Published in final edited form as:
J Neurosci Methods. 2009 November 15; 184(2): 375–379. doi:10.1016/j.jneumeth.2009.07.032.

Genetic White Matter Fiber Tractography with Global Optimization

Xi Wu[1,2,3], **Qing Xu**[3,4], **Lei Xu**[3], **Jiliu Zhou**[1], **Adam W. Anderson**[3,5], and **Zhaohua Ding**[3,4,5,6]

[1]College of Electronics and Information Engineering, Sichuan University, P.R. China

[2]Department of Electronic Engineering, Chengdu University of Information Technology, P.R.China

[3]Vanderbilt University Institute of Imaging Science, Vanderbilt University, USA

[4]Department of Electrical Engineering and Computer Science, Vanderbilt University, USA

[5]Department of Biomedical Engineering, Vanderbilt University, USA

[6]Chemical and Physical Biology Program, Vanderbilt University, USA

Abstract

Diffusion tensor imaging(DTI) tractography is a novel technique that can delineate the trajectories between cortical region of the human brain non-invasively. In this paper, a novel DTI based white matter fiber tractography using genetic algorithm is presented. Adapting the concepts from evolutionary biology which include selection, recombination and mutation, globally optimized fiber pathways are generated iteratively. Global optimality of the fiber tracts is evaluated using Bayes decision rule, which simultaneously considers both the fiber geometric smoothness and consistency with the tensor field. This global optimality assigns the tracking fibers great immunity to random image noise and other local image artifacts, thus avoiding the detrimental effects of cumulative noise on fiber tracking. Experiments with synthetic and in vivo human DTI data have demonstrated the feasibility and robustness of this new fiber tracking technique, and an improved performance over commonly used probabilistic fiber tracking.

Keywords

Diffusion Tensor Imaging; Fiber Tracking; Genetic Algorithm; Global Optimization

1 Introduction

Diffusion tensor imaging (DTI) has become a primary tool for non-invasive exploration of the structure of living tissue in vivo [1]. Since its first introduction a decade ago, this new imaging modality has been widely used to reconstruct neuronal fiber pathways in the human brain [2]. To date a variety of tracking algorithms have been proposed to infer fiber connections in the human brain [3][4][5], the basic principle of which is sequentially integrating local fiber directions from pre-defined seed point(s) to generate fiber connection pathways. Typically, these tracking algorithms “grow” fiber pathways by piecing a line segment to the end of the preceding segment. These methods can be broadly divided into two categories: deterministic fiber tractography [3] and probabilistic fiber tractography [4].

Corresponding Author: Zhaohua Ding, Ph.D. Vanderbilt University Institute of Imaging Science 1161 21st Avenue South, MCN AA-1105 Nashville, TN 37232-2310 Phone: (615)322-7889 Fax: (615)322-0734 zhaohua.ding@vanderbilt.edu.

A common drawback of the above streamline-like tracking methods, either deterministic or probabilistic, is cumulative errors arising from random image noise and/or partial volume averaging (PVA) along the tracking path [6][7], even with certain regularizations on the basis of geometric or other constraints. Furthermore, as the direction information that stream-like tracking methods rely on is only derived locally, the tracking results from these methods are not a globally optimized solution.

In this work, a novel fiber tracking technique based on well established genetic algorithms was proposed. The proposed technique allows globally optimized solutions to the fiber pathways to be obtained, and hence possesses superb immunity to local imaging artifacts. Additionally, the proposed technique provides optimal solutions for fiber connection pathways between two designated ROIs; this offers a great potential of applying it to studies of structure-function relations in the human brain, in which the structural connectivity between two functionally related regions is often sought.

2 Method

A common practice of fiber tracking is to track fiber pathways that connect to certain ROIs, among which a most useful application is to find connecting pathways between an ROI pair. Fiber tracking between a pair of ROIs can be cast as a path finding problem to which an optimal solution is found by genetic algorithm (GA). In this context, the goal is to evolve from initial solutions to produce a set of fibers with best fitness to the given data, which is presumably an optimal solution to the fiber path. Details of the implementation procedure for one iteration of the genetic tracking algorithm, called GeneTrack henceforth, are given below.

