Annals of the BMVA Vol. 2009, No. 6, pp 1–12 (2009)

Spectral Clustering and Label Fusion For 3D Tissue Classification: Sensitivity and Consistency Analysis

William R. Crum

Centre for NeuroImaging Sciences
Institute of Psychiatry (PO89), De Crespigny Park, London, SE5 8AF
⟨bill.crum@iop.kcl.ac.uk⟩

Abstract

Clustering algorithms have found application in tissue classification in MRI. Standard techniques such as k-means iteratively define intensity clusters based on the distribution of voxels in intensity space. Spectral clustering is potentially more powerful as it models voxel-to-voxel relationships rather than voxel-to-cluster relationships. Unfortunately, for images of n-voxels naive application leads to an $n(n-1)/2$ voxel comparison problem and an order $n \times n$ eigenvalue problem which has prevented these techniques being widely investigated in 3D medical imaging. In this paper we report an empirical evaluation of a stochastic sampling approach to modelling voxel-to-voxel relationships for spectral clustering. Stochastic sampling captures sufficient intensity structure to give plausible tissue classification in 3D brain MRI. We test the stability of our approach to similarity parameter choice, sample size and stochastic effects in simulated and real 3D MR images.

1 Introduction

Data-clustering techniques are an important corner-stone of medical image analysis and a key component of many techniques for classifying Magnetic Resonance (MR) images into tissue types [Liew and Yan, 2006]. The most common application is in separating Grey-Matter (GM), White-Matter (WM) and Cerebro-Spinal Fluid (CSF) in brain imaging. Model-based techniques such as FAST [Zhang et al., 2001] provide powerful alternatives by incorporating prior knowledge, but their performance can suffer when model-assumptions are broken. Clustering can be used as an unbiased alternative to model-based techniques, to initialise model-based techniques or for applications where adequate models do not exist. The simplest techniques, such as k-means [MacQueen, 1967], model the intensity relationship of each voxel to a set of global intensity clusters (scalars for single image classification or vectors for multi-spectral classification). Techniques which model voxel-to-voxel intensity

relationships directly via the construction of a voxel similarity matrix should in principle be more powerful classifiers as they analyse am enhanced description of the whole data rather than considering each voxel only with respect to a small number of global clusters.

One such technique which has attracted recent attention is spectral clustering [von Luxburg, 2007, Higham et al., 2007, Filippone et al., 2008] which is closely related to graph partitioning problems such as **RATIOCUT** [Hagen and Kahng, 2006] and **NCUT** [Shi and Malik, 2000]. Spectral methods generally derive classification features by analysing a complete, undirected, weighted graph representation of the voxel data. However, naive spectral clustering quickly become impractical when the number of voxels is high. Modelling all pairwise voxel relationships requires an eigen-analysis of an $n \times n$ matrix; a 3D MR brain image may be composed of $n = 8$ million or more voxels.

Despite the computational problems, there have been a small number of previous applications of spectral clustering in various forms in medical image segmentation. In [Ziyan et al., 2006] a hierarchical variant was applied to thalamic Diffusion Tensor Imaging to extract thalamic nuclei. Although there is little technical detail it appears that the data from each thalamus was first extracted and then analysed separately. In [Song et al., 2006] an approach related to spectral clustering was used for MRI brain tissue classification into Grey Matter and White Matter but applied slice-by-slice in 2D due to computational limitations. In [Carballido-Gamio et al., 2004] spectral clustering was used for spinal MRI segmentation and incorporated an eigenvector interpolation technique for clustering in 2D. The common problem in prior work in this area is the challenge of dealing with large 3D data-sets (or even large 2D data-sets).

In this paper we solve a reduced eigen-problem using sparse stochastic sampling of voxel-to-voxel relationships. The emphasis is on preserving the connectivity structure implicit in the data rather than adopting a formal matrix-completion or eigen-vector interpolation strategy such as the Nyström approximation [Fowlkes et al., 2004]. We assess the influence of user-supplied parameters and explore whether repeated stochastic classifications combined using label fusion can further improve results.

