

# A Statistical Shape Model of Individual Fiber Tracts Extracted from Diffusion Tensor MRI

Isabelle Corouge<sup>1,2</sup>, Sylvain Gouttard<sup>2,3</sup>, and Guido Gerig<sup>1,2</sup>

<sup>1</sup> Department of Computer Science, [gerig@cs.unc.edu](mailto:gerig@cs.unc.edu)

<sup>2</sup> Department of Psychiatry  
University of North Carolina, Chapel Hill, USA

[corouge@unc.edu](mailto:corouge@unc.edu)

<sup>3</sup> ESCPE Lyon, France  
[sylvain.gouttard@cpe.fr](mailto:sylvain.gouttard@cpe.fr) <sup>†</sup>

**Abstract.** Diffusion Tensor MRI has become the preferred imaging modality to explore white matter structure and brain connectivity *in vivo*. Conventional region of interest analysis and voxel-based comparison does not make use of the geometric properties of fiber tracts. This paper explores shape modelling of major fiber bundles. We describe tracts, represented as clustered sets of curves of similar shape, by a shape prototype swept along a space trajectory. This approach can naturally describe white matter structures observed either as bundles dispersing towards the cortex or tracts defined as dense patterns of parallel fibers. Sets of streamline curves obtained from tractography are clustered, parametrized and aligned with a similarity transform. An average curve and eigenmodes of shape variation describe a compact statistical shape model. Reconstruction by sweeping the template along the trajectory results in a simplified model of a tract. Feasibility is demonstrated by modelling callosal and cortico-spinal fasciculi of two different subjects.

**Keywords:** Diffusion Tensor Imaging, statistical shape modelling.

## 1 Introduction

Diffusion Tensor Imaging (DTI) of brain structures measures diffusion properties by the local probability of self-motion of water molecules. A tensor field characterizes amount and locally preferred directions of local diffusivity. While diffusion can be considered isotropic in fluid it appears highly anisotropic along neural fiber tracts due to inhibition of free diffusion of intra- and extra-cellular fluid. DTI has become the preferred modality to explore white matter properties associated with brain connectivity *in vivo*. Most research work has been dedicated to the calculation of the tensor field, its regularization, its visualization

---

<sup>†</sup> This research is supported by the NIH-NIBIB grant P01 EB002779, the NIH Conte Center MH064065, and the Stanley Medical Research Institute. We acknowledge Matthieu Jomier and Julien Jomier, both UNC-Chapel Hill, for their precious help with software development.

and subsequently to the design of fiber tracking algorithms [1], [2], [3], [4], [5], [6]. A few groups have investigated ways towards quantitative analysis of DT images. Alexander *et al.* discuss matching of tensor fields to characterize variations in white matter structure within subject populations [7]. Xu *et al.* combine tractography and spatial normalization to produce statistical maps of fiber occurrence [8]. Fillard *et al.* perform statistical analysis of diffusion properties along fibers [9] while Ding *et al.* quantify collective properties of an anatomical fiber bundle along its medial axis [10].

In this paper, we continue preliminary work in which we extracted fibers by tractography, clustered them into anatomical bundles and analyzed the variability of local shape properties (e.g. curvature and torsion) within bundles [11]. Here, we focus on the statistical shape modelling of individual white matter fiber tracts. Our approach estimates a prototype shape of the considered fiber tract, e.g. a mean shape, and characterizes statistical shape deviations from this template shape within the fiber tract. Ultimately, we aim at modelling fiber tracts not only by template shapes and statistical variation but also by a prototype and its space trajectory. This model would be particularly appropriate for fasciculi like dense callosal fibers, observed as a “sweeping” of a U-shaped template and which form a manifold. Such a model provides a simplified representation of a fiber tract and could be used in a wider framework handling inter-individual comparison or pathological changes, for example.

