

# Accurate Whole-Brain Segmentation for Alzheimer’s Disease Combining an Adaptive Statistical Atlas and Multi-atlas

Zhennan Yan<sup>1</sup>, Shaoting Zhang<sup>1</sup>, Xiaofeng Liu<sup>2</sup>, Dimitris N. Metaxas<sup>1</sup>,  
Albert Montillo<sup>2</sup>(✉), and The Australian Imaging Biomarkers and Lifestyle  
Flagship Study of Ageing<sup>3</sup>

<sup>1</sup> CBIM, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

<sup>2</sup> GE Global Research, Niskayuna, NY, USA

montillo@ge.com

<sup>3</sup> Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia

**Abstract.** Accurate segmentation of whole brain MR images including the cortex, white matter and subcortical structures is challenging due to inter-subject variability and the complex geometry of brain anatomy. However a precise solution would enable accurate, objective measurement of structure volumes for disease quantification. Our contribution is three-fold. First we construct an adaptive statistical atlas that combines structure specific relaxation and spatially varying adaptivity. Second we integrate an isotropic pairwise class-specific MRF model of label connectivity. Together these permit precise control over adaptivity, allowing many structures to be segmented simultaneously with superior accuracy. Third, we develop a framework combining the improved adaptive statistical atlas with a multi-atlas method which achieves simultaneous accurate segmentation of the cortex, ventricles, and sub-cortical structures in severely diseased brains, a feat not attained in [18]. We test the proposed method on 46 brains including 28 diseased brain with Alzheimer’s and 18 healthy brains. Our proposed method yields higher accuracy than state-of-the-art approaches on both healthy and diseased brains.

**Keywords:** Brain segmentation · Alzheimer’s · Adaptive atlas · Multi-atlas · MRF

## 1 Introduction

The neurologic study and clinical diagnosis of many brain diseases, such as Alzheimer’s Disease (AD) or hydrocephalus, often requires magnetic resonance

---

Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database ([www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at [www.aibl.csiro.au](http://www.aibl.csiro.au).

(MR) imaging of the brain. Segmentation of distinct brain structures from MR images is a vital process for objective diagnosis, treatment planning, therapy monitoring and drug development. Manual labeling of brain structures in MR images by a human expert can require up to 1 week per subject and is operator dependent. For large data sets, manual segmentation of individual structures is not practical; however automating the segmentation is difficult due to image artifacts, noise, complex textures, complex shapes and partial volume effect. In recent decades, many approaches have been proposed to segment human organs or tissues in MR images or other modalities [11], for example classification based methods [6], deformable model based methods [10], and atlas-guided approaches [1, 7, 9, 12, 14, 16]. Among these approaches, atlas based methods are the most commonly used approaches for brain image segmentation. In medical image segmentation (e.g. [15]), an atlas is defined as a pair of an MR intensity scan (e.g. T1) and its corresponding manual segmentation. Given several atlases, there are two ways to segment a new target image. The first is to learn a single statistical atlas that models the spatial priors for individual structures. A single probabilistic atlas is fit in a Bayesian framework for voxel classification [1, 14]. Single statistical atlas based methods are accurate when the target scan has similar anatomical characteristics as the atlas populations. The second approach is to register the set of atlases (multi-atlas) to the target image and then compute the final segmentation via a label fusion approach [7, 9, 12, 16]. A multi-atlas method tends to be computationally expensive due to the required multiple non-linear registrations [2, 17] and also has limited ability to handle diseased brains with anatomical characteristics that vary from the training atlases.

