Supervised Method to Build an Atlas Database for Multi-atlas Segmentation-propagation

Kaikai Shen[a,b], Pierrick Bourgeat[a], Jurgen Fripp[a], Fabrice Meriaudeau[b], David Ames[c], Kathryn A. Ellis[e,c], Colin L. Masters[e,g], Victor L. Villemagne[d,e], Christopher C. Rowe[g], and Olivier Salvado[a]

[a]Australian e-Health Research Centre, CSIRO ICT Centre, Herston, Australia;
[b]Université de Bourgogne, Le Creusot, France;
[c]National Ageing Research Institute, Parkville, Australia;
[d]Department of Nuclear Medicine and Centre for PET, and Department of Medicine University of Melbourne, Austin Hospital, Melbourne, Australia;
[e]The Mental Health Research Institute, University of Melbourne, Parkville, Australia;
[f]CSIRO CMHT, Parkville, Australia;
[g]Centre for Neurosciences, University of Melbourne, Parkville, Australia

ABSTRACT

Multi-atlas based segmentation-propagation approaches have been shown to obtain accurate parcelation of brain structures. However, this approach requires a large number of manually delineated atlases, which are often not available. We propose a supervised method to build a population specific atlas database, using the publicly available Internet Brain Segmentation Repository (IBSR). The set of atlases grows iteratively as new atlases are added, so that its segmentation capability may be enhanced in the multi-atlas based approach. Using a dataset of 210 MR images of elderly subjects (170 elderly controls, 40 Alzheimer's disease) from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study, 40 MR images were segmented to build a population specific atlas database for the purpose of multi-atlas segmentation-propagation. The population specific atlases were used to segment the elderly population of 210 MR images, and were evaluated in terms of the agreement among the propagated labels. The agreement was measured by using the entropy H of the probability image produced when fused by voting rule and the partial moment μ_2 of the histogram. Compared with using IBSR atlases, the population specific atlases obtained a higher agreement when dealing with images of elderly subjects.

Keywords: Image segmentation, multi-atlas segmentation-propagation

1. INTRODUCTION

Magnetic resonance (MR) imaging proves to be essential in the field of neuroscience. It is an important tool and provides important insights to the study of brain. In the research into the physiology and pathology of brain, such as the studies on Alzheimer's disease (AD), the processing and the analysis of MR image data is becoming increasingly demanding. In order to extract features and biomarkers from brain MR images, the anatomical structures need to be delineated in the first place. The task of automated segmentation remains challenging, and crucial to further functional and morphological analysis.

Various methods have been proposed to parcelate brain images, including multi-atlas based classifier fusion and labelling,[1] EM-based segmentation using dynamic brain atlas,[2] and model-based methods such as profile active appearance models[3] and Bayesian appearance models.[4] These methods were evaluated and compared in a recent study,[5] and the multi-atlas based classifier fusion and labelling method was shown to outperform other methods in terms of accuracy with respect to the manual labels as ground truth. Multi-atlas based segmentation-propagation transforms the labels of an atlas image into the space of the unlabelled target image. Multiple atlases are used in order to reduce the bias towards each single atlas. The labels propagated from the atlas set are combined in a fusion step. This method takes the advantage of *a priori* knowledge encoded in the atlas segmentation. However, to achieve a good performance over a large population with significant inter-subject variability, it requires a large atlas database in which the anatomical structures of interest are reliably

Medical Imaging 2010: Computer-Aided Diagnosis, edited by Nico Karssemeijer, Ronald M. Summers,
Proc. of SPIE Vol. 7624, 76241N · © 2010 SPIE · CCC code: 1605-7422/10/$18 · doi: 10.1117/12.844048

labelled. For instance, two recent studies[1,6] involved a repository of atlases consisting of more than 270 brain MR images with manual delineation of various structures. Atlases used for the classifier fusion are selected based on image similarity to the query or demographic information. Often due to extremely demanding time and cost of expert's labelling, multi-atlas based method, as a successful approach, becomes less practical when the manual segmentation over a dataset of such size is unavailable.

