Clustering Method for Estimating Principal Diffusion Directions

M. Nazem-Zadeh[1], K. Jafari-Khouzani[2], A. Babajani-Fermi[2], S. P. Nejad-Davarani[1], H. Soltanian-Zadeh[2,3], and Q. Jiang[1]
[1]Neurology, Henry Ford Hospital, Detroit, MI, United States, [2]Diagnostic Radiology, Henry Ford Hospital, Detroit, MI, United States, [3]Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Introduction. High Angular Resolution Diffusion Imaging (HARDI) is a non-invasive tool for investigating white matter structure in the brain. Using HARDI data, the fiber Orientation Distribution Function (ODF) is calculated, from which the principal diffusion directions (PDDs) are extracted. Fast and accurate estimation of PDDs is a pre-requisite for fiber tracking algorithms that deal with fiber crossings.
For estimating PDDs, the initial candidates are refined by searching for local maxima using Powell's method [1], spherical Newton's method [2] or sequential quadratic programming [3], then, the tiny ODF peaks are discarded. In addition to convergence, the problem of these methods is getting trapped in small local maxima. In [4], the normalized ODF is projected on a tessellated sphere with a fine mesh, then, the finite difference method is applied. Thresholding in order to diminish minor peaks and sensitivity to the mesh grid size are the problems of this method. In [5,6] the high order Cartesian tensor equivalent to the ODF representation is calculated, then, the stationary points are classified based on their curvature. Exploiting higher order tensors, although increases the angular resolution and the accuracy of estimated PDDs, it increases the model complexity and instability against noises. In [7] the ODF surface is smoothed using a Gaussian kernel followed by taking derivative. Smoothing the ODF degrades its angular resolution of its PDDs. Using finite difference method, *Camino toolbox* [8] locates local maxima by ascertaining all points at which the function is larger than all other points within a fixed search radius. It then removes duplicates and tiny peaks. This procedure is both inaccurate and time-consuming. Here we propose an accurate and fast clustering algorithm to estimate PDDs.

Methods. The main maxima of the ODF can be interpreted as the PDDs. However, in our problem definition, cluster centers are not points; instead they are directions around which ODF points are concentrated. We define the distance between each ODF data point P_i and cluster direction S_j as $d(P_i,S_j)=\sqrt{\|P_i\|^2-|P_i.S_j|^2/\|S_j\|^2}$. There are some exact orientations $(S_{j*},S_{k*},\dots)$ for which their overall distances to the data points is minimum. In two-cluster case $S_{j*},S_{k*}=\arg\min_{j,k}\sum_{i=1}^{N}d^2(P_i,S_j,S_k)=\arg\min_{j,k}\sum_{i=1}^{N}\left[\|P_i\|^2-\mathrm{Max}\left\{|P_i.S_j|^2/\|S_j\|^2,|P_i.S_k|^2/\|S_k\|^2\right\}\right]$. One way for finding the cluster directions is to consider each possible pair of data points as candidates and to examine if they better represent the ODF data in the sense of minimizing the overall distance. We name this method 'Search-All' which an exhaustive search. Instead, we propose a clustering approach 'Sph-FCM ' to estimate PDDs which is an extension of fuzzy c-means clustering for applying on orientation coordinates of the on-sphere points. We cluster the ODF data points on the unit sphere into distinct clusters. For each ODF point on the sphere, we code the ODF value as extra points added to ODF profile. Using Spherical law of cosines, the distance between each ODF data point x_i and cluster direction μ_j is calculated as: $d(x_i,\mu_j)=\mathrm{acos}\left\{\cos\theta_{x_i}.\cos\theta_{\mu_j}+\sin\theta_{x_i}.\sin\theta_{\mu_j}.\cos(\varphi_{x_i}-\varphi_{\mu_j})\right\}$, where θ and φ are the corresponding co-latitudes and co-longitudes spherical angles. The cost function to be minimized is: $L=\sum_{i=1}^{N}\sum_{j=1}^{N_c}U^b(\mu_j|x_i).d(x_i,\mu_j)$. Taking derivatives with respect to the cluster centers and memberships and setting them to zero leads to:

$$U(\mu_k|x_i)=\frac{\left\{\frac{1}{d(x_i,\mu_k)}\right\}^{b-1}}{\sum_{j=1}^{N_c}\left\{\frac{1}{d(x_i,\mu_j)}\right\}^{b-1}}\qquad \varphi_{\mu_k}=\tan^{-1}\frac{\sum_{i=1}^{N}\frac{U^b(\mu_k|x_i).\sin\theta_{x_i}.\sin\varphi_{x_i}}{\sin d(x_i,\mu_k)}}{\sum_{i=1}^{N}\frac{U^b(\mu_k|x_i).\sin\theta_{x_i}.\cos\varphi_{x_i}}{\sin d(x_i,\mu_k)}}\qquad \theta_{\mu_k}=\tan^{-1}\frac{\sum_{i=1}^{N}\frac{U^b(\mu_k|x_i).\sin\theta_{x_i}.\cos(\varphi_{x_i}-\varphi_{\mu_k})}{\sin d(x_i,\mu_k)}}{\sum_{i=1}^{N}\frac{U^b(\mu_k|x_i).\cos\theta_{x_i}}{\sin d(x_i,\mu_k)}}$$

