Combination of distance measures for optimal fiber clustering in diffusion tensor imaging

D. Güllmar[1,2], J. Langner[1], J. Haueisen[3], and J. R. Reichenbach[1]

[1]Medical Physics Group, University Hospital Jena, Jena, Thuringia, Germany, [2]Biomagnetic Center, University Hospital Jena, Jena, Thuringia, Germany, [3]Institute of Biomedical Engineering and Informatics, TU Ilmenau, Illmenau, Thuringia, Germany

Introduction:

Fiber tractography has become a serious tool in diffusion tensor post-processing. However, the use of the resulting tractograms in clinical routine is rare. This is certainly due to the fact that the tools for fiber tractography are computationally intensive or hard to handle or both. But even when this post processing step is performed quick and easy, the question arises: what to do with the tracts. Neurosurgeons probably want to know if a major fiber tract is damaged by a tumor, physiologist or neurologist want to know if and in to what degree certain areas are connected along the known major tracts. Usually these tracts are selected manually by an experienced observer placing ROI to define which tracts are of interest. However, the results of such a processing are individual, observer dependent and often not reproducible. Clustering of the fiber tracts can overcome this deficit by grouping similar tracts through comparison of similarity or distance measures. Several distance and similarity measures were publish, which can be used for clustering, but no optimal measure was found yet. The aim of the presented study is to combine different distance measures to handle the deficiencies of single distance or similarity measures.

Materials and Methods

The fiber tracts were derived from a diffusion tensor data set of a healthy volunteer. Diffusion sensitized encoding was performed in 30 direction (5 b0-volumes) with a b-value of 1000 s/mm² over 55 slices of 128×128 pixel resulting in a $2.5 \times 2.5 \times 2.5$ mm³ voxel spacing volume. Fiber tracts were generated using the Diffusion Toolkit [1] for the whole brain. Three different distance measures (Hausdorff distance, corresponding segment ratio[2], Euclidean distance) were calculated for all tracts which exceed a length of 70 mm. These matrices can be used individually to cluster the tracts. In our study we combine these matrices ($Y_{1..3}$) by a simple formalism $\ln(Y') = k_1 {*} \ln(Y_1) + k_2 {*} \ln(Y_2) + k_3 {*} \ln(Y_3)$. Additionally to theses measures an *a priori* measure was generated, which weights the distance of two tracts low if their end points are in the same hemisphere, medium if the endpoints of one tract are in one hemisphere and the endpoints of second tract are in both hemispheres and the distance was set to high if one tract has its endpoints in left and the other in the right hemisphere. The clustering of the single distance measures as well as of the combined distance measure was performed by creating a hierarchical cluster tree based on the unweighted average distance since this method has the largest Cophenetic correlation coefficient (CCC) - a measure which describes how faithful the tree represents the dissimilarities among the different tracts or observations and ranges from 0 to 1. For the combined distance measure all k-factors were set to 1. The clustering was limited to 20 clusters. The 10 largest cluster were selected for visualization using Trackvis[1].

Results

In Fig. 1 the ten largest clusters derived from the separate distance measures are shown. Thereby the color-coding order (blue, green, red, cyan, purple, ocher, dark grey, pink, white, yellow) was subjected to the number of tracts in each cluster. Hence, the blue tracts belong to the largest cluster and so on. The figure clearly shows that the corresponding segment ratio (second image) as well as the Euclidean distance (third image) is not suitable as single distance measure in order to get a proper clustering of the major nerve bundles. However, the results derived by the Hausdorff distance (first image, left) and the combined distance measure (fourth image, right) show reasonable results. But it is also shown, that the Hausdorff distance has problems in separating the pyramidal tracts from the corpus callosum (blue segment in the left image). On the other hand the clustering based on the combined measure shows a correct separation.

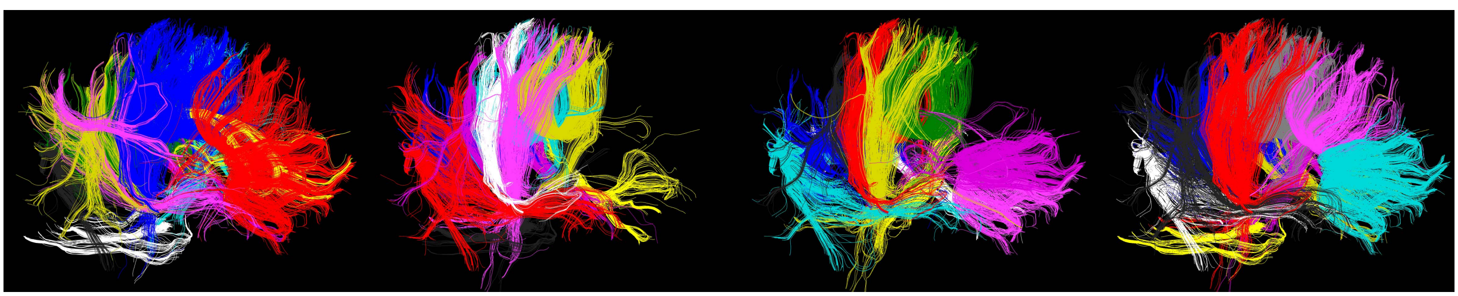

Fig. 1 Clustering of fiber tracts based on different distance measures (from left to right: Hausdorff distance, corresponding segment ratio, Euclidean distance, combined distance measure).

Discussion

In this study we have shown that the combination of different distance measures between tracts derived from deterministic tracking of Diffusion tensor data can overcome deficits of simple distance measures. Thereby the combination was not optimized yet. We anticipate that further investigation on this optimization will answer which measure contributes most in order to get a proper separation of the fiber bundles and which are inadequate for clustering studies.

References

[1] Ruopeng Wang and Van J. Wedeen, Diffusion Toolkit, http://www.trackvis.org, 2007
[2] Ding Z et al., Magn Reson Med. 2003 Apr ; 49(4): 716-21