HIERARCHICAL SPECTRAL CLUSTERING OF MRI FOR GLOBAL-TO-LOCAL SHAPE ANALYSIS: APPLIED TO BRAIN VARIATIONS IN ALZHEIMER'S DISEASE

D. Gors^(a), P. Suetens^(a), R. Vandenberghe^(b), P. Claes^(a)

^(a)KU Leuven, ESAT/PSI – UZ Leuven, MIRC, Belgium. ^(b)Laboratory for Cognitive Neurology, Department of Neurology, University of Leuven and University Hospitals Leuven Gasthuisberg, Leuven, Belgium.

ABSTRACT

In this paper a hierarchical brain segmentation from multiple MRIs is presented for a global-to-local shape analysis. The idea is to group voxels into clusters with high within-cluster and low between-cluster shape relations. Doing so, complementing voxels are analysed together, optimally wheeling the power of multivariate analysis. Therefore, we adapted hierarchical spectral clustering to volumetric image datasets. The outcome is a segmentation of the brain into regions at different levels of detail and anatomical overlap, which are tested in their ability to predict ADAS-cog (a score for cognitive function in Alzheimer's disease (AD) diagnosis) from brain shape. The results show a benefit for a global-to-local analysis compared to a typical voxel-based and whole brain-based analysis. Additionally, knowledge on brain variations in AD is perfectly confirmed and even elaborated on.

Index Terms— hierarchical spectral clustering, shape analysis, MRI, Alzheimer's disease, cognitive score

1. INTRODUCTION

The shape of the human brain is influenced by a range of factors, including growth and development, ageing, and neurodegenerative diseases. Currently, morphometric studies of the brain [1], where each voxel codes for a single shape feature, are performed either as numerous univariate analyses or a limited number of multivariate analyses. For example, statistical parametric mapping (SPM) [2] is an established approach for testing hypotheses about image data. In SPM, voxel-wise statistical inferences (e.g. student t-test) are performed, after which special attention is required for the multiple testing burden [3]. However, such an approach appears to be unnatural, since individual voxels are not independent from each other. Especially, in the context of shape analysis, neighbouring voxels typically exhibit similar shape features. E.g., it is known that AD causes grey matter atrophy with an increase in CSF as a consequence. Furthermore, shape interactions between voxels are overlooked. Alternatively, a multivariate approach combines a group of voxels, up to the whole brain, together in a single test, which lowers the multiple testing burden on statistical power and takes interactions into account. However, a multivariate analysis suffers from the

curse of dimensionality, making it hard to find patterns in large dimensional data, if not enough observations are provided. Especially, in brain research, the amount of variables (#voxels) always exceeds the amount of observations (#images). Therefore, adapted techniques are required to deal with the dimensionality, such as principal component analysis (PCA) and partial least squares (PLS). An often-ignored side effect, however, is the loss of localized shape variations. Therefore, there is a trade-off between localizing variations, following a voxel-wise approach, and incorporating voxel interactions, following a whole brain-based approach.

In this work we propose a hierarchical spectral clustering of MRI images to investigate voxel-based shape variations from a global-to-local perspective. The idea is to group voxels into clusters with high within-cluster and low between-cluster shape relations. Doing so, complementing voxels are analysed together, optimally wheeling the power of multivariate testing. The clustering is implemented hierarchically, resulting in a segmentation of the brain into regions at different levels of detail and anatomical overlap. We illustrate the work on shape variations in Alzheimer's disease (AD). Alzheimer's affects both the brain structure as well as the cognitive ability of the patient. Shape has already been correlated to cognitive function, either voxel-wise [4] or surface vertex-wise [5]. Opposing, however, we will predict the cognitive function from MRI images, similar to the work of Sing et al. [6]. Our results are contrasted against results from voxel-based and whole brain-based predictions. The remainder of this text is organized as follows: section 2 explains the materials and methods. Section 3 presents the results, which are discussed and concluded in section 4.

2. MATERIALS AND METHODS

Baseline T1-weighted structural MRI scans and demographical information of 593 subjects were selected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The cohort includes, 183 healthy volunteers, 284 mild cognitively impaired (MCI) patients and 126 AD patients. Cognitive ability was measured using the Alzheimer's Disease Assessment Scale-cognitive (ADAS-cog) score, which quantifies memory, language, praxis, attention, orientation and other cognitive abilities.