## 2 Preprocessing: Fiber Extraction and Filtering

The extraction of fiber tracts is performed with the fiber tracking tool described in [4]. The tensor field is computed from DTI data by solving the Stejskal-Tanner’s diffusion equation system as described in [5]. Streamlines following the principal diffusion tensor directions between source and target regions of interest are then extracted by tractography under local continuity constraints [8]. Except at branching or crossing points, these 3D curves are assumed to represent the most likely pathways through the tensor field. Note that the term “fibers” is used for streamlines in the vector field which do not represent real anatomical fibers. The tracking is performed backwards and with sub-voxel precision. Since the robustness of fiber tracking remains limited at junctions and in noisy low-contrast regions, the extracted fiber set contains outlier curves. Also, the set of reconstructed fibers might contain curves that are part of other anatomical tracts. We developed an iterative algorithm to reject outliers and to cluster curves to fiber bundles based on pairwise distance metrics measuring position and shape similarity of pairs of fibers [11].

## 3 Shape Modelling of Individual Fiber Tracts

The individual fiber tract previously extracted and filtered acts as a training set from which we estimate a template shape, the mean shape, and statistical deviations by learning its inherent shape variability. The resulting model is related

to what is commonly called a Point Distribution Model (PDM) [12]. Representation and matching of the training set relies on a data reparametrization and on the definition of a common origin from which we establish correspondences. Pose parameters are then estimated by a Procrustes analysis [13]. A principal component analysis is subsequently applied to characterize statistical shape variation. A simplified fiber tract model can finally be obtained by reconstruction based on the template shape and the set of individual pose parameters.

### 3.1 Parameterization and Correspondences

First, fibers represented as polylines are reparametrized by cubic B-spline curves. This ensures an equidistant sampling along each fiber as well as a consistent sampling for all fibers. We slightly oversample the observations in order to prevent any loss of shape information but also to avoid any undesirable increase of dimensionality. Second, for each fiber tract under analysis, we define a common origin which can be reliably identified across subjects. This is either a geometric criterion, e.g. a cross-section with minimal area, or anatomical information like intersection with the midsagittal plane. Points with the same curvilinear abscissæ along the fiber tract are defined as homologous. This correspondence scheme handles fibers with different overall lengths in a simplified way. Only points with common curvilinear abscissæ are matched; extreme pieces of individual fibers like the ones dispersing into various cortical regions for instance, are discarded. This explicit point to point matching has been proven relevant in [11] where we demonstrated that it properly aligns local shape features like curvature and torsion across all curves in a fiber bundle.

### 3.2 Pose Parameter Estimation: Procrustes Analysis

After establishing correspondence, we align all curves in a bundle by estimating pose parameters by Procrustes analysis.

Let  $\mathcal{F} = \{\mathbf{F}_n, 1 \leq n \leq N, \mathbf{F}_n \in \mathcal{M}_{k,m}\}$  be a set of  $N$  fibers, each defined by a set of  $k$  corresponding points in  $m = 3$  dimensions, and represented by a  $k \times m$  matrix. For  $N = 2$ , an Ordinary Procrustes Analysis (OPA) gives the optimal similarity transformation parameters in a least squares sense by minimizing

$$d_{\text{OPA}}^2(\mathbf{F}_1, \mathbf{F}_2) = \| \mathbf{F}_2 - (s\mathbf{F}_1\mathbf{R} + \mathbf{1}_k\mathbf{t}^t) \|^2, \tag{1}$$

where  $s \in \mathbb{R}^{+*}$  is a scaling parameter,  $\mathbf{R} \in SO(m)$  is a rotation,  $\mathbf{t}$  is a  $m \times 1$  translation vector and  $\mathbf{1}_k$  is a  $k \times 1$  vector of ones. Minimization of (1) over the similarity group has an algebraic solution when shapes are centered, i.e.  $\mathbf{1}_k^t \mathbf{F} = 0$ , and normalized to unit size, i.e.  $\| \mathbf{F} \| = \sqrt{\text{trace}(\mathbf{F}^t \mathbf{F})} = 1$ :  $\mathbf{t} = 0$ ,  $\mathbf{R} = \mathbf{U}\mathbf{V}^t$ ,  $s = \text{trace}(\mathbf{D})$  where  $\mathbf{V}\mathbf{D}\mathbf{U}^t = \mathbf{F}_2^t \mathbf{F}_1$  is the singular value decomposition of  $\mathbf{F}_2^t \mathbf{F}_1$ . In the actual case where  $N > 2$ , a Generalized Procrustes Analysis (GPA) estimates the similarity transformation parameters which minimize the sum of squared norms of pairwise differences

$$d_{\text{GPA}}^2(\mathbf{F}_1, \dots, \mathbf{F}_N) = \frac{1}{N} \sum_{n=1}^N \sum_{p=n+1}^N \|(s_n \mathbf{F}_n \mathbf{R}_n + \mathbf{1}_k \mathbf{t}_n^t) - (s_p \mathbf{F}_p \mathbf{R}_p + \mathbf{1}_k \mathbf{t}_p^t)\|^2. \quad (2)$$