2 Methods

There are many approaches to data-clustering of which K-means (KM) [MacQueen, 1967] is one of the best known. Classical K-means is applied directly to the supplied data and performs best when clusters are convex in the feature space. Spectral Clustering (SC) generates new features associated with each point in the supplied data and these derived features are then clustered using standard techniques such as KM.

2.1 K-means

Simple clustering techniques such as K-means (KM) are easily applied to either scalar or multi-spectral data. KM can be summarised as (a) define initial cluster centres, (b) assign each data-point to the nearest cluster based on a feature distance measure (c) recompute each cluster centre as the centre of mass of assigned data-points (d) repeat from (b). This process is guaranteed to converge but not necessarily to the globally optimal clustering. Multiple clustering attempts with randomised starting conditions can be used to find the optimal

clustering by minimising the distortion (cluster encoding error). In the most common case, the number of expected clusters is specified in advance.

2.2 Spectral Clustering

Rather than using voxel intensities as clustering features (or tuplets of intensities for multi-spectral voxel combinations), spectral clustering generates a new feature vector for each voxel from a model of the relationship between all pairs of voxels. These features, which are no-longer simply related to the intensities of individual points, are then clustered using standard techniques such as KM. The motivation is that spectral features characterise each voxel relative to the global ensemble rather than treating points in isolation. When all pairs of voxels are considered, the computational problems are immediately apparent: first the number of voxel similarity comparisons in the data-analysis stage is $n(n-1)/2$ and second, as will be described below, an eigen-analysis of an $n \times n$ symmetric matrix is required to generate suitable features. When $n >\sim 1$ million as for 3D MRI brain images this calculation is impractical.

Problems amenable to spectral clustering are described by an undirected complete similarity graph $G = (V, E)$ where V is a set of vertices v_i with one vertex for each of n points in the original data and E is a set of similarities $s_{ij} = s_{ji}$ between pairs of points. Complete graphs have connections between all pairs of points (i, j). To compute spectral features, G is expressed as a $n \times n$ Laplacian matrix. There are a number of definitions for the Laplacian [von Luxburg, 2007] but we use the simplest (unnormalized), $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{W}$ is the $n \times n$ symmetric matrix of pair-wise voxel-similarities, s_{ij} and $\mathbf{D}$ is the n-diagonal degree matrix formed from the row or column sums of $\mathbf{W}$. The degree matrix, $\mathbf{D}$, summarises the similarity of each voxel with respect to the other voxels as a whole, whereas the similarity matrix, $\mathbf{W}$, summarises pair-wise similarities between voxels. The matrix $\mathbf{L}$ is symmetric, positive semi-definite with smallest eigenvalue equal to 0. To construct spectral features, the eigenvectors associated with the next k smallest eigenvalues of $\mathbf{L}$, are formed into an $n \times k$ matrix $\mathbf{F}$. Spectral features associated with the i^{th} voxel, are taken from the i^{th} row of $\mathbf{F}$. The mathematical justification for this feature selection process can be framed in several different and suprising ways, either from a graph or matrix perspective; these justifications are discussed with clarity and insight in [von Luxburg, 2007]. The number of components, k, in the feature vector can be selected but is often related to the number of expected clusters; therefore in this work we choose $k = 3$ for GM, WM and CSF.

2.2.1 Stochastic Approach

MRI brain volumes are typically of size $n = 256 \times 256 \times 128 \sim 8$ million voxels total. We solve a reduced problem on the full-size 3D data, which retains global connectivity properties, by adopting a stochastic sparse sampling approach for building the Laplacian matrix. This approach is motivated by the way that connectivity properties define spectral features [von Luxburg, 2007]; solving a reduced problem with equivalent connectivity will result in spectral features which are suitable for classification.

The similarity matrix $\mathbf{W}$ is filled sparsely by computing pair-wise similarities between each voxel and m other randomly (uniformly) selected voxels. Each voxel is compared with $\sim 2m$ other voxels on average; (each voxel is explicitly compared with m other points and

on average another $\sim m$ voxels will be compared with it). Therefore, the total storage requirement for $\mathbf{W}$ is $\sim 2mn$ rather than n^2 for the naive case. We compute the n non-zero diagonal entries of the degree matrix, $\mathbf{D}$, directly from the row-sums of $\mathbf{W}$ to construct a sparsely sampled Laplacian.