In this study, we developed a method of building a database of atlases using publicly available Internet Brain Segmentation Repository (IBSR)*. Starting with the 18 atlases available in IBSR, this study aims to construct population specific atlases, for the purpose of multi-atlas based segmentation. Instead of labelling manually defined by experts, the database of atlas can be built up in a supervised manner iteratively. We apply this method to an elderly population of normal control (NC) and AD, producing an elderly specific atlas set.

2. MATERIALS AND METHOD

The image dataset used in this study consists of 210 T1-weighted MR brain images from participants enrolled in the Australian Imaging and Lifestyle Study (AIBL), 170 of which are NC, and the rest 40 are diagnosed as AD. All subjects underwent T1-weighted MRI scans. Sagittal T1-weighted MR images were acquired using a standard 3D MPRAGE sequence at 3T, with in-plane resolution 1×1mm, slice thickness 1.2mm, TR/TE/T1= 2300/2.98/900, flip angle 9° and field of view 240×256 and 160 slices. The average age of NC is 73.47 ($\sigma = 7.43$) years old, and for AD is 73.85 ($\sigma = 8.84$) years old.

2.1 Building the Atlas Database

The atlas database is a set of images well segmented, which can be used in multi-atlas based approach to segment unseen query images. Though a large database of manual segmentations may not be readily available, the set of atlases can be constructed by a supervised method. For a MR image dataset from a given population, the process to construct a population specific atlas database is listed below:

1. Initialize the atlas database with 18 segmented images in IBSR
2. Using multi-atlas segmentation propagation method, segment the image dataset with the current atlas database
3. Visually inspect the segmentation results
4. Well segmented images with high consistency between the image and labelling over structures of interest are qualified and added to the atlas database
5. Repeat the steps 2 to 4 until the size of atlas database reaches a predetermined threshold

The diagram of iterative process is illustrated in Figure 2. When performing the multi-atlas segmentation propagation in step 2, each image in the dataset was first segmented by propagating the labels from the IBSR atlases. All atlases were registered by affine transformation using a robust block matching approach[7] with 12 degrees of freedom followed by non-rigid (NR) registration using non-parametric diffeomorphic Demons algorithm,[8] which is based on Thirion's demons algorithm,[9] and transforms the atlases by diffeomorphic displacement fields. Demons algorithms were demonstrated to perform very well in aligning the medial temporal lobe regions in cross-participant registration.[10] The displacement field, from the registration algorithms warping the atlas image to the target image, maps the labels in the atlas space to the target image space by nearest neighbor interpolation.

When multiple atlases are propagated to the target image, the result segmentation is a combination of the propagated labels from the atlases. Majority vote rule and simultaneous truth and performance level estimation (STAPLE)[11] are used to combine segmentation results. In the multi-atlas based approach of brain image segmentation, the majority vote rule is more commonly used strategy. It has been shown[1,6] that the Dice similarity coefficient (DSC) score between the segmentation and the ground truth reaches the highest value when

*Available at `http://www.cma.mgh.harvard.edu/ibsr/`.