This work is based on tensor-based morphometry (TBM [1]). First, all MRI images are aligned using a non-linear diffeomorphic registration (DARTEL [7]) implemented in SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>). Then, for each voxel the determinant of the gradient of an image's deformation field, known as the Jacobian matrix field, is computed, which captures local volume differences (shrinking or expanding) relative to the average image. Non-brain voxels were removed with a mask. Subsequently, each Jacobian image is rescaled by its mean Jacobian value to eliminate differences in total brain volume. Those Jacobians serve as the voxel-wise shape features.

Spectral clustering (SC) [8] is a graph-partitioning technique, separating vertices into k clusters. The partitioning is one that maximizes/minimizes the affinity of vertices within/between clusters. Essential is an appropriate affinity matrix A , defining the connectivity between pairs of vertices. Subsequently, A is transformed into a graph Laplacian matrix L , which in turn is decomposed into its k selected eigenvectors, enhancing the (dis)similarities, and therefore relaxing the partitioning problem. In the eigenvector space, the Euclidean distance, serves as a proxy for the affinity, and drives a k -means clustering [9] to partition the vertices into k clusters. An extension to the conventional SC is implementing it in a hierarchical setting. In translation to using hierarchical SC onto a sample of MRI images, the image voxels simply constitute the vertices. Then, the goal is to create a global-to-local partitioning of all voxels, which involves three main challenges. First, a proper affinity for shape similarity of voxel-pairs is required. Second, the amount of voxels is huge such that computational tricks are essential. Third, partitioning of brain shape is not expected to be clear-cut. Therefore, small perturbations lead to different partitions.

In regard to the first challenge, total brain volume does not change, only internal shape differs between subjects. Therefore, if one region is expanding, another joined region will be shrinking. Although correlated, we prefer to separate such opposing shape variations as they typically occur at the boundary of structural regions in the brain. Additionally, the affinity between voxel-pairs is to be defined across multiple images, instead of a single object, which is commonly done in the original spectral graph theory. Therefore we define the affinity between voxels i and j as: $A(i, j) = 0.5(\text{corr}(\mathbf{J}_i, \mathbf{J}_j) + 1)$, with $\mathbf{J}_i$ the Jacobian-vector over all the different images at the i^{th} voxel. Negative/positive correlations now generate a low/high affinity between voxel pairs. Subsequently, a well-considered choice of the Laplacian is essential, since it relaxes the partitioning problem that is not clear cut. During experimentations, a normalized symmetric implementation of the Laplacian proposed by Ng, Jordon and Weiss [10] was chosen: $L = D^{-1/2}AD^{-1/2}$, from which the eigenvectors associated with the k highest eigenvalues needs to be taken, with D a diagonal matrix, $d_{i,i}$ represents the sum of the i^{th} column of the Affinity matrix.

Figure 1 Correlation-maps between the predicted and actual ADAS-cog of different transverse slices for the voxel-based (1st row) and 8th level of the cluster-based (2nd row) analysis. The voxels are colored according to the predictive correlation of the independent voxel or the cluster to which the voxel belongs.

Computationally the amount of voxels makes it impossible to build the Affinity and Laplacian matrices. E.g., the images used contain 453,650 voxels, roughly requiring 1500 GB under double precision. Therefore a straightforward eigendecomposition of L is impossible. As a solution we use Nyström [11], which is a technique that simultaneously estimates the Laplacian and the eigenvectors of a large-scaled matrix using only a random subset of its columns. Therefore the affinity between only a subset of voxels and all the other voxels is needed.

As for the third challenge, the clustering is sensitive to randomness, which is introduced by k -means and Nyström. Therefore, a two step-robustness procedure is implemented. The first step averages multiple Nyström estimations resulting from randomly different voxel subsets. In the second step k -means is repeated several times and the produced clusters are merged into a single partitioning with a normalized vote-weighting scheme [12].

Finally, SC generates a user-selected number (k) of non-overlapping clusters. Therefore the choice of k is important, but inconclusive defined. Alternatively, a hierarchical SC generates global-to-local non-overlapping clusters at the same hierarchical level and overlapping clusters across different hierarchical levels. Therefore, the choice of k is simplified (but not completely eliminated) by a choice in number of levels. A hierarchical SC is obtained by sequentially splitting a region into two disjoint regions, whereby each new region independently of the other is further partitioned into two and so on. Starting with all voxels at level 1, the L^{th} level will be a segmentation with $2^{(L-1)}$ non-overlapping regions. Interestingly, a split in just two disjoint regions tends to generate the strongest partitioning. Additionally, the parent-offspring relationship of clusters at different levels allows the propagation of analyses on a hierarchical tree.

