

White Matter Fiber Tract Segmentation Using Nonnegative Matrix Factorization

Xuwei Liang
Department of Computer Science
University of Kentucky
Lexington, KY 40506-0046, USA
Email: xuwei.liang@uky.edu

Jie Wang
Department of Computer Science
Minnesota State University at Mankato
Mankato, MN 56001, USA
Email: jie.wang@mnsu.edu

Zhenmin Lin and Jun Zhang
Department of Computer Science
University of Kentucky
Lexington, KY 40506-0046, USA
Email: jzhang@cs.uky.edu

Abstract—Accurate and efficient white matter fiber tract segmentation is an important step in clinical and anatomical studies that use diffusion tensor magnetic resonance imaging (DTI) tractography techniques. In this work, we present a novel technique to group white matter fiber tracts reconstructed from DTI into bundles using Nonnegative Matrix Factorization (NMF) of the frequency-tract matrix. A fiber tract is quantified by Fourier descriptors in terms of frequencies. Fourier descriptors derived from the shape signature, the central angle dot product, are used to construct the nonnegative frequency-tract matrix which is analogous to the term-document matrix in the document clustering context. In the NMF derived feature space, each basis vector captures the base shape of a particular fiber tract bundle. Each fiber tract is represented as an additive combination of the base shapes. The cluster label of each fiber tract is easily determined by finding the basis vector with which a fiber tract has the largest projection value. Preliminary experimental results with real DTI data show that this method efficiently groups tracts into plausible bundles. This indicates that NMF may be used in fiber tract segmentation with appropriate fiber tract encodings.

I. INTRODUCTION

The human brain's white matter contains complex neural pathways that are vital for the communication of information between functional gray matter regions. Diffusion tensor magnetic resonance imaging (DTI) is a type of diffusion weighted magnetic resonance imaging (MRI) technique based on measuring the Brownian motion of water molecules in isotropic or anisotropic media [1]. It is the only means currently available for *in vivo* investigation of human brain connectivity. With DTI datasets, the white matter fibers can be reconstructed using a class of techniques called tractography [2], [3], [4], [5]. DTI and white matter tractography play a crucial role in understanding anatomical connectivity and functional coupling between regions of the brain, and in clinical applications such as neurosurgery planning and brain disorder diagnosis [6], [7].

The most popular tractography techniques construct white matter fiber tracts as streamlines [2], [3]. A group of fiber tracts form a white matter fiber bundle that interconnects gray matter regions. Due to the spatial resolution of the currently available DTI datasets, the reconstructed fiber tract streamlines do not correspond to individual axons. Thus, individual streamlines are of no anatomical significance, unless they form a bundle that connects gray matter regions. Clustering reconstructed fiber tract streamlines into meaningful bundles has been an

important step in DTI-based white matter fiber analysis.

Over the past few years, a number of white matter tract segmentation schemes have been published. These approaches roughly fall into three categories, i.e., visual dissection, automatic clustering and atlas based segmentation. Conturo et al. [8] proposed a visual dissection method to interactively select fibers passing through one or more user defined ROIs in the process of DTI fiber tracking. This approach can also be used to remove anatomically implausible tracts. But it is time-consuming to segment large amounts of arbitrarily shaped fibers with complex structures. Good knowledge of brain white matter anatomy is necessary for this method.

In the second category methodologies, Ding et al. [9] bundled fiber tracts by finding the corresponding portion of a fiber that has pointwise correspondence to a portion of another fiber. Burn et al. [10] described a normalized cuts clustering algorithm. Gerig et al. [11] and Corouge et al. [12] introduced Hausdorff and similar distance metrics to implicitly model tract's shape characteristics. relationship Others [13], [14] emphasized their efforts mostly on the development of better clustering algorithms. Batchelor et al. [15] compared several mathematical tools including link, principal component analysis (PCA) and the Euclidean distance of Fourier descriptors to study the relative spatial configurations of trajectory pairs and indicated that these measures could be used in classifying and clustering the reconstructed fiber curves.

