

Content-based Retrieval of Brain Diffusion Magnetic Resonance Image

Siqi Liu¹, Nur Hadi¹, Sidong Liu¹, Sonia Pujol², Ron Kikinis², Fan Zhang¹,
Dagan Feng^{1,3}, and Weidong Cai^{1,2}

¹ Biomedical & Multimedia Information Technology (BMIT) Research Group, School
of Information Technologies, University of Sydney, Sydney, Australia

² Surgical Planning Laboratory (SPL), Brigham & Women’s Hospital, Harvard
Medical School, Boston, USA

³ Med-X Research Institute, Shanghai Jiaotong University, Shanghai, China

Abstract. The content-based retrieval of diffusion magnetic resonance (dMR) imaging data would enable a wide range of analyses on large databases with dMR images. This paper proposes a content-based retrieval framework for dMR images to explore the use of Diffusion Tensor Imaging (DTI) - derived parameters. The propagation graph algorithm is proposed for the query-centric retrieval of dMR subjects and the fusion of different features. The proposed framework was evaluated with ADNI database with 233 baseline dMR images. The preliminary results show that the proposed retrieval framework is able to retrieve subjects with similar neurodegenerative patterns.

1 Introduction

Diffusion MR imaging probes the diffusion of water molecules to investigate the architecture of the brain white matter [1, 2]. The introduction of the diffusion tensor model and the development of tractography techniques have allowed the reconstruction of the trajectory of major white matter bundles [16].

Content-based retrieval has been applied to many medical imaging modalities such as CT, MR and PET [14, 12, 5, 9]. The content-based retrieval of the dMR images [4] would be important for many applications, such as decision support, training and knowledge discovery. However the retrieval of brain dMR data is more challenging comparing to the traditional medical imaging modalities. Simply extracting the mean regional measurements, such as the mean diffusivity (MD) and the fractional anisotropy (FA), would be insufficient to represent the information in brain dMR, because the inter-region information should also be considered. Thus, a method is needed to combine the regional measurements as well as the brain connectome for brain dMR retrieval. Also, without proper feature encoding, the tasks related to raw dMR would pose a more intensive computational bottleneck in image querying than traditional MR due to the larger quantity of voxels required to index raw dMR images.

To explore the use of the DTI-derived parameters, we propose a content-based retrieval framework for brain dMR images which combines different DTI-derived

metrics. Elastic Net (EN) [21] is used to filter the extracted features for specific problem domains in a supervised way. We used the Propagation Graph Fusion (PGF) [20, 13] to construct the affinity graph to index the stored subjects for query-centric retrieval. The affinity graphs constructed with different features are fused into a global graph for the multi-view content-based retrieval of dMR.

2 Methods

2.1 Preprocessing and Feature Extraction for Diffusion MR Data

Automatic brain masking is firstly applied to the dMR images. Diffusion tensors are estimated for each masked dMR brain using the least square approach [11]. The Automated Anatomical Labelling (AAL) atlas is wrapped to the subject to obtain the predefined brain ROIs [15] by registering the DWI baseline template to the target baseline volume. The mean regional measurements, including the number of tracts, the tract lengths, the tracts volume, the fractional anisotropy (FA) and the apparent diffusion coefficient (ADC), are calculated within each predefined ROI from the whole brain to represent the regional tracts density and the average water diffusivity. An example FA index slice is shown in Fig. 1-(a). Whole-brain neural tractography is performed based on the estimated tensors using a deterministic algorithm [3]. The reconstructed fibres are filtered by the grey matter ROIs pair-wisely. The number of reconstructed fibres between each pair of AAL grey matter ROIs is stored in a symmetric matrix. The hyper-parameters used in the preprocessing pipeline are consistent across different subjects. An example of the reconstructed matrix is shown in Fig. 1-(b). For specific problem domains, such as disease determinant retrieval, Elastic Net is used to select features supervisely according to the pre-defined diagnostic labels [21] with only the normal cognitive (NC) and AD patients. It is reasonable to assume that affected brain ROIs of MCI patients are in a subset of the selected regions [7].

