Atlas selection strategy in multi-atlas segmentation propagation with locally weighted voting using diversity-based MMR re-ranking

Kaikai Shen[a,b], Pierrick Bourgeat[a], Fabrice Meriaudeau[b], Olivier Salvado[a] and the Alzheimer's Disease Neuroimaging Initiative*

[a]Australian e-Health Research Centre, ICT Centre, CSIRO, Herston, Australia;
[b]LE2I, CNRS UMR 5158, Université de Bourgogne, Le Creusot, France

ABSTRACT

In multi-atlas based image segmentation, multiple atlases with label maps are propagated to the query image, and fused into the segmentation result. Voting rule is commonly used classifier fusion method to produce the consensus map. Local weighted voting (LWV) is another method which combines the propagated atlases weighted by local image similarity. When LWV is used, we found that the segmentation accuracy converges slower comparing to simple voting rule. We therefore propose to introduce diversity in addition to image similarity by using Maximal Marginal Relevance (MMR) criteria as a more efficient way to rank and select atlases. We test the MMR re-ranking on a hippocampal atlas set of 138 normal control (NC) subjects and another set of 99 Alzheimer's disease patients provided by ADNI. The result shows that MMR re-ranking performed better than similarity based atlas selection when same number of atlases were selected.

Keywords: M RI, image segmentation, multi-atlas segmentation propagation, MMR, atlas selection, locally weighted voting

1. INTRODUCTION

Mutli-atlas based segmentation propagation is a segmentation approach which uses multiple atlases with the anatomical structures of interest delineated. It has been shown to be a successful approach to the segmentation of subcortical structures in brain MR images.[1,2] In this method, the atlases are registered to the image to be segmented producing a mapping from the coordinates in the atlas space to the target. The label maps of the atlases are thus propagated to the target image. The labeling of each voxel in the target image is produced by fusing the warped label maps of the atlases using a voting rule.

The relation between the segmentation accuracy and the increasing number of atlases fused has been studied. The segmentation accuracy measured by the Dice similarity coefficient (DSC) converges when more atlases are added. The convergence can be modeled as

$$\mathrm{DSC} \sim a - \frac{b}{\sqrt{n}}, \tag{1}$$

where a is the limit of the overlap between the ground truth and the segmentation result as the number of atlases fused n increases, and b controls the convergence rate.[3,4] In order to achieve better segmentation results, atlases are usually ranked and selected according to their closeness to the target image. The estimate of convergence (1) does not hold when the atlases are ranked according to the registration quality. The accuracy of multi-atlas segmentation reaches its peak when approximately the first 10 to 20 similarity ranked atlases are selected, and combined by the voting rule. When less similar atlases are fused to the segmentation result, the errors propagated to the result outweigh the information relevant to the segmentation of the target.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`www.loni.ucla.edu/ADNI`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf`

Medical Imaging 2011: Image Processing, edited by Benoit M. Dawant, David R. Haynor,
Proc. of SPIE Vol. 7962, 79621B · © 2011 SPIE · CCC code: 1605-7422/11/$18 · doi: 10.1117/12.877866

Locally based atlas selection strategies[5–7] were developed to reduce the error in the lower ranked atlases by selecting the regions or voxels locally more similar to the target image. The local weighted voting (LWV[8]) method is an alternative approach utilizing the intensity information of the target and the atlas images in the fusion step. It weights each atlas in the voting based on their local similarity to the target. We found that with the use of LWV the segmentation accuracy on the ranked atlases does not converge when more than 20 ranked atlases were fused, and keeps increasing as more atlases are added (see for instance Figure 2). Since LWV requires both the image and the label map of the atlases, it is computationally expensive in terms of both the computation time and the memory footprint, if we increase the number of atlases until the segmentation accuracy converges.

