Final version, in print SPIE 2012, to appear Feb. 2012

Measures for Validation of DTI Tractography

Sylvain Gouttard[a], Casey B. Goodlett[b], Marek Kubicki[c], Guido Gerig[a,d]

[a]Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT;
[b]Kitware, Inc., USA;
[c]Psychiatry Neuroimaging Laboratory, Harvard Medical School, Boston, MA;
[d]School of Computing, University of Utah, Salt Lake City, UT;

ABSTRACT

The evaluation of analysis methods for diffusion tensor imaging (DTI) remains challenging due to the lack of gold standards and validation frameworks. Significant work remains in developing metrics for comparing fiber bundles generated from streamline tractography. We propose a set of volumetric and tract oriented measures for evaluating tract differences. The different methodsdeveloped for this assessment work are: an overlap measurement, a point cloud distance and a quantification of the diffusion properties at similar locations between fiber bundles. The application of the measures in this paper is a comparison of atlas generated tractography to tractography generated in individual images. For the validation we used a database of 37 subject DTIs, and applied the measurements on five specific fiber bundles: uncinate, cingulum (left and right for both bundles) and genu. Each measurments is interesting for specific use: the overlap measure presents a simple and comprehensive metric but is sensitive to partial voluming and does not give consistent values depending on the bundle geometry. The point cloud distance associated with a quantile interpretation of the distribution gives a good intuition of how close and similar the bundles are. Finally, the functional difference is useful for a comparison of the diffusion properties since it is the focus of many DTI analysis to compare scalar invariants. The comparison demonstrated reasonable similarity of results. The tract difference measures are also applicable to comparison of tractography algorithms, quality control, reproducibility studies, and other validation problems.

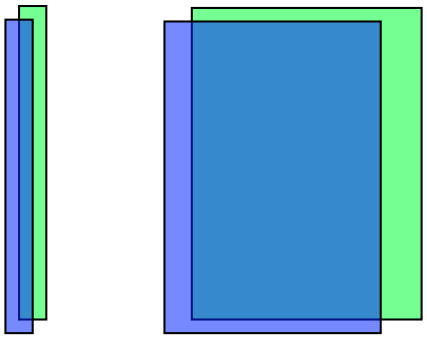
1. DESCRIPTION OF PURPOSE

Diffusion Tensor Imaging (DTI) has increasingly been used by clinical neuroimaging studies to study white matter properties in populations of subjects. Fiber tractography has been explored as a method for extracting white matter fiber bundles. Atlas building procedures for DTI intend to obtain automatic extraction of fiber bundles in a population by mapping to a reference coordinate system.[1–3] In all of these methods, evaluation of the quality and reliability of fiber bundle identification remains a significant challenge.

Several groups have proposed measures for evaluating tractography and DTI atlas building. Zhang et al. proposed several methods for evaluating their registration procedure differences in tensor parameters as well as evaluation of white matter fiber bundle differences.[1] The fiber bundle measure is similar to the one presented later in section 2.2 but does not account for the distribution of closest point distances between two fiber bundles. Ziyan et al. proposed a fiber match metric, FiT, to evaluate the agreement of a fiber bundle deformed into an image for the particular case of comparing tracts to a registered image.[4] This method, however, is unable to compare tracts produced by different algorithms and places an emphasis on the tangent vector of individual streamlines. We propose a set of measures for evaluating the difference between fiber bundles including both geometric measures and comparison of the diffusion statistics segmented by fiber bundles. These measures can be used for evaluating new tractography algorithms, quality control, measuring reproducibility, and comparing atlas based segmentation to manual tractography. In this paper we apply the measures to evaluate tractography mapped from an atlas to fiber bundles generated by tractography in native space.

2. METHOD FOR TRACTOGRAPHY COMPARISON

This section covers a set of measures which can be used to compare streamlines generated by fiber tractography. The motivation is to compare fiber bundles using measures that are robust to outliers, provide physical intuition, and focus on the global shape of the fiber bundle rather than individual streamlines.



(a) Overlap = 0.3 (b) Overlap = 0.9

Figure 1. Example of instability of overlap measures for thin structures.