We estimate the optimal number of clusters from the data distribution based on the Minimum Description Length (MDL) criterion. By increasing the number of clusters, if MDL decrease, we accept the new number of clusters and continue the algorithm with updating the memberships and the cluster centers. We modified the FACT algorithm [9] for dealing with more than one PDD and used to reconstruct the fiber tracts.

Results and Discussion.
Artificial Data. Artificial diffusion signals with three crossings with 45° isolation 55 gradient directions, were generated over a wide range of SNRs. The PDDs were then extracted by the proposed algorithm *Sph-FCM*, the Camino toolbox method and *Search-All* method as the gold standard. The similarity measure between the original PDDs and extracted PDDs is calculated. As shown in Fig. 1, *Sph-FCM* is considerably more accurate than the *Camino toolbox* and shows close performance to the *Search-All* method, especially at high SNRs.
Real Data. The HARDI data of a normal subject was acquired by a GE Signa Excite 3T MRI at Henry Ford Hospital (Detroit, MI) with 55 diffusion gradient directions and imaging voxel dimension of 0.975×0.975×2.6 mm^3. Fig. 2 and Fig. 3 show the color-coded FA at the intersection of corpus-callosum, corona radiata and longitudinal fasciculus tracts overlaid by PDDs in the crossing area extracted by *Camino toolbox and Sph-FCM,* respectively. For this dataset, it takes almost a week for *Camino toolbox* to estimate up to three PDDs for each voxel, while using *Sph-FCM,* this time is decreased to 3 hours using the same dataset and the same hardware. Fig. 4 shows the tractography results for these fiber tracts based on PDDs extracted by *Sph-FCM*. Note the accurate fiber tracts extracted at the mentioned intersection.

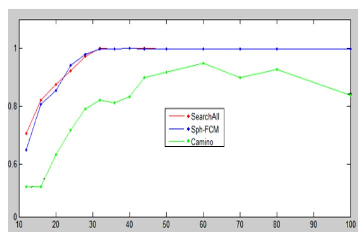

Fig. 1. Similarity Measures vs SNR. Showing close performance to the *Search-All*, Sph-FCM is more accurate than the *Camino*

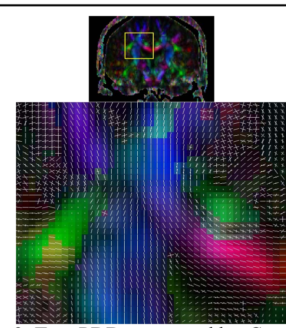

Fig. 2. Two PDDs extracted by *Camino*

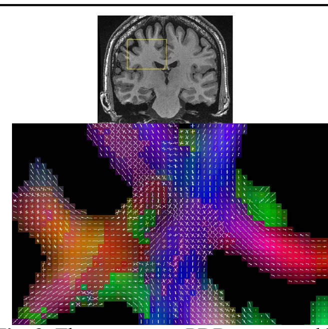

Fig. 3. Three accurate PDDs extracted by proposed *Sph-FCM*

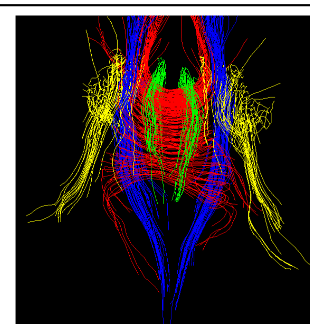

Fig. 4. Tractography results based on PDDs extracted by *Sph-FCM*

References. [1] K. Jansons, Inf Process Med Imaging 18:672-683, 2003. [2] J. Tournier, Neuroimage 23:1176-1185, 2004. [3] K. Sakaie, Neuroimage 34:169-176, 2007. [4] M. Descoteaux, Nice, France: University of Nice-Sophia Antipolis, 2008. [5] L. Bloy, Med Image Comput Comput Assist Interv 11:1-8, 2008. [6] A.Ghosh, Workshop on Computational Diffusion MRI, 2008. [7] S. Frey, Neurosci 28:11435-11444, 2008. [8] P. Cook , 14th Scientific Meeting of the ISMRM 2759, 2006. [9] S. Mori, Annals of neurology 45, 265-269, 1999.