Therefore, the aim of this paper is to propose a more efficient atlas ranking and selection strategy for the multi-atlas segmentation of hippocampus in brain MR images using LWV label fusion. We estimate the DSC based on the image similarity metrics and use the estimated DSC as measurement of similarity between the atlas and the target image. We re-rank the registration results according to the Maximal Marginal Relevance (MMR[9]) criterion to select the atlases to be fused. The re-ranking method is tested on two hippocampal atlas sets, one is consists of normal control (NC), and the other of Alzheimer's disease (AD).

2. MATERIALS AND METHODS

2.1 Materials

The hippocampus segmentations used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`www.loni.ucla.edu/ADNI`). The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research – approximately 200 cognitively normal elderly individuals 400 subjects with MCI, and 200 people with early AD. We use the hippocampal volumes segmented semi-automatically by SNT provided by ADNI. SNT hippocampal volumetry has been previously validated on the normal aging, MCI and AD subjects.[10] It first uses 22 control points manually placed on the individual brain MRI as local landmarks. Fluid image transformation is then used to match the individual brains to a template brain.[11] The segmentations were manually edited by qualified reviewers if the boundaries delineated by SNT were not accurate. The hippocampal volumes consist of 138 normal control (NC) subjects with average age 76.6(5.0) years old and 99 AD patients with average age 76.2(7.3) years old.

2.2 Multi-atlas segmentation propagation

In multi-atlas based segmentation, we use an atlas set $\{(I_k, L_k) : k = 1, 2, \cdots, n\}$, in which each atlas image I_k is labeled with known segmentation L_k, to segment the query image I. Each atlas image is first registered to I individually, producing the transformation $\mathcal{T}_k$ mapping the atlas I_k to the target image. The same transformation can be applied to L_k such that I can be segmented with the transformed label map $L_k \circ \mathcal{T}_k$. Using multiple atlases, the result segmentation L for the image I is obtained by combining the transformed label maps $\{L_k \circ \mathcal{T}_k : k = 1, 2, \cdots, n\}$.

Given the transformed atlas set $\{(I_k \circ \mathcal{T}_k, L_k \circ \mathcal{T}_k)\}$, the label of each voxel in the query image is determined by the consensus labeling on transformed segmentations. The voting rule is a simple but robust method to produce the consensus segmentation. In this paper, we use LWV instead of the simple voting rule in the fusion step to produce the consensus. With LWV, the vote from each atlas is weighted by the local similarity between the transformed atlas $I_k \circ \mathcal{T}_k$ and the image I within the neighborhood of the voxel to be labeled. Inverse of mean squared difference (MSD) measuring the image similarity was shown[8] to be the a good choice of weighting coefficient and is used in this paper. Since the local image similarity weighting suppresses the contribution from the regions with misalignment, adding atlases that are globally less well registered may still contribute positively to the segmentation details in specific regions where these atlases are better aligned.

2.3 Similarity based atlas ranking

The mis-alignment in the transformed atlases due to the registration error is propagated to the segmentation result, which can be reduced by selecting the atlases according to the registration accuracy measured by image similarity. In the similarity based atlas selection, the atlases are ranked by the image similarity $\text{sim}(I_k \circ \mathcal{T}_k, I)$

between the image I and the registration result $I_k \circ \mathcal{T}_k$. The atlases best registered to the target I are selected and combined in the fusion step. In practice, normalized mutual information (NMI[12]), cross correlation (CC), and mean reciprocal square difference (MRSD) are the common candidates for the image similarity metric $\mathrm{sim}(\cdot,\cdot)$.