2.1 Volumetric overlap

As a preliminary measure of volumetric overlap, the probabilistic overlap metric implemented in Valmet was adapted to tractography .[5] This measure is referred to as the binary tractography overlap (BTO) and is defined by

$$BTO = 1 - \frac{\sum_i |P_A(x_i) - P_B(x_i)|}{\sum_i P_A(x_i) + P_B(x_i)}. \quad (1)$$

Here $P_A(x_i)$ is a measure of the probability that voxel x_i is part of the fiber tract. This is approximated by dividing the number of streamlines in the voxel by the median number of streamlines over all voxels containing any streamlines and clamping to a maximum value of 1. This approximation is intended to label as high probability voxels of the tract containing a significant number of streamlines while tapering out the influence for voxels with only a few streamlines.

This method benefits from the similarity to existing volumetric overlap measures and its relative simplicity. Furthermore, it can compare streamline methods with proposed volumetric tractography methods.[6,7] However, volumetric overlap measures for tractography have several serious drawbacks. First, the measure is limited to grid based measurements which are significantly affected by partial voluming effects. Second, many fiber tracts in the human brain are long in one dimension and narrow in one or two orthogonal dimensions. For example, the cingulum is a long thin tube. As shown in Fig. 1, a misregistration of fiber bundles by less than one voxel in thin dimensions can result in overlap measures that are significantly smaller, while larger structures misaligned by similar physical amounts have much higher overlap measures. An additional drawback of the BTO measure is the lack of physical units that gives little intuition into how tracts differ.

2.2 Point Cloud Divergence

A second method of measuring tract differences can be considered that treats fiber bundles as sampled point clouds. This approach avoids some of the drawbacks of converting streamlines into a voxel grid. To compare two fibers bundles A and B, find the distance between each point p_i in A and the closest point q_i in B. For efficient lookup of the closest point a Delaunay triangulation of the points in tract B can be computed and used for fast lookup of the closest point to p_i. This produces a distribution of distances $d(p_i, q_i)$, from bundle A to bundle B. As shown in Fig. 3 these distributions are heavily weighted towards zero with a large percentage of points being very close. At the maximum of the distribution there are typically a small number of streamlines which diverge between the bundles that produce large distances. A graphical representation showing the closest point distance for each fiber bundle is shown in Fig. 2. Previous research has considered the minimum, mean, or maximum of such distributions.[1,8] However, the minimum and mean distance are heavily biased by the large percentage of closest point distances which are very close to zero. The maximum, on the other hand, is extremely sensitive to the outliers common in streamline tractography.

We propose a family of closest point distances between two fiber bundles A and B, $PC_\alpha(A, B)$, that is defined as the α quantile of the distribution of distances from A to B. Choosing α to be relatively close to 1 gives a measure that is resistant to outliers, but gives an intuition of how close the bulk of points are between the two

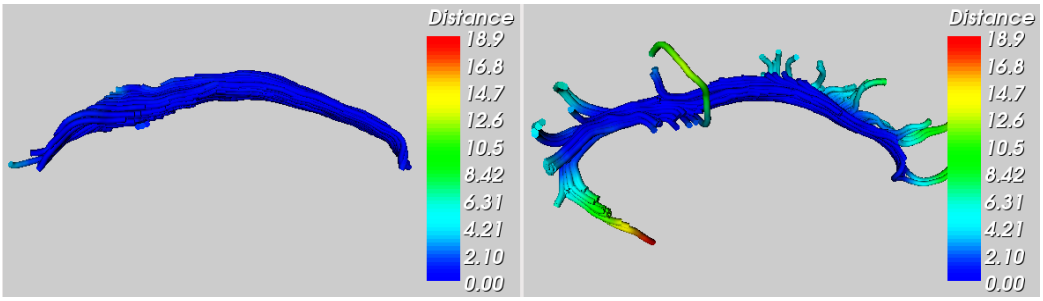


Figure 2. Closest point distances for atlas and individual tractography results for the right cingulum in one subject.

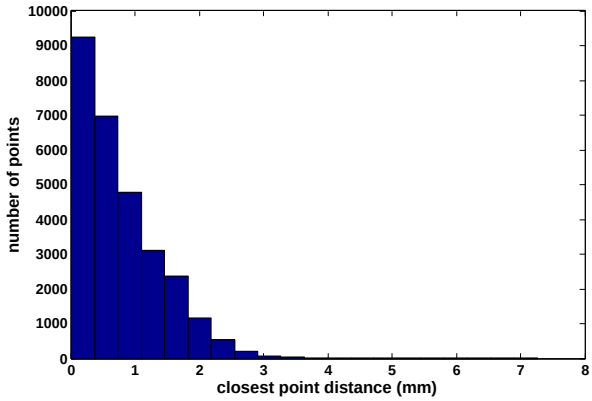


Figure 3. Histogram of closest point distances from atlas tract to individual tract.

bundles. For example, $CP_{.9}(A, B) = 2.0mm$ provides an upper confidence limit that 90% of points in A are within one voxel of B with the $2x2x2mm$ voxels common in DTI. PC_α is not symmetric with respect to the order of A, B and is therefore not a true metric. While the measure could be made symmetric by combining both the closest point distances from A to B and B to A, the asymmetry is left to enable measurements such as tract A being contained within a larger tract B. For this situation, $PC_\alpha(A, B)$ would be small, but $PC_\alpha(B, A)$ would be large.