To illustrate the applicability of the work for shape analyses, we took inspiration from Singh et al. [6]. Given voxel-based shape features $\mathbf{J}_i$, the goal is to predict the cognitive score of subjects. We employ, partial least squares regression (PLSR) of the scores on the feature-space enabling a shape-based prediction of the cognitive ability.

Figure 2 a) The first 5 levels of a 3D hierarchical clustering shown for a single transverse slice. The yellow colored voxels are the voxels within a cluster. b) The 16 clusters of the 5th level, each in a distinctive color. c) Correlation between predicted and actual ADAS-cog for all hierarchical clusters. The matrix is build up like a hierarchical tree as in a) and is limited to 9 levels for visualization purposes.

This prediction is evaluated using the correlation with the actual cognitive scores in a leave-one-out cross validation on all 593 images. The main differences with Sing et al. [6] include, 1) the use of TBM instead of momenta maps, 2) the use of linear PLSR instead of kernel PLSR, and 3) more importantly the use of the proposed clustering to obtain voxel groupings as predictors in PLSR in contrast to a whole brain-based prediction only. In fact we will contrast the proposed hierarchical SC (cluster-based analysis) against using each voxel separately (voxel-based analysis) and all voxels combined (whole brain-based analysis).

We also evaluate the extent to which different anatomical brain structures influence the cognitive score. To this end the voxel- and cluster-based analyses are linked to a fuzzy anatomical atlas. This atlas is obtained by diffeomorphic aligning 136 labelled functional structures (available in SPM12, provided by Neuromorphometrics, Inc. <http://neuromorphometrics.com>) from the MNI-space to the average space. The atlas is simplified by merging smaller subdivisions into 48 common brain structures. To account for possible misalignment of the atlas a fuzzy atlas is created using morphological dilatation (spherical mask of 3x3x3) of the structures into slightly overlapping regions. A cluster is associated with a structure if there is sufficient overlap. That is, when the overlap accounts for more than 10% of the total volume of the cluster or structure. Subsequently, a maximum and median value for each anatomical structure is obtained from its encapsulating voxel- or cluster-wise predictive correlation. Some interesting structures are e.g. amygdala (Amy), cingulate gyrus (CG), entorhinal area (EA), fusiform gyrus (FG), hippocampus (Hip), lateral ventricles (LV), parahippocampal gyrus (PhG), putamen (Put), thalamus (Tha), temporal pole (TP) and WM.

Before analysing the independent voxels and a hierarchical clustering with 12 levels, all shape features J_i and cognitive scores are corrected, using PLSR, for factors

like gender, age, genetic ancestry (as obtained from ADMIXTURE [13]) and years of education.

3. RESULTS

The predictive correlation-maps are depicted in Figure 1. Table 1 reports the results of the top 10 ranked anatomical structures (and the LV) associated with cognitive score according to the cluster-based analysis. The whole brain-based analysis returns a single correlation of 0.4076. This is lower than any of the best correlations found with the other analyses. From the voxel-based results it is obvious that certain areas in the brain are better at predicting the cognitive scores than others. The maximum voxel-based correlation is 0.6056 for a voxel located in the fuzzy Amy and WM. Anatomical structures such as Amy, Hip, EA, PhG and LV show a uniform distribution of correlations for all their encapsulated voxels. In contrast, other anatomical structures like WM, Tha, FG, TP, CG and Put show only localized associated voxels, which is noted by the discrepancy in maximum and median correlation in Table 1.