The third type of tract clustering techniques are based on a neural path atlas built in advance or drawn or selected manually by users [16], [17]. Knowledge of a fiber bundle is a requirement of this type of approaches.

These existing methods have their advantages and disadvantages and generate different clustering results of fiber tracts. The challenge that we face is to model a fiber tract with a representation method to separate fiber tracts belonging to distinct bundles while grouping tracts of the same bundle.

The contribution of this paper is a novel post processing technique for clustering or segmentation of DTI tractography reconstructed white matter fiber tracts. Especially, we employ Fourier descriptors to quantify a fiber tract's shape signatures and perform segmentation using Nonnegative Matrix Factorization (NMF). NMF has been widely used in many applications, including document clustering context. It has

been shown that NMF surpasses singular value decomposition (SVD) and the eigenvector based clustering methods in the sense that NMF produces both reliable and accurate document clustering results [18]. As a result, the advantage of our method lies in the facts that: (1) Fourier descriptors provide good representation and normalization of white matter fiber tracts; (2) each basis vector of NMF has a straightforward correspondence with each fiber tract bundle in the frequency domain and thereby fiber tract cluster label can be directly inferred without additional clustering operations. In detail, we construct the nonnegative frequency-tract matrix using Fourier descriptors derived from the shape signature, the central angle dot product. This frequency-tract matrix is analogous to the term-document matrix in the document clustering context. In the NMF derived feature space, each basis vector captures the base shape of a particular tract bundle. These basis vectors do not need to be orthogonal. This is different from the clustering methods based on SVD and the related spectral clustering methods. In our approach, each fiber tract is represented as an additive combination of the base shapes. The cluster label of each fiber tract is easily determined by finding the basis vector with which a fiber tract has the largest projection value.

The rest of this paper is organized as follows: Section II describes the methods used in this study. Especially, Subsection II-A illustrates a parameter fitting technique to sample fiber tracts in equal arc length. Subsection II-B introduces the fiber tract shape signature. In Subsection II-C, we describe the fiber tract encoding by Fourier descriptors derived from the fiber tract shape signature. Subsection II-D discuss the clustering method using NMF. Section III gives our preliminary results to show the performance of the NMF clustering method. Finally, Section IV concludes this paper.

II. METHODS

A. Fiber Tract Parameter Fitting

The backward streamline based DTI tractography technique proposed by Mori and van Zijl [3] is employed in this analysis to reconstruct white matter fiber tracts. This tractography technique produces fiber tracts as spatial curves in the three dimensional (3D) space with unequal separation distances between successive step points. Fourier transform approach is dependent on the parameterization of the curves. This requires the parameterization to be standardized. For this reason, all reconstructed tracts are fitted using a fixed geodesic arc length. Distinct fiber tracts have equal length unit but usually different numbers of step points.

B. Shape Signature

Each step point on a fiber tract has its 3D coordinates representing its spatial position and a principal eigenvector indicating the local diffusion orientation in a specific voxel it falls in. We define a fiber tract's global diffusion orientation as the accumulation of local diffusion orientations. Assume $e = \{e_1, e_2, \dots, e_l\}$ is the set of local diffusion orientations of a fiber tract, then the global diffusion orientation e_g is given by $e_g = \sum_{i=1}^l e_i$.

A voxel's diffusion orientation e_i is not equivalent to its associated principal eigenvector in the sense that a diffusion orientation is symmetrically bi-directional while a principal eigenvector is directional. In order to get the correct global orientation e_g , we need to check the directional consistency of the principal eigenvectors. In detail, we start from any one of the two end points of a fiber tract and move towards its unique successive step point. If the dot product between the principal eigenvector and this initial moving direction is negative, we reverse the the principal eigenvector of that point. Then we move forward following the direction of this checked principal eigenvector. At each intermediate step point, if the dot product of the two principal eigenvectors belonging to this step point and its previous neighbor are negative, we reverse the one belonging to this step point and repeat this procedure until reach the other end of this fiber tract.