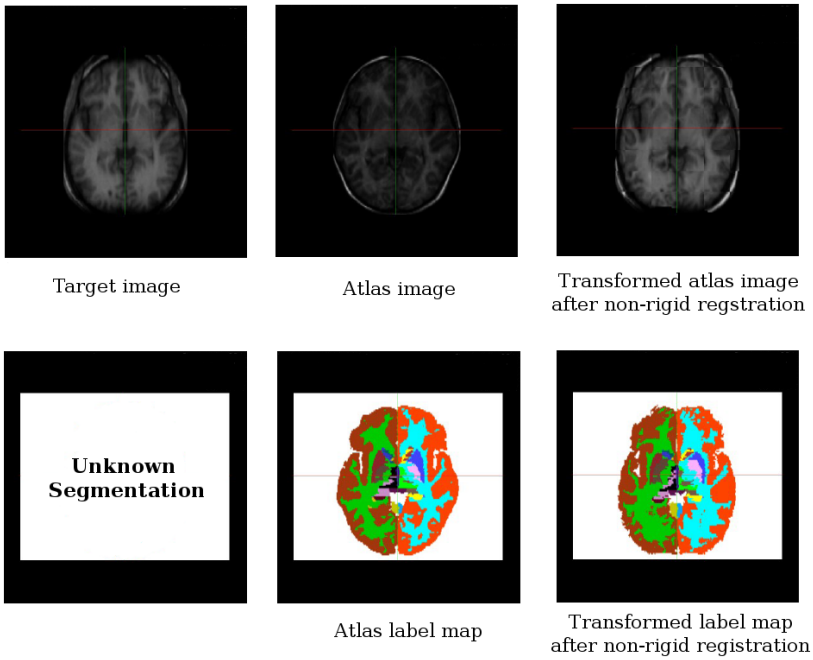

Figure 1. Propagation from an atlas to a target image. Top left: the target image; top middle: the atlas image; top right: the atlas image transformed by the non-rigid registration; bottom middle: the label map of the atlas image; bottom right: the atlas labels propagated to the target image space.

approximately 10 atlases are fused. Considering the fact that there 8 subjects in the IBSR set are under 18 years old, and to avoid tie votes, 9 atlases were selected in experiment for classifier fusion. The selection was based on image similarity in terms of normalised mutual information (NMI)[12] between the unlabelled target image and the NR registered atlases. Labels of selected atlases were fused through a voting rule. The segmentation results were visually inspected, with attention paid especially to the lateral ventricle and deep gray matter structures, such as hippocampus, thalamus, caudate, and putamen. Segmentations qualified in terms of their visual consistency between the image and the corresponding segmentation were added to the atlas database. As the size of atlas database grows, this step can be repeated so that more segmented MR images may be added to the atlas database, enhancing its capability in the multi-atlas based approach.

In this study, the atlas set was initialised with the IBSR data, in which anatomical structures are manually delineated by experts. The age of subjects in IBSR dataset ranges from juvenile to 71, including 4 juvenile subjects and another 4 under 18 years old. In the age-based atlas selection,[1] selecting atlases of subjects with similar age to the query provides a good estimation. The demographic information of atlases fused is relevant to the performance of multi-atlas based segmentation. Atlases of younger subjects in IBSR are likely to fail when being propagated to the brain images of a subject in elderly population. By adding to the atlas set well segmented brain images of subjects from elderly population, it may improve the performance in segmenting the images of elderly subjects.

2.2 Evaluation measurements

In the multi-atlas based segmentation method, the errors in the labelling of atlases may be propagated to the segmentation results. It is therefore necessary to evaluate the quality of atlas produced by the proposed method. Due to the lack of delineation of anatomical structures by experts, it is impossible to assess the accuracy of segmentation of these atlases quantitatively against the manual segmentation as ground truth. Measurements of segmentation's overlap with ground truth such as DSC score, and the boundary difference such as Hausdorff distance are thus not applicable.

In order to evaluate the atlases produced and selected as described in the previous section, the agreement among their propagated labels is taken into account. The assumption lying under this consideration is that