2.2 Subject-Centered Retrieval

Affinity Matrix Construction Assuming N_v types of features have been extracted from N_d images in the database \mathbb{D} and denoting x_i an image in the database and X the feature set, the neighbourhood of x_i in the n^{th} feature space is formed by itself and its k nearest neighbours, i.e., $X_i^{(n)} = x_i^{(n)}, x_{i_1}^{(n)}, \dots, x_{i_k}^{(n)}$. The Jaccard coefficient in Eq.(1) was used to establish the connections between subjects by measuring the neighbourhoods consistency as follow

$$w(x_i^{(n)}, x_j^{(n)}) = \frac{|X_i^{(n)} \cap X_j^{(n)}|}{|X_i^{(n)} \cup X_j^{(n)}|} \quad (1)$$

For a query x_q , its neighbourhood $X_q^{(n)}$ is constructed with the same procedure to determine its relative position in \mathbb{D} . The first iteration of neighbourhood

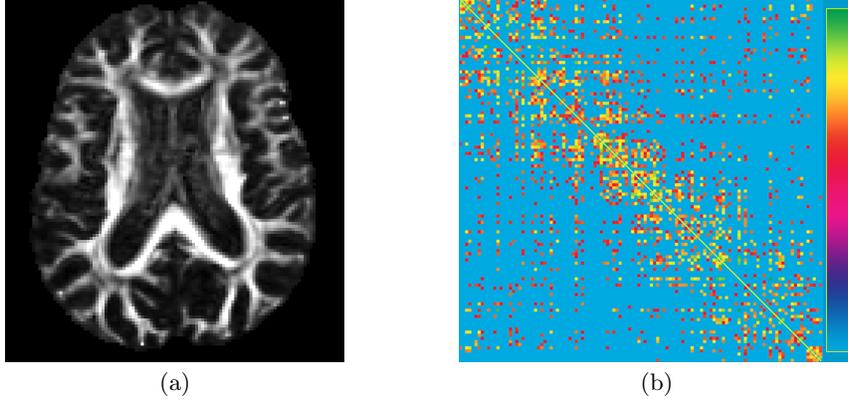


Fig. 1. (a) An example FA index slice which represents the mean water diffusivity within a voxel rendered with 3D Slicer 4.4 [8]; (b) An example of the symmetric brain inter-region matrix colour map with AAL grey matter (GM) ROIs. Each element in matrix is the number of tracts filtered by a pair of GM ROIs. The matrix is highly approximated for single subjects, due to the limitations of the dMR imaging resolution and the preprocessing pipeline. However the comparison between different subjects could be informative for the general dMR retrieval tasks.

searching of x_q is performed by propagating the paths from x_q to its k nearest neighbours $x_{q_1}^{(n)}, \dots, x_{q_k}^{(n)}$. The neighbourhood search is recursively continued from each reached node in $x_{q_1}^{(n)}, \dots, x_{q_k}^{(n)}$ to their own k nearest neighbours. The propagation stops until there is no other neighbour to be discovered. The lengths of the paths between the query and the subject indicate the relevance. The connection weights with regard to a specific query are updated as

$$w'(x_i^{(n)}, x_j^{(n)}) = \alpha^{t_q(x_i^{(n)}, x_j^{(n)})} \cdot w(x_i^{(n)}, x_j^{(n)}) \quad (2)$$

where α is a weight decay parameter to control the damping effect of the walk and $t_q(x_i^{(n)}, x_j^{(n)})$ is the number of iterations to reach the link $(x_i^{(n)}, x_j^{(n)})$. If a link is visited multiple times, we select the smallest t_q for it.

The updated weights are saved in a $N_d \times N_d$ sparse affinity matrix, $A(i, j) = w'(x_i^{(n)}, x_j^{(n)})$. The Laplace smoothing is applied to A by adding $1/N_d$ to each element to guarantee the non-zero values, because it reduces the coverage of the discoverable relevant subjects after the fusion of different features. A is then normalised to be row-stochastic.

Affinity Matrix Fusion For each query with N types of features, N_v affinity matrices $A^{(1)}, \dots, A^{(N_v)}$ can be obtained. The geometric mean is used to fuse the N_v matrices as follow

$$A^*(i, j) = \sqrt[N_v]{\prod_{n=1}^{N_v} A^{(n)}(i, j)} \quad (3)$$

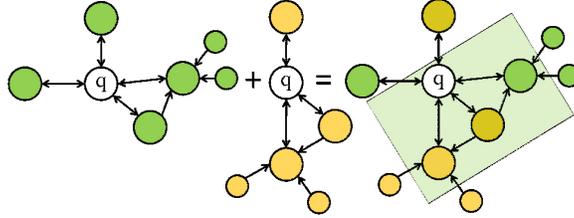


Fig. 2. A simple illustration of the fusion of each feature graph into a single graph.