Since the segmentation accuracy is directly determined by the registration result in atlas based segmentation approach, we can estimate the DSC for the structure of interest based on similarity measurements such as NMI, CC and MRSD via a linear model

$$\mathrm{DSC} = \beta_0 + \beta_1 \mathrm{NMI} + \beta_2 \mathrm{CC} + \beta_3 \mathrm{MRSD} + \varepsilon, \tag{2}$$

where $\beta_m (m = 0, 1, 2, 3)$ are the coefficients of the model. In order to determine the coefficients β_m, we register each atlas $\{(I_k, L_k)\}$ to the rest in the atlas set. The segmentation accuracy

$$\mathrm{DSC}_{ij} = \mathrm{DSC}(L_i \circ \mathcal{T}_{j\to i}, L_j) \quad i \neq j \tag{3}$$

can be evaluated on the warped label map, and the image similarity measurements can be calculated on the pair $(I_i \circ \mathcal{T}_{j\to i}, I_j)$

$$\mathrm{NMI}_{ij} = \mathrm{NMI}(I_i \circ \mathcal{T}_{j\to i}, I_j) \tag{4}$$
$$\mathrm{CC}_{ij} = \mathrm{CC}(I_i \circ \mathcal{T}_{j\to i}, I_j) \tag{5}$$
$$\mathrm{MRSD}_{ij} = \mathrm{MRSD}(I_i \circ \mathcal{T}_{j\to i}, I_j) \tag{6}$$

Once the coefficients $\{\beta_m\}$ are fixed by fitting the linear model (2) to the observed data $\{(\mathrm{DSC}_{ij}, \mathrm{NMI}_{ij}, \mathrm{CC}_{ij}, \mathrm{MRSD}_{ij})\}$, we can use the predicted $\widehat{\mathrm{DSC}}$ as the new similarity measurement between an atlas and a target image. With this new measurement, the similarity between the atlas I_k and the target image I can be defined as

$$\mathrm{Sim}_A(I_k, I) = \widehat{\mathrm{DSC}} = \widehat{\beta}_0 + \widehat{\beta}_1 \mathrm{NMI}(I_k \circ \mathcal{T}_k, I) + \widehat{\beta}_2 \mathrm{CC}(I_k \circ \mathcal{T}_k, I) + \widehat{\beta}_2 \mathrm{MRSD}(I_k \circ \mathcal{T}_k, I). \tag{7}$$

Since this Sim_A is asymmetric, we can symmetrize it to make it a symmetric metric Sim_S between two atlases

$$\mathrm{Sim}_S(I_i, I_j) = \frac{1}{2}\left(\mathrm{Sim}_A(I_i, I_j) + \mathrm{Sim}_A(I_j, I_i)\right), \tag{8}$$

in which the atlas images I_i and I_j are registered to each other.

2.4 Re-ranking atlases using MMR

MMR criterion was introduced in the field of information retrieval to reduce the redundancy in document summarization by taking diversity into consideration, while still maintaining the relevance to the query.[9] Since MMR uses only the similarity measurement in the re-ranking, it can be easily translated into the context of atlas ranking and selection in multi-atlas based segmentation.

We initialize the set of selected atlases S to be empty and select one atlas each iteration. At iteration i, the atlas (I_{k_i}, L_{k_i}) is selected according to MMR, such that

$$(I_{k_i}, L_{k_i}) = \underset{(I_k, L_k)\notin S}{\arg\max} \left(\lambda \mathrm{Sim}_A(I_k, I) - (1-\lambda) \max_{(I_j, L_j)\in S} \mathrm{Sim}_S(I_j, I_k)\right), \tag{9}$$

until the selected atlas set S reaches the a threshold. The parameter $\lambda \in (0, 1]$ controls the the similarity measurement $\mathrm{Sim}_A(I_k, I)$, introduces diversity by penalizing the redundancy $\max_{(I_j, L_j)\in S} \mathrm{Sim}_S(I_j, I_k)$ within the selected set S. When $\lambda = 1$, MMR is equivalent to the similarity ranking using $\mathrm{Sim}_A(\cdot,\cdot)$.