Recently, several adaptive statistical atlas-based expectation maximization (EM) algorithms were proposed to deal with the above limitation of atlas based methods. One by Shiee et al. [13] was applied to segment brains with ventriculomegaly. Another by Cardoso et al. [3] was applied to measure cortical thickness. In their work the brain is segmented into only 4–6 coarse structures: white matter (WM), gray matter (GM), ventricles, cerebrospinal fluid (CSF). However, the subcortical GM structures (e.g. hippocampus), which are critical in clinical diagnoses are not handled. In [18], we proposed an extended adaptive statistical atlas (EASA) using spatially varying adaptivity prior to segment many structures from whole brain MR scans, and use a strategy to combine the statistical atlas with multi-atlas to enhance the accuracy of segmenting diseased brains. This method can fail for some structures such as the cerebral cortex which is critical for disease detection (e.g. cortical thinning in AD). One reason is the use of a simplistic Markov Random Field (MRF) model that assumes every pair of voxel label arrangements has the same probability. This can lead to leaking, an unregulated, undesirable form of atlas adaptivity.

In this work, we segment the whole brain into 34 anatomical structures simultaneously and accurately by improving the adaptive atlas methods [13, 18]. Our methodological contributions include: (1) structure specific relaxation along with the spatial adaptivity prior, (2) an isotropic pairwise class-specific MRF model, (3) a complete hybrid framework of our proposed method, denoted **EASA++**,

with a multi-atlas approach which improves segmentation accuracy on diseased brains, especially for the cerebral cortex and ventricles. We evaluate our hybrid method *on 46 brains*. We *qualitatively* evaluate our results on all 46 brains including 19 with moderately enlarged ventricles, and *quantitatively* evaluate on 27 brains including 18 normal brains, and 9 AD brains with severely enlarged ventricles. Finally we compare our performance against state-of-the-art approaches such as FreeSurfer [5].

## 2 Methods

**Background for the Extended Adaptive Single Statistical Atlas (EASA) and Weighted Majority Voting (WMV) multi-atlas methods.** In [13] an EM-based adaptive single statistical atlas (ASA) brain segmentation method was proposed to address the common clinical situation in which the target brain to be segmented is poorly represented by the brains in the training set. This method assumes that the brain consists of  $K$  structures ( $k = 1, \dots, K$ ), the number of voxels in the T1 MR image is  $N$  ( $i = 1, \dots, N$ ), and that the intensity distribution of each structure follows a Gaussian distribution. The observed image is modeled by a  $K$ -component Gaussian Mixture Model (GMM) [14] with unknown parameters: the mixing coefficients  $\pi_{ik}$  and  $\theta_k = (\mu_k, \sigma_k^2)$ , where  $\mu_k$  and  $\sigma_k^2$  are the means and variances, respectively. The true label for voxel  $i$  is denoted as  $z_i$  (a  $K \times 1$  binary-valued vector), while the prior probability that voxel  $i$  belongs to structure  $k$  is written as  $p_i = (p_{i1}, \dots, p_{iK})$ , and its posterior probability as  $w_{ik}$ . Since brain structure labels are piecewise constant, a MRF prior on  $z_i$ 's is incorporated to the complete model:

$$f(Z, X | \boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{1}{Norm} \prod_{i=1}^N \prod_{k=1}^K (\pi_{ik} G(x_i; \theta_k))^{z_{ik}} \exp \left\{ -\beta \sum_{j \in N_i} \sum_{l=1, l \neq k}^K z_{ik} z_{jl} \right\}$$

where  $x_i$  is the intensity in target image at voxel  $i$ ,  $G(x_i; \theta_k)$  is the Gaussian model of structure  $k$ ,  $Norm$  is the MRF normalizer term and  $N_i$  is the 6-connected neighborhood of voxel  $i$ . The EM algorithm is used to solve the maximum a posteriori (MAP) estimation of parameters  $\pi_{ik}$  and  $\theta_i$ . Assuming  $\pi_i$  follows a Dirichlet distribution and applying a Gaussian smooth filter on  $w_{ik}^t$ , they used  $\pi_{ik}^{t+1} \approx (1 - \kappa)p_{ik} + \kappa(G * w_{ik}^t)$  to trade off spatial prior fidelity and the current EM estimate. The method in [13] is limited in that it can only parcellate 4 coarse anatomical classes (WM, GM, exterior CSF and ventricular CSF).