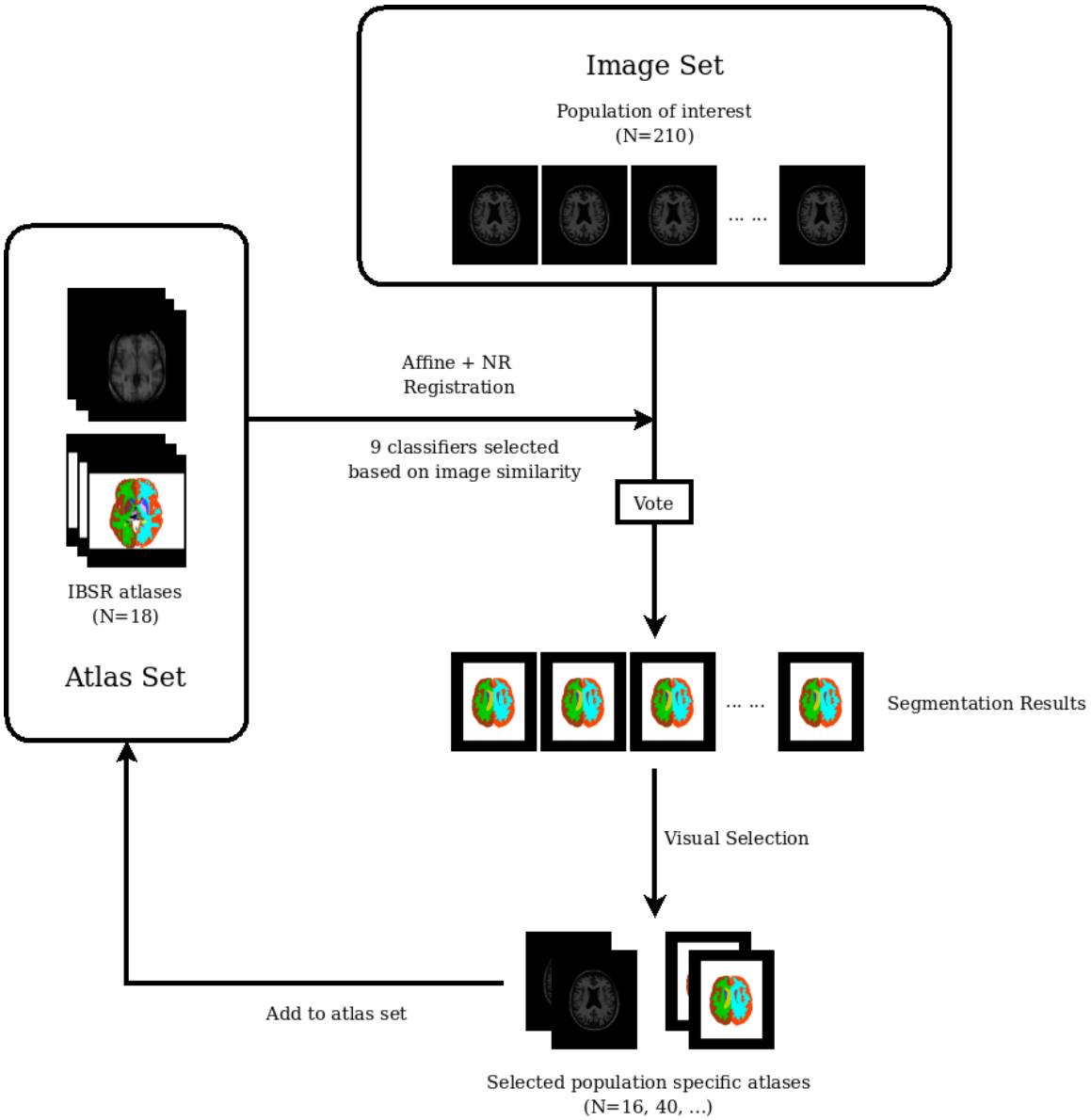

Figure 2. Diagram demonstrating the process building a set of population specific atlases.

the performance of classifier fusion may be affected by the disagreement among propagated segmentations. The agreement among the fused atlases is more likely resulted from the reduction of random error, even though the accuracy of the propagated labels is intrinsically limited by registration algorithm. It is preferable to deal with individual segmentations agreeing with each other in the fusion of labels. A higher agreement among the atlases means more overlap between the segmentation results, i.e. the consensus region with majority votes, and each individual segmentation propagation. It becomes a measurement of accuracy when the segmentation result is back-propagated to the atlas, and overlap is transformed into the atlas space.

