Tract-Based Probability Densities of Diffusivity Measures in DT-MRI

Çağatay Demiralp and David H. Laidlaw

Brown University, USA

Abstract. We evaluate probability density functions of diffusivity measures in DTI fiber tracts as biomarkers. For this, we estimate univariate and bivariate densities, such as joint probability densities of the tract arc length and FA, MD, RD, and AD, in the transcallosal fibers in the brain. We demonstrate the utility of estimated densities in hypothesis testing of differences between a group of patients with VCI and a control group. We also use the estimated densities in classifying individual subjects in these two groups. Results show that these estimates and derived quantities, such as entropy, can detect group differences with high statistical power as well as help obtain low classification errors.

1 Introduction

Tract-based approaches to analysis of diffusion brain data sets are attracting interest because they can geometrically localize quantities while reducing sensitivity of associated statistics to computational errors and noise (e.g., [1,2,3]). In this context, we propose using probability densities in characterizing diffusion along diffusion-tensor MRI (DTI) fiber tracts generated from diffusion-weighted imaging data sets. We experiment with well known diffusivity measures such as fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD) [4]. We estimate univariate and bivariate diffusion densities of these four measures nonparametrically, using histogram and kernel density estimation methods. We use the estimated densities for hypothesis testing of differences between a group of patients with vascular cognitive impairment (VCI) and a control group. Results indicate that probability density estimates can detect group differences with high statistical significance.

Contributions. Our main contribution is to introduce tract-based probability density functions, including the joint density of tract arc length and scalar diffusivity measures, as potential metrics for characterizing diffusivity in DT-MRI brain data sets. Our approach is a simple and effective addition to existing methods. We also show that probability density entropy itself is a useful biomarker.

2 Related Work

Using voxel or integral curve (as in tractography) representations of fiber tracts, previous research has proposed tract-based analysis of diffusion-derived measures in order to increase the specificity and robustness of related statistics [1,2,3]. A typical approach in these earlier studies is to find a representative skeleton (or curve) first and then project the diffusivity measures of individual subjects on this representative. For example, tract-based spatial statistics (TBSS) has been proposed to overcome some of the disadvantages

T. Jiang et al. (Eds.): MICCAI 2010, Part I, LNCS 6361, pp. 542–549, 2010.

of approaches using voxel-based morphometry (VBM) [1]. TBSS reduces the sensitivity of statistics to alignment problems by projecting subject data onto a voxel skeleton of the white matter as characterized by FA. Using tractography makes the incorporation of fiber tract arc length into the analysis easier. One of the joint densities evaluated here is the probability of a diffusivity value given an arc-length distance. Parametrization of tracts based on arc-length distance has already been shown to be useful [5], and tract arc length by itself has been proposed as a metric [6]. As in TBSS, previous studies using tractography have also been able to detect differences between groups of subjects by analyzing diffusivity measures projected onto an arc-length parametrized representative tract [2,3].

Recently, probability density functions (PDFs) have been used as shape descriptors in cortical folding analysis and in quantifying FA change with respect to the thickness of the surface sheet covering the corticospinal tract [7,8].

While earlier tract-based analysis methods essentially collapse bundles into a single representative subset (skeleton or curve) and run the subsequent analysis through this subset, our approach considers each bundle directly and separately, modeling the full spectrum of individual bundle's data as a sample from a probability distribution characterizing the corresponding subject. Also, our use of PDFs is simple, general (it can be applied to any tracts and combine any number of variables), and truly spatial, since we orient tracts.

3 Data Collection

The underlying premise of using tract-based probability densities in clinical research is that they have different values in patients with and without known white-matter injury. In order to verify this, we generated tractograms of a group of patients with vascular cognitive impairment (VCI) and a group of healthy control subjects and compared the group differences using probability densities. VCI is a general term for vascular cognitive deficits caused by injuries to the white matter from cerebrovascular diseases. VCI typically manifests itself with problems in speech, language, and the ability to organize thoughts and track actions. Its effects on memory, however, are considered mild in comparison to Alzheimer's disease. Often, deficits in cognitive domains caused by VCI are similar to those common in individuals of advanced age. Below we give only a summary of our data collection process; details can be found in [6].

Subjects. Our subjects were 19 patients with VCI and 20 healthy individuals who served as a control group. The subjects were age-matched. Diffusion-weighted MRI (DWI) sequences for each subject's brain were acquired on a 1.5T Siemens Symphony scanner with the following parameters in 12 bipolar diffusion encoding gradient directions: thickness = 1.7 mm, FOV = 21.7 cm $\times$ 21.7 cm, TR = 7200 ms, TE = 156 ms, b = 1000, and NEX = 3.