Initialization

In a typical GA, solutions are encoded as strings of binary bits, which are analogous to chromosomes in biology. Binary strings can be extended to continous strings, which are chosen as initial solutions in our design. More specifically, we represent an arbitrary fiber pathway between a pair of ROIs with a space curve (f), and express the curve f as Fourier series (of continuous values) in Cartesian system, i.e.:

$$f^d(t) = \sum_{n=1}^{N} \left[a^d{}_n \cos(nt) + b^d{}_n \sin(nt) \right], \quad (1)$$

where superscript d denotes the x, y, or z direction in the Cartesian system, n is the order of Fourier series, t is the number of points in curve f, the a_n and b_n are coefficients of the cosine and sine components respectively, and N is the maximum order of Fourier series for approximation with reasonable accuracy (t = 50 and N=10 in this work). According to the equation above, every fiber curve can be represented by $2N$ Fourier coefficients in each direction, which consists of N coefficients for both a_n and b_n. The $6N$ coefficients in all three directions, which correspond to the genes of the chromosomes, encode the solutions for the fiber pathways between an ROI pair and will be used for recombination and mutation to produce offspring in later steps.

Selection

In designing objective functions for selecting best fit solutions among the candidates, two primary considerations are taken into account. First, the solutions should possess best geometric smoothness along the entire fiber path. Second, the solutions should be those best fit the diffusion tensor field among all candidates. Simultaneously considerations of these

(with certain trade-off) may yield solutions that are globally optimal in the sense that the curves are both smooth and reasonably fit the data.

Globally optimal solutions may be obtained by using the classical Bayesian theory [8]. Let **T** denote the diffusion tensor field and C denote spatial curves that cover all possible fiber pathways connecting the designated ROI pair. According to the Bayes decision rule, the optimal solution C_{opt} is the one with maximum *a posteriori* (MAP) probability $p(C|\mathbf{T})$:

$$p(C|\mathbf{T}) = \frac{p(C)\, p(\mathbf{T}|C)}{p(\mathbf{T})}. \quad (2)$$

Since $p(\mathbf{T})$ is independent of C, maximizing $p(C|\mathbf{T})$ reduces to maximizing the product of $p(C)$ and $p(\mathbf{T}|C)$. The term $p(C)$ is a prior probability of curve C, which is the probability of the existence of curve C without any measurement data; the term $p(\mathbf{T}|C)$ is a conditional probability, which defines the probability of the existence of tensor field **T** given curve C.

To find the MAP solution, $p(C)$ and $p(\mathbf{T}|C)$ need to be modeled. In our design, these probabilities are modeled so that constraints on curve smoothness and consistency with tensor field are imposed. First, let us assume the tensor field to be a Markovian random field (MRF), and v_t be a unit vector representing the local tangential direction at the t^{th} point of the curve. According to the MRF theory, v_t is a random realization of the vector field in the neighborhood of t, which observes a Gibbs distribution [9]:

$$p(C) = \frac{1}{Z_1} \prod_t e^{-p(v_t)} = \frac{1}{Z_1} e^{-\sum_t p(v_t)}, \quad (3)$$

where Z_1 is a normalization constant. When only the preceding point along the curve is considered, a simple definition of $P(v_t)$ is:

$$p(v_t) = \arccos(v_t \cdot v_{t-1}). \quad (4)$$

This definition of the prior probability gives preference to the curves with low curvature, thus imposing a smoothness constraint to the fiber.

Second, assuming that the tensor measurement in voxel t depends only on the local fascicle direction v_t, the conditional probability $p(\mathbf{T}|C)$ can be rewritten as a second Gibbs distribution below:

$$p(\mathbf{T}|C) = \frac{1}{Z_2} \prod_t e^{-p(\mathbf{T}_t|v_t)} = \frac{1}{Z_2} e^{-\sum_t p(\mathbf{T}_t|v_t)}, \quad (5)$$

where Z_2 is a normalizing constant.