2.3 Functional Difference

The previous two measures all focus on establishing geometric distances between two fiber bundles. The final proposed measure instead describes differences in the diffusion parameters sampled by the fiber bundles. This provides a more explicit measure of differences for studies focused on the statistical analysis of scalar invariants. Using the methodology described by Corouge et al., an arc length function for FA and MD is compared between fiber bundles.[9] To summarize the functional difference, FD, the mean difference between the function for bundle A, $f_A(t)$, and the function for bundle B, $f_B(t)$, is computed by

$$FD = \frac{1}{t_n - t_m} \int_{t_m}^{t_n} |f_A(t) - f_B(t)| dt. \tag{2}$$

Eq. 2 is computed for both FA and MD.

For this study, all tracts were mapped into a template atlas space to compare the functional differences. A single origin was used for each bundle to compare the tract generated in the atlas to the native space tract mapped to the atlas. To ensure that functions for both bundles have the same domain, the values for t_m and t_n are restricted to the interval that contains at least an adequate percent of the total streamlines. For this study the interval was restricted to contain at least 30% of the streamlines for both fiber bundles. An example of $f_A(t) - f_B(t)$ for a population is shown in Fig. 4. The summary measure, Eq. 2 is the average absolute difference over the domain.

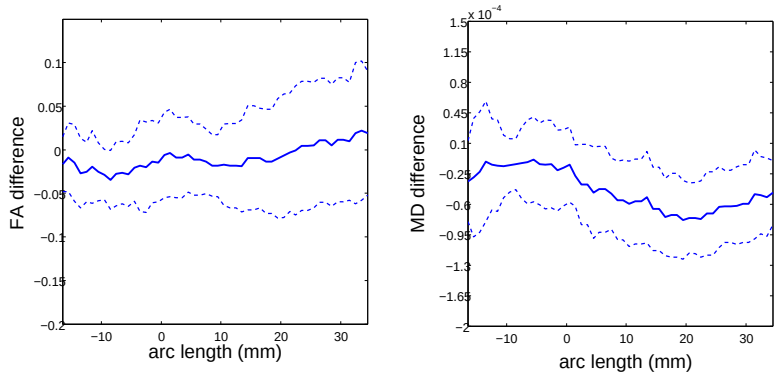


Figure 4. Mean and standard deviation of differences between functions produced by atlas tractography and individual tractography mapped to atlas for the right uncinate.

3. RESULTS

The measures from the previous section were used to evaluate tractography based on an atlas built from images of 37 subjects in a study of schizophrenia in adults. Each subject was imaged using a protocol with 8 non-diffusion weighted images and 51 diffusion weighted gradients at a voxel resolution of $1.6667x1.6667x1.7mm^3$. A b-value of 900 was used for the diffusion weighted images. The purpose of this application was to use the proposed measures to evaluate differences between tractography produced by mapping from an atlas to tractography generated in an individual. As is true for most DTI studies, there is no ground truth for the true geometry of fiber bundles.

Instead this evaluation bounds differences of atlas mapped tractography to native space analysis. An atlas was computed using the method described by Goodlett et al.[3] Fiber bundles were extracted in the mean atlas image using a Runge-Kutta streamline tractography algorithm. Manual clustering and cutting of the tract was performed to obtain an anatomically appropriate set of streamlines for several tracts. The five extracted tracts are the genu, left and right cingulum, and left and right uncinate. They are shown in Fig. 5. Tractography was then computed in each individual using seeding regions mapped from the atlas. After testing a variety of FA thresholds for the individual tractography, a global threshold of $FA = 0.15$ was chosen to give a reasonable approximation of tract geometry.