In [18] we extended ASA (denoted EASA) to parcellate many more (30+) structures throughout the brain. This was achieved through two steps. First we took the global invariant relaxation parameter  $\kappa = 0.5$  and made it spatially *variant*. The spatially variant adaptivity map  $\kappa(\mathbf{x}): \mathbb{R}^3 \mapsto \mathbb{R}$  depends on the coordinates of voxel  $\mathbf{x} \in \mathbb{R}^3$ . Second to segment diseased brains more accurately, we combined EASA with a multi-atlas label fusion approach called intensity weighted majority voting (WMV) [9]. WMV extends canonical majority voting label fusion by incorporating the intensity of the target subject to further guide

the final label selection. This approach, which we denote simply as [18], applies the multi-atlas WMV to the target. This (1) generates a rough initial parcellation for subsequent EM-based EASA, and (2) enables the creation of ASA priors that are target subject specific since the multi-atlas WMV maps the training atlases to the target.

While [18] successfully extended ASA to segment 30+ structures the method’s limitations include an inaccurate segmentation of structures, such as the cerebral-cortex and neighboring WM. There was also limited quantitative evaluation in [18]. In this paper we propose a new ASA approach, **EASA++**, and two new hybrid approaches, **Hybrid2** and **Hybrid3**, which address these methodological limitations and we perform a thorough quantitative evaluation of the impact of our new methods.

**Our proposed EASA++ and hybrid approaches.** Though the IBSR<sup>1</sup> atlas is extensively used in the brain segmentation literature, its manual labels contain errors including an over segmentation of cortical gray (exterior CSF voxels labeled as cortex). Consequently multi-atlas approaches, such as WMV, can fail to achieve good cortical GM parcellation because the propagated atlases consistently over segment the target’s cortex. [18] does not improve cortical GM segmentation because all registered atlases are incorrect in the same way, causing low label entropy (high confidence) which would have otherwise triggered relaxation and subject-specific adaptivity. This motivates our first improvement to improve segmentation accuracy for cortex, external CSF, and WM, by making the spatially varying relaxation map  $\kappa(\mathbf{x})$  dependent on the anatomical structure. This yields greater adaptivity for structures whose manual segmentation is problematic. *Formally, our new relaxation map is a function of location  $\mathbf{x}$  and structure  $k$ :*

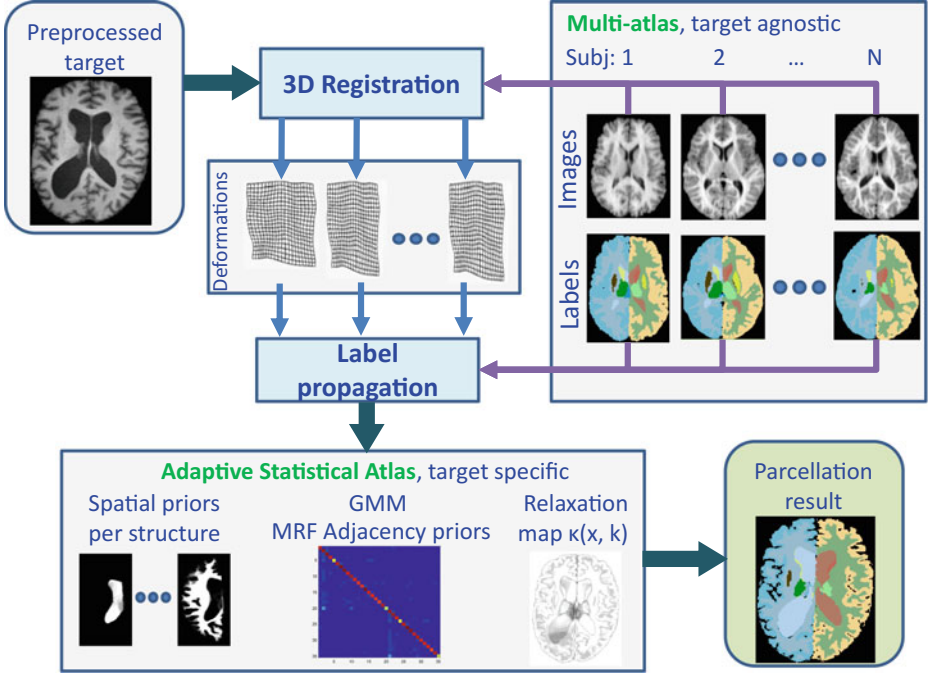
$$\kappa(\mathbf{x}, \mathbf{k}) = \begin{cases} \alpha_1 \cdot \kappa(\mathbf{x}) & \text{if } (k = k_{ECSF}) \\ \alpha_2 \cdot \kappa(\mathbf{x}) & \text{if } (k \in \{k_{WM's}\}) \\ \kappa(\mathbf{x}) & \text{otherwise} \end{cases}$$