In order for the probability $p(\mathbf{T}_t|v_t)$ to decrease with the discrepancy between the local fiber direction v_t and the major eigenvector of the local diffusion tensor e_t, we propose the following conditional probability model:

$$p(\mathbf{T}_t|v_t) = \arccos(v_t \cdot e_t). \quad (6)$$

Combining all the previous models (Equ.3 and Equ.5) leads to an expression for $p(C|\mathbf{T})$ which turns out to be a new Gibbs distribution:

$$p(C|T) = \frac{1}{Z_3} e^{-\sum_t p(v_t) + \alpha p(T_t|v_t)}, \quad (7)$$

where $Z_3 = Z_1 \bullet Z_2$.

With this new Gibbs distribution, an optimal solution can be reached by minimizing the following cost function:

$$f_{cost} = \sum_t p(v_t) + \alpha \sum_t p(\mathbf{T}_t|v_t). \quad (8)$$

The first term in the above cost function imposes a smoothness constraint on the fiber pathway, and the second term encourages a consistency between the fiber and tensor dominant directions. The relative weights of these two terms are determined by the parameter α, which regulates the trade-off between the smoothness of the fiber and consistency with the data. In this work, the range of α is chosen to be between zero and ten.

Recombination

The subset of optimal fibers selected above are recombined to form a new set of fibers. This is implemented by randomly selecting a value for each of the $6N$ coefficients in Equ. 1 from the parent fibers, so that the coefficients for each new fiber come from different parents. This new set of fibers, along with a small number of best fit parents, constitute the offspring for the next generation.

Mutation

The mutation process perturbs the coefficients of the above fibers so that each solution contains a new set of values for the coefficients. To maintain the stability of this algorithm, the amount of perturbation for each coefficient is at the order of one standard deviation of the coefficients in the existing fibers. This provides stability to the solution, and in the meantime offers an opportunity to generate new solutions that better fit the cost function.

3 Tracking Experiments and Results

To evaluate comprehensively the performance of the GeneTrack algorithm proposed, we have carried out a series of fiber tracking experiments on synthetic and in vivo human DTI datasets. Fiber tracking on synthetic datasets was performed using both the GeneTrack and a probabilistic fiber tracking method by Friman et al. [4], and comparisons between them were made with respect to tracking accuracy and validation. The reproducibility of GeneTrack was demonstrated from the tracking result in five different in vivo datasets.

3.1 Fiber tracking with synthetic dataset

Evaluation of the performance of the algorithm began with a synthetic tensor field, which was designed to contain five curves of different geometric complexity. To mimic closely physiological conditions *in vivo*, synthetic tensors in the curves were constructed to have a trace of 2.1×10^{-5} cm^2/s and a fractional anisotropy (FA) of 0.8. These diffusion parameters were similar to those in normal brain parenchyma. Diffusion weighted imaging was simulated along 32 non-collinear directions with a b value of 1000 s/mm^2, and the diffusion weighted data were corrupted with zero mean Gaussian noise at a standard deviation of 0.05.

The synthetic dataset had one curve with two extrema and four parabolic curves with increasing curvature, as shown in Fig. 1a. To better visualize the details of the synthetic

fibers, the boxed region in Fig. 1a is shown in an enlarged view in Fig. 1b. Each voxel is visualized by an ellipsoid whose principal axes are the three orthogonal eigenvectors of the tensor and the radii of the ellipsoid along the axes are proportional to their corresponding eigenvalues.

The proposed GeneTrack algorithm was first used to track fibers between two seed points in each curve (denoted as a diamond and a square in Fig. 1c). For each synthetic curve, a total of 1000 random curves were generated initially by assigning random values to the coefficients a_n and b_n in Eq. 1, followed by 60 iterations of selection, recombination and mutation. In each iteration, 100 parents were sampled, recombined and then mutated to generate 900 children as offspring. These children along with 100 parents with best cost function values were fed to the next generation as parents. Meanwhile, a probabilistic fiber tracking method by Friman et al. [4] was also implemented on the same dataset, with the same seed point near one end of the curve (denoted as a diamond in Fig. 1d). Parameter settings of this method were the same as those described in Friman et al. [4]. A total of 1000 repeated trials were performed with a step size of one voxel.

