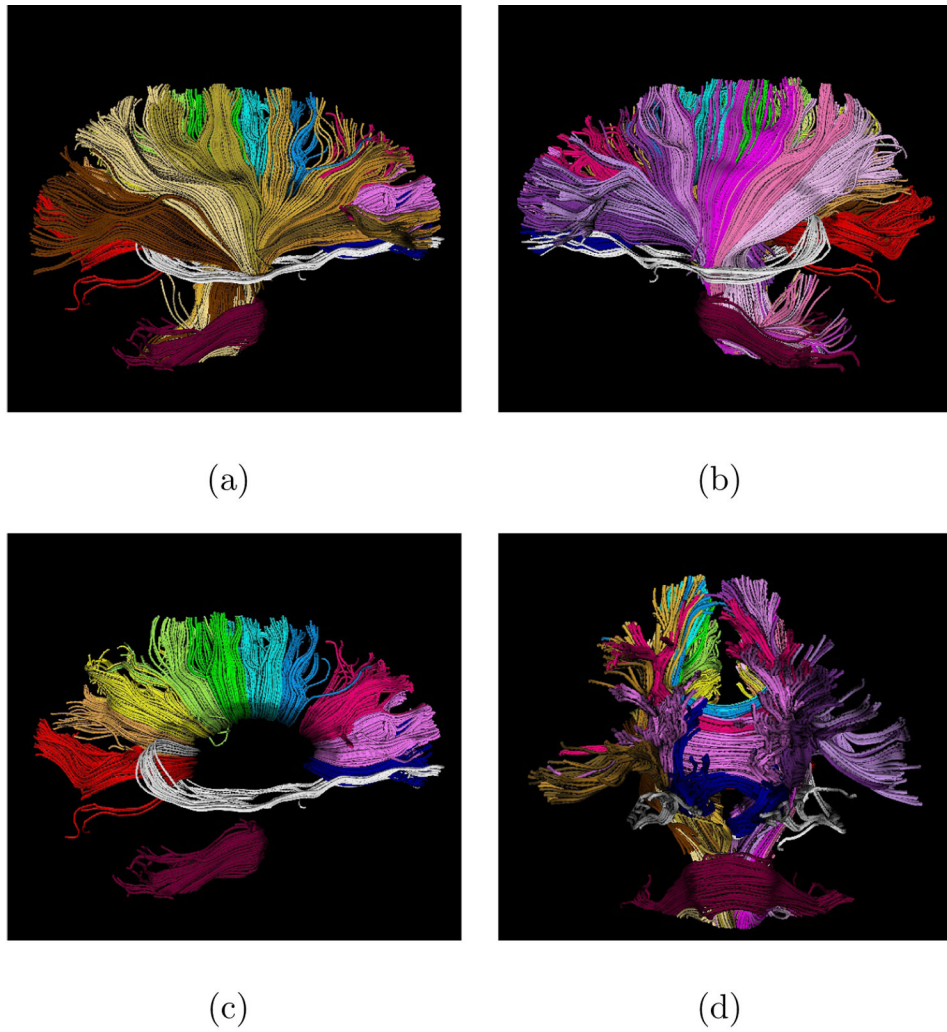


Fig. 3. About 3000 trajectories clustered to 25 user-initialized bundles. Clusters include different segments of the corpus callosum, tapetum, middle cerebellar peduncle, corticobulbar and corticospinal tracts, and different portions of thalamic radiation.

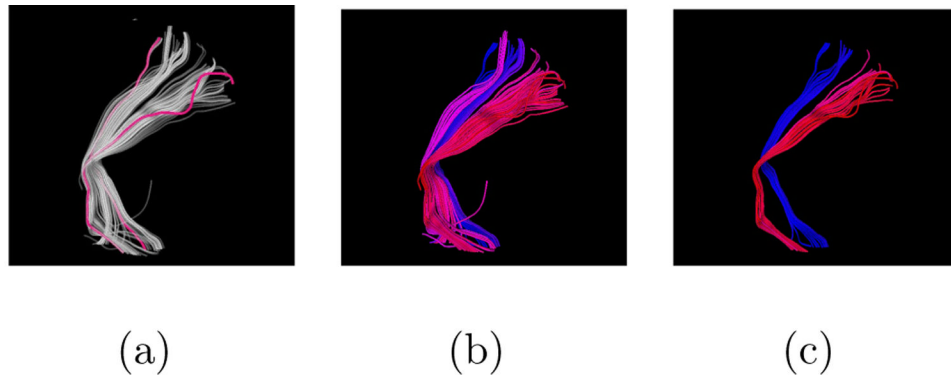


Fig. 4. Probabilistic visualization of corticospinal fiber trajectories (a) clustered into two sub-bundles (b). When a threshold is imposed on the minimum likelihood membership, a hard clustering (c) is obtained. The cluster centers are shown in pink in part (a).