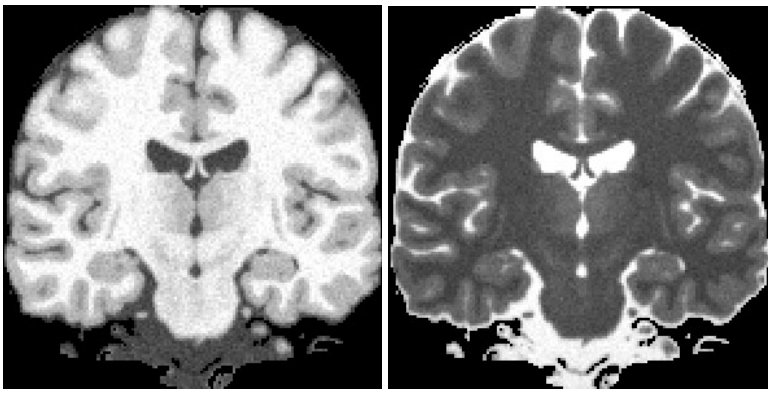
The number of samples must be sufficient to preserve the connectivity structure of the data. In particular, spurious disconnections (i.e. bogus cluster structure) must be avoided. A necessary condition to ensure this is that $\mathbf{W}$ must represent a connected graph i.e. there must be a path via non-zero similarities from any element to any other element. Theory suggests that $m \sim f(n)$ will ensure a connected graph with $f(n) \sim> \log n$ [von Luxburg, 2007]. The similarity should be a positive symmetric function which falls away rapidly for non-similar points. We adopt a parametric exponential similarity $s_{ij} = \exp\left(-x_{ij}^2/2p^2\right)$ where x_{ij} is the Euclidean intensity distance between voxels i and j and p is an intensity scaling parameter. For multi-channel data such as T1 and T2 image channels, $s_{ij} = s_{ij}(T1) \cdot s_{ij}(T2)$. For each voxel, we stochastically sample $l > m$ other voxels and use the m largest similarities in $\mathbf{W}$ to minimise the chance that any voxel has only m low-similarity connections and is effectively disconnected. As $l \to n-1$, this approach becomes equivalent to m-Nearest-Neighbour selection. In this paper we arbitrarily set the total sampled points at each voxel to be $l = 5m$. We use a Jacobi-Davidson method which has been optimised for generalised symmetric eigenvalue problems and embedded in a sparse-matrix framework [Geus, 2002] to compute the first $k+1$ eigenvalues of $\mathbf{L}$.

2.3 Stochastic Consistency and Label Fusion

The sparse eigenvector decomposition and therefore the resulting classification could be affected by stochastic sampling if the connectivity properties of the data are not captured in the reduced representation. Therefore, in this work we consider consistency of SC and whether combining repeated stochastically sampled classifications using label-fusion [Heckemann et al., 2006] can further improve results. Label fusion is a technique which has been used to reduce the effect of random labelling and registration errors in atlas-based segmentation techniques. In brief, if a labelled (i.e. segmented or tissue-classified) image can be accurately warped (i.e. registered) to an unlabelled image then the labels can also be warped, resulting in a new labelled image. In practice, labelling and registration errors reduce the achievable accuracy, however if multiple labelled images are each independently warped to an unlabelled image, their associated warped labels can be combined, which reduces the effects of the random sources of error. In this work, there is no registration error, but each independent classification attempt in general samples a different sub-set of voxels which may affect the final classification. We hypothesise that combining independent classification attempts will improve the overall classification accuracy.

3 Experiments

We performed three experiments. The first two experiments assessed sensitivity to parameter selection and classification accuracy in simulated data where the underlying classification is known. The third experiment assessed consistency of classification in real data using fused classifications generated from varying numbers of independent classifications.