In addition, the number of fibers going through each voxel was recorded. These numbers were normalized against the total number of trials to yield a density map that indicates the probability of fiber connection along the fiber path between the two seed points (for the GeneTrack) or to the starting seed point (for the Friman method).

Figure 1c and 1d show the tracking results obtained using our method and the Friman method, respectively. For each figure, 100 “fibers” (out of 1000) were randomly selected for demonstration. It can be seen that, at zero mean Gaussian noise with a standard deviation of 0.05, fibers reconstructed by both methods grossly follow the synthetic curves. However, close inspections reveal that the Friman method tends to generate fibers that diverge distally along the path, which results in some fibers with geometric structure that is inconsistent with that of the “ground truth”; also, some fibers show premature termination, and as such, there is a large difference in the length of the fibers. These problems render subsequent quantitative analysis rather difficult. On the other hand, fibers from our method exhibit good convergence along the fiber path and no premature termination at all. This is attributable to the fact that our method finds optimal fiber connections between two designated regions, rather than tracking fibers from one seed region as with the Friman method.

Density maps of the middle slice that depict the connection probability are shown in Fig. 1e and 1f for our method and the Friman method, respectively. As can be observed, the probability of connection is relatively uniform along the entire curves for our method, albeit there is a slight drop-off around the middle portion of some curves. However, the connection probability for the Friman method exhibits gradual decreases along the curves, which is caused by deviations of some fibers from the curves along the path. This situation is most pronounced for the 1st bundle whose density map quickly drops to 0.3 or less.

Mean values of the cost functions of the tracked fibers and their variations with the number of iterations are shown in Fig. 1g. For each curve, the mean value of the cost functions of 1000 fibers is calculated for each of the 60 iterations performed. It is shown that the mean values of the cost functions are high for all curves at the beginning, but decrease rapidly during the first a few iterations. They tend to stabilize to low values around 20 iterations, beyond which further declines are quite small. This indicates that the algorithm starts to converge around the 20^{th} iteration, at which the tracked fibers are close to the final optimal pathways. This also demonstrates, although indirectly, the validity of the GeneTrack since if there exist genuine fibers between two ROIs, the cost function will decrease and converge to a very low value.

To demonstrate the advantage of the GeneTrack algorithm over the Friman method in terms of tracking efficiency, a synthetic phantom was constructed that contained two thick parallel "fiber bundles" bridged by a thin "fiber bundle" (see Fig. 1h). Fibers were tracked using the GeneTrack algorithm and the Friman method with the same pair of ROIs (denoted by a diamond and square in Fig. 1i and 1j). For the GeneTrack, the same parameters as before were used; for the Friman method, 5000 trials with the same parameter settings as in [4] were launched from each of the ROIs. The results of the two tracking methods are shown in Fig. 1i and 1j respectively. It can be appreciated from Fig. 1i that the fibers from the GeneTrack algorithm converge nicely to the intended fiber pathways. On the other hand, it took the Friman method a considerably larger number of trials to reach the results desired (Fig. 1j). This discrepancy in tracking efficiency comes from the fact that the GeneTrack algorithm takes the ROI pair as prior knowledge, whereas the Friman method tracks fibers without any targeted regions, and hence only resorts to post-hoc filtering with a second ROI to reach intended results. Evidently, the lack of consideration of designated targets in the tracking process makes the Friman method much more computationally expensive, and also puts it at the risk of missing small fiber structures that have low connection probability.