3. RESULTS AND DISCUSSION

In the experiment, the atlases were registered first by affine transformation using a robust block matching approach[13] with 12 degrees of freedom followed by non-rigid registration using non-parametric diffeomorphic Demons algorithm,[14] which is based on Thirion's demons algorithm,[15] and transforms the atlases by diffeomorphic displacement fields. Each NC atlas was registered to all other cases in NC set, and each AD atlas was registered to all the others in the AD set. In total $138 \times 137 + 99 \times 98 = 28608$ registrations were performed, in which 235 failed. The relation between the accuracy DSC_{ij} of the transformed label map and the image similarity NMI_{ij} based on the rest successful 28373 registrations is plotted in Figure 1. The linear model (2) is fitted using the results of the cross-registrations on both atlas sets

$$\widehat{\mathrm{DSC}} = -2.43 + 4.61\mathrm{NMI} + 2.99\mathrm{CC} - 4.48\mathrm{MRSD}, \tag{10}$$

and the estimated $\widehat{\mathrm{DSC}}$ is used as the similarity measurement Sim_A between two images.

The atlases are selected by image similarity ranking and MMR re-ranking. Both NMI and the estimated DSC are used as the asymmetric similarity measurement Sim_A between image and atlas for MMR.

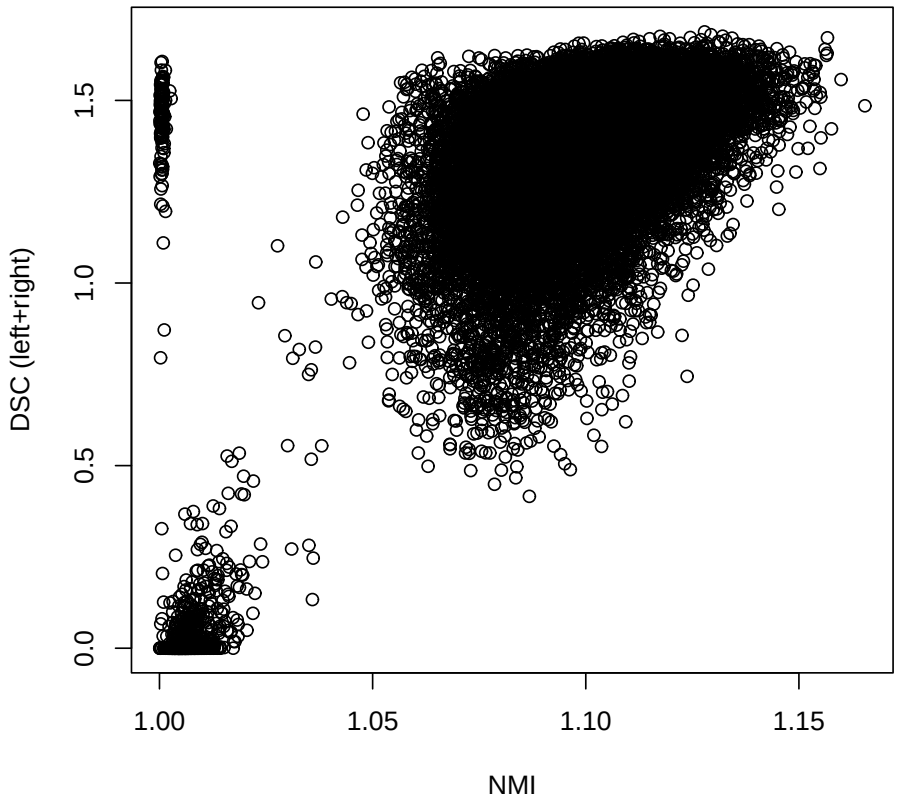
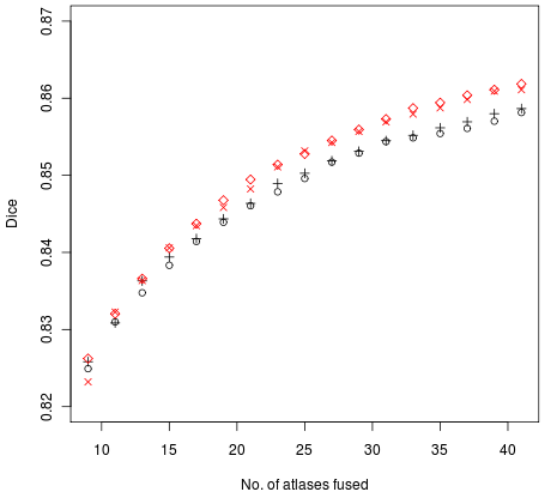


Figure 1. Relation between the accuracy DSC_{ij} and the image similarity NMI_{ij}, based on 28373 registrations.