An example of the affinity matrix fusion is illustrated in Fig 2. Our method requires $A^{(n)}(i, j) \neq 0$, since otherwise the link $(x_i^{(n)}, x_j^{(n)})$ will be disconnected regardless of the connections in other types of features. In this way, some potential candidates will be blocked. This problem is solved by Laplace smoothing during the affinity matrix construction. The weights in A^* reflect the overall white matter affinity to the query. The PageRank algorithm is applied to derive the equilibrium distributions in A^* [20] to re-rank the retrieved subjects according to their probabilities of relevance.

3 Experiments and Results

To evaluate the proposed dMR retrieval framework, 233 subjects with 3D axial brain dMR images were recruited from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database with ages ranging between 48 and 90 [10]. The subjects were divided into three groups with 61 normal cognitive (NC) subjects, 123 mild cognitive impairment (MCI) subjects and 49 Alzheimer’s Disease (AD) subjects. For each image, the preprocessing of the dMR images, including tensor estimation, tractography and fibre filtering, were conducted with DSI studio (<http://dsi-studio.labsolver.org>) [19, 17, 18]. Some parameters used for the fibre tracking are listed in Table 1.

Table 1. The parameters used in DT fibre tracking.

Parameter	Value
FA Threshold	0.12
Angular Threshold	60
Step Size	0.68
Seed Number	10000

Leave-one-out was used for evaluating the retrieval results. The tested case in each trial of leave-one-out was excluded in the feature select with EN. For

each trial in leave-one-out, the Mean Average Precision (MAP) was obtained as

$$MAP = \frac{\sum_{q=1}^Q \sum_{k=1}^{K_q} (p_q(k) \cdot rel_q(k)) / T_q}{Q} \quad (4)$$

where q is the query index, Q is the total number of queries, k is the rank in sequence of retrieved subjects, K_q is the total number of retrieved subjects, $p_q(k)$ is the precision at cut-off k , $rel_q(k)$ is the relevance score of the k^{th} retrieval result, and T_q is the number of relevant subjects associated with the query. The relevance criteria were designed as shown in Table 2.

Table 2. MAP Relevance Criteria used for retrieval evaluation [6].

Group	NC	MCI	AD
NC	1	0.25	0
MCI	0.25	1	0.25
AD	0	0.25	1

Grid Search was used to obtain the hyper-parameters in the propagation graph fusion algorithm (the neighbourhood size k and the dampening factor in PageRank). The single-label and the overall MAP are presented in Table 3. With the multiple types of features (the regional mean statistics and the fibre tracking-based features), we compared MAP of the simple feature vector concatenation (CONCAT) and the proposed propagation graph fusion algorithm (PGF). The features with/without EN filtering are also compared.

Table 3. The MAP results of the 233 dMR subjects obtained by the feature concatenation (CONCAT) and the propagation graph (PGF) algorithm using features with/without EN filtering (EN+*).

Fusion Method	MAP			
	NC	MCI	AD	Average
CONCAT	34.04	56.39	74.26	54.90
PGF	24.90	52.07	56.51	44.49
EN+CONCAT	39.84	55.66	74.77	56.76
EN+PGF	72.57	52.27	74.54	66.46

The PGF algorithm with features filtered by EN (EN+PGF) performed the best MAP overall (66.46). The MAP results yielded by EN+PGF on individual labels were also more evenly distributed than EN+CONCAT which was challenging to achieve because of the imbalanced dataset. It was noticeable that without EN filtering, the PGF algorithm did not outperform the simple concatenation (CONCAT). The retrieval of dMR is highly dependent on the quality of the extracted features, which could be effected by many components along

the preprocessing pipeline, such as imaging, correction, tensor estimation, tractography, template wrapping, etc. Thus the fibre tracking based features were approximated. The feature selection would filter out the noisy and biased neural patterns falsely revealed by the dMR features, enabling PGF to construct the local neighbourhoods with larger confidence. In Table 3, the feature selection only used the data with NC and AD labels, since there were big overlaps between MCI and the other two groups.

