Evaluation of the Effect of Doubling Atlases using Midsagittal Plane on Multi-Atlas based Segmentation of Brain Structures

Subrahmanyam Gorthi

Abstract— **Normal human brain exhibits approximately bifold symmetry with respect to its midsagittal plane (MSP). The objective of this work is to investigate the effect of doubling atlases (i.e., reference images) used in multi-atlas fusion methods by exploiting the inherent bilateral symmetry of human brain. To this end, we perform automated segmentation of 15 subcortical structures using Local Weighted Voting (LWV) fusion method with varying number of atlases. We consider three specific scenarios for atlases while performing fusion: (i) fusion with original OASIS atlases, (ii) with atlases obtained by flipping the original atlases based on their MSP, and (iii) with both original and flipped atlases. Evaluations are performed on the publicly available OASIS dataset of 20 normal human brain MR images. One of the key findings of this study is that when the number of atlases available for fusion is less than 10, fusion by combining both the original and flipped atlases provided more accurate segmentations than using only the original atlases, or only the flipped atlases.**

I. Introduction

Atlas (i.e., reference image) based methods are widely used in medical imaging for automated segmentation of anatomical structures. Multiple atlases based segmentation methods (i.e., atlas fusion methods) are empirically proven to provide more accurate segmentations than single-atlas based methods [1], [2]. One of the main reasons for such improved accuracy is because multiple atlases could model and account for a wider anatomical variability of structures than single atlas based methods. Another key reason is that segmentation errors associated with the propagation of segmentations from individual atlases are averaged out while combining results from multiple atlases.

However, creation of large number of atlases could be a very challenging task for many applications. For example, the average expert-time required for performing manual labeling of brain structures from scratch on an image of 1 mm^3 resolution is about 2-3 days [3]. Another common problem encountered in many applications is that it is difficult to get appropriate datasets along with the consent from participants to use those datasets in the preparation of atlases, and thereby limiting the number of atlases.

A common hypothesis in case of normal human brain is that it exhibits approximately bilateral symmetry with respect to its midsagittal plane (MSP). A canonical correlation analysis of brain structures presented in [4] also shows a strong correlation between each pair of corresponding structures in the left and right hemispheres. Symmetry analysis of human brain with respect to the MSP has been used in the literature for detection and segmentation of brain pathologies like brain tumors [5]. In view of the aforementioned difficulties in the creation of multiple atlases, in this work, we double the existing atlases by exploiting the inherent symmetry, and evaluate its impact on the accuracy of automated segmentation of brain structures.

There are only few works in the literature ([6], [7], [8]) where the existing atlases are flipped and doubled based on MSP. However, no evaluations have been reported so far that analyze the effect of doubling atlases on segmentation accuracy *with varying number of atlases*. In addition to a detailed evaluation with varying number of atlases, in this work, we also present a detailed evaluation of segmentation accuracy while using (i) only the original atlases, (ii) only the flipped atlases, and (iii) both the original and flipped atlases.

The author is with Samsung R&D Institute, Bangalore, India. Email: s.gorthi@samsung.com.

II. Methods

In this section, we first present the details of the method we use to automatically extract the MSP, and then the multi-atlas fusion method we use to merge segmentation results obtained from multiple atlases.