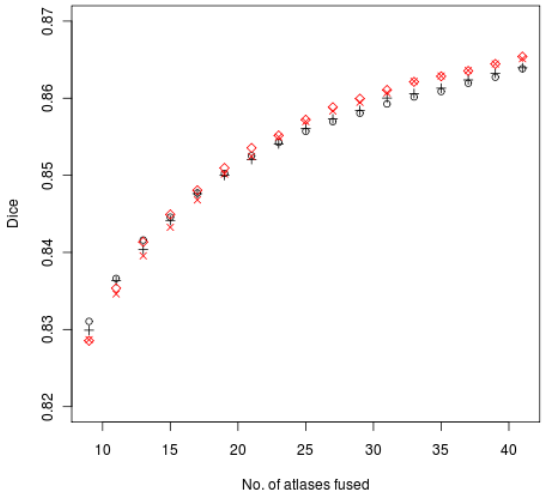
In the experiment, the parameter λ controlling similarity and diversity is set to 0.5. The relation between the segmentation accuracy in terms of DSC and the number of atlases fused using different atlas selection strategies are shown in Figure 2.

Re-ranking atlases by MMR criteria improves the segmentation accuracy as compared to the atlases selected based only on image similarity. The behavior curves of the segmentation accuracy against the number of atlases

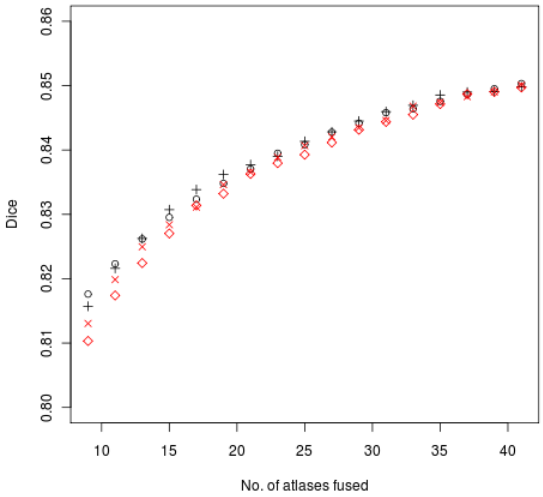
fused are similar when different similarity measurements are used. Introducing diversity to re-rank the atlas brought more impact upon the result than using a better estimation of image similarity. In order to achieve the same segmentation accuracy, more atlases were generally required if they are selected by similarity ranking. Reducing the redundancy redundancy within the atlas set by MMR is therefore preferable to the optimization of the ranking based solely on similarity.



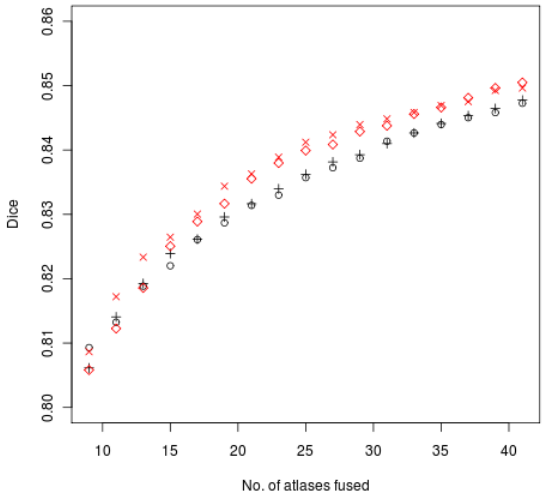
(a) Left hippocampus, NC



(b) Right hippocampus, NC



(c) Left hippocampus, AD



(d) Right hippocampus, AD

Figure 2. Cross-validation results on AD atlas set. ○: NMI; +: $\widehat{\text{DSC}}$; ×: MMR re-ranking with NMI as similarity metric; ◇: MMR with $\widehat{\text{DSC}}$ as similarity metric.