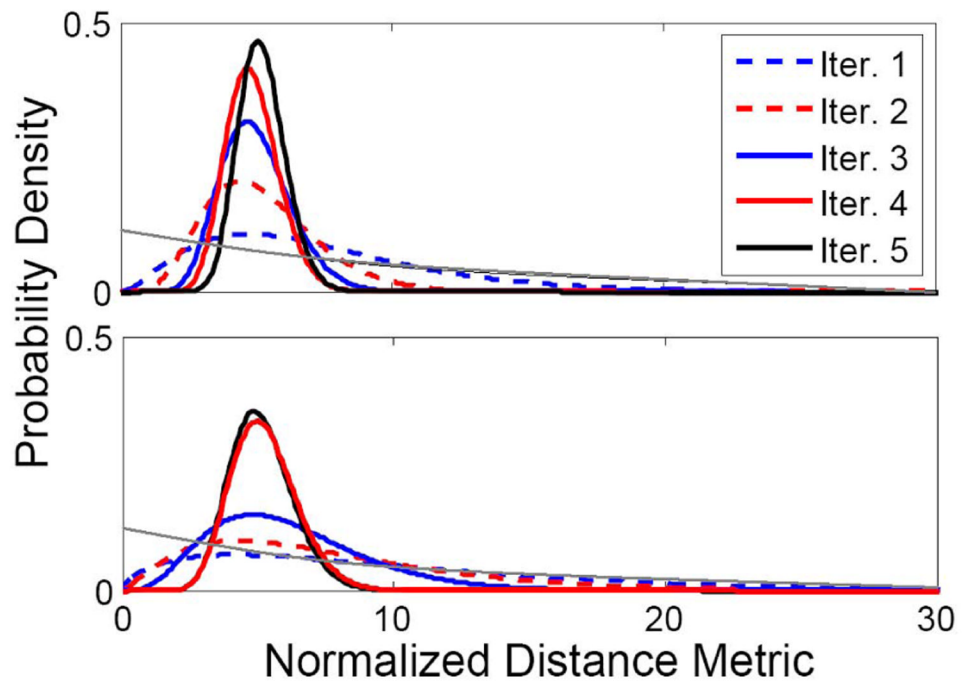


Fig. 5. Evolution of the Gamma distributions that describe the normalized distance metric for the two clusters shown in Fig. 4. After just five iterations, the distribution converges to a narrow distribution with small but non-zero mode.

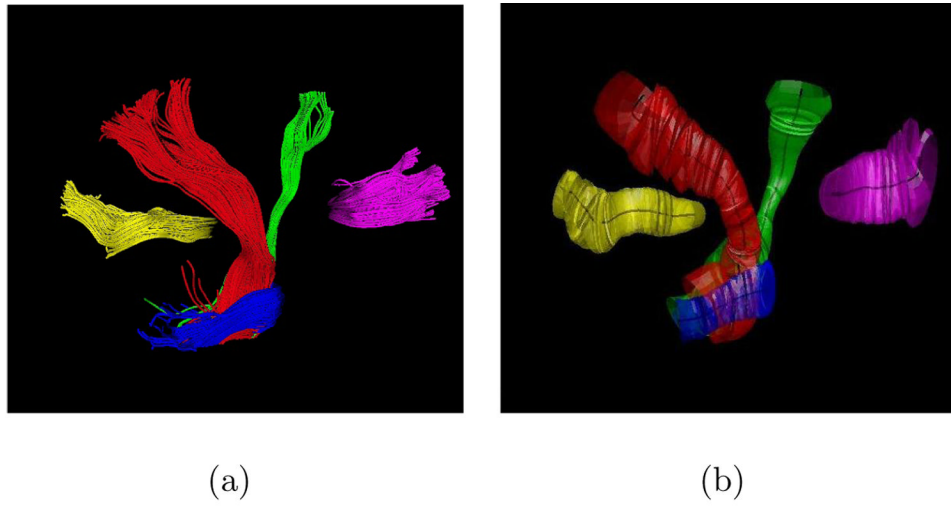


Fig. 6. (a) Trajectories of 5 different clusters used for quantitative analysis: splenium (yellow), corticospinal (red), corticobulbar (green), middle cerebellar peduncle (blue), and genu (magenta). (b) A model representation of the bundles as the mean trajectory and the isosurfaces corresponding to spatial variation of the clusters.