A probability image can be created for each structure when fusing the label maps propagated from atlases. For a given label, each pixel in the probability image counts the number of votes it receives, and is normalised by the total number of atlases. The entropy of this probability image, and the partial moment of its histogram are used to measure the performance of the atlases.

2.2.1 Entropy of probability image

The entropy of image is a statistical measure of its randomness. As far as the probability image is concerned, it can be defined as

$$H = \sum_{i=1}^{n} -p_i \log p_i, \tag{1}$$

where i is the number of votes agree on the structure, n is the total number of atlases used in the classifier fusion step of the segmentation, and p_i the probability mass function (pmf) of voxels which have i votes on the given structure. In the ideal case when all the atlases agree unanimously, the entropy $H = 0$.

2.2.2 Partial moment of histogram

In addition to the entropy, the second order partial moments of histogram with respect to the reference point 1 is used. It measures the overall deviation of the distribution of the votes from unanimous agreement. It can be defined as follows,

$$\mu_2 = \sum_{i=1}^{n} p_i \cdot \left(1 - \frac{i}{n}\right)^2. \tag{2}$$

In the ideal case of unanimous agreement, $\mu_2 = 0$.

3. RESULTS

In the experiment, 210 brain MR images from an elderly population enrolled in the AIBL study, consisting of 170 NC and 40 AD subjects, were segmented using the multi-atlas based segmentation method with the IBSR images used as the initial set of atlases. 16 segmentations were qualified and added to the database of atlases after the first iteration. Another 24 were added after the second iteration. Thus an atlas database of 40 segmented images was built on a population of elderly subjects. The demographics of the selected atlases are shown in Table 1, all of which are NC subjects.

	IBSR	Iter. 1	Iter. 2
No. of atlases	18	16	24
Male/Female	14/4	6/10	10/14
Age	min: Juv. max: 71	mean: 74.54 min: 63.68 max: 86.10	mean: 75.73 min: 62.35 max: 88.26
MMSE	N/A	27.81	28.88

Table 1. Demographics of selected atlases

In the experiment, the labels in the 18 IBSR atlases and 40 population specific atlases were propagated to 210 brain MR images. The atlases were selected based on image similarity. For each query, 9 atlases are selected from IBSR atlases and atlases of elderly subjects respectively. Two examples of probability image of hippocampus produced from IBSR atlases and atlases selected from elderly population are shown in Figure 3. The probability images fused by IBSR atlases and atlases of elderly population are evaluated quantitatively in terms of their entropy H and the second order partial moment μ_2 of their histograms. For the IBSR and the population specific atlases, the results fusing 9 atlases selected according to image similarity are presented in Table. 2 and 3.

4. DISCUSSION

The results show that, the agreement among the elderly population specific atlases is in general higher than that from IBSR, when being propagated to query images in the dataset of elderly subjects. For both NC and AD cases, the entropy is higher when probability images are fused with atlases of the IBSR than that of atlases selected from elderly population, indicating a higher randomness in the probability images produced using IBSR