(a) T1-weighted (b) T2-weighted

Figure 1: MNI BrainWeb simulated image with Rician noise

Software for KM and SC was written in Python including use of NumPy and SciPy[1], and PySparse[2]. The experiments below required $1-2$ hours compute time per SC classification on 2.8GHz AMD processors. Construction of the Laplacian is currently a computational bottleneck in Python which will soon be reduced by recoding in 'C'. The eigen-analysis makes use of optimized functions for handling sparse matrices in PySparse. We also made use of tools from the FMRIB Software Library[3], in particular, BET (brain extraction), FLIRT (rigid registration) and FAST (model-based brain tissue classification). To measure overall classification accuracy and consistency in each brain we computed the Total Accumulated Dice Overlap (TAO) [Crum et al., 2006] over GM, WM and CSF. The standard Dice overlap between a single pair of tissue labels is defined as the ratio of the coincident volume of tissue to the mean volume of tissue. The TAO for multiple pairs of tissue labels is defined as the ratio of the total coincident tissue volume summed over all pairs to the total mean tissue volume summed over all pairs. The TAO is in the range $[0, 1]$ which reduces to the standard Dice overlap for a single label comparison but allows the overall agreement of multiple labels to be summarised by a single index.

3.1 Parameter Sensitivity in Simulated Data

We used the original MNI BrainWeb[4] digital brain phantom T1- and T2-weighted 1mm isotropic voxel images [Cocosco et al., 1997]. We added independent Rician (2-channel Gaussian) noise of 2% maximum intensity to each image The images were pre-processed to leave 4 underlying clusters (background, CSF, grey-matter, white-matter) and a mask which identified the background cluster. As pure intensity clustering does not depend on the spatial location of voxels a more efficient computation can be applied in data which have been masked as part of a pre-processing step. Packing each image into a vector containing only

[1]http://www.scipy.org/
[2]http://sourceforge.net/projects/pysparse/
[3]http://www.fmrib.ox.ac.uk/fsl
[4]http://www.bic.mni.mcgill.ca/brainweb

voxels which are non-zero in the mask reduces both the size of the clustering problem and the number of clusters to be determined by excluding background voxels. We initially set the number of stochastic samples high at $m = 40$ to ensure global connectivity and examined the sensitivity of the classification accuracy to the intensity parameters P_1 and P_2 associated with the T1- and T2- weighted images respectively. We first experimented with the T1-weighted image alone and set $p = P_1 = \{1000, 1500, 2000, 2500, 3000\}$ based on the intensity range in the T1-weighted image ($[0, 4095]$). From these results we selected an appropriate fixed value for P_1 and repeated the experiment on T1- and T2-weighted multi-spectral images. Finally, we repeated the first two experiments for the selected values of P_1 and P_2 and varied the number of samples $m = \{20, 30, 40\}$. We then compared the effect on classification accuracy of increasing numbers of stochastic samples against fusing multiple independent classification attempts.

3.2 Classification Accuracy in Simulated Data

We next applied FAST and SC to 10 subjects chosen at random from the 20 available new BrainWeb simulations of normal brains generated with a revised simulation model [Aubert-Broche et al., 2006] with the brain and CSF extracted from background as before and with added Rician noise of 2% maximum intensity. Accuracy was evaluated by comparison with the supplied underlying tissue-maps and summarised using the TAO.