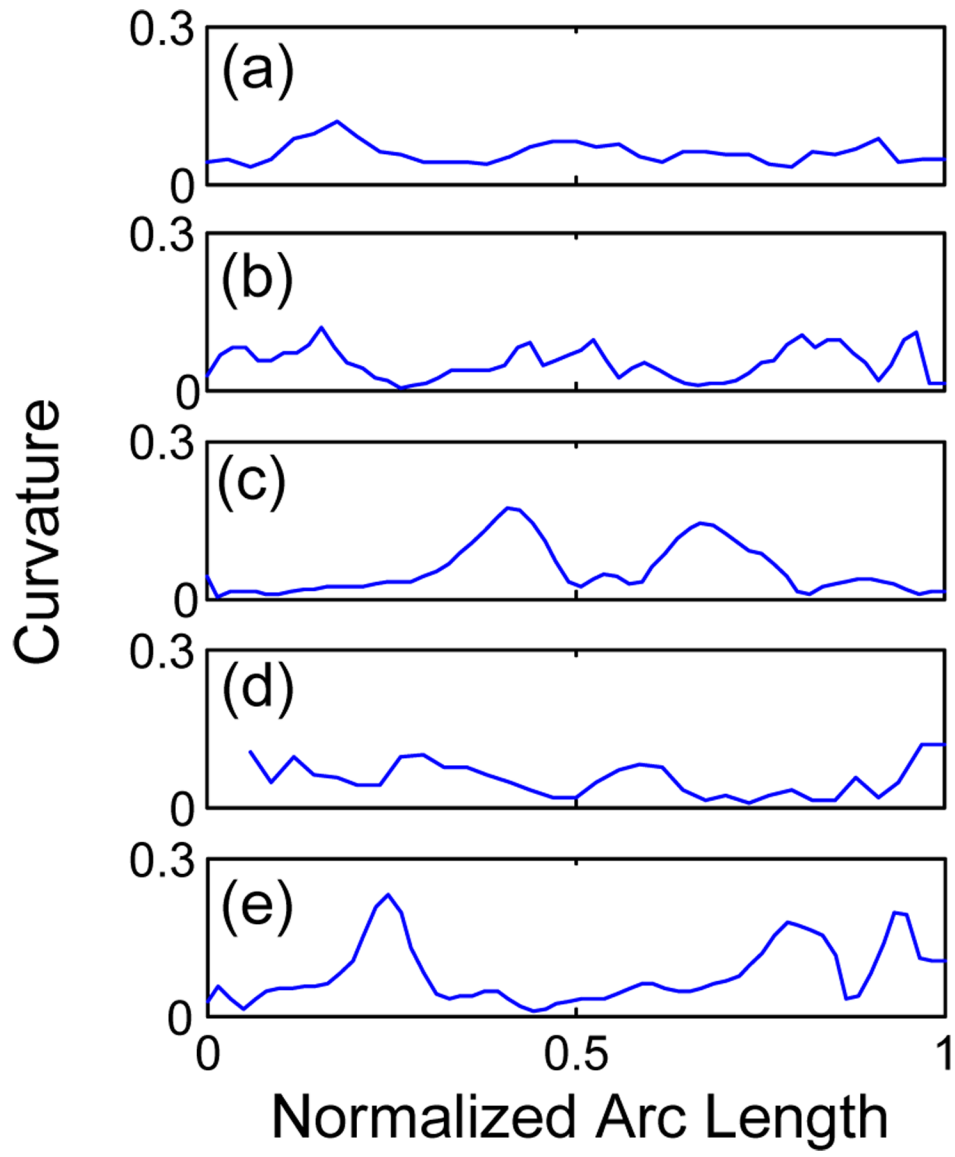


Fig. 7. Curvature of the cluster center along its normalized arc length for fiber bundles shown in Fig. 6: (a) splenium, (b) genu, (c) middle cerebellar peduncle, (d) corticospinal, and (e) corticobulbar fiber tracts.

