Automatic Classification of High Angular Resolution Diffusion Data

J. D. Carew[1,2], G. Wahba[1,2], C. G. Koay[3], Y-C. Wu[4], A. L. Alexander[4], M. E. Meyerand[4]

[1]Statistics, University of Wisconsin, Madison, WI, United States, [2]Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, United States, [3]Department of Physics, University of Wisconsin, Madison, WI, United States, [4]Medical Physics, University of Wisconsin, Madison, WI, United States

INTRODUCTION

High angular resolution diffusion-weighted imaging (HARD) data can provide information about diffusion in voxels that intersect one or more white matter fibers. The diffusion measurements can be interpreted as noisy samples from the fiber orientation distribution function (ODF), which is defined on the surface of a sphere. Since the ODF can be complicated, it is difficult to visualize on a large scale. For many voxels a single tensor model may be adequate, but it is difficult to automatically identify these voxels. It is also difficult to characterize how the HARD measurements differ between populations of subjects (e.g., patients vs. controls). To address these problems we propose a method to automatically classify HARD data based on the shape of the ODF.

METHODS

THEORY The K-means clustering algorithm [1] is used to automatically classify the HARD data. Specifically, we wish to classify a set of diffusion measurements $D_i = \{\mathbf{d}_{i1}, \ldots, \mathbf{d}_{iN}\}$ along *N* directions from voxel *i*, as belonging to one of *K* classes that characterize the shape of the diffusion. The apparent diffusion along direction *j* is given by $\mathbf{d}_{ij} = (-1/b)\log(S_j/S_0)$, where S_j and S_0 are from images with and without diffusion weighting, respectively. The K-means algorithm requires specification of an appropriate distance measure. A distance measure $m(D_i, D_k)$ that is sensitive to shape must be both rotationally and scale invariant. Scale invariance can be satisfied by dividing each of the $\mathbf{d}_{ij}$ by $\sum_{j=1}^{N} \mathbf{d}_{ij}$. We define a rotationally invariant distance measure to be $m(D_i, D_k) = \sqrt{\min_M \sum_{j=1}^{N} ||\mathbf{d}_{ij} - M\mathbf{d}_{kj}||^2}$, where *M* is a rotation matrix. A solution to finding the *M* that minimizes the sum in the distance measure is given in [2].

EXPERIMENT A series of 252 diffusion weighted images was acquired on a 3T scanner in a single coronal slice from a normal volunteer that was positioned to bisect the corpus callosum. The scan parameters were: TR/TE: 1000/80 ms, FOV: 20 cm, 5 mm slice thickness, 64 by 64 pixel matrix, and diffusion weighting 1492.34 s/mm^2. An icosahedral gradient encoding scheme was used [3]. To reduce noise in the diffusion measurements, a locally weighted smoothing filter was applied. This filter weighted the diffusion in the center direction by 2/3 and the diffusion in the six adjacent directions by 1/18. An 8-by-16 pixel ROI (Figure A) was selected for classification. Four classifications were specified *a priori* to have freedom to accommodate the following types of diffusion: isotropic, anisotropic, two classes of fiber crossings. Since the K-means algorithm is sensitive to the choice of the initial class means, the clustering was repeated 100 times with random initial selection of the cluster means. Of the 100 classifications, the one that minimized the mean within-class variance was selected.

RESULTS

The K-means algorithm converged quickly for the diffusion data. Of the 100 runs the maximum number of iterations before convergence was five. The total computer time was 2 minutes on a 2.5 GHz G5. Figure B shows the classification that minimized the within-cluster variance. The means of the clusters are shown in Figure C. The voxels that are classified according to one of the shapes in C are closer to some rotation and scaling of the mean diffusion of that class than scaling and rotations of the other three classes.

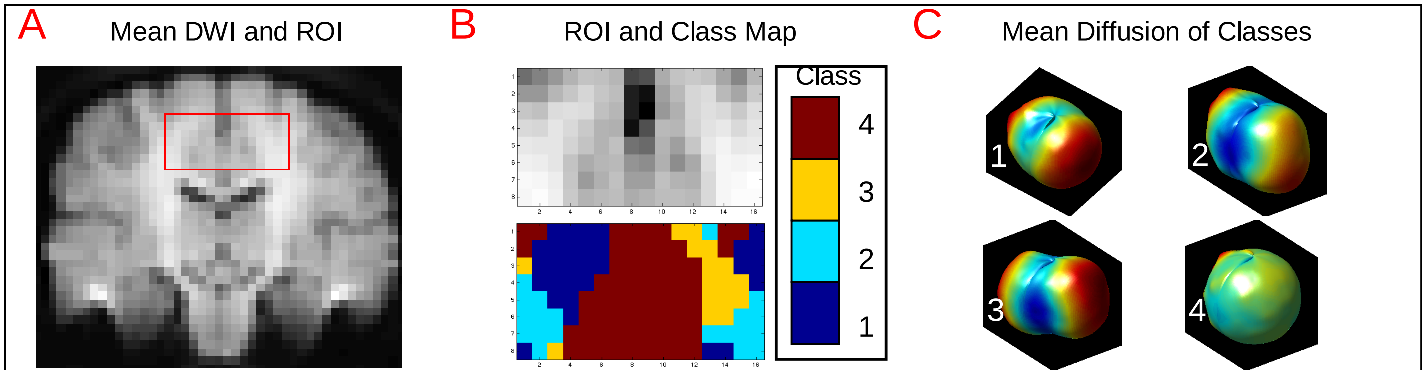

Figure: (A) Mean of the 252 DWIs and the ROI selected for classification. (B) (Top) zoom of ROI; (Bottom) Classification map; voxels with the same color belong to the same class of diffusion. (C) Mean diffusion shapes for each of the four classes. Class 1 represents anisotropic diffusion. Classes 2 and 3 likely represent crossing fibers at angles of roughly 60^0 and 45^0, respectively. Class 4 likely represents isotropic diffusion. Note that the voxels mapped to class 4 are located in gray matter.

DISCUSSION

This study demonstrates that it is possible to classify HARD diffusion data based on the shape of the diffusion. The ROI was purposely selected to contain voxels with a variety of diffusion types, including those with fiber crossings. The K-means algorithm has correctly identified voxels (classes 2 and 3) where there are likely to be fiber crossings. The shape of diffusion in class 1 is anisotropic. All voxels in this class are located in white matter. Gray matter voxels were correctly identified as having isotropic diffusion (class 4). One limitation of the K-means algorithm is the need to specify *a priori* the number of classes. One method for selecting the number of classes from the data is the gap statistic [4]. The shape of the ODF will be influenced by the choice of the diffusion weighting. Future work involves a study of the dependence of diffusion weighting on the resulting classification and how well the diffusion can be classified. Preliminary results from data acquired with a higher diffusion weighting of 2998.84 s/mm^2 showed that the proposed classification did not work as well as with the data reported here. Applications of this method include visualization, identifying voxels where a single tensor model may be inadequate, and testing for group differences in the shape distribution between patients and controls.

References

[1] Hastie T, Tibshirani R, Friedman J. (2001) *The Elements of Statistical Learning*. Springer, New York.

[2] Farrell JL, Stuelpnagel JC. (1966) *SIAM Review* 8(3):384-386.

[3] Hasan KM, Parker DL, Alexander AL. (2001) *J Magn Reson Imaging* 13:769-780.

[4] Tibshirani R, Walther G, Hastie T. (2001) *J R Statist Soc B* 63(2):411-423.