Fiber-tract Generation. From acquired sequences, we calculate tensors and then derive the three principal eigenvalues and eigenvectors for each image voxel after interleaving the three sets of DWIs. We then generate fiber-tract models of the whole brain for each subject by integrating (by second-order Runge-Kutta integration) the major

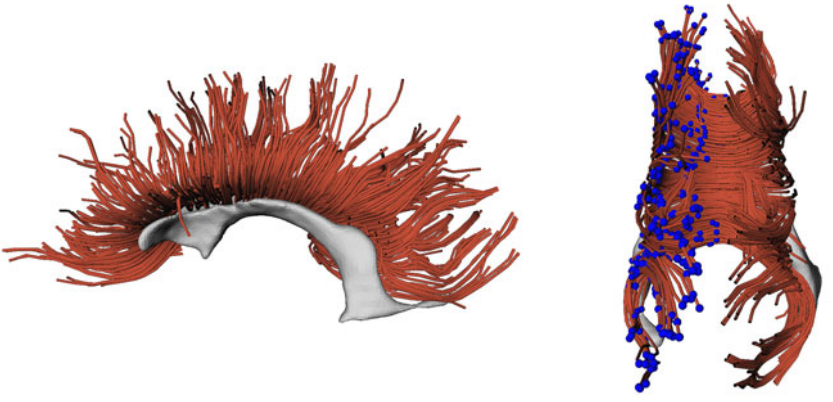

Fig. 1. Transcallosal fiber tracts obtained from a healthy subject. The gray surface representing the ventricles is displayed for anatomical correspondence. Left: Sagittal view. Right: Tilted coronal view, where spheres (blue in color) indicate the "beginning" of each tract.

eigenvector field of the diffusion tensor field bidirectionally at seed points (see Figure 1). Transcallosal fibers, defined as all trajectories passing through the corpus callosum, for each participant were selected manually by a rater using a custom interactive visualization program. Selected models were checked by two experts for anatomical correctness. Since tract of interest (TOI) selection was based on each participant's own anatomy, registration was not necessary.

Orienting Tracts. Since we compute the joint probabilities of arc length and the diffusivity measures, we need to assign start- and end-point designations consistently within and across fiber tracts and across subjects. We achieve this by simply computing the start-to-end vector for each curve and iteratively reorienting the curves until all the vectors are in the same half-space; we repeat this interactively between data sets.

4 Methods

Density Estimation. Since we do not assume a particular distribution of diffusivity measures along tracts, we take a nonparametric approach to estimation. We use two widely studied estimators, histogram and kernel estimators [9,10]. Let $X_1^N=\{X_1,\ldots,X_N\}$ be independent and identically distributed (i.i.d.) random variables drawn from a density $f(x)$. The histogram estimator of f is obtained simply by dividing the range of the samples into bins of width h and counting the number of samples falling into each bin.

Kernel density estimation (KDE) addresses some of the obvious disadvantages of histograms, including non-differentiability and reference-point dependency. The kernel estimator of f using X_1^N is given by $\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K(\frac{x-X_i}{h})$, where $K(x)$ is some kernel satisfying $\int_{-\infty}^{\infty} K(x)dx = 1$ and h is the kernel *bandwidth*, analogous to the histogram bin width. We omit here multivariate versions of the equations; they can be found in any good multivariate statistical analysis book (e.g., [9,10]).

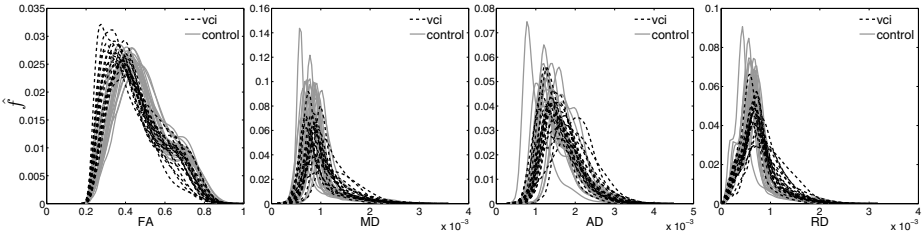

Fig. 2. Univariate density estimates of FA, MD, AD, and RD measures for control subjects and subjects with VCI. They suggest an increase in variances of MD, AD, RD densities with VCI. Also, the shifts in the densities of the two groups are consistent with previous reports: while FA decreases, MD, RD and, to a lesser extent, AD increase with atrophy in the brain.