The optimization is performed iteratively:

1. Translation. Fibers are centered with respect to their center of mass,  $\mathbf{g}_n$ :  $\mathbf{F}_n^c = \mathbf{F}_n - \mathbf{g}_n$ .
2. Scaling. Centered data is normalized to unit size:  $\mathbf{F}_n^{\text{cs}} = \mathbf{F}_n^c / \|\mathbf{F}_n^c\|$ .
3. Rotation. Let  $\mathbf{F}_n^{\text{old}} = \mathbf{F}_n^{\text{cs}}$ . The  $N$  shapes are rotated in turn. For each  $n$ ,  $1 \leq n \leq N$ :
  - a)  $\bar{\mathbf{F}}_n = \frac{1}{N-1} \sum_{p \neq n} \mathbf{F}_p^{\text{old}}$ ,
  - b)  $s_n = 1$ ,  $\mathbf{t}_n = 0$ ,  $\mathbf{R}_n = \arg \min_{\mathbf{R}} d_{\text{OPA}}^2(\mathbf{F}_n^{\text{old}}, \bar{\mathbf{F}}_n)$ ,
  - c)  $\mathbf{F}_n^{\text{new}} = \mathbf{F}_n^{\text{old}} \mathbf{R}_n$  and  $\mathbf{F}_n^{\text{old}} = \mathbf{F}_n^{\text{new}}$ .

Step 3 is iterated until the Generalized Procrustes distance  $d_{\text{GPA}}^2(\mathbf{F}_1^{\text{old}}, \dots, \mathbf{F}_N^{\text{old}})$  can not be reduced further. The alignment of the training set is achieved by applying the estimated rotations to the centered but non unit-scaled initial shapes  $\mathbf{F}_n^c$ , resulting in the set of aligned fibers  $\mathcal{F}^{\mathcal{A}} = \{\mathbf{F}_n^{\mathcal{A}}, 1 \leq n \leq N\}$ . Indeed, the scaling is needed to optimally estimate the rotation but a size normalization is not desirable since the training fibers belong to the same individual fiber tract.

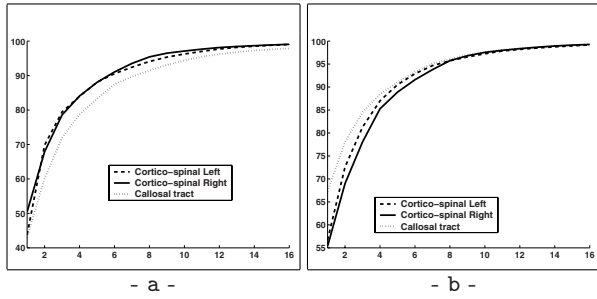
### 3.3 Estimation of the Mean Shape

Given the set of aligned shapes  $\mathcal{F}^{\mathcal{A}}$ , the estimated mean shape  $\hat{\mathbf{F}}$  is given by

$$\hat{\mathbf{F}} = \frac{1}{N} \sum_{n=1}^N \mathbf{F}_n^{\mathcal{A}}. \quad (3)$$