The central angle dot product (CADP) shape signature is defined as the absolute value of the dot product between a step point's local diffusion orientation and the global diffusion orientation of the tract

$$\frac{|e_i^T \cdot e_g|}{\|e_i\| \|e_g\|}.$$

C. Fourier Descriptors of the Shape Signature

We compute the discrete Fourier transform of a tract as

$$FD(i) = \frac{1}{l} \sum_{k=1}^l s(k) \exp\left(\frac{-j2\pi ik}{l}\right), \quad i = 1, 2, \dots, l, \quad (1)$$

where $s(k)$, $k = 1, 2, \dots, l$, is the CADP shape signature at the k th point, and $j = \sqrt{-1}$. The coefficients $FD(i)$, $i = 1, 2, \dots, l$, are called Fourier descriptors of the shape of a fiber tract. Since the shape signatures that we introduced are all real values, there are only $l/2$ different frequencies in the discrete Fourier transform. If l is an odd number, the number of different frequencies is $(l+1)/2$. In the following discussion, we assume that l is even for convenience. Therefore, only a half of the Fourier descriptors are needed.

Fourier descriptors transform the fiber tract shape from the spatial domain into a frequency domain. This eliminates the difficulty to establish matching correspondence between two randomly organized fiber tracts in the spatial domain. The number of coefficients from Fourier transform is usually large, but a small subset of these coefficients is sufficient to capture the general shape features of a fiber tract. The coefficients corresponding to very high frequencies are not helpful in fiber tract shape differentiation. These high frequency components can be ignored without significant accuracy loss for fiber tract clustering. The lower order components also help filter out noise dependent perturbations. As a result, the dimension of Fourier descriptors used for fiber tract clustering are significantly reduced. In addition, two fiber tracts compared based on the Fourier descriptors do not have to have the same numbers of step points. Consequently, Fourier descriptors can be employed in matching of fiber tracts with unequal length.

To encode a fiber tract with nonnegative values, we discard Fourier descriptors' phase information and keep only their

magnitudes. These magnitudes are all nonnegative. This also makes Fourier descriptors rotation invariant without any loss of encoding accuracy in the context of this study. In the following discussion, we use $f(i)$ to represent a Fourier descriptor's magnitude, i.e., $f(i) = |FD(i)| \geq 0$. We also refer $f(i)$ as a Fourier descriptor for simplicity though it is actually a Fourier descriptor's magnitude. Now, each fiber tract can be encoded by an encoding vector, f , composed of a set of Fourier descriptors $f = [f(1), \dots, f(i), \dots, f(m)]$ with all $f(i) \geq 0$ for $i = 1, 2, \dots, m$, where $m (\leq l/2)$ is the number of truncated Fourier descriptors.

D. Clustering based on Nonnegative Matrix Factorization

To construct the nonnegative frequency-tract matrix, we put each fiber tract's encoding vector, e.g., for the j th tract, $f_j = [f_j(1), \dots, f_j(i), \dots, f_j(m)]$ into the j th column of a two dimensional (2D) matrix V . Assume n is the number of fiber tracts under consideration, the constructed nonnegative matrix V thus has the dimension $m \times n$.

We have expressed a set of fiber tracts as an $m \times n$ nonnegative frequency-tract matrix V . Each column V_j of V is an encoding of a fiber tract in the frequency domain and each entry v_{ij} of vector V_j is the significance of frequency i with respect to the fiber tract representation of V_j .

The NMF is defined as finding two low rank nonnegative matrix factors W and H of a given nonnegative matrix V such that $V \approx WH$ [19]. Each column of W is a basis vector and each column of H contains an encoding of the linear combination of the basis vectors that approximates the corresponding column of V . W and H each has the dimension $m \times r$ and $r \times n$ respectively. r is the number of clusters selected. Usually, $r \ll \min(m, n)$. Finding or estimating the approximate value of r depends on applications.

