

# Segmentation of Thalamic Nuclei from DTI Using Spectral Clustering

Ulas Ziyani<sup>1</sup>, David Tuch<sup>2</sup>, and Carl-Fredrik Westin<sup>1,3</sup>

<sup>1</sup> MIT Computer Science and Artificial Intelligence Lab, Cambridge MA, USA  
ulas@mit.edu

<sup>2</sup> Novartis Pharma AG, Basel, Switzerland  
david.tuch@novartis.com

<sup>3</sup> Laboratory of Mathematics in Imaging, Brigham and Women's Hospital,  
Harvard Medical School, Boston MA, USA  
westin@bwh.harvard.edu

**Abstract.** Recent work shows that diffusion tensor imaging (DTI) can help resolving thalamic nuclei based on the characteristic fiber orientation of the corticothalamic/thalamocortical striations within each nucleus. In this paper we describe a novel segmentation method based on spectral clustering. We use Markovian relaxation to handle spatial information in a natural way, and we explicitly minimize the normalized cut criteria of the spectral clustering for a better optimization. Using this modified spectral clustering algorithm, we can resolve the organization of the thalamic nuclei into groups and subgroups solely based on the voxel affinity matrix, avoiding the need for explicitly defined cluster centers. The identification of nuclear subdivisions can facilitate localization of functional activation and pathology to individual nuclear subgroups.

## 1 Introduction

All of the sensory pathways of the human brain (with the exception of the olfactory pathway) project to the cortex via relay neurons in the thalamus. These relay neurons are clustered into discrete clusters or nuclei that can be delineated based on histological or functional criteria. Delineation of the nuclei is important for surgical treatment of the thalamus and for localization of functional brain activation to specific nuclei.

Recently, it has been shown that the thalamic nuclei can be resolved using a type of magnetic resonance imaging (MRI) called diffusion tensor imaging (DTI) [1, 2]. DTI measures the molecular diffusion, i.e., Brownian motion, of the endogenous water in tissue. In fibrous biological tissues such as cerebral white matter, water diffusion is anisotropic. In nerve tissue, the diffusion anisotropy is thought to be due to the diffusion barrier presented by the cell membrane, although the physical mechanisms for diffusion anisotropy in nerve tissue are not entirely understood [3].

On DTI, the thalamus shows distinct clusters of diffusion orientation. These clusters correspond in anatomic location and fiber orientation to the thalamic

nuclei [2]. Diffusion orientation thus provides a new anatomic criterion for distinguishing thalamic nuclei.

Although DTI can resolve thalamic nuclei, a segmentation algorithm is required to explicitly delineate the nuclei from the DTI data. Several methods have been proposed to resolve the thalamic nuclei, including the use of k-means [2], level-sets [4] and tract reconstructions from manually defined regions on the cortical surface [1]. A weakness of the k-means algorithm is its geometric bias towards ellipsoidal clusters. Further, both k-means and level sets show sensitivity to initialization, and susceptibility to local minima.

The present report describes a new approach for segmentation of thalamic nuclei using the spectral segmentation algorithm described by Shi and Malik [5] with some modifications. The spectral segmentation algorithm has attracted considerable interest in the pattern recognition community due its computational simplicity, strong empirical performance, and rich underlying theory [6, 7, 8]. The algorithm is based on a classical result from spectral graph partitioning which relates the second smallest eigenvector of the Laplacian matrix of a graph to optimal partitions. The spectral clustering algorithm has a number of desirable features. For example, the algorithm consists entirely of direct matrix operations and is therefore computationally efficient. Furthermore, the algorithm does not require a geometric representation of the clusters and therefore contains no explicit geometric biases.

DTI data from 10 healthy participants were collected and segmented using the modified spectral clustering to demonstrate feasibility of the proposed method.

## 2 Theory

Spectral clustering reduces segmentation into a graph partitioning problem. The nodes of the graph are chosen to be the data points, and the links connecting the nodes, commonly referred as edges, are assigned weights based on similarities of the data points being linked together. In the case of DTI data, the nodes can be chosen as the diffusion tensors, and the edge weights can be related to the similarities of the neighboring tensors. In Section 2.1, we will describe the details of the graph construction for DTI data, and in Section 2.2, we will describe a procedure to solve the graph partitioning problem via explicit minimization of an objective function defined on the graph. Once the graph partitioning is completed, the algorithm produces a hierarchical tree, from which different levels of segmentations can be extracted.