The cluster-based analysis has the advantage of directly generating a single correlation for an area, which results in a less fluctuating correlation-map. In comparison to the voxel-based analysis, it is clear that combining multiple voxels is in general better than the results of all the voxels individually. Anatomical structures like Tha, FG and CG clearly improve due to voxel interactions. The clusters of the hierarchical segmentation, shown in Figure 2(a,b), are spatially coherent (although not explicitly enforced) and quite similar to anatomical structures, which is reflected in the high overlap between clusters and structures. This resemblance to anatomical structures implies a homogenous deformation of such structures. Yet, the clusters are not exactly equal to them either, indicating that shape variations are not necessarily bound by the brain's anatomy. The global-to-local correlations are shown in Figure 2(c), where

it is easy to follow the change in prediction accuracy across different levels: starting from the global brain (1st level) an increase is observed when the brain is sequentially split into more localised clusters leading to a maximum correlation of 0.6239 for cluster 29 at level 8. However, when this cluster is split further the correlation decreases again. This pleasantly reveals the optimal trade-off between global and local shape variations. Interestingly, cluster 29 at level 8 overlaps with Amy, Hip, EA, PhG and WM.

Table 1 Median and maximum correlation between predicted and actual ADAS-cog in anatomical structures. For the cluster-based analysis the max correlation found in a cluster linked to a specific structure from level 8 to 12 is reported, together with the median correlation of all the clusters linked with that structure for the optimal level.

	voxel-based		cluster-based		
	median	max	level	median	max
Amy	0.4684	0.6056	8	0.5070	0.6239
WM	0.0775	0.6056	8	0.1789	0.6239
Hip	0.3317	0.5969	8	0.3206	0.6239
EA	0.3883	0.5896	8	0.4821	0.6239
PhG	0.2187	0.5779	8	0.2223	0.6239
Tha	0.1185	0.5137	10	0.1983	0.5944
FG	0.1087	0.5297	11	0.1670	0.5929
TP	0.1667	0.5357	12	0.2610	0.5861
CG	0.1026	0.4579	11	0.1792	0.5438
Put	0.1181	0.5504	12	0.1290	0.5092
LV	0.2047	0.4706	8	0.2249	0.3900

4. DISCUSSION AND CONCLUSION

In this paper a hierarchical brain segmentation from multiple MRIs is presented for a global-to-local shape analysis. The ability to predict cognitive function from shape was tested and compared to a voxel-based and whole brain-based analysis. A voxel-based analysis, does not take voxel-interactions into account, and struggles to generalize towards anatomical structures. A whole brain-based analysis performed the worst, overlooking localized shape variations due to the curse of dimensionality. Finally, the cluster-based analysis using the proposed segmentation performed better, clearly investigating a wider range of voxel-interactions without overlooking local shape variations.

Singh et al. [6] reported a correlation of 0.6 using a whole brain-based analysis. This is in contrast to our result of 0.4076, and is probably due to the difference in shape features and the use of kernel PLSR. Yet, with the hierarchical segmentation we were able to improve the correlation to 0.6239, which is an increase of 53%. Singh et al. [6] identifies local deformation patterns, by looking at the partial regression coefficients of the voxels. However, the dimension reduction onto latent variables in PLS, takes all the voxels into account, leading to an averaging effect and a dilution of localized variations. This is especially so when the amount of variables (#voxels) is larger than the

amount of observations (#images). In a hierarchical segmentation the clusters are becoming more localised at each level, better focusing on localised variations.

All the analyses performed flagged the same anatomical structures: Amy, Hip, EA and PhG. This is expected, because together they make up the medial temporal lobe and the temporal lobe (also including TP and FG) is responsible for memory, language and emotion association. Hip and Amy are also affected first by AD. Galton et al. [14] investigated the atrophy for the different temporal structures in AD. They found significant differences in the volumes of Amy, Hip and PhG (including the EA). The volumes of the TP and FG were smaller for AD patients, but not significantly. De Jong et al. [15] correlates the volumes of Hip, Put, Tha, Amy, caudate and pallidum against different cognitive scores. Only the first three structures significantly correlated. In our study using shape instead of volume, we identified most of the investigated structures in [14] and [15], with the exception of the caudate and pallidum, which is in line with de Jong et al. [15]. Additionally, we also reveal that the medial temporal lobe deforms as a whole, rather than having independent structure changes.