### 3.4 Characterization of Shape Statistical Variability

We now perform a principal components analysis (PCA) on  $\mathcal{F}^{\mathcal{A}}$ . This linear analysis characterizes the variations within a given training population and extracts the principal modes of deformation relative to the mean shape. Briefly, PCA expresses the observations in a new orthogonal basis, with the mean fiber as the origin and eigenvectors or modes of the observations covariance matrix  $\mathbf{C}$  as axes. In our case, an observation  $\mathbf{f}$  is represented by the concatenation vector of the 3D fiber coordinates  $\mathbf{f} = (x_1, y_1, z_1, \dots, x_k, y_k, z_k)^t$  and the corresponding estimated mean fiber is denoted  $\bar{\mathbf{f}}$ . An approximation of these observations can be obtained by truncating certain number of modes. The reconstructed observation is then written as  $\mathbf{f} = \bar{\mathbf{f}} + \Phi_{\mathbf{m}} \mathbf{b}_{\mathbf{m}}$  where  $m$  is the number of remaining modes,  $\Phi_{\mathbf{m}}$  the matrix of truncated modes and  $\mathbf{b}_{\mathbf{m}} = (b_i)_{i=1, \dots, m}$  the  $m$ -dimensional vector representing the original observation in the truncated modal basis. Reconstruction quality can be measured by  $\tau = \sum_{i=1}^m \lambda_i / \lambda_T$  where  $\lambda_i$  is the  $i^{\text{th}}$  eigenvalue, in decreasing order, of matrix  $\mathbf{C}$  and corresponds to the variance explained by the  $i^{\text{th}}$  mode, and where  $\lambda_T$  is the total variance.



**Fig. 1.** Percentage  $\tau_p$  of cumulative variance as a function of the number of modes for three fiber tracts (left and right cortico-spinal tract, callosal fiber tract) of two subjects (a and b);  $\tau_p = \sum_{i=1}^m \lambda_i / \lambda_T \times 100$ .

Under the assumption that the distribution of the elements of  $\mathcal{F}^A$  is Gaussian, the variation of  $b_{i,i=1,\dots,m}$  in an interval such as  $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$  explains the variability of the set of objects. Calculating these statistics with a sufficient number of representative instances of shapes, these variations will cover the major variability of the population.

### 3.5 Fiber Tract Reconstruction

Qualitative views of 3D rendering suggest that sets of fibers might be described as a replication of a prototype curve along a space trajectory, simulating the sweeping of a space curve to form a manifold. In an initial attempt to test this hypothesis, we reconstruct an approximation to the initial fiber tract by applying the inverse rigid body transform per fiber to the template.

Let  $\mathcal{T}$  be the set of transformations computed from the Procrustes analysis to align the training set,  $\mathcal{T} = \{(-\mathbf{g}_n, \Gamma_n), 1 \leq n \leq N\}$  with  $\mathbf{g}_n$  the translation vector defined by the center of mass of the  $n^{\text{th}}$  fiber and  $\Gamma_n$  the resulting rotation for fiber  $n$ :  $\Gamma_n = \Pi_i \mathbf{R}_n^{(i)}$ ,  $\mathbf{R}_n^{(i)}$  being the rotation computed in the  $i^{\text{th}}$  step 3 iteration of the Generalized Procrustes Analysis. The reconstructed fiber tract  $\tilde{\mathcal{F}}$  is given by

$$\tilde{\mathcal{F}} = \{\tilde{\mathbf{F}}_n = \hat{\mathbf{F}} \Gamma_n^t + \mathbf{g}_n, 1 \leq n \leq N\} \tag{4}$$

where  $\hat{\mathbf{F}}$  is the estimated mean fiber. The fiber set  $\tilde{\mathcal{F}}$  is a simplified representation of the initial fiber tract defined as the template shape  $\hat{\mathbf{F}}$  and the set  $\mathcal{T}^{-1} = \{(\mathbf{g}_n, \Gamma_n^t), 1 \leq n \leq N\}$  of inverse transformations. This model might have advantages for inter-subject matching and comparison since the template represents a robust estimate of the core shape whereas individual variability is encoded in the shape eigenmodes.

**Table 1.** Mean point to point distance (in voxels) between original and reconstructed fibers averaged over the bundle for the three tracts of each subject. Voxel size is  $2 \times 2 \times 2$  mm<sup>3</sup>.

