

STATISTICAL MODELING AND EM CLUSTERING OF WHITE MATTER FIBER TRACTS

Mahnaz Maddah¹, W. Eric L. Grimson¹, Simon K. Warfield^{1,2,3}

¹ Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

² Computational Radiology Laboratory, Brigham and Women's Hospital,

³ Children's Hospital and Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

ABSTRACT

A statistical model of the fiber bundles is calculated as the average and standard deviation of a parametric representation of the fiber tracts, using the coefficients of the 3D quintic B -spline representation of the tracts. An atlas of the fiber tracts is constructed by averaging the bundle models over a population with the fiber tracts mapped onto the atlas coordinate. Using the model representation and with the atlas as the prior map, expectation-maximization (EM) is performed to cluster the fiber tracts in a mixture model framework. As an application, the method is applied to cluster the corpus callosum fiber tracts into its subdivisions and to calculate quantitative parameters for each region.

1. INTRODUCTION

Diffusion tensor MR imaging (DT-MRI) is considered a powerful tool to identify pathologies in white matter well before any structural change is visible in other imaging modalities [1]. It measures the local diffusivity of water molecules within the tissue, and thus provides some information about the density and orientation of white matter fiber tracts as the water diffusion is restricted in the direction normal to the fibers. Using a tractography scheme, pathways of the fiber tracts can be extracted from the DT data and visualized as fiber bundles to give a sense of the degree of connectivity between different functional regions. Any disease or pathology that affects the density or orientation of the fiber tracts can be detected via fiber tracking and comparison with an atlas of fiber bundles in normal cases. However, most clinical studies performed so far are limited to the analysis of the local parameters measured in a manually defined region of interest (ROI) and averaged over a set of healthy and patient cases. Such methods are

not time-efficient and their accuracy is limited by the reliability of specifying the ROIs by the expert. Others performed a voxel-based analysis of a registered DTI dataset, which requires non-linear wrapping of the tensor field and is prone to its associated problems [2, 3].

An alternative approach is to construct an atlas of the fiber tracts as 3D curves extracted from a set of DTI dataset and averaged over a population [4]. In its simplest form, an atlas can be a labeled set of fiber tracts extracted from DT-MR images of a group of healthy subjects and registered onto a given coordinate system [5]. However, it would be computationally expensive to compare a given tract with such an atlas. Also, the presence of any outlier in the atlas would deteriorate the quality of such comparisons. A more rigorous approach is to build a statistical model for each fiber bundle.

In this work we calculate a statistical model for each fiber bundle and construct an atlas as an aggregate set of such models averaged over a population. Furthermore, we employ this modeling framework along with the atlas as the prior map for mixture-model clustering of the fiber tracts. Closely related prior work are those on the shape modeling of the fiber bundles [6, 7], where each model is represented by a prototype and its trajectory in the space. The prototype is represented by the rotation and translation invariant parameters (curvature and torsion) of the fiber tracts as 3D curves plotted against the arc length and averaged over each bundle. However, the common origin of the plots is defined either manually or an anatomical landmark is chosen if possible (e.g. midsagittal points for corpus callosum). This complicates the application of the method in population studies. Other attempts on building an atlas of the bundles construct a probabilistic map of the occurrence of different fiber bundles at each voxel [8, 4].

We start by calculating a statistical model for each fiber bundle as the average and standard deviation of a parametric representation of the fiber tracts. The coefficients of the 3D quintic B -spline representation of the fibers is used as the parameters [9]. An atlas of the fiber tracts is constructed by averaging the bundle models over a population with the fiber tracts mapped onto the atlas coordinate. The transform pa-

The authors would like to thank W.M. Wells, Y. Ivanov, and P. Golland for helpful discussions. This work is supported in by NSF ITR 0426558, a research grant from the Whitaker Foundation, a research grant from CIMIT, grant RG 3478A2/2 from the NMSS, and by NIH grants R21 MH67054, R01 LM007861, R01 CA109246, P41 RR13218, 1U54 EB005149, and P01 CA67165.

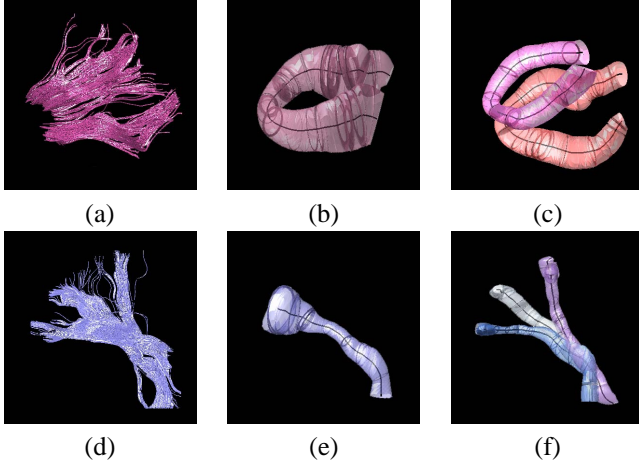


Fig. 1. Depending on the application, a set of fiber tracts can be modeled into a desired number of bundles. Modeling of the splenium fiber tracts (a) as either one (b) or two (c) bundles, and the capsula interna anterior (d) as either one (e) or three (f) bundles. The surfaces represents the mean tract plus the standard deviation.