The approach using NMF to obtain factor matrices W and H in this study is to minimize the Frobenius norm of the difference $V - WH$, i.e., minimize $\|V - WH\|^2$ with respect to W and H with $W_{ij} \geq 0$ and $H_{ij} \geq 0$ for each i and j [19]. The updating rule to produce W and H is based on the multiplicative method proposed by Lee and Seung [19]. In the multiplicative method, we first initialize W and H with nonnegative values and then perform iterations for each α , i , and j until convergence. The updating formulas are

$$H_{\alpha j} \leftarrow H_{\alpha j} \frac{(W^T V)_{\alpha j}}{(W^T W H)_{\alpha j}} \quad (2)$$

$$W_{i\alpha} \leftarrow W_{i\alpha} \frac{(V H^T)_{i\alpha}}{(W H H^T)_{i\alpha}} \quad (3)$$

At each iteration, W and H remain nonnegative and the columns of W or the basis vectors are normalized to unity. Updating W and H simultaneously generally yields better results than updating each matrix factor individually.

This multiplicative method is related to the expectation maximization (EM) technique used in image processing context and can be classified as a diagonally scaled gradient descent method. It is proved that the Euclidean distance

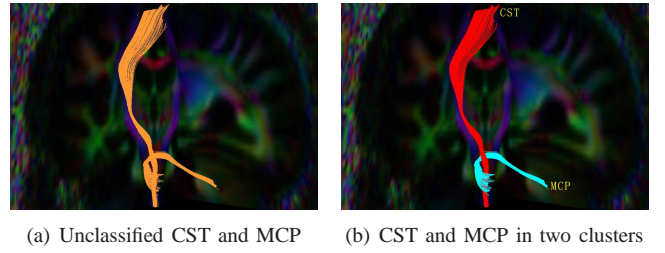


Fig. 1. Illustration of unclassified and classified CST-MCP tracts.

$\|V - WH\|$ is monotonically nonincreasing under the above multiplicative method updating rules, and that the convergence of the iteration is guaranteed [19].

We then use matrix H to identify the cluster membership of a fiber tract. In detail, for the j th fiber tract, we compare entries across rows on the j th column of H . Its cluster label is x if $x = \arg \max_i \{H_{ij}\}$.

III. PRELIMINARY RESULTS

The real dataset used to assess the performance of this method is a fiber tract collection consisting of the corticospinal tract (CST) bundle and the medial cerebellar peduncle (MCP) bundle. It is denoted as CST-MCP dataset and shown in Figure 1(a). This dataset has totally 310 fiber tracts. Figure 1(b) illustrates the segmented two subbundles, CST and MCP. Among the 310 fiber tracts, 199 and 111 of them are grouped into CST and MCP respectively.

In this test, the first 30 Fourier descriptors derived from the CADP shape signature of each fiber tract are used to form the nonnegative frequency-tract matrix V which thus has the dimension 30×310 . The cluster number r is selected as 2. After applying NMF, the dimensions of the yielded nonnegative matrices W and H are 30×2 and 2×310 respectively. For the j th ($j = 1, 2, \dots, 310$) fiber tract, we then compare the values of the two rows on the j th column of matrix H . Its cluster label is identified as the row index i ($i = 1, 2$) if this row has the larger value than that of another row.

Figure 2 graphically demonstrates the entry value comparisons between each two rows of the first 10 columns of the matrix H . Among these 10 fiber tracts, the first and last 3 (colored in red) are identified as CST. For each of them, the entry value on the first row (marked as +) is greater than that of the second row (marked as o), and vice versa for the middle four fiber tracts (colored in blue). Figure 3 shows the corresponding first 10 fiber tracts colored according to their cluster labels.