### 2.1 Graph Construction

Our approach to nuclei segmentation is to partition the DTI data into compact regions with homogeneous diffusion properties, since similar diffusion properties are likely to be indicative of belonging to the same nuclei. For that purpose we are looking to construct a graph in which the nodes within a homogeneous region are well connected and the nodes in different homogeneous regions are not. This

graph is constructed based on the spatial distances between the voxels as well as the tensor dissimilarities.

The first step in the construction is to choose an appropriate metric to quantify the tensor dissimilarities. There are a number of possible metrics for this purpose, such as angular difference between the principle eigenvector directions,

$$f_1(\mathbf{T}_i, \mathbf{T}_j) = \arccos(|\mathbf{v}_i \cdot \mathbf{v}_j|) \quad , \quad (1)$$

where  $\mathbf{v}_i$  is the principal eigenvector of tensor  $\mathbf{T}_i$ . The absolute value of the dot product solves the problem with the sign ambiguity of the eigenvectors (if  $\mathbf{v}_1$  is an eigenvector, so is  $-\mathbf{v}_1$ ). Another option is to use the full tensor information,

$$f_2(\mathbf{T}_i, \mathbf{T}_j) = \sqrt{\text{trace}((\mathbf{T}_i - \mathbf{T}_j)^2)} \quad . \quad (2)$$

$f_2(\mathbf{T}_i, \mathbf{T}_j)$  is explored in several DTI studies under names such as generalized tensor dot product [9], Frobenius norm [2] and Euclidean distance metric [10].

We also experimented with a recently proposed information theoretic measure, symmetric K-L divergence [11],

$$f_3(\mathbf{T}_i, \mathbf{T}_j) = \sqrt{\text{trace}(\mathbf{T}_i^{-1}\mathbf{T}_j + \mathbf{T}_j^{-1}\mathbf{T}_i) - 2n} \quad , \quad (3)$$

where  $n = 3$  for diffusion tensors. After choosing a tensor dissimilarity metric, we need to incorporate spatial relations of the voxels into the graph. One way to do so is to linearly combine the tensor dissimilarity metric,  $f(\mathbf{T}_i, \mathbf{T}_j)$ , with a spatial distance,  $s(\mathbf{x}_i, \mathbf{x}_j)$ , as in [2]:

$$d((\mathbf{x}_i, \mathbf{T}_i), (\mathbf{x}_j, \mathbf{T}_j)) = f(\mathbf{T}_i, \mathbf{T}_j) + \gamma s(\mathbf{x}_i, \mathbf{x}_j) \quad , \quad (4)$$

where  $\gamma$  is a weighting factor to control the trade-off between the tensor dissimilarity and the spatial distances. However, selecting an appropriate weighting factor is complicated by the fact that the spatial distances and the tensor dissimilarities are inherently different types of features. To avoid this problem, we use Markovian relaxation, which provides a natural way to incorporate the spatial information. This is done by constructing a sparse affinity matrix,  $\mathbf{W}_s$ , with non-zero similarities only between face neighbors:

$$\mathbf{W}_s(i, j) = \begin{cases} \exp\left(\frac{-f(\mathbf{T}_i, \mathbf{T}_j)^2}{\sigma^2}\right), & \text{if } s(\mathbf{x}_i, \mathbf{x}_j) \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The parameter  $\sigma$  is chosen to be the sample variance of  $f(\mathbf{T}_i, \mathbf{T}_j)$  for neighboring voxels. We then calculate the full affinity matrix,  $\mathbf{W}$ , using Markovian relaxation [12]. Specifically, we first convert the sparse weight matrix into a one-step transition probability matrix whose rows and columns add up to one:

$$\mathbf{P}_1(i, j) = \frac{1}{\max_l \mathbf{d}(l)} \times \begin{cases} \mathbf{W}_s(i, j), & \text{if } i \neq j \\ \max_l \mathbf{d}(l) - \mathbf{d}(i), & \text{if } i = j \quad , \end{cases} \quad (6)$$

where  $\mathbf{d}$  is a vector containing row sums of  $\mathbf{W}_s$ ,  $\mathbf{d}(i) = \sum_j \mathbf{W}_s(i, j)$ . The diagonal adjustment ensures that the underlying random walk has a uniform steady state distribution, and therefore every part of the corresponding Markov field is explored at the same speed during the relaxation. Once the one-step transition probability matrix is computed, it is possible to calculate the  $n$ -step transition probabilities through Markovian relaxation, which is equivalent to raising the one-step transition probability matrix to the power  $n$ :  $\mathbf{P}_n = \mathbf{P}_1^n$  [12].  $n$  is chosen to be smallest integer that will produce a non-zero transition probability between every node. Finally the diagonal elements of  $\mathbf{P}_n$  are set to zero to produce the full affinity matrix  $\mathbf{W}$ .