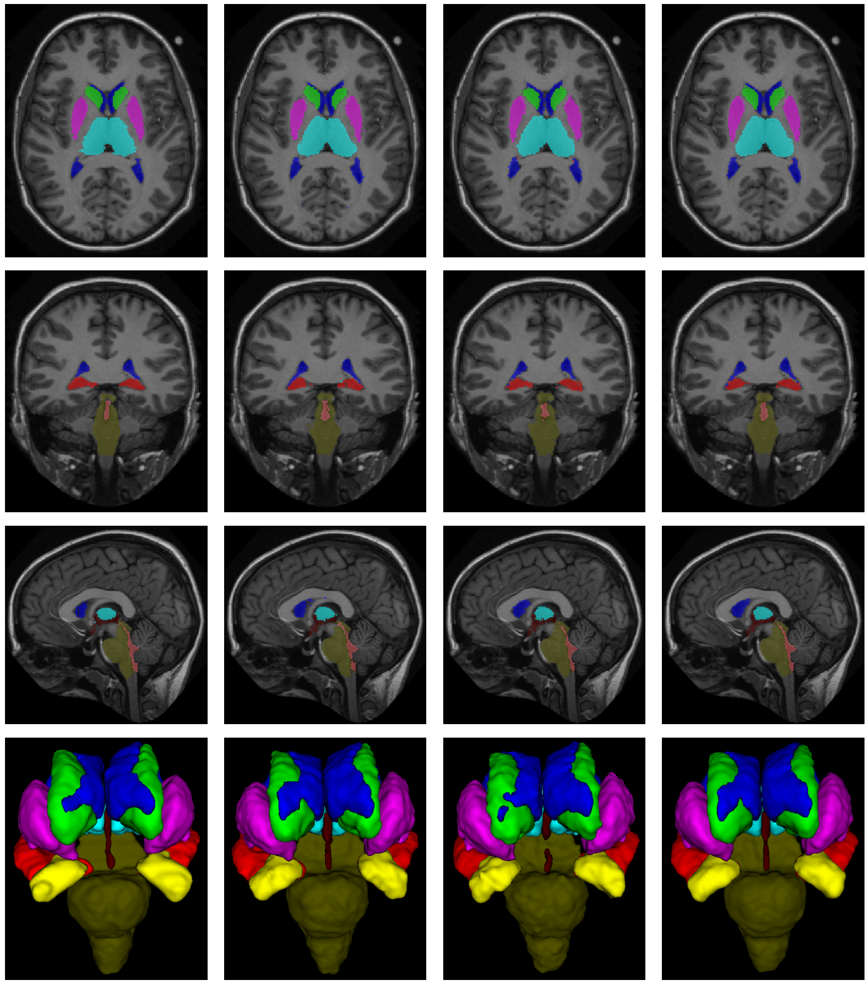
A. MSP Extraction Method

The algorithm presented in [9] is used in this paper to automatically extract the MSP. It is a fast and accurate method that finds out the MSP based on bilateral symmetry maximization.

The key steps involved in the algorithm of [9] are as follows. Bilateral symmetry measure is based on edge features extracted from image. A 3D Sobel edge detector is first applied on the input image, followed by thresholding. Note that applying Sobel edge detector is a fast and effective way to enhance edges, and it also removes high frequency noise, improving the robustness of the method. It was observed [9] that 4% of the brightest voxels in the enhanced image are good enough to represent edge features, and hence a binary image is created by applying the corresponding threshold on the output of the Sobel edge detector.

Given a candidate plane, the degree of symmetry of the input neuro-image with respect to that plane is evaluated by computing the correlation between the aforementioned binary image, and its flipped copy with respect to the candidate plane. A plane can be uniquely represented by a set of three points in the corners of the image. By varying these points, it is possible to represent any sagittal plane. Hence, these three points are systematically varied in a 3-stage multi-scale approach, such that the algorithm converges to a

978-1-4577-0220-4/16/$31.00 ©2016 IEEE



(a) Ground Truth (b) Original Atlases (c) Flipped Atlases (d) Original+Flipped

Fig. 1: Screen-shot of segmentation results for 15 brain structures for one of the images in the OASIS dataset. Number of atlases used is 4 for this result. Ground truth segmentations are shown in column-(a). Segmentation results obtained based on original atlases, flipped atlases, and combined atlases are shown in columns (b), (c) and (d) respectively.

solution (i.e., sagittal plane) that provides the best correlation between the original and flipped binary images.

B. Fusion Method

Local Weighted Voting (LWV) based fusion is used in this work as it is one of the best and widely used methods [1], [10]. LWV method quantifies *locally* how similar each atlas is to the target image to be segmented, and accordingly weighs labels propagated from each atlas. It then selects the label that has higher cumulative weight.

Let N be the total number of atlases. Let X_p^j and w_p^j respectively represent the propagated label and similarity weight corresponding to the j^{th} atlas at the p^{th} voxel. Let δ be a Kroneker delta function. Let l represent any of the possible labels that can be assigned to the target image. Let Y_p be the output label assigned by LWV method to the p^{th} voxel in the target image, and it can be computed using the following equation:

$$Y_p = \arg\max_{l} \sum_{j=1}^{N} w_p^j \; \delta\left(X_p^j, l\right). \tag{1}$$

Similarity weights (i.e., w_p^j) are computed in this work using Normalized Cross Correlation (NCC). All NCC values are rescaled voxel-wise such that the minimum and maximum values of w_p^j are equal to 0 and 1 respectively. Furthermore, in order to save the computational time, instead of computing NCC over the entire image, we a create a mask that contains only those voxels where there is a disagreement between at least two atlases regarding the output label to assigned, and we compute NCC values only at those voxels.

