Robust atlas-based brain segmentation using multi-structure confidence-weighted registration

Ali R. Khan[1], Moo K. Chung[2], and Mirza Faisal Beg[1]

[1] School of Engineering Science, Simon Fraser University, 8888 University Drive,
Burnaby BC, V5A 1S6, Canada
`akhanf@sfu.ca` , `mfbeg@ensc.sfu.ca`
[2] Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin,
Madison, WI 53706, USA
`mkchung@wisc.edu`

Abstract. We present a robust and accurate atlas-based brain segmentation method which uses multiple initial structure segmentations to simultaneously drive the image registration and achieve anatomically constrained correspondence. We also derive segmentation confidence maps (SCMs) from a given manually segmented training set; these characterize the accuracy of a given set of segmentations as compared to manual segmentations. We incorporate these in our cost term to weight the influence of initial segmentations in the multi-structure registration, such that low confidence regions are given lower weight in the registration. To account for correspondence errors in the underlying registration, we use a supervised atlas correction technique and present a method for correcting the atlas segmentation to account for possible errors in the underlying registration. We applied our multi-structure atlas-based segmentation and supervised atlas correction to segment the amygdala in a set of 23 autistic patients and controls using leave-one-out cross validation, achieving a Dice overlap score of 0.84. We also applied our method to eight subcortical structures in MRI from the Internet Brain Segmentation Repository, with results better or comparable to competing methods.

1 Introduction

Developing robust, automated tools for brain MR image registration and segmentation is challenging due to many factors including the high degree of neuroanatomical variability in both healthy controls and patients. Registration and segmentation of medical images can be aided by using expert-derived features, but this manual intervention step can become costly in very large studies and can also suffer from rater drift. Automated computation of such features can eliminate the need for manual intervention, however the reliability and accuracy of automatically generated features is also influenced by the variability in image quality and neuroanatomy, in addition to the systemic bias, if any, present in this automated method. Hence, using a small set of manually labeled training images, learning the accuracy of the automatically generated features can be used to improve the overall utility of these features.

We have extended the large deformation atlas-based brain MRI segmentation approach of [1], which used Freesurfer segmentation labels to initialize the

region of interest (ROI) based registration, to instead use the Freesurfer labels as anatomical constraints simultaneously during registration. The main motivation for this is that the simultaneous usage of the automated segmentations as separate cost terms allows the overall MR image matching to help avoid local minima, while providing flexibility in setting weights for different channels to emphasize certain properties, such as larger weight for smaller structures, or smaller weight where the channel data is known to be less reliable. We accomplish this by using a multi-cost registration framework, with each additional data term utilizing the matching of one automatically-generated segmentation label.

Our approach is also similar in spirit to [2], which presented multi-channel registration with a few semi-automatically defined subcortical structures that were quality controlled and corrected manually prior to their use in the registration. However, instead of correcting the automated segmentations manually for each image, we attempt to learn the errors made by the automatic segmentation method using the segmentation confidence maps from a small set of manually labeled images, and account for these for segmentation of all other images in the database. To avoid computation of SCMs for each cohort atlas and to work with cohort datasets that do not have some manually segmented scans to construct the SCMs, we show how to transfer SCMs from another database atlas. We generate and apply the SCMs to weight the automated segmentations that are used as "features" in our atlas-based segmentation.

In single atlas propagation, errors or bias in the atlas segmentation, perhaps due to manual rater error, can also lead to the bias being propagated in all atlas-propagation derived segmentations. Furthermore, if the atlas contains anatomical variability, which the registration is not able to accommodate fully, then the propagated segmentations will also possess this template-dependent anatomical bias. To account for this additional source of variability, we present a supervised atlas correction procedure, which involves performing atlas-based segmentation on a manually labeled training set to learn the systematic bias present in the atlas, and correcting the atlas segmentation correspondingly. In this paper we describe our method for the generation of SCMs, multi-structure registration and supervised atlas correction, apply these techniques to two datasets, and compare our results with current brain segmentation methods.