## 2.2 Graph Cuts

Once the graph is constructed, one can calculate the  $k$ -way graph partition using Normalized Cuts criteria. Formally, the problem is now reduced to the following:

Given a graph,  $\mathbf{G} = (\mathbf{V}, \mathbf{E})$  and its edge weight matrix  $\mathbf{W} = \{\omega_{ij}\}$ , find a set of disjoint sets  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_k$ , such that  $\bigcup_{i=1}^k \mathbf{V}_i = \mathbf{V}$ , which minimizes the normalized cut (NCut):

$$\text{NCut}(\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_k) = \sum_{i=1}^k \frac{\text{asso}(\mathbf{V}_i, \mathbf{V}) - \text{asso}(\mathbf{V}_i, \mathbf{V}_i)}{\text{asso}(\mathbf{V}_i, \mathbf{V})}, \quad (7)$$

where  $\text{asso}(\mathbf{A}, \mathbf{B}) = \sum_{i \in \mathbf{A}, j \in \mathbf{B}} \omega_{ij}$  [5].

Finding the normalized cuts, even for the case where  $k = 2$ , is NP-complete due to the combinatorial nature of the discrete permutations, therefore, we need a polynomial time algorithm that will provide a reasonable solution. Shi and Malik suggest a number of possibilities for such an algorithm. In this paper, we use a combination of the methods suggested.

Specifically, we first normalize the weight matrix;  $\mathbf{M} = \mathbf{D}^{-1}\mathbf{W}$ , where  $\mathbf{D}$  is a diagonal matrix with  $\mathbf{D}(i, i) = \sum_j \omega_{ij}$ . We then calculate the two largest eigenvectors of the normalized weight matrix,  $\mathbf{M}$ . The largest eigenvector is a vector of constants due to the normalization and therefore discarded. We then consider cutting  $\mathbf{W}$  into two clusters by thresholding the second eigenvector of  $\mathbf{M}$  at every possible threshold level, looking for the 2-way cut with the minimum NCut value. Once the cut is made, we repeat this process for each of the two newly created clusters, until a very high threshold of a 2-way NCut value is reached. This threshold is typically close to 1, which is the level at which every point is clustered by itself regardless of the weights. This splitting is followed by a greedy merging algorithm, minimizing NCut at every merging step, to reduce the number of clusters to  $k$ . During this merging, the algorithm produces a hierarchical tree, from which different levels of segmentation can be extracted by varying  $k$ .

At the final stage of our algorithm we allow single data node swaps between clusters to further reduce the NCut value. At every iteration of this stage, the algorithm considers the possibility of reassigning a single node to another cluster. If the resulting cut has a smaller NCut value than before, that node is reassigned to its new cluster. The algorithm stops if there are nodes that can be reassigned.

### 3 Methods

#### 3.1 Image Acquisition and Pre-processing

DTI data were acquired using a twice-refocused spin-echo EPI sequence [13] on a 3 Tesla Siemens Trio MRI scanner using an 8-channel head coil. The sequence parameters were TR/TE=8400/82 ms,  $b=700$  s/mm<sup>2</sup>,  $g_{\max}=26$  mT/m, 10 T2 images, 60 diffusion gradient directions, 1 average, total acquisition time 9 min 59 seconds. The field-of-view was  $256 \times 256$  mm and the matrix size was  $128 \times 128$  to give  $2 \times 2$  mm in-plane resolution. The slice thickness was 2 mm with 0 mm gap.

Correction for motion and residual eddy current distortion was applied by registering all of the scans to the first acquired non-diffusion-weighted scan for each participant. The registration used a 12 degree-of-freedom global affine transformation and a mutual information cost function [14]. Trilinear interpolation was used for the resampling. The diffusion tensor, the tensor eigensystem, and the FA metric were calculated for each voxel using the formulas of [15] and [16].

The diffusion tensor and FA volumes were normalized to MNI-space (Montreal Neurological Institute) by registering each participant's T2 volume to a skull-stripped version of the MNI 152-subject T2 template [17] and then applying the transformation to the diffusion tensor and FA volumes with 12 degree-of-freedom global affine transformation and a mutual information cost function [14].

The registration transformation was then applied separately to the FA and tensor volumes. The FA volume was resampled using trilinear interpolation and the tensor volume was resampled using nearest neighbor interpolation. The tensors were reoriented using the rotational portion of the atlas transformation.