rameters for this mapping is calculated by registering the T2 baseline images of each case to the atlas coordinate. Using the model representation and with the atlas as the prior map, we perform expectation-maximization to cluster the fiber tracts in a mixture model framework. As an application, the method is applied to cluster the callosal fiber tracts into corpus callosum subdivisions.

2. MODELING OF A BUNDLE OF FIBER TRACTS

To represent a bundle of fiber tracts with a limited number of parameters, the following issues must be considered: The tracts in a bundle are not aligned with each other, and they even do not have the same length. Furthermore, due to the presence of local noise and imaging imperfections, the extracted fiber tracts exhibit variations. There is more variation at the extreme parts of the tracts where the fractional anisotropy is rather low and thus the tractography algorithm is more sensitive to local noise and possibly its termination criteria. In other words, the medial part of the tracts are more reliable and should contribute more to the average shape of the bundle. Considering the above issues, we used the following scheme to model a bundle of fiber tracts, where each tract is represented with an equally-spaced sequence of control points from its quintic B -spline representation. We first apply a 3D string matching technique as described in [10] to find the correspondence between the control points on each tract to those on the longest tract in the bundle or a given reference tract. The mean and the standard deviation of the tracts in the bundle are then readily computed. We introduce a weighting factor as a measure of the reliability of each point

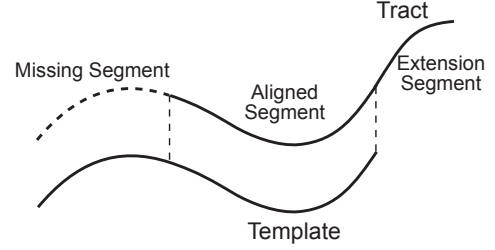


Fig. 2. Each tract is aligned to the template tract (current mean tract) of the cluster to either contribute to the mean and variance or to assign its membership likelihood. Curve matching is performed to find a single longest matched segment. So, the tract might have extra segments not aligned (extension) or missing segments.

in the model sequence. This is obtained by counting the number of data points contributing to the mean tract, normalized to the total number of tracts. Fig. 1 illustrates the obtained models for two bundles of fiber tracts.

3. MIXTURE MODEL CLUSTERING

Being able to construct a statistical model of the bundles, we use a mixture model approach to cluster the fiber tracts. Here, we invoke the expectation maximization framework [11] with modifications required to handle the issues specific to 3D curves as the input data set. Three-dimensional curve matching is required to align the tracts in each cluster at each EM iteration. Furthermore, provisions should be made to deal with the variable length of the input data.

3.1. Estimating the Cluster Parameters

To estimate the mean and the variance of the cluster k , we first need to align each tract \mathbf{r}_i to the current mean tract of the cluster, $\boldsymbol{\mu}_k$ (Fig. 2). If we denote the aligned tract with \mathbf{r}_{ik} , and the probability of the tract to belong to the cluster with p_{ik} , the mean and variance can be calculated as:

$$\boldsymbol{\mu}_k = \sum_i p_{ik} \mathbf{r}_{ik} / \sum_i p_{ik} \quad (1)$$

and

$$\boldsymbol{\lambda}_k = \sum_i p_{ik} (\mathbf{r}_{ik} - \boldsymbol{\mu}_k)^2 / \sum_i p_{ik}. \quad (2)$$

Furthermore, for each point j of the mean tract, $\boldsymbol{\mu}_k$, a weight, w_{kj} , is calculated as the number of tracts that contributed to that point.

3.2. Aligning the Tracts

Each tract needs to be aligned with the current mean tract of the cluster, $\boldsymbol{\mu}_k$, before either it contributes to the calcu-

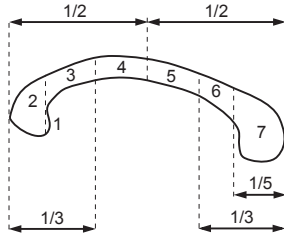


Fig. 3. A schematic of Witelson corpus callosum subdivisions [12] based on the midsagittal slice: (1) rostrum (2) genu (3) rostral body (4) anterior midbody (5) posterior midbody (6) isthmus, and (7) splenium. We further divide the splenium to its upper and lower parts to have a finer model.

lation of the new mean and variance or its new membership likelihood is assigned (Fig. 2). A string matching algorithm [10] is performed to find a single longest match. The string elements are the coefficients of the 3D quintic B -spline representation of the curves. The algorithm performs a search to find a match that gives the minimum average distance between the matched segments. More elaborate algorithms can be implemented to speed up the 3D curve matching process. However, as the spatial position of the tracts bears meaningful information, care must be taken when using those algorithms that rely on rotation- and translation-invariant parameters by assigning proper penalty to such transforms.