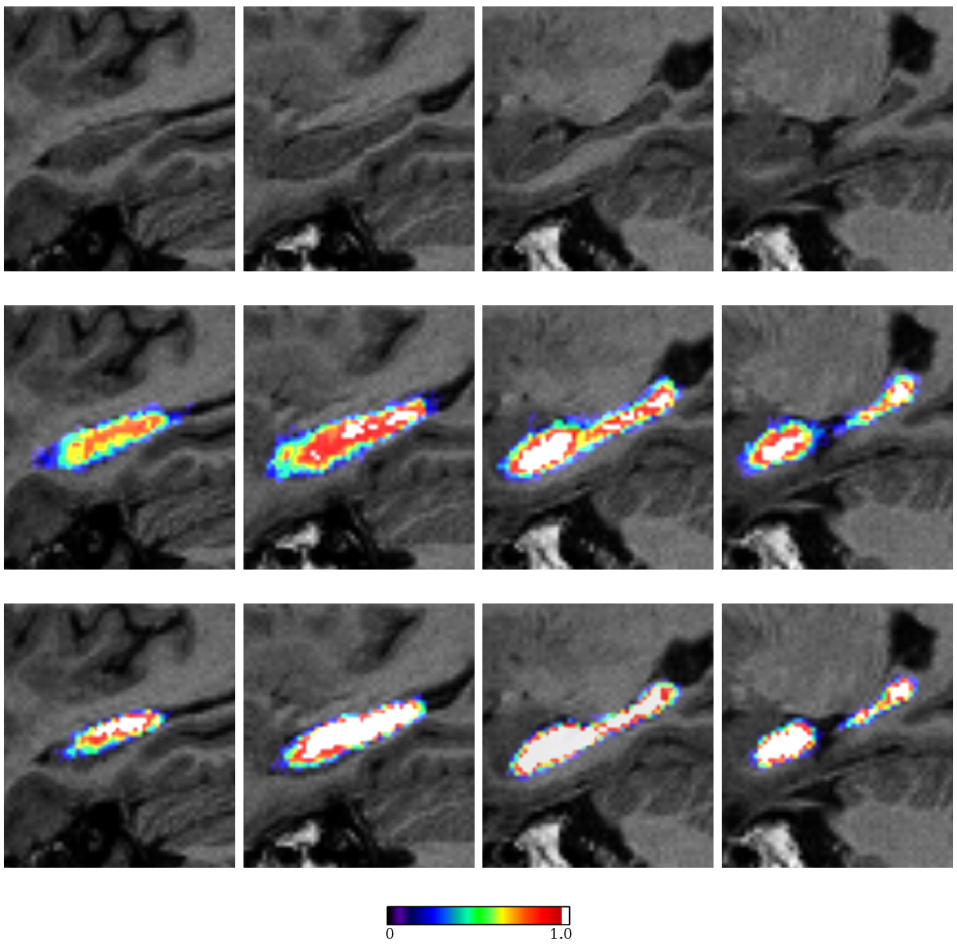

Figure 3. Probability images of hippocampus of the same NC subject, with the white colour indicating unanimous agreement. Top: T1-weighted MR images of hippocampus; middle: the probability images fusing 9 propagated atlases from IBSR, selected according to image similarity; bottom: the probability images fusing 9 atlases propagated from population specific atlases, selected according to image similarity.

atlases. In terms of the second order partial moment of histograms, segmentation propagation with IBSR atlases results in more dissidence than using the atlases selected.

The partial moment μ_2 of the histogram of probability images produced using similarity selected atlases are lower than those randomly selected, indicating more voxels are labelled with fewer votes when fusing a random set of atlases. A similar trend is observed in the results of probability image entropy using population specific atlases. Another observation made on the effect of atlas selection is that this effect is more significant when dealing with AD cases (Table. 3).

Considering the size of the population specific atlas set, selecting based on image similarity from a larger pool of atlases produces the label map with a higher degree of agreement. In terms of probability image entropy H, the label maps of hippocampus, thalamus and lateral ventricle are produced with less randomness when selecting atlases from 40 atlases, with the exception of caudate and putamen. While in terms of the partial moment μ_2, the distributions of the votes using 40 atlases are consistently closer to the unanimous agreement.

All the 40 atlases visually inspected and selected are from NC subjects. Since IBSR atlases are from NC subjects, the well segmented images using IBSR set are more likely to come from NC population. This effect is reflected in the measurement of agreement. The NC atlases are more easily to agree upon NC cases compared to AD ones. The trend is consistently present as measure by the partial moment μ_2, while it is not very clearly shown in terms of entropy H.

	H Entropy of probability image					
	NC			AD		
	IBSR	16	40	IBSR	16	40
Hippocampus	3.01	2.82	**2.79**	3.00	2.82	**2.81**
Thalamus	2.71	2.61	**2.53**	2.73	2.56	**2.54**
Caudate	2.97	**2.87**	2.89	2.92	**2.84**	2.89
Putamen	2.99	**2.90**	2.92	3.02	**2.94**	2.97
Lateral Ventricle	2.96	2.73	**2.55**	2.92	2.73	**2.49**

Table 2. Entropy H of probability images, the bold indicates the best case in NC and AD respectively, for each structure.