where the white matter structures,  $k_{WM's}$  include left and right cerebral WM,  $\alpha_1, \alpha_2 \in (0..1)$  are empirically determined structure specific relaxation coefficients.  $\kappa(\mathbf{x})$  is computed via voxel label entropy:  $H(\mathbf{x}) = \sum_{k=1}^K -r_k(\mathbf{x}) \log(r_k(\mathbf{x}))$  where  $r_k(\mathbf{x})$  is the rate that voxel  $\mathbf{x}$  is labeled as  $k$  in training data. More relaxation is allowed at voxels with larger entropy and less where entropy is lower.

Studying the evolution of labels of [18], we observe that if the initial segmentation had miss-labeled a voxel as another similar-intensity class to the true label, then incorrect label can be propagated to neighboring voxels with similar intensity values. To suppress such error propagation, we replace the isotropic MRF of [18], with a pairwise class-specific MRF model:

$$f(Z, X | \boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{1}{Norm} \prod_{i=1}^N \prod_{k=1}^K (\pi_{ik} G(x_i; \theta_k))^{z_{ik}} \exp \left\{ - \sum_{j \in N_i} \sum_{l=1, l \neq k}^K \beta(l, k) z_{ik} z_{jl} \right\}$$

<sup>1</sup> <http://www.cma.mgh.harvard.edu/ibsr/>



**Fig. 1.** Proposed **Hybrid3** algorithm first applies a multi-atlas to robustly initialize the segmentation and then applies an adaptive single statistical atlas for precise subject specific parcellation.

where  $\beta(l, k)$  is the  $K \times K$  MRF parameter matrix. These parameters are estimated in 3 steps. First for neighboring voxels throughout the whole atlas, we compute the pairwise class probability. Specifically we compute the probability that a voxel with label  $c2$  appears next to the voxel with label  $c1$  as  $P_{pair}(c1, c2) = \frac{\# \text{ of pairs}(c1, c2)}{\# \text{ of voxels with label } c1}$  and average  $P_{pair}(c1, c2)$  over all atlases. Finally we map  $P_{pair}(c1, c2)$  to  $\beta(c1, c2)$  values, where the larger the value of  $P_{pair}(c1, c2)$  the smaller  $\beta(c1, c2)$  should be to make the model have a larger probability of connecting classes. We model the relation as follows:

$$\beta(c1, c2) = \begin{cases} \frac{\gamma}{P_{pair}(c1, c2)} & \text{if } (P_{pair}(c1, c2) > \gamma) \\ 1 & \text{otherwise} \end{cases}$$

where  $\gamma \in (0..1)$ .