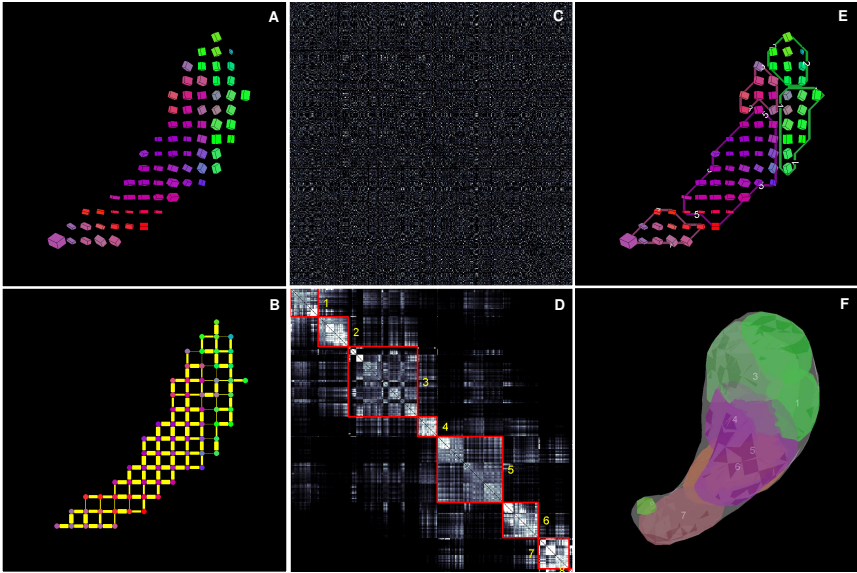
#### 3.2 Segmentation

Thalamus masks were drawn manually for each individual by a trained neuro-anatomist. The masks were drawn for each hemisphere on each individual's MNI-normalized FA map. Each hemisphere further segmented into its seven nuclei on the corresponding tensor map by the neuro-anatomist. The masked thalamus DTI data was later segmented separately for each hemisphere using the spectral algorithm described in Section 2.

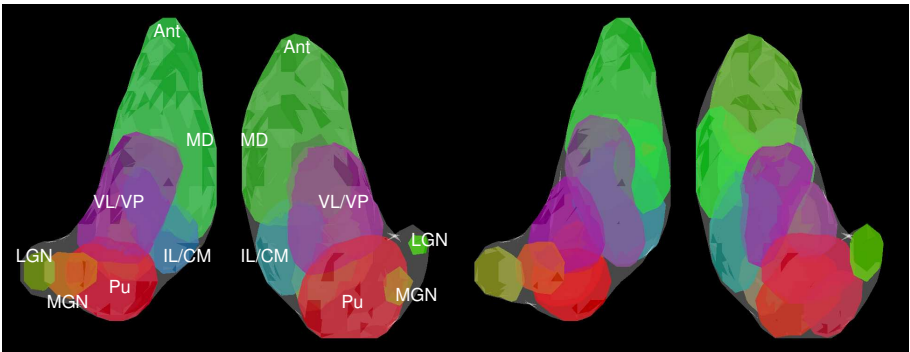
### 4 Results

For all thalamic data sets, spectral reordering of the voxel affinity matrix revealed significant clustering structure (Figure 1D). The network created by the algorithm provides necessary information for the clustering and the spectral ordering followed by the modified spectral clustering algorithm was able to identify the clusters. These clusters are presented in Figure 2 (Right) for one of the individuals along with the expert labels (Left). The automatically segmented clusters are similar in both hemispheres which were processed independently, and they match well with the expert segmentation.

All 10 subjects were segmented using the modified spectral clustering algorithm at two different levels (of the hierarchical tree) with  $k = 7$  and  $k = 12$ . The



**Fig. 1.** A schematic outline of spectral segmentation algorithm. (A) A single slice tensor data (B) Initial graph corresponding to  $W_s$  (C) Unordered  $W$  (D) Ordered and clustered  $W$  (E) Clusters on the slice (F) Clusters in 3D.



**Fig. 2.** Left: 3D rendering of expert segmentation of both hemispheres from one subject. Right: The same subject segmented by spectral clustering with  $k = 12$ .

experiment is repeated with  $W_s$  and  $W$  to quantify the improvement achieved by the Markovian relaxation. The resulting clusters are then named in accordance with the expert labels. At this stage several clusters are allowed to inherit the same labels for a quantitative comparison. These final segmentations are then compared with the expert labels using the Dice volume overlap measure [18]. Mean and standard deviation of the volume overlaps from the 10 subjects are presented in Table 1.