3.2 Fiber tracking with *in vivo* human DTI data

To assess the performance of the GeneTrack algorithm on *in vivo* data, diffusion weighted images (DWI) were acquired from five healthy human volunteers with a 3T Philips Intera Achieva MR scanner (Best, The Netherlands) and an eight-element SENSE coil. A single shot, echo-planar pulsed gradient spin-echo imaging sequence was used, and diffusion weighting was performed along 32 non-collinear directions with a *b* value of 1000 s/mm^2. A total of 64 contiguous, 2- *mm*-thick slices with a matrix size of 128×128 were acquired from a field of view of 256×256 mm^2, yielding an in-plane pixel size of 2×2 mm^2. Three repeated scans were obtained from each subject, which were motion and distortion corrected and then averaged using Philips diffusion registration PRIDE tool (Release 0.4). Diffusion tensors were estimated from the averaged DWI data using a linear least-square fitting procedure [1], from which fractional anisotropy (FA) maps were computed. T_1-weighted images were acquired as well during the same session for the purpose of image registration. Prior to the study, the subject gave informed consent for the study protocol that was approved by the local ethics committee.

To define the regions of interest for the *in vivo* experiments, two different approaches were used. First, the left and right thalami were defined using the WFU PickAtlas tool in SPM2 (http://www.fil.ion.ucl.ac.uk/spm/software/spm2/). To do so, the left or right thalamus atlas was picked in the MNI space, which was normalized to the T_1 weighted image in the native space. The T_1 image along with the thalamus atlas were then co-registered with the FA map, so that the ROI of the thalamus was brought to the native space of the DWI. In addition to the thalami, the left and right Brodman's areas (BA) 17 were also defined using the WFU PickAtlas tool. Second, the Broca's and the pre-motor regions were defined using functional MRI signals obtained from the subject during the performance of designated language tasks, as reported earlier [10]. Fiber tracking was performed with the three sets of ROIs defined above, using the GeneTrack with the same parameter settings as in the synthetic experiments.

Reproducibility of the GeneTrack algorithm is demonstrated in Fig. 2, which shows 100 fibers tracked between left/right thalamus and left/right BA 17 in the occipital lobe, and between the left Broca's area and the pre-motor region in five different in vivo datasets. The data acquisition, fiber tracking and visualization parameters are the same among these datasets. It is evident that, although there are small differences among the five subjects, which is likely due to inter-subject anatomical variations, the pathways reconstructed are

largely consistent across these subjects, all exhibiting good agreement with known neuro-anatomy.

4 Discussion and Conclusion

In this paper, we proposed a new genetic tracking technique, GeneTrack, for DTI based white matter fiber tractography. In contrast to most other fiber tracking methods, the proposed algorithm searches for globally optimal fiber pathways between a given pair of ROIs. Using the concept from evolutionary biology that includes selection, recombination and mutation, the GeneTrack algorithm iteratively finds fiber pathways that have optimal trade-off between fiber geometric smoothness and consistency with the tensor field. Experiments with synthetic and in vivo human DTI data have demonstrated the feasibility and robustness of this new fiber tracking technique, and an improved performance over commonly used probabilistic fiber tracking.

Acknowledgments

This study was supported by science innovation grant of Sichuan NO.29 (Mingyuan Xie).