	μ_2 Histogram moment of probability image					
	NC			AD		
	IBSR	16	40	IBSR	16	40
Hippocampus	0.61	0.60	**0.54**	0.63	0.61	**0.55**
Thalamus	0.52	0.56	**0.49**	0.53	0.61	**0.50**
Caudate	0.60	0.60	**0.56**	0.63	0.63	**0.58**
Putamen	0.55	0.54	**0.52**	0.57	0.56	**0.54**
Lateral Ventricle	0.51	0.52	**0.54**	0.48	0.52	**0.43**

Table 3. Partial moment μ_2 of probability image histogram, the bold indicates the best case in NC and AD respectively, for each structure.

5. CONCLUSION AND FUTURE WORK

In this study, we used a supervised approach to produce a set of population specific atlases from elderly subjects using multi-atlas based segmentation-propagation. Starting with 18 IBSR atlases, 16 images from elderly population well segmented were added to the atlas set. More images were added in the seconde iteration. The result 40 population specific atlases were evaluated in terms of their agreement when propagated and fused to target images. Comparing to the 18 images in the IBSR atlases, the population specific atlas set built from the elderly population reaches a higher level of consensus generic IBSR atlases. The result segmentations are produced by a majority vote rule with higher certainty. Although as a supervised approach our method requires visual inspection, it is still less time consuming and costs less than manually segment images. In the future work it would be interesting to extend the currently work using IBSR atlas to the image data collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI)†. It provides semi-automated segmentations of hippocampus, which is of interest to the study of Alzheimer's disease.

REFERENCES

[1] Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., and Rueckert, D., "Classifier selection strategies for label fusion using large atlas databases," in [*Medical Image Computing and Computer-Assisted Intervention MICCAI 2007*], 523–531 (2007).

[2] Murgasova, M., Dyet, L., Edwards, D., Rutherford, M., Hajnal, J., and Rueckert, D., "Segmentation of brain MRI in young children," *Academic Radiology* **14**, 1350–66 (Nov. 2007). PMID: 17964459.

[3] Babalola, K., Petrovic, V., Cootes, T., Taylor, C., Twining, C., Williams, T., and Mills, A., "Automatic segmentation of the caudate nuclei using active appearance models," in [*Workshop on 3D Segmentation in the Clinic*], 57–64 (Oct. 2007).

[4] Patenaude, B., Smith, S., Kennedy, D., and Jenkinson, M., "Bayesian shape and appearance models," Tech. Rep. TR07BP1, Oxford (2007).

[5] Babalola, K., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., Smith, S., Cootes, T., Jenkinson, M., and Rueckert, D., "Comparison and evaluation of segmentation techniques for subcortical

†Information about ADNI available at `http://www.loni.ucla.edu/ADNI/`.

structures in brain MRI," in [*Medical Image Computing and Computer-Assisted Intervention MICCAI 2008*], 409–416 (2008).

[6] Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., and Rueckert, D., "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage* **46**, 726–738 (July 2009).

[7] Ourselin, S., Roche, A., Subsol, G., Pennec, X., and Ayache, N., "Reconstructing a 3D structure from serial histological sections," *Image and Vision Computing* **19**(1-2), 25–31 (2001).

[8] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N., "Non-parametric diffeomorphic image registration with the demons algorithm," in [*Medical Image Computing and Computer-Assisted Intervention MICCAI 2007*], 319–326 (2007).

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

[10] Yassa, M. A. and Stark, C. E., "A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe," *NeuroImage* **44**(2), 319–327 (2009).

[11] Warfield, S., Zou, K., and Wells, W., "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *Medical Imaging, IEEE Transactions on* **23**(7), 903–921 (2004).

[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).