3.3 Consistency in Real Data

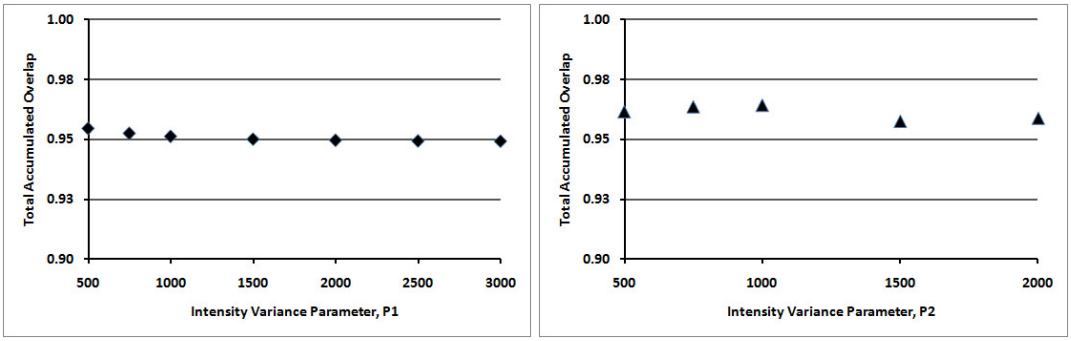
We applied KM and SC to multi-spectral 3D MRI from 10 normal volunteers acquired as part of an imaging protocol evaluation on a General Electric 1.5T Signa HDx scanner. The image channels acquired were axial T1-weighted (T1w) MPRAGE ($256 \times 256 \times 180$) and T2-weighted (T2w) ($512 \times 512 \times 36$). The T1w scans were corrected for intensity inhomogeneity using N3 [Sled et al., 1998]. The T2w scans were rigidly registered and resampled into the space of the T1w scan using FLIRT resulting in all scans having voxel dimensions of $0.9375 \times 0.9375 \times 1.2$mm compared with the $0.46875 \times 0.46875 \times 4.0$mm original dimensions of the T2w images. All images were preprocessed using BET to remove non-brain/CSF voxels. We examined the consistency of SC when running repeated independent classifications with $m = 30$. For comparison we produced classifications using KM as described above and using FAST, a model-based tissue classifier.

4 Results

4.1 Parameter Sensitivity in Simulated Data

Figure 2(a) shows the Total Accumulated Overlap for CSF, GM and WM as a function of P_1 for SC of the T1-weighted MNI BrainWeb image. We chose $P_1 = 1000$ for subsequent experiments. Figure 2(b) shows the TAO as a function of P_2 for SC of the T1- and T2- weighted BrainWeb images and we chose $P_2 = 1000$ also. Figure 3 shows the TAO as a function of the number of stochastic samples, m, and number of fused classifications for T1w alone, and T1w and T2w BrainWeb images. The additional T2w information channel makes the most difference to the classification accuracy but there is also a weak linear relationship of classification accuracy with number of fused classifications. Interestingly the best results

for two-channel classification are obtained with the fewest samples (20). This suggests that the optimum clustering structure inherent in the data is not quite aligned with the optimum tissue classification defined by the gold-standard. (This is analagous to the case in image registration where maximimising image-similarity does not always result in the best anatomical match between images.)



(a) T1w image results showing P_1 dependence.

(b) T1w and T2w image results with $P_1 = 1000$ showing P_2 dependence.

Figure 2: The dependence of spectral clustering accuracy on intensity parameters in simulated BrainWeb images.

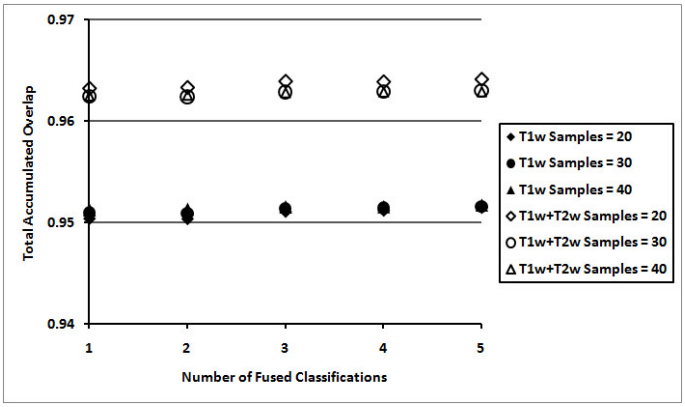


Figure 3: The dependence of spectral clustering accuracy on number of stochastic samples and fusion of repeated classifications in original simulated BrainWeb images.