III. Results

Evaluations are performed on the publicly available OASIS dataset [11] of 20 normal human brain MR images. Fifteen brain structures are considered for the evaluation.

TABLE I: Average Dice Similarity Metric (DSM) values for segmentation of 15 subcortical structures, and a dataset of 20 OASIS images. Results with (i) original, (ii) flipped, and (iii) combined atlases are presented while using 3, 10 and 19 atlases.

Label No.	Structure Name	DSM - 3 Atlases			DSM - 10 Atlases			DSM - 19 Atlases		
		Original	Flipped	Combined	Original	Flipped	Combined	Original	Flipped	Combined
1	L-Hippocampus	79.43%	79.08%	**81.60%**	83.90%	83.09%	**84.02%**	84.07%	83.72%	**84.12%**
2	R-Hippocampus	80.20%	78.98%	**81.44%**	83.68%	83.12%	**83.77%**	**84.04%**	83.28%	83.92%
3	L-Lateral Ventricle	83.54%	82.75%	**85.43%**	87.53%	86.97%	**87.77%**	87.79%	87.41%	**87.93%**
4	R-Lateral Ventricle	82.27%	81.88%	**84.25%**	**86.70%**	85.89%	86.91%	**87.05%**	86.30%	86.93%
5	L-Caudate	83.02%	81.99%	**84.31%**	87.11%	86.16%	**87.24%**	**87.15%**	85.78%	86.95%
6	R-Caudate	82.44%	82.08%	**84.11%**	87.16%	86.64%	**87.48%**	86.96%	86.64%	**87.16%**
7	L-Amygdala	70.72%	70.27%	**73.22%**	77.15%	76.92%	**77.67%**	77.71%	77.20%	**77.80%**
8	R-Amygdala	70.84%	69.04%	**73.25%**	77.74%	76.97%	**78.13%**	78.52%	77.98%	**78.58%**
9	L-Putamen	85.81%	85.42%	**86.85%**	89.73%	89.32%	**89.88%**	**89.68%**	89.24%	89.58%
10	R-Putamen	85.58%	84.96%	**86.64%**	89.44%	89.04%	**89.51%**	**89.47%**	89.15%	89.43%
11	L-Thalamus	87.17%	86.42%	**88.09%**	**90.32%**	88.58%	89.91%	**90.68%**	89.13%	90.29%
12	R-Thalamus	87.09%	85.95%	**87.83%**	89.77%	89.07%	**89.91%**	**90.08%**	89.26%	90.01%
13	Brain Stem	88.39%	87.92%	**89.26%**	91.53%	91.38%	**91.66%**	**91.67%**	91.57%	91.65%
14	3rd Ventricle	69.00%	68.40%	**73.27%**	77.58%	76.74%	**78.39%**	78.57%	78.11%	**78.86%**
15	4th Ventricle	77.12%	76.03%	**79.69%**	82.67%	82.31%	**83.22%**	83.11%	82.85%	**83.19%**
	AVERAGE	80.84%	80.08%	**82.62%**	85.47%	84.81%	**85.70%**	**85.77%**	85.18%	85.76%
	Standard Deviation	6%	6%	6%	5%	5%	5%	5%	5%	5%

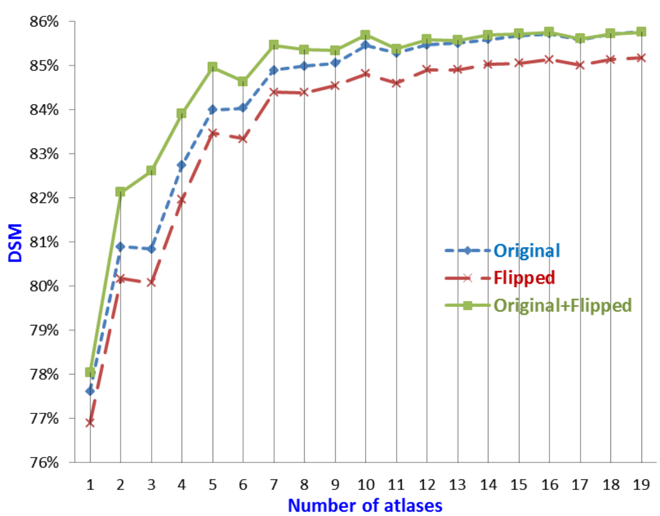


Fig. 2: Overall Dice Similarity Metric (DSM) values for all 3 approaches, with varying number of atlases. DSM presented here are values averaged across all 15 subcortical structures and a dataset of 20 OASIS images.