The hierarchical segmentation possibly results in the definition of clusters across anatomical boundaries. This allows for the analysis of shape variations from interacting (sub-)anatomical structures. E.g. a cluster significantly overlapping with the LV, Amy and Hip, laying in the proximity of Tha gives a correlation of 0.39. This is perfectly in line with, Ferrarini et al. [16] who discovered that for the LV the shape of the lateral horns (reflecting hippocampal atrophy) and areas adjacent to Amy, Tha and caudate were significantly affected by AD. This sub-anatomical area is identical to the cluster. Other examples where structures profit of those boundary-voxel interactions are e.g. Tha and CG. The shape of those structures correlates the best with cognitive changes when joining parts of the Tha, Hip, PhG and WM; and parts of the CG, Hip, Tha and WM in clusters, resulting in respectively a correlation of 0.5944 and 0.5438. Table 1 also reveals that the optimal clusters for the LV, Tha, CG and Put are located at respectively the 8th, 10th, 11th and 12th level of the 12-level hierarchical clustering. Revealing that the choice of the used number of levels is inconclusive. In conclusion, our results clearly illustrate the need for global as well as local shape analyses.

Towards the future, the proposed TBM-based segmentation can be applied to other (Jacobian derived) shape analyses. Statistical testing within the clusters can be contrasted with voxel-based t-tests [4, 17] for group differences between healthy controls and Alzheimer's patients. Another application is the prediction of the AD vs. healthy prognoses or the risk of converting to AD (whole brain-based: [5, 6, 17]). Furthermore, other shape features can be derived from the Jacobians within the clusters (e.g. PCA or volumes) instead of the latent variable of the PLSR [5, 17].

5. REFERENCES

- [1] J. Ashburner, and K. Friston, *Chapter 6: Morphometry*, Human Brain Function (Second Edition), Academic Press, 21 pag., 26 jan. 2004.
- [2] K.J. Friston, A.P. Holmes, K.J. Worsley, et al., “Statistical parametric maps in functional imaging: A general linear approach”, *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.
- [3] A.P. Holmes, “Statistical Issues in Functional Brain Mapping”, PhD thesis, University of Glasgow, dec. 1994.
- [4] X. Hua, A.D. Leow, S. Lee, et al., “3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry”, *NeuroImage*, vol. 41, pp. 19-34, 2008.
- [5] X. Tang, D. Holland, A.M. Dale, et al., “Baseline Shape Diffeomorphometry Patterns of Subcortical and Ventricular Structures in Predicting Conversion of Mild Cognitive Impairment to Alzheimer's Disease”, *J Alzheimers Dis*, vol. 44(2), pp. 599–611, 2015.
- [6] N. Singh, P.T. Fletcher, J.S. Preston, et al., “Quantifying anatomical shape variations in neurological disorders”, *Medical Image Analysis*, vol. 18, pp. 616-633, 2014.
- [7] J. Ashburner, “A fast diffeomorphic image registration algorithm”, *NeuroImage*, vol. 38, pp. 95-113, 2007.
- [8] U. von Luxburg, “A tutorial on spectral clustering”. *Statistics and Computing*, vol. 17(4), pp. 395-416, 2007.
- [9] S.P. Lloyd, “Least squares quantization in PCM”. *IEEE Trans Inf Theory*, vol. 28(2), pp. 129-137, march 1982.
- [10] A. Ng, M. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm”, *Advances in neural information processing systems*, vol. 2, pp. 849-856, 2002.
- [11] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, “Spectral Grouping Using the Nyström Method”, *IEEE Trans Pattern Anal Mach Intell*, vol. 26(2), pp. 888-905, febr. 2004.
- [12] H.G. Ayad, and M.S. Kame, “Cumulative voting consensus method for partitions with variable number of clusters”, *IEEE Trans Pattern Anal Mach Intell*, vol. 30(1), pp. 160-173, 2008.
- [13] D.H. Alexander, J. Novembre, and K. Lange, “Fast model-based estimation of ancestry in unrelated individuals”, *Genome Research*, vol. 19, pp. 1655-1664, 2009.
- [14] C.J. Galton, K. Patterson, K.S. Graham, et al., “Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia”, *Neurology*, vol. 57, pp. 216–225, 2001.
- [15] L.W. de Jong, K. van der Hiele, I.M. Veer, et al., “Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study”, *Brain*, vol. 131, pp. 3277-3285, 2008.
- [16] L. Ferrarini L, W.M Palm, H. Olofsen, et al., “Shape differences of the brain ventricles in Alzheimer's disease”, *NeuroImage*, vol. 32, pp. 1060-1069, 2006.
- [17] R. Wolz, V. Julkunen, J. Koikkalainen, et al. “Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease”, *PLoS One*, vol. 6 (10), e25446, 2011.