3.3. Assigning the Membership Probabilities

The next step in the EM algorithm is to re-calculate the membership probability of each tract, r_i , to a cluster k . It uses the cluster parameters calculated in the first step, along with the prior probability given by an atlas. Bayes rule is used to assign the posterior probability:

$$p_{ik} = f_{ik}g_{ik} / \sum_k f_{ik}g_{ik}, \quad (3)$$

where f_{ik} and g_{ik} are the likelihood and *a priori* probability of the tract r_i to belong to the cluster k , respectively. Assuming that the points on each tract are independent, the likelihood function can be calculated by multiplying the likelihood functions of the individual points, $f_{ik} = \prod_j f_{ikj}$, where

$$f_{ikj} = \frac{1}{\sqrt{2\pi\lambda_{kj}}} \exp\left(-\frac{(r_{ikj} - \mu_{kj})^2}{2\lambda_{kj}}\right), \quad (4)$$

for aligned points, and $f_{ikj} = c \leq 1$, is the penalty for extension and missing points.

4. CORPUS CALLOSUM MODELING

The corpus callosum (CC) is the largest fiber bundle that connects the two hemispheres of the brain. Different pathologies

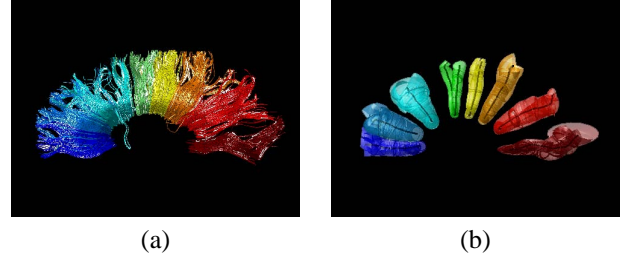


Fig. 4. Modeling of the corpus callosum fiber tracts. (a) The tracts are clustered into Witelson subdivisions based on their spatial position (Fig. 3). (b) The tracts in each cluster are then 3D aligned and statistical models are calculated.

such as multiple sclerosis [13] and schizophrenia [14] can affect different regions of the CC. Moreover, it is one of the first bundles discernible in early stages of the neonatal brain development. In addition to the quantitative analysis being performed based on the evolution of parameters such as FA, the study of the shape evolution of the CC during brain development is of interest.

Following the notion of Witelson [12], we divide the corpus callosum to seven subdivisions. We further divide the splenium into two bundles to obtain a finer model, as shown in Fig. 3. Note that any other definition for subdividing process could be used to construct the atlas which would be used in our clustering method. Fig. 4 shows the calculated model for a normal subject under study. The fiber tracts are extracted using a stochastic tractography method, starting from the ROI specified by an expert [5]. The CC tracts are then clustered into the subdivisions based on the location of their midsagittal point. The tracts in each bundle are then aligned to a template and the mean and standard deviation are calculated.

Having performed the above procedure on a set of five normal cases, an average model of the fiber bundles is calculated by mapping the tracts onto the atlas coordinate. The transformation parameters are calculated by registering the baseline image of each case to that of the atlas. Such an atlas is used as the prior map for mixture model clustering of the callosal fibers of another normal case as shown in Fig. 5. Once the model is constructed, any quantitative parameter can be statistically analyzed in one subject or over a population. As an example, Fig. 6 shows the variation of the curvature along the tracts for two of the CC bundles for a healthy subject.

5. CONCLUSION

A novel approach was presented to build a statistical model for a bundle of fiber tracts. This allows to construct an atlas of the fiber tracts by averaging the bundle models over a population. Using the model representation and with the atlas as the prior map, expectation-maximization was performed to

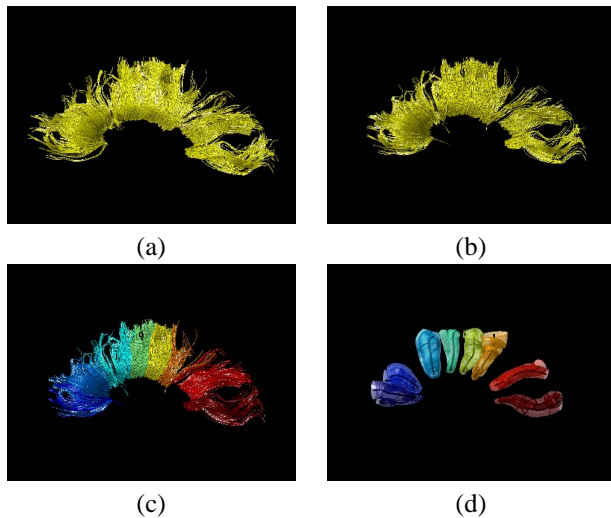


Fig. 5. Clustering of the corpus callosum from one of the subjects into its subdivisions: the original fiber tracts (a) are mapped into the atlas coordinate by registering the T2 baseline image of the subject to that of the atlas. EM clustering is used to cluster the resulting fiber tracts (b) into the bundles (c) using the atlas model as the prior. The statistical model of the bundles (d) can be used for quantitative analysis of the subject.