**Table 1.** Dice volume overlap comparison of spectral clustering results with expert labels for three tensor distance metrics  $f_1$ ,  $f_2$ ,  $f_3$ . The experiment is repeated with the sparse affinity matrix,  $\mathbf{W}_s$ , and with the product of Markovian relaxation,  $\mathbf{W}$ , at clustering levels of  $k = 7$  and  $k = 12$ . A score of 100 would indicate a perfect match.

	$\mathbf{W}_s$			$\mathbf{W}$			$\mathbf{W}_s$			$\mathbf{W}$		
	$k = 7$			$k = 7$			$k = 12$			$k = 12$		
Tensor Dist.	$f_1$	$f_2$	$f_3$	$f_1$	$f_2$	$f_3$	$f_1$	$f_2$	$f_3$	$f_1$	$f_2$	$f_3$
Mean	75.3	60.6	62.9	79.6	67.2	65.9	80.5	68.3	68.6	83.4	72.7	74.2
St. Dev.	5.7	4.9	5.0	5.6	4.8	5.6	4.6	5.4	3.9	4.1	4.9	5.1

## 5 Conclusions

In this paper, we presented a modified spectral clustering method for segmenting the thalamic nuclei using DTI data. Our method can automatically identify nuclear groups and subgroups solely based on the voxel affinity without any prior knowledge of cluster centers. Our experiments indicate that the angular distances between principle eigenvector directions of tensors produce more consistent segmentations compared to full tensor based measures.

## Acknowledgements

The authors would like to thank Jonathan J Wisco for drawing the thalamus masks and assigning anatomical labels to the clusters. This work was supported by NINDS NS46532, NCRR RR14075, NCI CA09502, P41-RR13218 (NAC), GlaxoSmithKline, the Athinoula A. Martinos Foundation, the Mental Illness and Neuroscience Discovery (MIND) Institute, and the National Alliance for Medical Image Computing (NAMIC) (NIBIB U54 EB005149) which is funded through the NIH Roadmap for Medical Research.

## References

- [1] Behrens, T.E., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., Thompson, A.J., Brady, J.M., Matthews, P.M.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* **6**(7) (2003) 750–757
- [2] Wiegell, M.R., Tuch, D.S., Larsson, H.B., Wedeen, V.J.: Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *Neuroimage* **19** (2003) 391–401
- [3] Beaulieu, C.: The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed* **15**(7-8) (2002) 435–455
- [4] Jonasson, L., Hagmann, P., Richero Wilson, C., Bresson, X., Pollo, C., Meuli, R., Thiran, J.P.: Coupled, region based level sets for segmentation of the thalamus and its subnuclei in DT-MRI. In: ISMRM. (2005) 731

- [5] Shi, J., Malik, J.: Normalized cuts and image segmentation. *PAMI* **22**(8) (2000) 888–905
- [6] Higham, D., Kibble, M.: A unified view of spectral clustering. *Mathematics Research Report* (2004) 1–17
- [7] Kannan, R., Vempala, S., Vetta, A.: On clusterings: Good, bad and spectral. *J ACM* **51**(3) (2004) 497–515
- [8] Weiss, Y.: Segmentation using eigenvectors: A unifying view. In: *ICCV*. (1999) 975
- [9] Jones, D.K., Horsfield, M.A., Simmons, A.: Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *MRM* **42**(3) (1999) 515–525
- [10] Alexander, D.C., Gee, J.C., R., B.: Similarity measures for matching diffusion tensor images. In: *BMVC*. (1999)
- [11] Wang, Z., Vemuri, B.C.: An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In: *CVPR* (1). (2004) 228–233
- [12] Tishby, N., Slonim, N.: Data clustering by markovian relaxation and the information bottleneck method. In: *NIPS*. (2000) 640–646
- [13] Reese, T.G., Heid, O., Weisskoff, R.M., Wedeen, V.J.: Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *MRM* **49**(1) (2003) 177–182
- [14] Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**(2) (2002) 825–841
- [15] Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B* **103**(3) (1994) 247–254
- [16] Pierpaoli, C., Basser, P.J.: Toward a quantitative assessment of diffusion anisotropy. *MRM* **36**(6) (1996) 893–906
- [17] Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: theory and rationale for its development. the international consortium for brain mapping (icbm). *Neuroimage* **2**(2) (1995) 89–101
- [18] Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26** (1945) 297–302