Having described **EASA++**, we combine it with multi-atlas method forming the new hybrid algorithm **Hybrid3**, which is illustrated in Fig. 1. After removing the skull using ROBEX [8], the preprocessed target (shown in the top left) is then non-linearly registered to each training scan using a diffeomorphic symmetric normalization [2]. The same transforms are used to propagate training subject label maps to the target. We combine these target specific label maps:



**Fig. 2.** Our development of  $\kappa(\mathbf{x}, \mathbf{k})$  and isotropic pairwise class-specific MRF for the proposed **Hybrid3** approach visibly improves cortical GM segmentation.

(1) to form statistical probability maps  $p_{ik}$  for each structure, (2) to estimate isotropic pairwise class-specific MRF parameters, and (3) to compute the relaxation map  $\kappa(\mathbf{x})$ . We initialize  $w_{ik}^0$  by the  $p_{ik}$  (rather than the hard initialization  $w_{ik}^0 = z_{ik}^{label\ fus}$  in [18]) and then apply adaptive segmentation using the structure specific  $\kappa(\mathbf{x}, \mathbf{k})$  in using in an EM framework. Similar to **Hybrid3** method, we define **Hybrid2** method with only the structure specific relaxation enhancement.

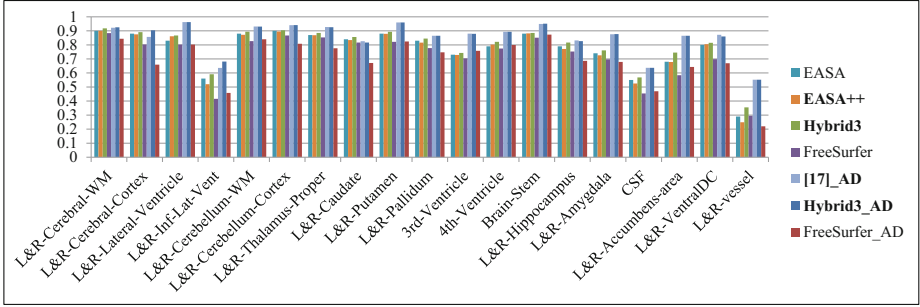
Figure 2 shows, for the same subject, the impact of the addition of structure specific relaxation and isotropic pairwise class-specific MRF. We observe that [18] mislabels many exterior CSF voxels as cortical GM. **Hybrid2** improves cortical GM segmentation but has some exterior to CSF voxels labeled as lateral ventricle (purple). For the **Hybrid3** method both enhancements are used and this yields the most precise parcellation.

### 3 Experiments

In this section, we evaluate the proposed **Hybrid3** method and compare it to state-of-the-art methods. We use two datasets. One is the IBSR data set which has 18 healthy subjects with T1 intensity volumes and medical expert delineated ground truth. Each brain volume has  $256 \times 256 \times 128$  voxels, 1 mm in-axial plane resolution and 1.5 mm between-axial plane resolution. Each segmentation has 30+ labels including the left and right WM, GM, ventricles, and the subcortical GM structures (e.g. putamen, hippocampus, etc.). The other dataset consists of 28 Alzheimer’s disease subjects randomly selected from the AIBL database<sup>2</sup>. Each brain volume has about  $240 \times 256 \times 160$  voxels, 1 mm in-slice resolution and 1.2 mm between-slice resolution. The AIBL dataset consists of 9 subjects with severely enlarged lateral ventricles (denoted as Severe\_AD) and 19 subjects with moderately enlarged lateral ventricles (denoted Moderate\_AD). We manually labeled each subject volume in Severe\_AD to provide ground truth for diseased brains.

In our first experiment we compared **EASA++**, **Hybrid3** to EASA and FreeSurfer on the IBSR dataset. From leave-one-out cross-validation tests, we observe overall best performance from the proposed **Hybrid3** method (see Fig. 3). The bar graph shows results from the first experiment and the second experiment (denoted as \*\_AD).