References

1. Basser PJ, Matiello J, Le Bihan D. Estimation of the effective self diffusion tensor from the NMR spin echo. J. Magn. Reson. 1994; 103:247–254. Ser. B.
2. Mori S, van Zijl PCM. Fiber tracking: principles and strategies — A technical review. NMR Biomed. 2002; 15:468–480. [PubMed: 12489096]
3. Lu Y, Aldroubi A, Gore JC, Anderson AW, Ding Z. Improved fiber tractography with Bayesion tensor regularization. NeuroImage. 2006; 31:1061–1074. [PubMed: 16563804]
4. Friman O, Farneback G, Westin C. A bayesian approach for stochastic white matter tractography. IEEE Trans. Med. Imag. 2006; 25(8):965–978.
5. Zhang F, Hancock ER, Goodlett C, Gerig G. Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling. Med Image Anal. 2009; 13(1):5–18. [PubMed: 18602332]
6. Alexander AL, Hasan KM, Lazar M, Tsuruda JS, Parker DL. Analysis of partial volume effects in diffusion-tensor MRI. Magn. Reson. Med. 2001; 45:770–780. [PubMed: 11323803]
7. Anderson AW. Theoretical analysis of the effects of noise on diffusion tensor imaging. Magn. Reson. Med. 2001; 46:1174–1188. [PubMed: 11746585]
8. Duda, RO.; Hart, PE. Pattern Classification and Scene Analysis. John Wiley and Sons; New York: 1973. p. 10-15.
9. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell. 1984; PAMI-6(6):721–741.
10. Newton, AT.; Mishra, A.; Gore, JC.; Ding, Z.; Morgan, VL. Investigation of the Structural Connectivity - Functional Connectivity Relationship in the Human Language System. 13th Annual Meeting of the Organization for Human Brain Mapping; Chicago, Illinois. 2007. p. 285

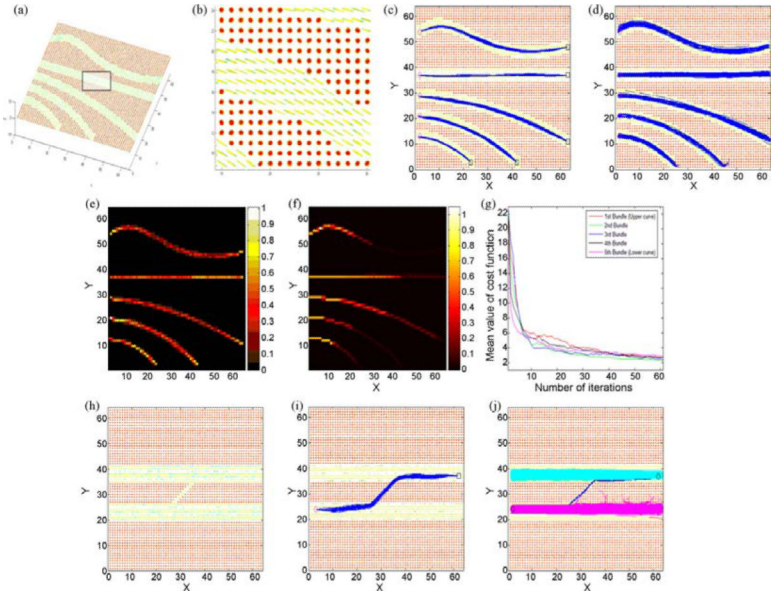

Fig. 1.

Fibers tracking with synthetic tensor data. (a) A 3D view of one slice of the synthetic dataset. (b) Enlarged 2D view of the boxed region in (a). (c) Results from the GeneTrack algorithm. (d) Results from the Friman method. (e) Density map of connection probability from GeneTrack algorithm. (f) Density map of connection probability from Friman method. (g) Variations of the mean value of cost function with the number of iterations. (h) A synthetic dataset for comparing the tracking efficiency between the GeneTrack and Friman method. (i) Fiber tracts resulting from the GeneTrack. (j) Fiber tracts resulting from the Friman method. Magenta denotes fibers tracked from the ROI at the left; cyan denotes fibers tracked from the ROI at the right; fibers connected to both ROIs are denoted in red or blue.

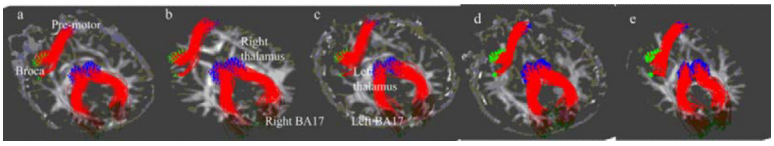

Fig. 2.
GeneTrack resulting fibers between left/right thalamus (blue dots) and left/right BA 17 (green dots) in the occipital lobe, and between the left Broca's area (green dots) and the pre-motor region (blue dots) in five different in vivo datasets.