4.2 Classification Accuracy in Simulated Data

Figure 4 shows the TAO (evaluated for WM, GM and CSF) for each BrainWeb subject when SC, KM (applied directly to the images) and FAST tissue classifications are compared with the known tissue maps. The TAOs are comparable between techniques but lower than those reported for the original standard BrainWeb digital phantom. There are two factors which

will affect accuracy. First as supplied, the gold-standard tissue-maps for these subjects require interpolation into the image space before they can be compared voxel-by-voxel; this interpolation stage will introduce error in the TAO. Second the images are qualitatively different from the original BrainWeb image (due presumably to the revised simulation model) and tissue classes exhibit more heterogeneity than in the original BrainWeb digital phantom. It is also worth noting that an intensity-based clustering (minimum-distance clustering) is used in the simulation process; the resulting intensity distributions may affect different clustering techniques in subtly different ways. The result for SC applied to subject 45 is particularly poor and visual inspection showed a clear failure to separate GM and WM adequately. Simply increasing the number of random initial cluster positions for the final KM classification stage from 25 to 50 (i.e. increasing the chance of initialising close to the optimal clustering but not changing the way the spectral features used for the classification were derived) gave a more plausible classification and raised the TAO for that case from 0.80 to 0.90.

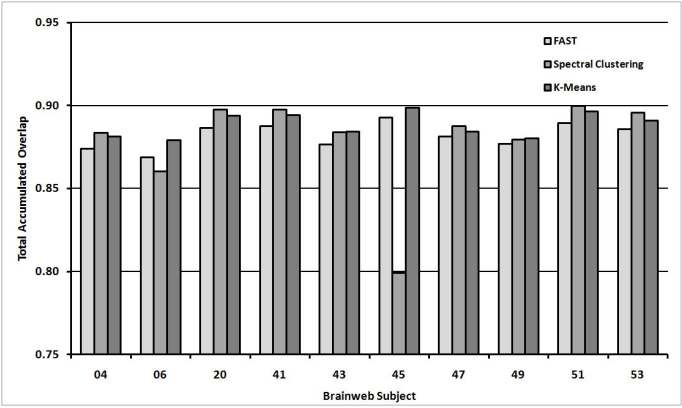
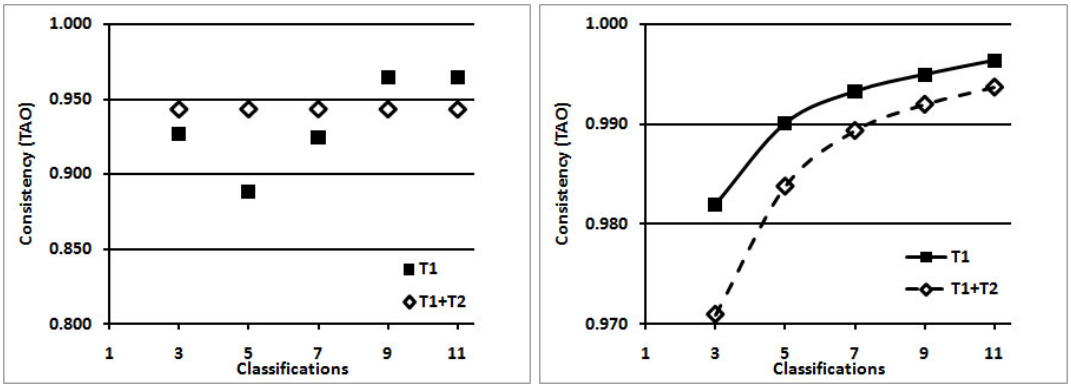


Figure 4: Classification accuracy for spectral clustering, KM and FAST applied to selected new T1-weighted BrainWeb images

4.3 Consistency in Real Data

We assessed stochastic sampling and label fusion consistency by computing (a) overlaps of independent spectral classifications of the same data and (b) overlaps of classifications formed by fusing increasing multiples of independent classifications. Figure 5 shows that independent spectral classifications vary in consistency from one attempt to the next (a) but that fused classifications converge towards a limiting result (b). The extra information available by including T2w images with independent noise stabilises the performance of individual spectral classifications. However fusing multiple independent classifications using either single-channel (T1w) or two-channel (T1w, T2w) images results in much more consistent classifications. Interestingly multiple classification attempts of single-channel (T1w) images gives the most consistent results overall. Figure 6 gives a visual example of the classification obtained for one subject by KM and SC.