2   Method and Materials

2.1   Brain MRI datasets

The Internet Brain Segmentation Repository (IBSR) dataset consists of 18 T1-weighted MR scans, with some manually segmented structures. This is a public database used by many groups to test segmentation methods ([3–7]). The amygdala dataset consisted of T1-weighted scans from 24 subjects (12/12 autism/control), aged 10-24 [8], along with manual segmentations.

2.2   Segmentation confidence maps

In using automated segmentations for registration, segmentation errors could result in correspondence errors. We wish to learn the errors that an automated

segmentation method makes so that if a given region in an automated segmentation consistently exhibits lower accuracy, we would like to reduce its contribution to the registration. To this end we have defined a segmentation confidence map (SCM), α^j, for each anatomical structure, j, as the probability of accuracy:

$$\alpha^j(x) = P(f^j_{error}(x) < \epsilon), \tag{1}$$

where $f^j_{error}(x)$ is the distribution of segmentation errors at spatial location x, and ϵ is a distance bound placed on the confidence map. To find $f^j_{error}(x)$, we require a map of segmentation errors between a manual gold standard, M^j, and an automated segmentation, A^j. Because correspondence between M^j and A^j is not known, we approximate this using the signed distance transforms of the binary segmentations, denoted as $DT(\cdot)$, to obtain the closest boundary distances between the manual and automated contours, so that:

$$f^j_{error}(x) \approx d^j_{M,A}(x) = \begin{cases} 0 & \text{if } M^j(x) = A^j(x), \\ (|DT_{M^j}(x)| + |DT_{A^j}(x)|)^2 & \text{if } M^j(x) \neq A^j(x). \end{cases}$$

We also used grayscale dilation followed by Gaussian smoothing ($\sigma = 1.0$) to widen the affected neighborhood. After computing this approximation for f^j_{error}, we can determine α^j by evaluating $P(f^j_{error}(x) < \epsilon)$ over a manually labeled training set.

Supervised training/learning For a given small set of M training images A_k, we first compute error maps f^j_{error,A_k} for each structure j in each image A_k. To learn the combined confidence map, we then spatially transform these to a chosen template space, B, using the large deformation diffeomorphic metric mapping (LDDMM) transformation between individual automated segmentations, A^j_k and B^j. To compute the confidence map on B, we use the transformed error maps $f^j_{error,A_k} \circ \phi_{B^j,A^j_k}$, to generate sample histograms at each voxel using a $3 \times 3 \times 3 \times M$ neighborhood. We used a distance bound, $\epsilon = 1$mm, and Equation 1 to compute α^j_B for each segmented structure j. Note that for computational reasons we computed confidence maps for each structure in each hemisphere separately, then combined them to create a confidence map relating to both hemispheres of a given structure. Figure 1 shows the SCM for the caudate nucleus in which the highest variability in automated segmentation is found to be around its tail. This is reasonable as the narrow caudate tail is where automated segmentation algorithms are likely to yield the highest variability.

Propagation to cohort atlas If cohorts are structurally similar such as matched for age and pathological state, then the SCMs learned from one cohort can be propagated to the other cohort atlas. We propagate SCMs from one cohort atlas, α^j_B, to another, α^j_C, by spatially transforming the maps, defined on B, to the space of C using the LDDMM transformation between their automated segmentations, B^j and C^j. By performing this step for each structure, j, we can estimate the SCMs for any cohort atlas given a previously trained SCM.