IV. DISCUSSION AND CONCLUSION

In this work, we presented a novel technique to group white matter fiber tracts reconstructed from DTI into bundles. This approach is based on NMF. The construction of the nonnegative frequency-tract matrix and the encoding of fiber tracts were described. Preliminary experimental results have

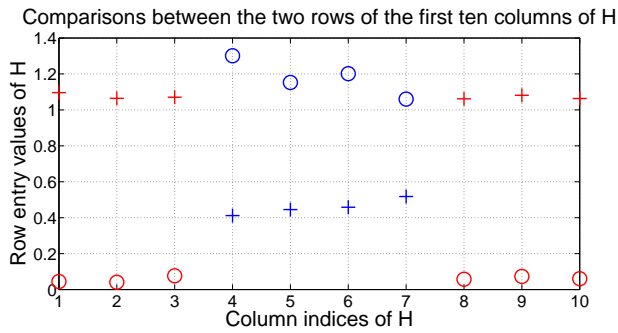


Fig. 2. Value comparisons of the first 10 columns of the matrix H . Red: labeled as CST; Blue: labeled as MCP; +: first row; o: second row.

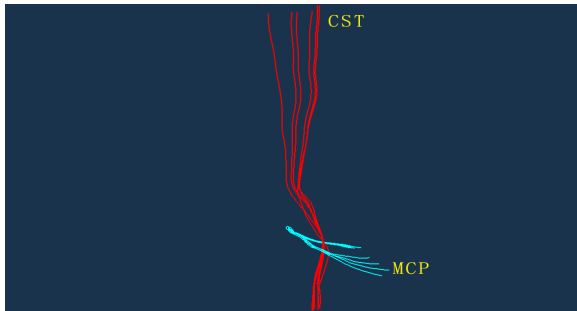


Fig. 3. Classified first 10 fiber tracts.

been obtained and show that this technique can efficiently separate fascicles into plausible bundles.

In summary, NMF has mostly been applied to text mining and image analysis. Medical imaging, especially the white matter fiber tract analysis, may benefit from this technique as well. Problems such as identifying significant features in the encoded fiber tracts are natural candidates for applications of NMF. In such contexts, fiber tracts can be treated as analogous to text documents and the quantified weights of features to term frequencies. The basis feature vectors can be viewed as analogous to the base topics. With further development, it may be possible to apply NMF in quantifications of the deformation of white matter fibers under certain pathologies and brain diseases, e.g., tumor, or for clinical studies.

ACKNOWLEDGMENT

Technical Report CMIDA-HiPSCCS 012-08, Department of Computer Science, University of Kentucky, Lexington, KY, 2008.

The authors acknowledge the support of funding agencies and the collaborators. The research work of J. Zhang was supported in part by the US National Science Foundation under grant CCF-0527967 and CCF-0727600, in part by the National Institutes of Health under grant 1R01HL086644-01, in part by the Kentucky Science and Engineering Foundation under grant KSEF-148-502-06-186, and in part by the Alzheimer's Association under grant NIRG-06-25460.

REFERENCES

[1] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI," *Journal of Magnetic Resonance Series B*, vol. 111, no. 3, pp. 209–219, 1996.

[2] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, "In vivo fiber tractography using DT-MRI data," *Magnetic Resonance in Medicine*, vol. 44, no. 4, pp. 625–632, 2000.

[3] S. Mori and P. C. M. van Zijl, "Fiber tracking: principles and strategies - a technical review," *NMR in Biomedicine*, vol. 15, no. 7-8, pp. 468–480, 2002.

[4] J. Zhang, N. Kang, and S. E. Rose, "Approximating anatomical brain connectivity with diffusion tensor MRI using kernel-based diffusion simulations," in *Proceedings of Information Processing in Medical Imaging (IPMI 2005)*, *Lecture Notes in Computer Science*, 2005, vol. 3565, pp. 64–75.