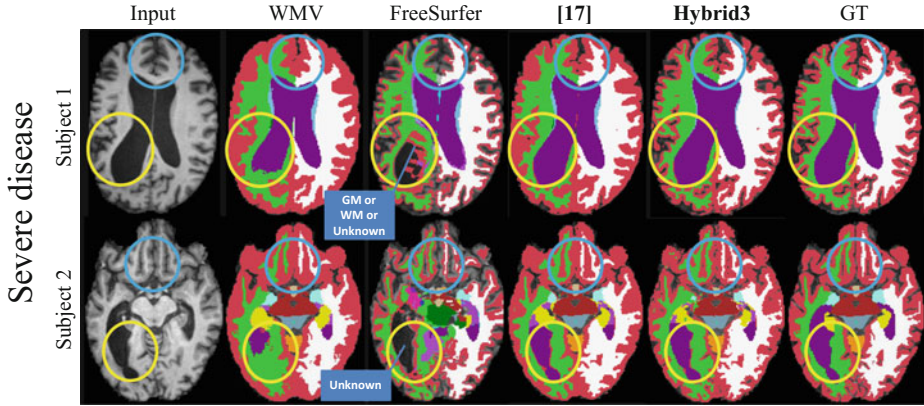
<sup>2</sup> Data was collected by the AIBL study group. AIBL study methodology has been reported previously ([4]).



**Fig. 3.** Comparison of dice score per structure on 18 healthy subjects and 9 severe AD subjects. Proposed algorithm out performs benchmark methods such as FreeSurfer for most structures. ‘L&R’ stands for average result of structure in left and right spheres.

In our second experiment, we used the 18 healthy brains of IBSR as a training dataset and used the 28 subjects from the AIBL Alzheimer’s disease dataset for testing. We evaluated and compared results of **Hybrid3** with FreeSurfer v5.1.0 and our implementation of WMV and [18]. For the **Hybrid3** approach we use empirically chosen parameters  $\alpha_1 = 0.7$ ,  $\alpha_2 = 0.9$ ,  $\gamma = 0.001$ . For quantitative evaluation, we compute average dice scores for 30+ structures in Severe\_AD data set based on the ground truth. As shown in Fig. 3, the proposed approach yields better results than FreeSurfer in 27 out of 34 structures and **16 of these 27 are statistically significant differences by t-test**, ( $p \leq 0.05$ ). These 16 structures include the Lateral Ventricles, which are known biomarkers for early detection of AD and can distinguish the stages of AD. FreeSurfer is **not** statistically significantly better for any structure. Comparing with [18], the proposed **Hybrid3** performs better for cerebral WM, cerebral cortex and ventricles. For general analysis, we recommend the **Hybrid3** approach as it works well for diseased **and** healthy appearing brains. Its computation time depends on the non-linear registrations performed to register atlases to target image and the number of training atlases. Once these registrations are performed (in parallel they take 30 min), the proposed algorithm only requires 7 minutes to segment one brain volume. For clinical practices where the subject is known to be similar in morphology to the set of training atlases, then our EASA or **EASA++** algorithm could be used alone (without the multi-atlas initialization step) in which case the entire segmentation requires just 7 minutes.

Qualitative results for two of Severe\_AD subjects are shown in Fig. 4, while qualitative results for all 28 subjects are shown in the supplemental file. In Fig. 4, each row is for an axial slice of one subject in Severe\_AD. The first column is the T1 input image; the second column is the result by WMV method; the third is FreeSurfer result; the fourth column comes from [18]; the fifth is the proposed **Hybrid3** result and the last column is ground truth (GT). The GT for each subject in Severe\_AD is obtained by taking the WMV label fusion results as the starting point and manually re-labeling it. As shown by yellow circles in Fig. 4,



**Fig. 4.** Method comparison on AD patients. Proposed **Hybrid3** provides visibly improved parcellation (circled regions) throughout the brain

neither FreeSurfer, a statistical atlas method, nor WMV a multi-atlas method are able to reliably segment diseased ventricles. Blue circles indicate the areas where voxel based FreeSurfer under-segments the cortex, while yellow circles show where WMV and [18] over-segment the cortex. In our tests, *only Hybrid3 is able to accurately segment the ventricles and cortex* of these diseased brains.