4. CONCLUSION

LWV label fusion gives better accuracy by utilizing the local image similarity information, which makes it possible to improve the segmentation quality by increasing the number of atlases. It is necessary to select the registration

results when the number of atlas to be fused is limited. As a combination optimization problem, selection by exhaustive search is not tractable. Simple similarity ranking selects the atlases most close to the query image while does not consider the inter-atlas redundancy. Selecting atlases re-ranked according to MMR criteria is more efficient compared to image similarity selection when labels are fused by LWV. MMR re-ranking provides more accurate results when the same number of atlases are selected and fused. This method is advantageous when the number the atlases to be fused is limited by the computation time, memory constraint and/or the size of atlas set.

REFERENCES

[1] Rohlfing, T., Brandt, R., et al., "Quo vadis, atlas-based segmentation?," in [*Handbook of Biomedical Image Analysis*], Micheli-Tzanakou, E., Suri, J. S., et al., eds., 435–486, Springer (2005).

[2] Babalola, K., Patenaude, B., et al., "Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI," in [*MICCAI 2008*], *LNCS* **5241**, 409–416 (2008).

[3] Heckemann, R. A., Hajnal, J. V., et al., "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage* **33**, 115–126 (Oct. 2006).

[4] Aljabar, P., Heckemann, R., et al., "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage* **46**, 726–738 (July 2009).

[5] Commowick, O., Warfield, S. K., and Malandain, G., "Using Frankenstein's creature paradigm to build a patient specific atlas," in [*MICCAI 2009*], *LNCS* **5762**, 993–1000 (2009).

[6] van Rikxoort, E. M., Isgum, I., et al., "Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus," *Medical Image Analysis* **14**, 39–49 (Feb. 2010).

[7] Coupé, P., Manjón, J. V., et al., "Nonlocal patch-based label fusion for hippocampus segmentation," in [*MICCAI 2010*], *LNCS* **6363**, 129–136, Springer (2010).

[8] Artaechevarria, X., Munoz-Barrutia, A., and Ortiz-de-Solorzano, C., "Combination strategies in Multi-Atlas image segmentation: Application to brain MR data," *Medical Imaging, IEEE Transactions on* **28**(8), 1266–1277 (2009).

[9] Carbonell, J. and Goldstein, J., "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in [*SIGIR '98*], 335–336, ACM, Melbourne, Australia (1998).

[10] Hsu, Y., Schuff, N., Du, A., Mark, K., Zhu, X., Hardin, D., and Weiner, M. W., "Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia," *Journal of Magnetic Resonance Imaging* **16**(3), 305–310 (2002).

[11] Christensen, G., Joshi, S., and Miller, M., "Volumetric transformation of brain anatomy," *Medical Imaging, IEEE Transactions on* **16**(6), 864–877 (1997).

[12] Studholme, C., Hill, D., and Hawkes, D., "Incorporating connected region labelling into automated image registration using mutual information," in [*Mathematical Methods in Biomedical Image Analysis, 1996., Proceedings of the Workshop on*], 23–31 (1996).

[13] Ourselin, S., Roche, A., et al., "Reconstructing a 3D structure from serial histological sections," *Image and Vision Computing* **19**(1-2), 25–31 (2001).

[14] Vercauteren, T., Pennec, X., et al., "Non-parametric diffeomorphic image registration with the demons algorithm," in [*MICCAI 2007*], *LNCS* **4792**, 319–326 (2007).

[15] Thirion, J., "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis* **2**, 243–260 (Sept. 1998).