$M^{j=caudate}_{Template}, A^{j=caudate}_{Template}$ $f^{j=caudate}_{error,Template}$ $\alpha^{j=caudate}_{Template}$
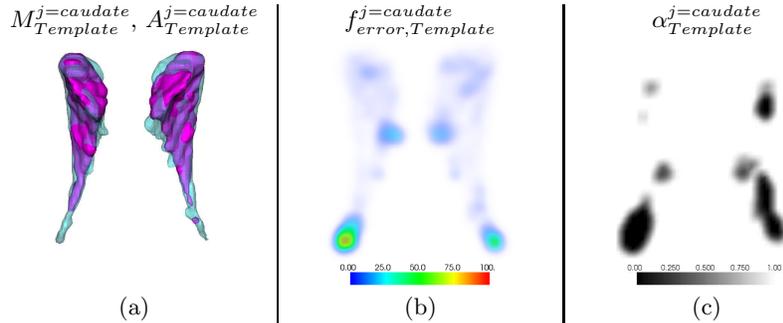
(a)   (b)   (c)

Fig. 1. Visualizations showing the generation of the segmentation confidence map (SCM) for the caudate nucleus: (a) manual (magenta) and initial automated (blue) segmentations for the template brain, (b) distance error map $f^{j=caudate}_{error,Template}$, (c) caudate SCM, $\alpha^{j=caudate}_{Template}$, computed using distance error maps from all images in the training set. Note that the SCM identifies regions of highest segmentation variability to be near the tail of the caudate.

2.3 Multi-structure confidence-weighted registration

To introduce multiple structures into a diffeomorphic registration scheme, we extended the large deformation diffeomorphic metric mapping (LDDMM) [9] method to use multiple data terms, each weighted with a SCM. Let the pair A^{MR} and B^{MR} of brain ROI MR images be given to be registered, where B is the designated template, and let their N automated segmentations, $A^j, j \in [1, \ldots, N]$ and $B^j, j \in [1, \ldots, N]$ be available. The diffeomorphic transformation matching A and B is given by $\varphi = \phi_1 : \Omega \to \Omega$ such that $A(\phi_1^{-1}) \approx B$. This transformation ϕ_1^{-1} results from velocity $\dot{\phi}_t = v_t(\phi_t), v_t \in V, t \in [0, 1]$ where V is a space of smooth vector fields on Ω. The energy for the extended confidence-weighted multi-structure registration to be minimized is therefore:

$$\int_0^1 \|v_t\|_V^2 dt + \|A^{MR}(\phi_1^{-1}) - B^{MR}\|_{L^2}^2 + \sum_{j=1}^N \|\sqrt{\alpha_B^j}\left(A^j(\phi_1^{-1}) - B^j\right)\|_{L^2}^2, \quad (2)$$

which uses the confidence map for a given structure, α_B^j to weight the mismatch $A^j(\phi_1^{-1}) - B^j$. Note that because the cost is computed in the coordinate frame of template B, the SCM need only be specified for B. Figure 2 shows the multi-structure registration images for the the left amygdala ROI registration, along with the corresponding SCMs.

2.4 Supervised atlas correction

Suppose atlas-based segmentation between an atlas, A and a target image T gives an invertible transformation $\phi_{A,T}$ which transforms the atlas labels, A_M to T via $\phi_{A,T}(A_M)$. In the ideal case of perfect registration, if T_M are known target labels, then $\phi_{T,A}(T_M) = A_M$, but due to errors in registration and manual
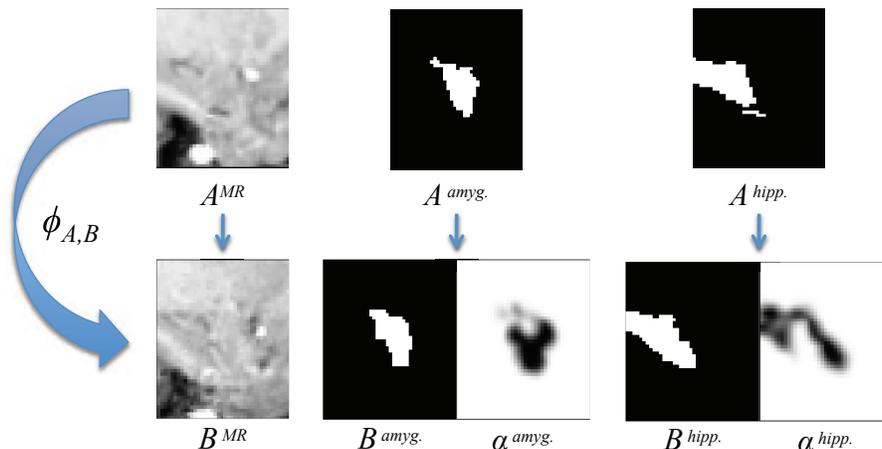
Fig. 2. Illustration of the multi-structure confidence-weighted registration for the left ROI in the amygdala segmentation, showing the MR images, initial segmentations, and SCMs. The multi-structure registration used the left hippocampus, amygdala and lateral ventricle along with the MRI images to find the optimal ROI transformation.