Both bin width and kernel size can affect the results significantly. Therefore we use a data-driven rule originating in the L_2 theory of histograms, minimizing the mean integrated square error (MISE) [10,11]. We set the histogram bin width $h = 3.5\hat{\sigma}N^{-\frac{1}{d+2}}$, where d is the dimensionality of the data (e.g., $d = 2$ for a bivariate distribution) and N is the number of samples. We select $\hat{\sigma} = min\{s, IQR/1.349\}$ as discussed in [12], where s is the standard deviation of samples and IQR is the inter-quantile range. We use a Gaussian kernel with both univariate and bivariate estimators. As for the kernel bandwidth h, we apply the "normal scale" rule introduced in [9], which is also asymptotically optimal.

Non-parametric Multivariate Hypothesis Testing. Both histogram and kernel estimates (kernel estimators in their discretized forms) essentially provide a vector of values identifying the underlying distribution. These "density vectors" can be used for testing differences between groups, which requires a multivariate test statistic. We use the permutation Hotelling's T^2 test [13]. Hotelling's T^2 statistic is a multivariate extension of Student's t statistic and permutation is a common method for running non-parametric tests using using scores obtained with parametric statistics.

Entropy. One of the usual quantities computed over probability densities is entropy (or Shannon entropy) [14]. While entropy has different interpretations in different domains, it can be viewed here as a measure of randomness in data (i.e. uniformity of its distribution). We compute the entropy of the density f using the estimated density as $H(\hat{f}) = -\sum_x x\hat{f}(x)$ and evaluate it in group comparison of our subjects.

Classification. Classification is now a standard approach to exploring medical data sets. We use a support vector machine (SVM) classifier to assess the usefulness of quantities derived from tract-based diffusivity probability densities. SVM is a maximum-margin classifier minimizing the classification error while maximizing the geometric margin between the classes [15]. It is particularly suitable for two-class problems.

5 Results

We estimate univariate densities of FA, MD, AD, and RD as well as joint densities of these measures with fiber tract arc length, where arc length is normalized to 1, in each

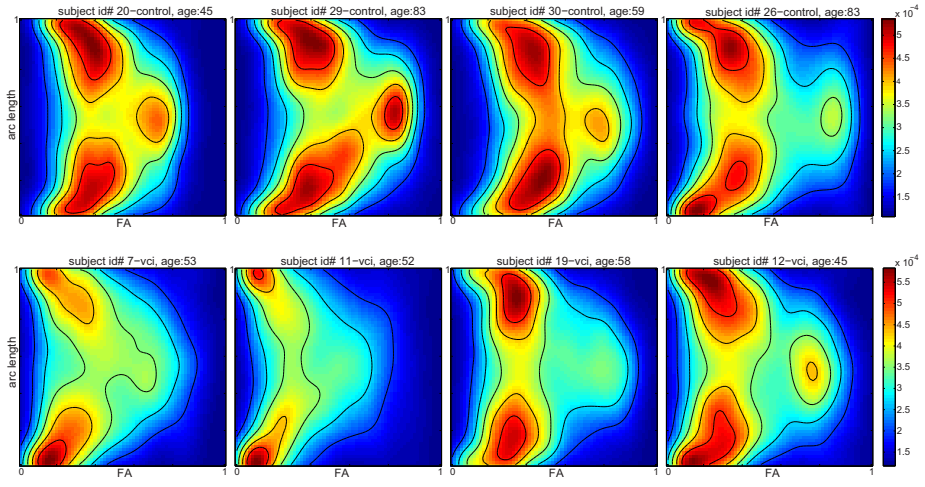

Fig. 3. Kernel density estimates of FA and arc length joint PDF for 8 subjects–4 controls (first row) and 4 patients (second row). The third local maximum (on the mid-right) starts disappearing with VCI. Also, observe, in the last column, the similarity between a healthy subject of advanced age and a young patient.

subject's transcallosal fiber tract. Overall, results show that the tract-based probability densities help detect differences between groups and classify individual subjects with high accuracy. Figure 2 shows kernel estimates for univariate FA, MD, AD, and RD probability densities. Even in this simple univariate form, estimated PDFs are rich in information. They suggest an increase in variances of MD, AD, RD densities with VCI. Also the shifts in the estimated densities of the two groups support previous reports that, while FA decreases, MD, RD and, to a lesser extent, AD increase with atrophy in the brain. Figure 3 shows kernel estimates for the joint probability of FA with the arc length. These 2D density results are visually informative and can help generate hypotheses. For example, it is clear from the figures that the third local maximum (on the mid-right) starts disappearing with VCI and age. Similarly, the decrease in MD with VCI is clear in MD and the arc length joint density shown in Figure 4.