## 4 Conclusions and Future Work

We develop a new adaptive statistical atlas, **EASA++**, using structure specific relaxation priors, refined by a non-stationary relaxation map and isotropic pairwise class-specific MRF model. We propose a new hybrid approach **Hybrid3** that combines **EASA++** with the multi-atlas method. Our hybrid approach simultaneously segments 34 structures throughout the brain with state-of-the-art accuracy. We evaluate our method on normal brains and those with moderate and severe Alzheimer's. Qualitative and quantitative results demonstrate the superior performance of the proposed **Hybrid3** method to FreeSurfer for multiple structures including the cortex and ventricles. Since modern large datasets, such as ADNI and AIBL, have thousands of subjects, our fully automated proposed method is well suited for segmenting these images automatically across a wide range of disease severity. Future work entails improving accuracies for other small structures simultaneously and utilizing volumetry and thickness measures from our segmentations to quantify disease stage and help guide treatment.

## References

1. Ashburner, J., Friston, K.: Unified segmentation. *NeuroImage* **26**, 839–851 (2005)
2. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**(1), 26–41 (2008)



3. Cardoso, M., Clarkson, M., Ridgway, G., Modat, M., Fox, N., Ourselin, S.: Load: a locally adaptive cortical segmentation algorithm. *NeuroImage* **56**(3), 1386–1397 (2011)
4. Ellis, K., et al.: The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer’s disease. *Int. Psychogeriatr.* **21**(04), 672–687 (2009)
5. Fischl, B., Salat, D., Busa, E., Albert, M., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**(3), 341–355 (2002)
6. Gao, Y., Liao, S., Shen, D.: Prostate segmentation by sparse representation based classification. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 451–458. Springer, Heidelberg (2012)
7. Han, X., Hibbard, L., Oconnell, N., Willcut, V.: Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. In: *MICCAI*, pp. 297–304 (2010)
8. Iglesias, J., Liu, C., Thompson, P., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. *TMI* **30**(9), 1617–1634 (2011)
9. Liu, X., Montillo, A., Tan, E., Schenck, J.: iSTAPLE: improved label fusion for segmentation by combining STAPLE with image intensity. In: *SPIE Medical Imaging* (2013)
10. Mitchell, S., Bosch, J., Lelieveldt, B., van der Geest, R., Reiber, J., Sonka, M.: 3-d active appearance models: segmentation of cardiac MR and ultrasound images. *TMI* **21**(9), 1167–1178 (2002)
11. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. *Annu. Rev. Biomed. Eng.* **2**(1), 315–337 (2000)
12. Rousseau, F., Habas, P., Studholme, C.: A supervised patch-based approach for human brain labeling. *TMI* **30**(10), 1852–1862 (2011)
13. Shiee, N., Bazin, P.-L., Cuzzocreo, J.L., Blitz, A., Pham, D.L.: Segmentation of brain images using adaptive atlases with application to ventriculomegaly. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011. LNCS*, vol. 6801, pp. 1–12. Springer, Heidelberg (2011)
14. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *TMI* **18**(10), 897–908 (1999)
15. Wang, H., Suh, J.W., Das, S.R., Pluta, J., Altinay, M., Yushkevich, P.A.: Regression-based label fusion for multi-atlas segmentation. In: *CVPR*, pp. 1113–1120 (2011)
16. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *TMI* **23**(7), 903–921 (2004)
17. Wu, G., Kim, M., Wang, Q., Shen, D.: Hierarchical attribute-guided symmetric diffeomorphic registration for MR brain images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 90–97. Springer, Heidelberg (2012)
18. Yan, Z., Zhang, S., Liu, X., Metaxas, D., Montillo, A., AIBL: accurate segmentation of brain images into 34 structures combining a non-stationary adaptive statistical atlas and a multi-atlas with applications to Alzheimer’s disease. In: *ISBI* (2013)