labeling, this is not observed. However, if, given the manually labelled target T_M, the atlas labels were 'corrected' to be $\phi_{T,A}(T_M)$, then label propagation would result in perfect segmentation correspondence. We use this insight to average the back-propagated labels, $\phi_{T,A}(T_M)$, for all images in a training set, and denote this as the 'corrected' atlas segmentation. The corrected segmentation accounts for both manual labeling inconsistencies and systematic correspondence errors, thus improving the overall label propagation.

2.5 Experimental Procedure

All brain MR images were processed with the Freesurfer image analysis suite (version 4.1.0), using the subcortical processing stream [10], which labels 37 volumetric structures; these segmentations were used as the initial automated segmentations for our method. In preparation for atlas-based ROI segmentation, the MR images underwent pre-processing including affine registration, definition of a bounding box for each hemisphere and histogram-based intensity normalization. Thus for each target image, a cropped region of interest (ROI) for each hemisphere, containing the structures to be segmented, was linearly aligned and intensity normalized to the corresponding ROI in the template MRI.

For the IBSR dataset, we used two disjoint sets for training and testing; nine brains were used to generate Freesurfer SCMs for the left and right caudate, putamen, pallidum, nucleus accumbens, thalamus, hippocampus, amygdala, and lateral ventricles, with the other nine used as the test data. The multi-structure registration used all eight structures along with the MRI to find the diffeomorphic transformation for each hemispheric ROI. For the amygdala dataset, we propagated the IBSR SCMs to an arbitrarily chosen control subject, and performed

multi-structure registration using ROIs containing the hippocampus, amygdala and lateral ventricles. Supervised atlas correction was tested with a leave-one-out cross-validation scheme on the above described test data. Spatial overlap was measured with the Dice similarity coefficient, $DSC(A, M) = \frac{2V(A \cap M)}{V(A)+V(M)}$, where $V(A)$ and $V(M)$ refers to the volume of the automated and manual segmentations respectively.

3 Results

Spatial overlaps for the amygdala dataset are shown in top row of Figure 3, with the multi-structure segmentation outperforming the Freesurfer segmentations used in the multi-structure registration, and the supervised atlas correction further improving the results. For all methods performance is better for the control subjects, with the highest mean DSC being 0.85 and 0.83 for control and autism subjects respectively. Bottom panel shows representative MRI slices of an autism subject, with manual and our automated (multi-structure, atlas-corrected) segmentation outlines. Table 1 compares the amygdala DSCs to competing methods, and Table 2 summarizes results on the IBSR database.

4 Conclusions and Discussions

As evident in Table 1, performance of our amygdala segmentations compares very favorably among the current state-of-the-art methods; only the method in [11], which performs hierarchical parcellation of the brain, reports slightly higher but comparable numbers. Multiple atlas propagation and fusion is used in [12] with good results, a technique that can also be applied with the proposed method to further improve performance at the cost of additional registrations. The best results for the IBSR database, shown in Table 2, are emphasized in bold, with our method showing the highest spatial overlap for the majority of structures (lateral ventricles, caudate, putamen, thalamus, nucleus accumbens, and hippocampus), and within 0.02 of the highest for the pallidum and amygdala.