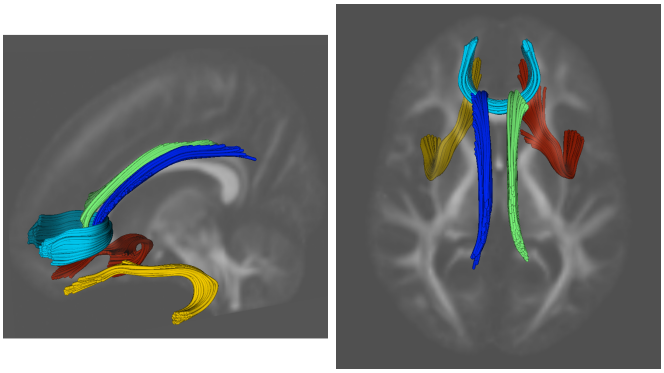


Figure 5. Fiber bundles in population atlas. Left and right cingulum in dark blue and green. Left and right unincate in yellow and red. Genu in light blue.

To be as close as a DTI analysis based on individual subject tractography, an extra step was added to the processing of the individual fiber bundles. Each tractography, for the 37 subjects and 5 bundles, was manually

cleaned based on anatomical criteria. The cleaning, based on the fiber geometry, included removing outlayer streamlines and cropping of streamlines. The goal was to obtain a set of streamlines with a similar geometry as the corresponding bundle tracked in the atlas space. An example of the cleaning steps is shown in figure 6.

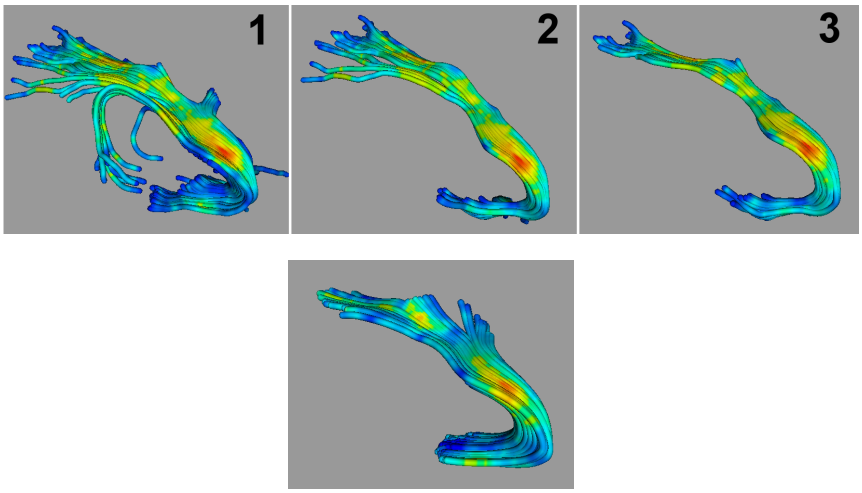


Figure 6. Example of cleaning individual set of streamlines (top row) to match atlas bundle (bottom row). The individual set of fibers is cleaned in two steps: first the obvious outlayers with wrong geometry are removed ($1 \rightarrow 2$), second the bundle is refined to have a tighter set of streamlines and a better match with the atlas fibers ($2 \rightarrow 3$).

A summary of the mean and standard deviation of proposed geometrical measures over the population is presented in Tbl. 1. The binary overlap metrics are significantly lower than those typically encountered in segmentation studies. However, as mentioned in Sec. 2.1 this is likely due to the thin shape of fiber bundles. The CP measure indicates that for most bundles 90% of points in the atlas tract are within slightly about one voxel of the native space tract. This lends evidence that the atlas mapped tracts are in reasonable agreement with tracts produced by individual tractography. Functional measures showed the atlas tract to be about 20% lower for FA and 6% higher for MD. (in Tbl. 2) This difference may be due to partial voluming effects combined with the use of taking the mean value at corresponding arc length values. The atlas tract often appears slightly larger than the individual tract and using the mean FA at each arc length point biases the atlas tract lower. As no gold standard exists further evaluation is needed to determine a preference for the atlas or individual tract.

Tract	BTO	$CP_{.5}$	$CP_{.9}$
genu	0.52 (0.08)	0.57 (0.31)	2.02 (1.22)
uncinate-left	0.39 (0.09)	0.88 (0.52)	3.23 (2.19)
uncinate-right	0.43 (0.08)	0.82 (0.48)	2.78 (1.96)
cingulum-left	0.55 (0.08)	0.43 (0.22)	1.48 (0.87)
cingulum-right	0.56 (0.07)	0.37 (0.11)	1.22 (0.39)

Table 1. Mean and standard deviation of geometric distance measures between warped atlas tract and individual tract over the population.