4 Conclusion

In this paper we proposed a content-based retrieval framework for brain diffusion MR imaging data which retrieves subjects from the diffusion MR database regarding the affinity in the DTI-derived parameters. Different DTI features were derived and selected from the dMR data. The propagation graph fusion (PGF) algorithm was used to fuse the neighbourhoods obtained from different dMR features. The proposed framework was evaluated with the dMR data recruited from the ADNI database. We compared different possible settings of this dMR retrieval framework. The PGF algorithm with features selected by Elastic Net achieved the best performance according to the MAP of the AD diagnostic labels. This work indicated that DTI-derived parameters can be used to index the dMR databases for content-based retrieval in order to perform further analysis in regard to the neural changes.

References

1. Assaf, Y., Pasternak, O.: Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *Journal of Molecular Neuroscience* 34(1), 51–61 (2008)
2. Basser, P.J.: Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR in Biomedicine* 8(7-8), 333–344 (1995)
3. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine* 44(4), 625–632 (2000)
4. Ben Ahmed, O., Benois-Pineau, J., Allard, M., Catheline, G., Ben Amar, C.: Diffusion tensor imaging retrieval for Alzheimer’s disease diagnosis. In: *Content-Based Multimedia Indexing (CBMI), 2014 12th International Workshop on*. pp. 1–6 (June 2014)
5. Cai, W., Kim, J., et al.: Content-based medical image retrieval. In: Feng, D. (ed.) *Biomedical Information Technology*, pp. 83–113. Elsevier (2008)
6. Che, H., Liu, S., Cai, W., Pujol, S., Kikinis, R., Feng, D.: Co-neighbor multi-view spectral embedding for medical content-based retrieval. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. pp. 911–914. IEEE (2014)
7. Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., Initiative, A.D.N., et al.: Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 39(4), 1731–1743 (2008)

8. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al.: 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging* 30(9), 1323–1341 (2012)
9. Hanbury, A., Müller, H., Langs, G., Weber, M., Menze, B., Fernandez, T.: Bringing the algorithms to the data: Cloudbased benchmarking for medical image analysis. In: *Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics*, vol. 7488, pp. 24–29. Springer Berlin Heidelberg (2012)
10. Jack, C., Bernstein, M., et al.: Update on the magnetic resonance imaging core of the Alzheimer’s disease neuroimaging initiative. *Alzheimer’s & Dementia* 6(3), 212–220 (2010)
11. Koay, C.G., Chang, L.C., Carew, J.D., Pierpaoli, C., Basser, P.J.: A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *Journal of Magnetic Resonance* 182(1), 115–125 (2006)
12. Liu, S., Cai, W., et al.: Multi-channel brain atrophy pattern analysis in neuroimaging retrieval. In: *ISBI*. pp. 202–205. IEEE (2013)
13. Liu, S., Liu, S., et al.: Propagation Graph Fusion for Multi-Modal Medical Content-based Retrieval. In: *ICARCV*. IEEE (2014)
14. Müller, H., Michoux, N., et al.: A review of content-based image retrieval systems in medical applicationsclinical benefits and future directions. *International journal of medical informatics* 73(1), 1–23 (2004)
15. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* 15(1), 273–289 (2002)
16. Westin, C.F., Maier, S.E., Mamata, H., Nabavi, A., Jolesz, F.A., Kikinis, R.: Processing and visualization for diffusion tensor MRI. *Medical Image Analysis* 6(2), 93–108 (2002)
17. Yeh, F.C., Tseng, W.Y.I.: Ntu-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *NeuroImage* 58(1), 91–99 (2011)
18. Yeh, F.C., Verstynen, T.D., Wang, Y., Fernández-Miranda, J.C., Tseng, W.Y.I.: Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS One* 8(11), e80713 (2013)
19. Yeh, F.C., Wedeen, V.J., Tseng, W.Y.: Generalized-sampling imaging. *Medical Imaging, IEEE Transactions on* 29(9), 1626–1635 (2010)
20. Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.N.: Query specific fusion for image retrieval. In: *The proceedings of ECCV 2012*, pp. 660–673 (2012)
21. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67(2), 301–320 (2005)