Supervised training in this setting, used in both the SCM generation and atlas correction steps, can be problematic if differences exist between the training set and the test set, such as manual segmentation protocols, scanner differences, or pathological differences. For the supervised atlas correction these differences could lead to degraded performance, since errors in the corrected segmentation would correspond directly with final segmentation errors, which is why we chose to use cross-validation on the test data for this purpose. For the SCM generation, however, we did use a single training set for both the IBSR and amygdala segmentation; satisfactory results were obtained likely because the Freesurfer segmentations had consistent bias or errors for both datasets. We plan to further study the consistency and applicability of SCMs generated from different training sets. To conclude, we have proposed a novel two-fold strategy for improving performance of atlas-based brain segmentation using multi-structure confidence-weighted registration, and supervised atlas-correction. Results show promise for improved segmentation of many subcortical structures, including the amygdala, with performance better than or comparable to the leading current methods.

References

1. Khan, A.R., Wang, L., Beg, M.F.: Freesurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping. Neuroimage **41**(3) (Jul 2008) 735–46

2. Magnotta, V.A., Bockholt, H.J., Johnson, H.J., Christensen, G.E., Andreasen, N.C.: Subcortical, cerebellar, and magnetic resonance based consistent brain image registration. Neuroimage **19**(2 Pt 1) (Jun 2003) 233–45

3. Akselrod-Ballin, A., Galun, M., Gomori, J.M., Brandt, A., Basri, R.: Prior knowledge driven multiscale segmentation of brain MRI. Medical image computing and computer-assisted intervention **10**(Pt 2) (Jan 2007) 118–26

4. Gouttard, S., Styner, M., Joshi, S., Smith, R., Hazlett, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. Proceedings of SPIE **6512** (2007) 65122J

5. Joshi, A., Shattuck, D., Thompson, P., Leahy, R.: Surface-constrained volumetric brain registration using harmonic mappings. IEEE Transactions on Medical Imaging **26**(12) (2007) 1657–1669

6. Ciofolo, C., Barillot, C.: Brain segmentation with competitive level sets and fuzzy control. Information processing in medical imaging **19** (Jan 2005) 333–44

7. Zhou, J., Rajapakse, J.: Segmentation of subcortical brain structures using fuzzy templates. Neuroimage **28**(4) (2005) 915–924

8. Nacewicz, B.M., Dalton, K.M., Johnstone, T., Long, M.T., McAuliff, E.M., Oakes, T.R., Alexander, A.L., Davidson, R.J.: Amygdala volume and nonverbal social impairment in adolescent and adult males with autism. Archives of general psychiatry **63**(12) (Dec 2006) 1417–28

9. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision **61**(2) (2005) 139–157

10. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S.: Whole brain segmentation automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3) (2002) 341–355

11. Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M.: A hierarchical algorithm for MR brain image parcellation. IEEE Transactions on Medical Imaging **26**(9) (Sep 2007) 1201–12

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

13. Chupin, M., Mukuna-Bantumbakulu, A., Hasboun, D., Bardinet, E., Baillet, S., Kinkingnéhun, S., Lemieux, L., Dubois, B., Garnero, L.: Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: method and validation on controls and patients with Alzheimer's disease. Neuroimage **34**(3) (2007) 996–1019

14. Shen, D., Davatzikos, C.: Hammer: hierarchical attribute matching mechanism for elastic registration. Medical Imaging, IEEE Transactions on **21**(11) (2002) 1421–1439

15. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. general methods and intrasubject, intramodality validation. Journal of Computer Assisted Tomography **22**(1) (Jan 1998) 139–52
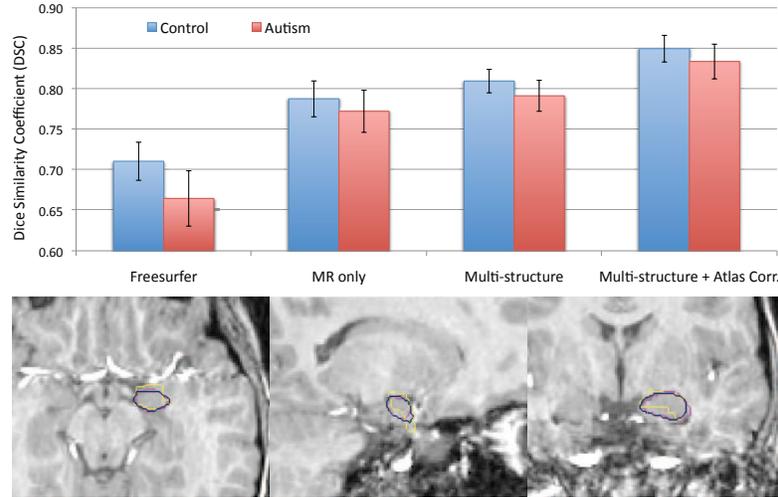
Fig. 3. Top: Mean DSC for the amygdala dataset, where the height of the error bars is equal to the standard deviation. Bottom: Representative axial (left), sagittal (center), and coronal (right) slices showing amygdala segmentations for the manual rater (pink), Freesurfer (yellow) and the multi-structure atlas-corrected method.

Method	Cohort	Age range	DSC (L ; R)
Our method	healthy/autistic	13-23/10-24	0.85 ± 0.033 / 0.83 ± 0.043
Fischl et al. [10]	healthy/autistic	13-23/10-24	0.71 ± 0.047 / 0.66 ± 0.068
Pohl et al. [11]	schiz.+healthy	18-41	0.86 ± 0.028; 0.85 ± 0.030
Chupin et al. [13]	healthy/Alzheimer's	< 35/66-81	0.81 ± 0.04 / 0.76 ± 0.07
Heckemann et al. [12]	healthy	20-54	0.80 ; 0.81

Table 1. Dice similarity coefficients for amygdala segmentation for our method and competing methods using various datasets, with standard deviations shown where available. Note that both methods in the first two rows use the same amygdala dataset.

Method	Lat. Vent.	Caud.	Put.	Thal.	Pall.	Nuc. Acc.	Hipp.	Amyg.
Our method	**0.85** (0.06)	**0.83** (0.03)	**0.87** (0.02)	**0.89** (0.01)	0.72 (0.09)	**0.61** (0.10)	**0.76** (0.03)	0.66 (0.08)
Fischl et al. [10]	0.78 (0.07)	0.82 (0.05)	0.81 (0.02)	0.86 (0.02)	0.71 (0.13)	0.58 (0.08)	0.75 (0.02)	**0.68** (0.06)
Akselrod-Ballin et al. [3]	-	0.80	0.79	0.84	**0.74**	-	0.69	0.63
Gouttard et al. [4]	**0.85**	0.76	0.78	-	0.72	-	0.67	0.64
Joshi et al. [5]	-	0.54	0.49	0.60	-	-	0.41	-
Ciofolo et al. [6]	-	0.65	0.70	0.77	0.58	-	-	-
Zhou et al. [7]	-	0.80 (0.08)	0.81 (0.06)	0.84 (0.06)	-	-	0.70 (0.11)	0.64 (0.15)
Shen et al. [14] (in [5])	-	0.54	0.45	0.74	-	-	0.30	-
Woods et al. [15] (in [5])	-	0.40	0.36	0.65	-	-	0.50	-

Table 2. DSCs on the IBSR dataset, with **bold** entries denoting the highest performing for each structure, with standard deviation shown in parentheses where available.