5.1 Group Comparisons

Using Entropies. We expect neurological diseases to affect the entropy of probability distributions of diffusivity-related measures. We test this hypothesis for all the measures using their univariate and bivariate densities. The results show that MD, RD, and AD entropies increase significantly while FA decreases (see Figure 5 and Table 1a). This outcome is in line with previous findings that AD is more sensitive to atrophy than FA [16].

Density Vectors. Using vectorized "raw" density values, we compare the group of subjects with VCI with our control group, employing the multivariate permutation

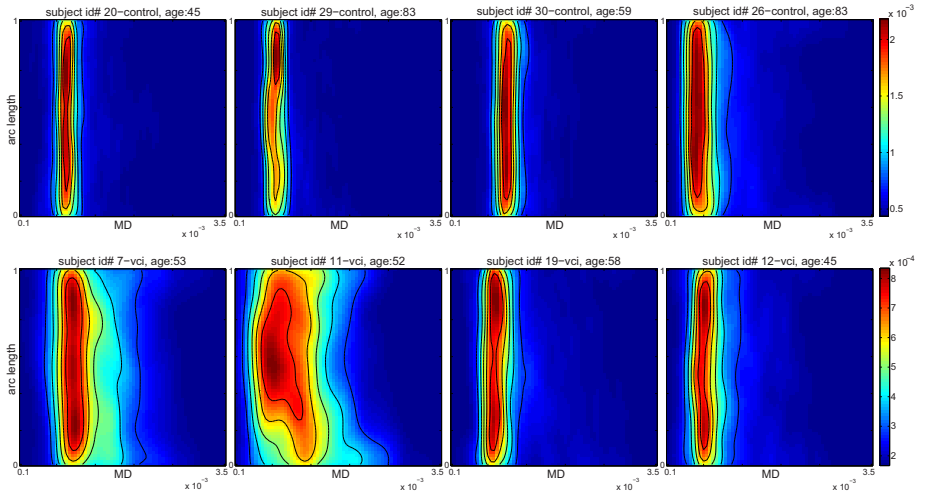

Fig. 4. Kernel density estimates of MD and arc length joint PDF for 8 subjects–4 controls (first row) and 4 patients (second row). The joint PDFs tend to spread out (i.e., increase in variance) with VCI. As seen in the last column, aging can have an effect similar to VCI.

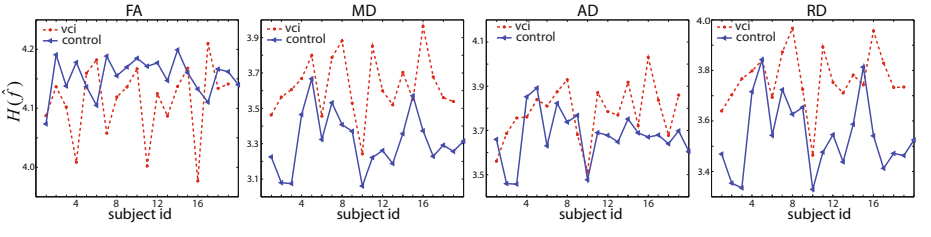

Fig. 5. Entropies for univariate probability density estimates of FA, MD, AD, and RD for all subjects. MD, AD, and RD density entropies significantly increase with VCI, but the FA entropy decreases, though not significantly (see also Table 1b).

Table 1. (a) p-values for the group comparison (VCI vs. control) using entropies of the estimated probability densities. Except for FA, the entropy measure was able to detect the group difference for all densities. (b) p-values for the group comparison (VCI vs. control) using probability density vectors. Histogram estimates of univariate densities were able to detect the group differences.

(a)

	univariate		bivariate	
	histogram	kernel	histogram	kernel
FA	0.36797	0.01239	0.0297	0.8379
MD	0.00000	0.00000	0.00000	0.00000
AD	0.00106	0.00878	0.0042	0.00016
RD	0.00000	0.00000	0.00000	0.00000

(b)

	univariate
FA	0.00078
MD	0.00000
AD	0.00005
RD	0.00002

Table 2. Classification results for our 39 subjects, obtained using an SVM classifier with 10-fold cross-validation. Data points used for classification are univariate histogram density estimates.

density	sensitivity	specificity	error
FA	83.3%	81%	17.9%
MD	90%	94.7%	8%
AD	77.2%	88.2%	17.9%
RD	89.5%	90%	10.2%

Hotelling's test [13,17]. We reduce the dimensionality of the density vectors using principal component analysis (PCA) before running the test, while preserving 99% of the variance. The results show that the density vectors of univariate estimates are able to detect group differences (see Table 1b).