[5] N. Kang, J. Zhang, E. S. Carlson, and D. Gembris, "White matter fiber tractography via anisotropic diffusion simulation in the human brain," *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1127–1137, 2005.

[6] S. C. Patridge, P. Mukherjee, J. I. Berman, R. G. Henry, S. P. Miller, Y. Lu, O. A. Glenn, D. Ferriero, A. J. Barkovich, and D. B. Vigneron, "Tractography-based quantitation of diffusion tensor imaging parameters in white matter tracts of preterm newborns," *Journal of Magnetic Resonance Imaging*, vol. 22, no. 4, pp. 467–474, 2005.

[7] X. Liang and J. Zhang, "White matter integrity analysis along cingulum paths in mild cognitive impairment - a geodesic distance approach," in *Proceedings of 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2008)*, vol. 1, Shanghai, China, June 2008, pp. 510–513.

[8] T. E. Conturo, N. F. Lori, T. S. Cull, E. Akbudak, A. Z. Snyder, J. S. Shimony, R. C. McKinstry, H. Burton, and M. E. Raichle, "Tracking neuronal fiber pathways in the living human brain," *Neurobiology*, vol. 96, no. 18, pp. 10 422–10 427, 1999.

[9] Z. Ding, J. C. Gore, and A. W. Anderson, "Classification and quantification of neuronal fiber pathways using diffusion tensor MRI," *Magnetic Resonance Medicine*, vol. 49, pp. 716–721, 2003.

[10] A. Brun, H. Knutsson, H.-J. Park, M. E. Shenton, and C.-F. Westin, "Clustering fiber tracts using normalized cuts," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004)*, *Lecture Notes in Computer Science*, 2004, vol. 3216, pp. 368–375.

[11] G. Gerig, S. Gouttard, and I. Corouge, "Analysis of brain white matter via fiber tract modeling," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, California, September 2004, pp. 4421–4424.

[12] I. Corouge, S. Gouttard, and G. Gerig, "Towards a shape model of white matter fiber bundles using diffusion tensor MRI," in *Proceedings of International Symposium on Biomedical Imaging (ISBI) 2004*, vol. 1, no. 5, Arlington, Virginia, April 2004, pp. 344–347.

[13] V. E. Kouby, Y. Cointepas, C. Poupon, D. Riviere, N. Golestani, J.-B. Poline, D. L. Bihan, and J.-F. Mangin, "MR diffusion-based inference of a fiber bundle model from a population of subjects," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2005)*, *Lecture Notes in Computer Science*, 2005, pp. 196–204.

[14] L. Jonasson, P. Hagmann, J.-P. Thiran, and V. J. Wedeen, "Fiber tracts of high angular resolution diffusion MRI are easily segmented with spectral clustering," in *Proceedings of 13th Annual Meeting of International Society for Magnetic Resonance in Medicine*, Miami, Florida, May 2005, p. 1310.

[15] P. G. Batchelor, F. Calamante, J.-D. Tournier, D. Atkinson, D. L. G. Hill, and A. Connelly, "Quantification of the shape of fiber tracts," *Magnetic Resonance Medicine*, vol. 55, pp. 894–903, 2006.

[16] L. J. O'Donnell and C.-F. Westin, "Automatic tractography segmentation using a high-dimensional white matter atlas," *IEEE Transaction on Medical Imaging*, vol. 26, no. 11, pp. 1562–1575, 2007.

[17] M. Maddah, W. E. Grimson, S. K. Warfield, and W. M. Wells, "A unified framework for clustering and quantitative analysis of white matter fiber tracts," *Medical Image Analysis*, vol. 12, no. 2, pp. 191–202, 2008.

[18] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada, July 28-August 01 2003, pp. 267 – 273.

[19] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, vol. 13, Vancouver, British Columbia, Canada, December 2001, pp. 556–562.