(a) Individual spectral classifications of the same data showing the effect of stochastic sampling.

(b) Comparison of classifications obtained by fusing increasing numbers of individual classifications.

Figure 5: Consistency of spectral classification in 10 subjects.

5 Conclusions

We presented a stochastic sampling approach which makes spectral clustering analysis of full 3D MRI volumes tractable. We found that results were consistent across a range of parameter values. Classification accuracy improved with increasing numbers of stochastic samples but larger gains came from including both T1- and T2-weighted images in the classification. In real data spectral clustering resulted in biologically plausible classifications which compared well with a model-based classification algorithm well-suited to the data. Interestingly, adding T2-weighted images into the classification of real data seemed to reduce consistency. We attribute this to partial volume effects from the 4mm T2-weighted slices confounding the classification; (the *simulated* T2-weighted images were of isotropic 1mm voxel dimensions). Partial volume effects in general have not been explicitly modelled in this work but could be considered by using (a) fuzzy c-means rather than k-means to generate partial cluster-membership maps (b) inclusion of more clusters to explicitly model partial-volume effects. Another possible confound, intensity inhomogeneity can be corrected as a pre-processing step. Alternatively, the space from which stochastic samples are acquired could be restricted to a volume surrounding each voxel. We will investigate these approaches in future work.

We are not the first to apply spectral clustering to MRI data but we believe we are the first to explicitly classify from complete 3D volumes. Others have approximated or solved other kinds of reduced problem. For instance, in [Archip et al., 2005], 15×15 pixel windows in ultrasound images were classified independently and the results were combined *post hoc*. In [Carballido-Gamio et al., 2004] voxel-similarity features were computed in 3D but then projected onto a 131×256 slice for clustering using the Nyström approximation. One obvious extension of this work is to assess how spectral clustering using the Nyström approximation compares with the stochastic approach on the large 3D data-sets used in this work.