The details of structure names and their corresponding label numbers are presented in the first two columns of Table I.

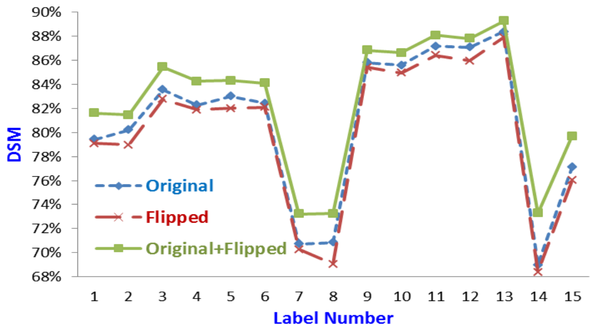
All 20 images in the dataset are first flipped based on their MSP, as described in Section II-A. Each of the original 20 images are registered to the remaining 19 images and their corresponding flipped images in a leave-one-out manner. In other words, each of the 20 images are registered to 38 images, and thus, in total, 760 registrations are performed. Each registration in turn is performed in 3 levels with increasing degrees of freedom, by first performing rigid registration, and it is then followed by affine and diffeomorphic Demons registrations [12] respectively.

While studying the effect of varying the number of atlases, notice that R atlases from a dataset of N images can be selected in ${}^{N}P_{R}$ ways. For example, 10 atlases from a dataset of 19 images can be selected in (${}^{19}P_{10} =$) 92,378 ways. Since evaluation of so many combinations is not practically feasible, in this work, we randomly select the required number of atlases from the entire dataset, and then repeat the whole evaluation 5 times in order to avoid any bias in the random atlas selection process. All those repeated evaluations have shown similar trends in accuracy, and resulted in same conclusions. Hence, we present here results from just one of those five evaluations.

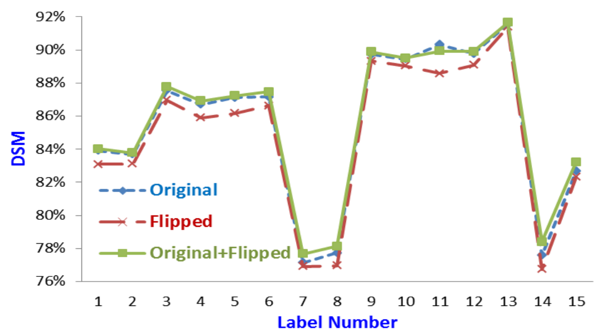
Multi-atlas segmentation results with varying number of atlases is evaluated particularly in 3 specific scenarios: (i) fusion based on atlases from the original OASIS dataset, (ii) fusion based on flipped versions of atlases used in the first scenario, and (iii) fusion based on both original and flipped atlases used in the preceding two scenarios. Notice that when we say the number of atlases as 'R', it means that 'R' original and their corresponding 'R' flipped atlases are used in scenarios (i) and (ii) respectively, while both original and flipped atlases (i.e., $2R$ atlases in total) are used in scenario-(iii). Quantitative evaluations are performed using Dice Similarity Metric (DSM). DSM is a commonly used statistical metric that provides a measure average percentage of overlap between ground truth and automated segmentation results.

Fig. 1 shows segmentation results for one of the images when using 4 atlases. Ground truth and automated segmentation results from all 3 approaches, in one of the axial, coronal and sagittal slices are presented in this figure along with the volumes of each structure. From these images, improvements from combined approach (with doubling of atlases) can be visually observed for some of the structures. However, number of slices presented here are not enough to visually notice the differences for all structures, and also with varying number of atlases.

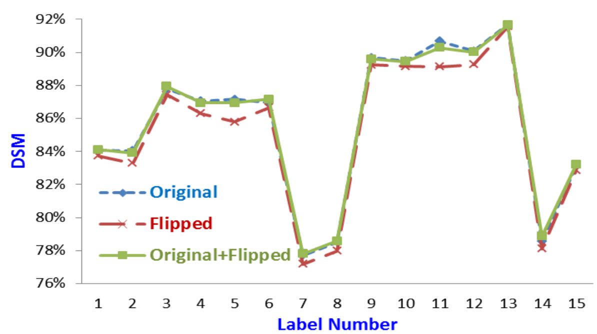
Fig. 2 presents overall DSM values computed across all



(a) 3-Atlases



(b) 10-Atlases



(c) 19-Atlases

Fig. 3: Average Dice Similarity Metric (DSM) values for 15 structures with (i) original, (ii) flipped, and (iii) combined atlases, while using 3, 10 and 19 atlases. Structure names corresponding to these label numbers are mentioned in Table I.