5.2 Classifying Normal and Pathology

We ran an SVM with a polynomial kernel on 39 data sets using 10-fold cross-validation. Using density estimates of MD yields the best classification. Table 2 summarizes the results.

6 Discussion and Conclusions

We have proposed a simple and practical addition to existing tract-based multi-subject diffusion data analysis techniques, which are essentially based on collapsing bundles into a representative subset and sampling diffusion measures along this subset. Our approach considers all the tracts in a given bundle and treats the associated diffusivity measures as samples from a PDF characterizing the subject. We estimate this PDF using the histogram and kernel estimators. While using the histogram density estimator is fast and simple, the kernel estimator provides smoother results with useful differential properties.

Given a reasonable choice of variables, estimated PDFs are likely to be descriptive and rich in information, and therefore well-suited for analysis of individual and group differences in diffusion data sets. In the case examined here, both PDFs themselves and their entropy measures appear to be good indicators of group differences, although additional work is required to understand how this generalizes to other cases. Similarly, our initial results on the classification of subjects based on densities suggest the potential utility of PDFs in detecting individual differences. Note that using natural distance measures on PDFs such as the Bhattacharyya distance and Kullback-Leibler (KL) divergence could further improve the analysis discussed above. Finally, although we have not dealt with alignment issues, the probabilistic approach discussed here can also increase the robustness of the analysis to registration errors in fully automated settings.

Acknowledgments. This work has been supported by NIH grant 1R01EB00415501A1.

References

1. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 31(4), 1487–1505 (2006)
2. O'Donnell, L.J., Westin, C.F., Golby, A.J.: Tract-based morphometry. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 161–168. Springer, Heidelberg (2007)
3. Goodlett, C.B., Fletcher, P.T., Gilmore, J.H., Gerig, G.: Group statistics of DTI fiber bundles using spatial functions of tensor measures. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 1068–1075. Springer, Heidelberg (2008)
4. Song, S.K., Sun, S.W., Ramsbottom, M.J., Chang, C., Russell, J., Cross, A.H.: Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. NeuroImage 17(3), 1429–1436 (2002)
5. Corouge, I., Fletcher, P.T., Joshi, S., Gouttard, S., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. Medical Image Analysis 10(5), 786–798 (2006)
6. Correia, S., Lee, S.Y., Voorn, T., Tate, D.F., Paul, R.H., Zhang, S., Salloway, S.P., Malloy, P.F., Laidlaw, D.H.: Quantitative tractography metrics of white matter integrity in diffusion-tensor MRI. NeuroImage 42(2), 568–581 (2008)
7. Awate, S.P., Yushkevich, P.A., Song, Z., Licht, D.J., Gee, J.C.: Multivariate high-dimensional cortical folding analysis, combining complexity and shape, in neonates with congenital heart disease. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) Information Processing in Medical Imaging. LNCS, vol. 5636, pp. 552–563. Springer, Heidelberg (2009)
8. Zhang, H., Awate, S.P., Das, S.R., Woo, J.H., Melhem, E.R., Gee, J.C., Yushkevich, P.A.: A tract-specific framework for white matter morphometry combining macroscopic and microscopic tract features. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 141–149. Springer, Heidelberg (2009)
9. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. In: Monographs on Statistics and Applied Probability. Chapman and Hall, London (1986)
10. Scott, D.W.: Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley Interscience, Hoboken (1992)
11. Freedman, D., Diaconis, P.: On the histogram as a density estimator: L_2 theory. Probability Theory and Related Fields 57(4), 453–476 (1981)
12. Wand, M.P.: Data-based choice of histogram bin width. The American Statistician 51(1), 59–64 (1997)
13. Good, P.I.: Permutation, Parametric, and Bootstrap Tests of Hypotheses. Springer Series in Statistics. Springer, New York (2004)
14. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley Interscience, New York (1991)
15. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
16. Metwalli, N., Benatar, M., Usher, S., Bhagavatheeshwaran, G., Hu, X., Carew, J.: Sensitivity of axial and radial diffusivity to the neuropathology of amyotrophic lateral sclerosis. NeuroImage 47 (Suppl. 1), S128–S128 (2009)
17. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, New York (1993)