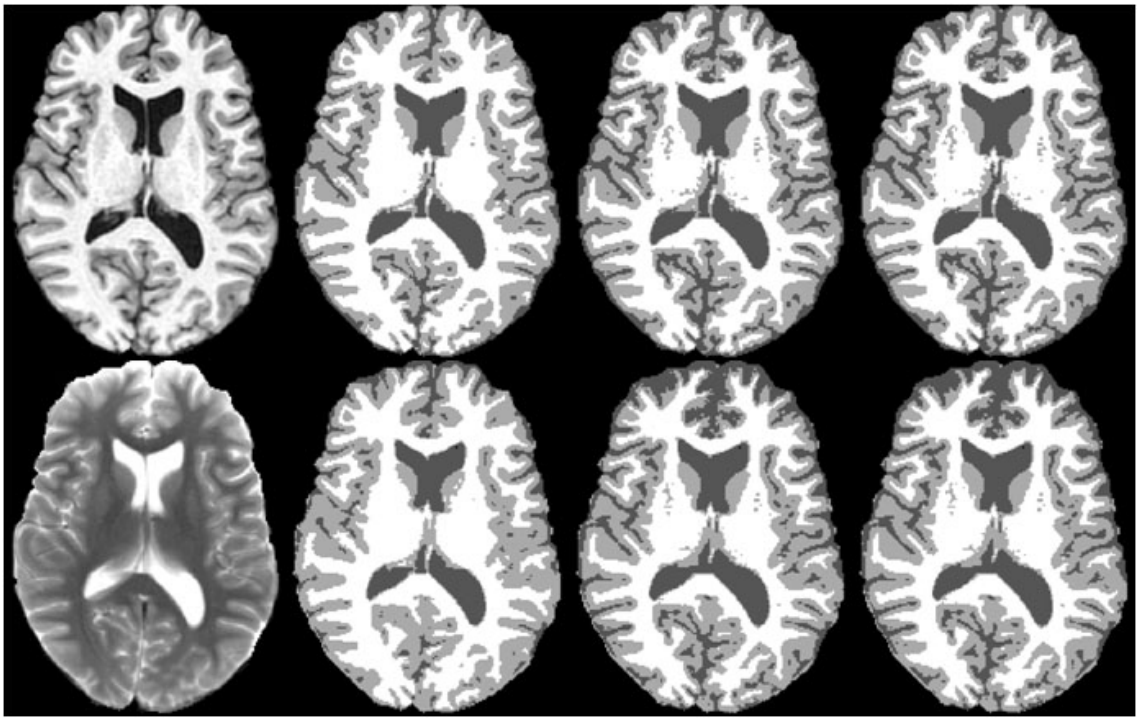


Figure 6: Tissue classification of subject 2. A single slice is shown but the classification was fully 3D. Top Row (L-R): T1-weighted image, KM of T1-weighted image, Single SC classification, Fusion of 9 SC classifications. Bottom Row (L-R): T2-weighted image, KM of T1+T2 images, Single SC classification of T1+T2, Fusion of 9 SC classifications of T1+T2

Importantly, the stochastic approach and Nyström approximation now give researchers the tools to explore the potential of this class of techniques without exceptional computing power or over-simplification of their imaging application. Spectral clustering is a relatively new and potentially powerful clustering approach but there are other methods (see [Jain et al., 1999] for a review) such as hierarchical clustering which can also model pairwise relationships. Future work will include a quantitative comparison of alternative methods.

Acknowledgements

The author acknowledges financial support from the Department of Health via the National Institute for Health Research (NIHR) Specialist Biomedical Research Centre (BRC) for Mental Health award to South London and Maudsley NHS Foundation Trust (SLaM) and the Institute of Psychiatry at King's College London. The author also thanks Andrew Simmons and the BRC for the volunteer data and Paul Aljabar of Imperial College London for useful discussion on spectral clustering.

References

N. Archip, R. Rohling, P. Cooperberg, and H. Tahmasebpour. Ultrasound image segmentation using spectral clustering. *Ultrasound in Medicine and Biology*, 31(11):1485–1497, 2005.

B. Aubert-Broche, A.C. Evans, and L. Collins. A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1):138–145, 2006.

J. Carballido-Gamio, S.J. Belongie, and S. Majumdar. Normalized cuts in 3-D for spinal MRI segmentation. *IEEE Transactions on Medical Imaging*, 23(1):36–44, 2004.

C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans. BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4), 1997.

W.R. Crum, O. Camara, and D.L.G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 2006.

M. Filippone, F. Camastra, F. Masulli, and S. Rovetta. A survey of kernel and spectral methods for clustering. *Pattern Recognition*, 41:176–190, 2008.

C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the Nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):214–225, 2004.

R. Geus. *The Jacobi-Davidson algorithm for solving large sparse symmetric eigenvalue problems with application to the design of accelerator cavities*. Doctor of technical sciences, Swiss Federal Institite of Technology, Zurich, 2002.

L. Hagen and A. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design*, 11(9):1074–1085, 2006.

R.A. Heckemann, J.V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, Oct 2006.

Desmond J Higham, Gabriela Kalna, and Milla Kibble. Spectral clustering and its use in bioinformatics. *Journal of Computational and Applied Mathematics*, 204:25–37, 2007.

A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999.

A.W.C. Liew and H. Yan. Current methods in the automatic tissue segmentation of 3D magnetic resonance brain images. *Current Medical Imaging Reviews*, 2(1):91–103, 2006.

J.B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. University of California Press, 1967.

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A non-parametric method for automatic correction of intensity non-uniformity in mri data. *IEEE Transactions on Medical Imaging*, 17(1): 87–97, 1998.

Z. Song, N. Tustison, B. Avants, and J. Gee. Adaptive graph cuts with tissue priors for brain MRI segmentation. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI), Arlington, VA, April 2006*, pages 762–765, 2006.

U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–47, 2001.

U. Ziyan, D. Tuch, and C.-F. Westin. Segmentation of thalamic nuclei from dti using spectral clustering. In *Ninth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '06)*, pages 807–814, 2006.