15 structures and 20 images, with varying number of atlases. It can be noticed from this graph that when the number of original atlases is less (< 5), improvements due to doubling of atlases are very significant compared to the remaining 2 approaches. One can also notice that DSM values are typically improving with the increasing number of atlases, and reaching a saturation point when the number of atlases is around 10. Another interesting observation here is that results when using only the flipped atlases are consistently inferior to the results from original atlases; it requires further investigations and inputs from clinical experts for finding out the exact reason for this behavior.

Fig. 3 presents average structure-wise DSM values while using 3, 10 and 19 atlases. These three values of atlases respectively represent low, moderate and, using all the available atlases. Table I presents the exact DSM values plotted in Fig. 3, with the best DSM values among the three approaches for a given number of atlases marked in bold for easy reference. It can be noticed from these results that the trend of DSM values observed at individual structure level is very similar to the overall DSM trend observed in Fig. 2.

IV. Conclusions

In this paper, we have studied the effect of doubling the number of atlases in multi-atlas segmentation of MR brain images by exploiting the bilateral symmetry of the human brain. We are particularly interested in performing this study with varying number of atlases as that can shed light on when to include (or when there is no need to include) flipped atlases during fusion. Evaluations are performed for segmentation of 15 subcortical structures in the publicly available OASIS dataset of 20 normal brain MR images. It is found that when the number of atlases is around 10 or more, there is no significant improvement in accuracy with doubling of atlases. On the other hand, if the number of available atlases is less (e.g., $<= 5$), there is a significant improvement in segmentation accuracy with doubling of atlases.

References

[1] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solórzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1266–1277, 2009.

[2] S. Gorthi, A. Akhondi-Asl, and S. K. Warfield, "Optimal MAP parameters estimation in STAPLE using local intensity similarity information," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1589–1597, 2015.

[3] A. Klein and J. Tourville, "101 labeled brain images and a consistent human cortical labeling protocol," *Frontiers in Neuroscience*, vol. 6, no. 171, pp. 10–3389, 2012.

[4] A. Rao, K. Babalola, and D. Rueckert, "Canonical correlation analysis of sub-cortical brain structures using non-rigid registration," in *Biomedical image registration*. Springer, 2006, pp. 66–74.

[5] S. X. Liu, "Symmetry and asymmetry analysis and its implications to computer-aided diagnosis: A review of the literature," *Journal of Biomedical Informatics*, vol. 42, no. 6, pp. 1056–1064, 2009.

[6] M. J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N. C. Fox, S. Ourselin, A. D. N. Initiative, *et al.*, "STEPS: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation," *Medical Image Analysis*, vol. 17, no. 6, pp. 671–684, 2013.

[7] K. K. Leung, J. Barnes, G. R. Ridgway, J. W. Bartlett, M. J. Clarkson, K. Macdonald, N. Schuff, N. C. Fox, S. Ourselin, A. D. N. Initiative, *et al.*, "Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and alzheimer's disease," *Neuroimage*, vol. 51, no. 4, pp. 1345–1359, 2010.

[8] D. Ma, M. J. Cardoso, M. Modat, N. Powell, J. Wells, H. Holmes, F. Wiseman, V. Tybulewicz, E. Fisher, M. F. Lythgoe, *et al.*, "Automatic structural parcellation of mouse brain mri using multi-atlas label fusion," *PloS one*, vol. 9, no. 1, p. e86576, 2014.

[9] G. Ruppert, L. Teverovskiy, C.-P. Yu, A. X. Falcao, and Y. Liu, "A new symmetry-based method for mid-sagittal plane extraction in neuroimages," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2011, pp. 285–288.

[10] S. Gorthi, M. Bach Cuadra, P. Tercier, A. S. Allal, and J.-P. Thiran, "Weighted shape-based averaging with neighborhood prior model for multiple atlas fusion-based medical image segmentation," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1036–1039, 2013.

[11] http://www.oasis-brains.org/.

[12] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.