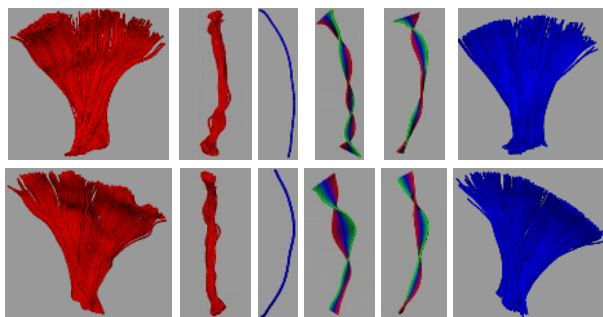
Tract	Subject 1	Subject 2
Left cortico-spinal tract	0.60	0.71
Right cortico-spinal tract	0.66	0.63
Callosal tract	0.53	0.64

## 4 Experiments and Results

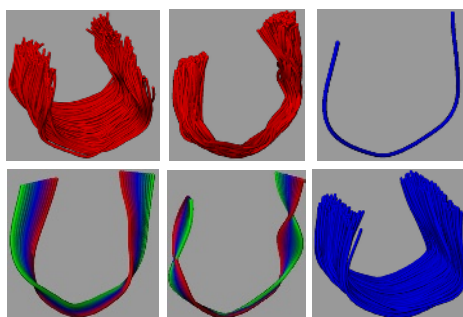
We selected two cases out of a 3Tesla high resolution ( $2 \times 2 \times 2$  mm<sup>3</sup>) DT MRI database of healthy controls and extracted three major fiber tracts. The first two represent the major projection tracts between the internal capsule and the upper cortex of the left and right hemisphere (see Fig. 2, left and Fig. 4, top left) while the third corresponds to a dense set of callosal fibers (see Fig. 3, top left and Fig. 4, bottom left). The common origin of all fibers for the cortico-spinal tracts is chosen as the location of minimal area cross-section between pons and internal capsule. For the callosal fiber bundle, a natural choice is the intersection with the midsagittal plane. For each tract, Figures 2, 3 and 4 present the set of aligned shapes and the estimated mean shape which defines the core shape of the tract. In addition to Fig. 1 which presents the percentage of cumulative variance as a function of the number of modes, Figures 2, 3 and 4 illustrate the variations around the mean fiber for the the first two modes. For all tracts and all subjects, less than 10 modes are needed to explain more than 90% of the variance. The figures also show the reconstructed fiber tract associated to each of the original data sets. The reconstructed tracts visually appear to be a very good approximation to the original data, representing the *major characteristics* of the bundle without small scale variability. This observation is corroborated by computing the mean point to point distance between original and reconstructed fiber tracts averaged over the bundle (see Table 1). We also observe that the results are quite consistent for both subjects, which suggests the processing scheme might achieve a good reproducibility of tract extraction and tract modelling. A quantitative validation of reproducibility and validity in 10 controls, including anatomical landmarking by clinical experts, is in progress.

## 5 Discussion and Perspectives

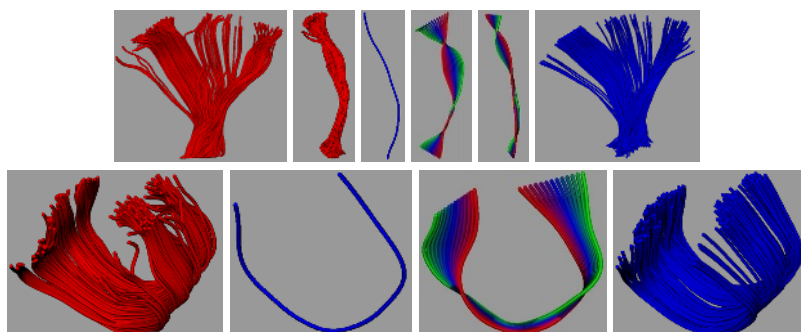
We propose a statistical shape model of individual white matter fiber tracts extracted from Diffusion Tensor MRI. Correspondences are derived from a reparametrization of the streamline curves and the definition of a common origin, and alignment is achieved by a Procrustes fit. Results obtained from two healthy control subjects seem consistent and suggest that these models might be reproducibly obtained for each subject. We currently test the new processing scheme on 10 healthy controls which got imaged with the best available DTI protocol on a new 3Tesla MRI scanner (Siemens Allegra head-only).



**Fig. 2.** Left (top) and right (bottom) cortico-spinal tracts of Subject 1. From left to right: original data sets (more than 600 fibers each); data sets after alignment; estimated mean fiber; variations around the mean shape according to first and second mode; reconstructed fiber tract.