cluster the fiber tracts in a mixture model framework. Once the model is constructed, any quantitative parameter can be statistically analyzed in one subject or over a population. The method was successfully applied to cluster the callosal fibers into Witelson subdivisions and to calculate quantitative parameters.

6. REFERENCES

- [1] D. J. Werring, C. A. Clark, G. J. Barker, A. J. Thompson, and D. H. Miller, "Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis," *Neurology*, vol. 52, pp. 1626–1632, 1999.
- [2] D. C. Alexander, C. Pierpaoli, P. J. Basser, and J. C. Gee, "Spatial transformations of diffusion tensor magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 1131–1139, 2001.
- [3] H. J. Park, M. Kubicki, M. E. Shenton, A. Guimond, S.E. McCarley, R. W. ad Maier, R. Kikinis, F. A. Jolesz, and C. F. Westin, "Spatial normalization of diffusion tensor MRI using multiple channels," *NeuroImage*, vol. 20, pp. 1995–2009, 2003.
- [4] S. Wakana, H. Jiang, L. M. Nagae-Poetscher, P. C. M. van Zijl, and S. Mori, "Fiber tract-based atlas of human white matter anatomy," *Radiology*, vol. 230, pp. 77–87, 2004.
- [5] M. Maddah, A. U. J. Mewes, S. Haker, W. L. E. Grimson, and S. K. Warfield, "Automated atlas-based clustering of white matter fiber tracts from DTMRI," in *MICCAI*, 2005, pp. 188–195.
- [6] I. Corouge, S. Gouttard, and G. Gerig, "Towards a shape model of white matter fiber bundles using diffusion tensor MRI," in *IEEE Int. Symp. Biomed. Imag.*, 2004, pp. 344–347.
- [7] I. Corouge, S. Gouttard, and G. Gerig, "A statistical shape model of individual fiber tracts extracted from diffusion tensor MRI," in *MICCAI*, 2004, pp. 671–679.
- [8] D. Xu, S. Mori, M. Solaiyappan, P. C. M. van Zijl, and C. Davatzikos, "A framework for callosal fiber distribution analysis," *NeuroImage*, vol. 17, pp. 1131–1143, 2002.
- [9] F. S. Cohen, Z. Huang, and Z. Yang, "Invariant matching and identification of curves using B -splines curve representation," *IEEE Trans. Image Proc.*, pp. 1–10, 1995.
- [10] H. J. Wolfson, "On curve matching," *IEEE Trans. PAMI*, vol. 12, pp. 483–489, 1990.
- [11] I. T. Nabney, *Netlab: Algorithms for Pattern Recognition*, Springer-Verlag, 2001.
- [12] S. F. Witelson, "Hand and sex differences in the isthmus and genu of the human corpus callosum," *Brain*, vol. 112, pp. 799–835, 1989.
- [13] N. Evangelou, D. Konz, M. M. Esiri, S. Smith, J. Palace, and P. M. Matthews, "Regional axonal loss in the corpus callosum correlates with cerebral white matter lesion volume and distribution in multiple sclerosis," *Brain*, vol. 123, pp. 1845–1849, 2000.
- [14] J. Foong, M. Maier, C. A. Clark, G. J. Barker, D. H. Miller, and M. A. Ron, "Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study," *J. Neurol. Neurosurg. Psychiatry*, vol. 68, pp. 242–244, 2000.

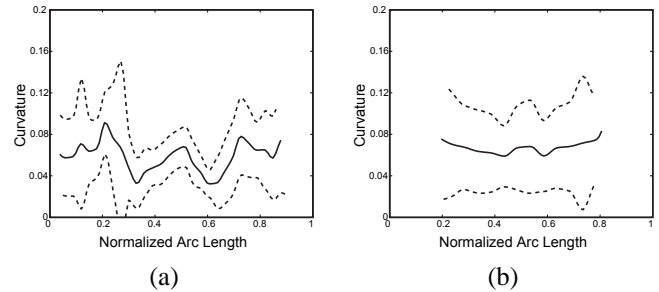


Fig. 6. The mean and variation of the curvature along the tracts calculated for (a) splenium and (b) genu in a normal case.