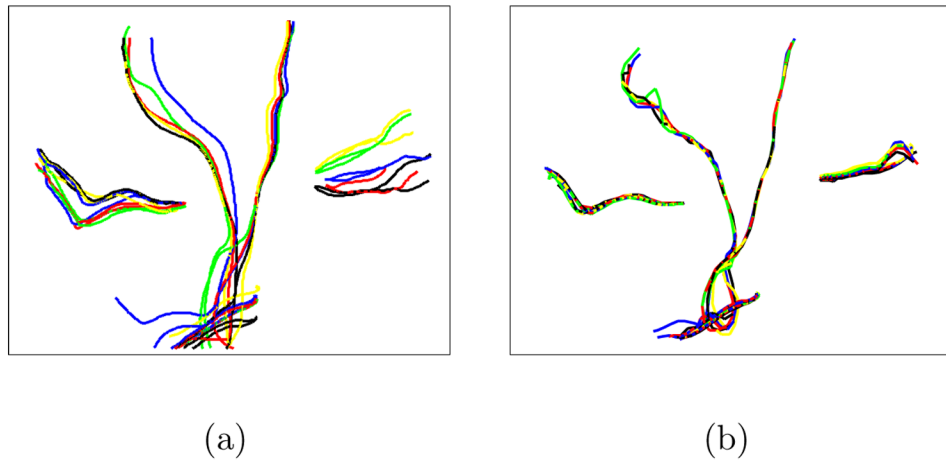


Fig. 8. Robustness of the EM algorithm with respect to the initial cluster centers. The algorithm was run 5 times with different initial centers (a) to cluster the trajectories in Fig. 6. Final cluster centers collected in (b) show little dependence on initial centers.

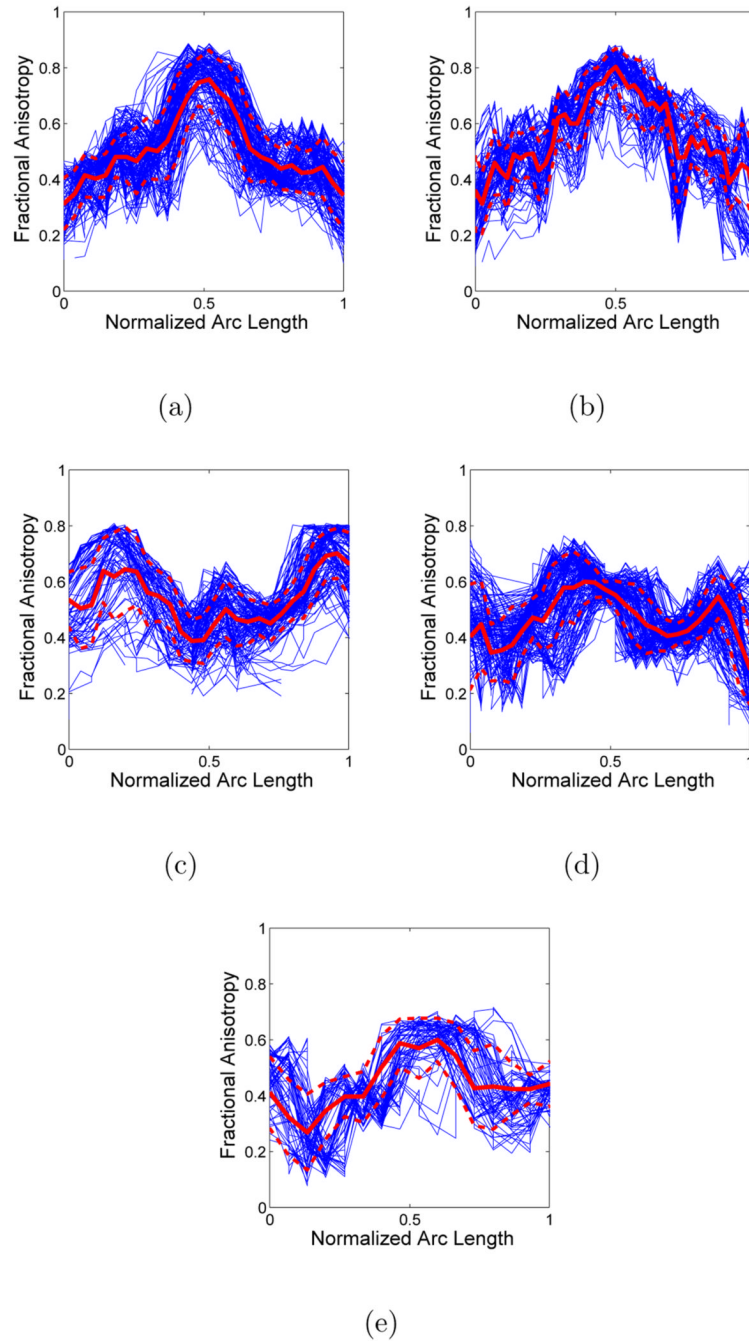


Fig. 9. Fractional anisotropy vs. normalized arc length for fiber bundles in Fig. 6: (a) splenium, (b) genu, (c) middle cerebellar peduncle, (d) corticospinal, and (e) corticobulbar fiber tracts.