**Fig. 3.** Callosal fibers of Subject 1. From top left to bottom right: original data sets (363 fibers); data sets after alignment; estimated mean fiber; variations around the mean shape according to first and second mode; reconstructed fiber tract.



**Fig. 4.** Results for Subject 2. Top: left cortico-spinal tract with original data sets (more than 250 fibers); data sets after alignment; estimated mean fiber; variations around the mean shape according to first and second mode; reconstructed fiber tract. Bottom: callosal fiber tract with original data sets (128 fibers); estimated mean fiber; variations around the mean shape according to first mode; reconstructed fiber tract.

This paper focuses on the discussion of the calculation of the template and its major shape variation. Reconstructed bundles using the template  $\hat{\mathbf{F}}$  and the inverse set of pose transformations  $\mathcal{T}^{-1}$  closely represent the original bundle and suggest that this type of modelling might be a viable concept for the modelling of various major fiber tracts. The set of transformations  $\mathcal{T}^{-1}$  is still unstructured with arbitrary ordering. The major focus of future research will be devoted to develop a concept for ordering the set of fibers and hence the transformations within the shape manifold and for describing this set as a continuous sweeping curve. Ultimately, each fiber tract would be characterized by the prototype and its sweeping along this curve. Preliminary results and our analysis of the major large and small tracts of interest (cortico-spinal, callosum, cingulum, uncinate, splenium and genu) suggest that modelling by “ribbon-cables” indeed might be an appropriate representation.

Modelling will potentially serve for improved inter-individual registration and comparison of diffusion tensor properties along and across fiber tracts. Clinical research is interested in a quantitative analysis which finally might lead to answer questions in regard to fiber integrity or fiber disruption and its effect on brain connectivity. Moreover, modelling of fiber tracts in healthy controls will help to study geometric and diffusion changes of white matter tracts in the presence of pathology, e.g. tumor and edema or white matter lesions.

## References

1. P.J. Basser, S. Pajevic, C. Pierpaoli, and A. Aldroubi. Fiber tract following in the human brain using DT-MRI data. *IEICE Trans. on Information and Systems*, E85-D(1):15–21, 2002.
2. M. Björnemo, A. Brun, R. Kikinis, and C.-F. Westin. Regularized stochastic white matter tractography using diffusion tensor MRI. In *Proc. of MICCAI*, vol. 2488 of *LNCS*, pages 435–442, 2002.
3. O. Coulon, D.C. Alexander, and S.R. Arridge. A regularization scheme for diffusion tensor magnetic resonance images. In *Proc. of IPMI*, LNCS, pages 92–105, 2001.
4. P. Fillard and G. Gerig. Analysis tool for diffusion tensor MRI. In *Proc. of MICCAI*, vol. 2879 of *LNCS*, pages 967–968, 2003.
5. C.-F. Westin, S.E. Maier, H. Mamata, A. Nabavi, F.A. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6:93–108, 2002.
6. L. Zhukov and A.H. Barr. Oriented tensor reconstruction: tracing neural pathways from diffusion tensor MRI. In *Proc. IEEE Visualization*, 2002.
7. D.C. Alexander and J.C. Gee. Elastic matching of diffusion tensor images. *Computer Vision and Image Understanding*, 77:233–250, 2000.
8. D. Xu, S. Mori, M. Solaiyappan, P.C.M. van Zijl, and C. Davatzikos. A framework for callosal fiber distribution analysis. *NeuroImage*, 17:1131–1143, 2002.
9. P. Fillard, J. Gilmore, W. Lin, and G. Gerig. Quantitative analysis of white matter fiber properties along geodesic paths. In *Proc. of MICCAI*, vol. 2879 of *LNCS*, pages 16–23, 2003.
10. Z. Ding, J.C. Gore, and A.W. Anderson. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Magn. Res. Med.*, 49:716–721, 2003.



11. I. Corouge, S. Gouttard, and G. Gerig. Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *Proc. of IEEE ISBI*, pages 344–347, 2004.
12. T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.
13. C. Goodall. Procrustes methods in the statistical analysis of shape. *J.R. Statist. Soc. B*, 53(2):285–239, 1991.