Tract	FD_{FA}	$FD_{FA\%}$	FD_{MD}	$FD_{MD\%}$
genu	0.05 (0.02)	11.53%	4.87e-05 (1.29e-05)	5.49%
uncinate-left	0.05 (0.02)	15.64%	2.30e-05 (1.50e-05)	2.67%
uncinate-right	0.07 (0.03)	24.50%	4.32e-05 (1.96e-05)	4.42%
cingulum-left	0.03 (0.01)	6.26%	3.53e-05 (1.63e-05)	4.49%
cingulum-right	0.09 (0.03)	24.43%	3.37e-05 (1.36e-05)	4.05%

Table 2. Mean and standard deviation of average absolute difference in FA and MD between atlas tract and warped individual tract. Percent differences are expressed as the ratio of the difference to the value from individual tractography.

4. CONCLUSIONS

We have presented a set of metrics that can be used to evaluate the similarity of tractography results. Our application of these metrics is the comparison of atlas based tractography to tractography generated in the individual space. Volumetric overlap proved to be hard to evaluate given that many tracts are narrow in at least one dimension resulting in relatively low overlap measures with even subvoxel differences in registration. The point cloud divergence served to be particularly useful because of the physical units involved as well as resistant to outliers in unstable streamline tractography. Functional diffusion differences is a useful tool for evaluating variability of statistics but does not provide a geometric evaluation. Together, these methods can be used to improve quality control and validation of DTI analysis. Other uses of these metrics could include the comparison of different tractography routines, the evaluation of reproducibility on repeated scans of the same anatomy, and generation of variance measures to be used for power analysis for future clinical studies. Future work using an expert segmentation of individual tracts could further evaluate atlas based tractography.

5. ACKNOWLEDGMENTS

The authors would like to acknowledge the following grants that founded the research of this paper: the National Institutes of Health under Grant RO1 HD055741 (Autism Center of Excellence, project IBIS), and the National Alliance for Medical Image Computing (NA-MIC), funded by the NIH through Grant U54 EB005149.

REFERENCES

[1] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, "Deformable registration of diffusion tensor MR images with explicit orientation optimization," *Medical Image Analysis* **10**(5), pp. 764–785, 2006.

[2] W. Van Hecke, A. Leemans, E. D'Agostino, S. De Backer, E. Vandervliet, P. M. Parizel, and J. Sijbers, "Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information," *Medical Imaging, IEEE Transactions on* **26**, pp. 1598–1612, Nov. 2007.

[3] C. B. Goodlett, P. T. Fletcher, J. H. Gilmore, and G. Gerig, "Group analysis of DTI fiber tract statistics with application to neurodevelopment," *NeuroImage* **45**(1, Supplement 1), pp. S133 – S142, 2009. Mathematics in Brain Imaging.

[4] U. Ziyan, M. R. Sabuncu, L. J. O'Donnell, and C.-F. Westin, "Nonlinear Registration of Diffusion MR Images Based on Fiber Bundles," in *Medical Image computing and Computer Assisted Intervention (MICCAI)*, *LNCS* **4791**, pp. 351–358, Springer-Verlag, 2007.

[5] G. Gerig, M. Jomier, and M. Chakos, "Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation," in *Medical Image computing and Computer Assisted Intervention (MICCAI)*, *LNCS* **2208**, pp. 516–523, Springer-Verlag, 2001.

[6] P. T. Fletcher, R. Tao, W.-K. Joeng, and R. Whitaker, "A Volumetric Approach to Quantifying Region-to-Region White Matter Connectivity in Diffusion Tensor MRI," in *Information Processing in Medical Imaging*, *LNCS* **4584**, pp. 346–358, 2007.

[7] J. Melonakos, V. Mohan, M. Niethammer, K. Smith, M. Kubicki, and A. Tannenbaum, "Finsler Tractography for White Matter Connectivity of the Cingulum Bundle," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, *LNCS* **4791**, pp. 36–43, Springer-Verlag, 2007.

[8] I. Corouge, S. Gouttard, and G. Gerig, "Towards a shape model of white matter fiber bundles using diffusion tensor mri," in *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pp. 344–347, (Washington DC), April 2004.

[9] I. Corouge, P. T. Fletcher, S. Joshi, S. Gouttard, and G. Gerig, "Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis," *Medical Image Analysis* **10**(5), pp. 786–798, 2006.