Improving whole-brain segmentations through incorporating regional image intensity statistics

Christian Ledig[a*], Rolf A. Heckemann[bc], Alexander Hammers[bc] and Daniel Rueckert[a]

[a]Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK;
[b]The Neurodis Foundation, CERMEP, Lyon, France;
[c]Centre for Neuroscience (Hammersmith), Imperial College London, UK

ABSTRACT

Multi-atlas segmentation methods are among the most accurate approaches for the automatic labeling of magnetic resonance (MR) brain images. The individual segmentations obtained through multi-atlas propagation can be combined using an unweighted or locally weighted fusion strategy. Label overlaps can be further improved by refining the label sets based on the image intensities using the Expectation-Maximisation (EM) algorithm. A drawback of these approaches is that they do not consider knowledge about the statistical intensity characteristics of a certain anatomical structure, especially its intensity variance. In this work we employ learned characteristics of the intensity distribution in various brain regions to improve on multi-atlas segmentations. Based on the intensity profile within labels in a training set, we estimate a normalized variance error for each structure. The boundaries of a segmented region are then adjusted until its intensity characteristics are corrected for this variance error observed in the training sample. Specifically, we start with a high-probability "core" segmentation of a structure, and maximise the similarity with the expected intensity variance by enlarging it. We applied the method to 35 datasets of the OASIS database for which manual segmentations into 138 regions are available. We assess the resulting segmentations by comparison with this gold-standard, using overlap metrics. Intensity-based statistical correction improved similarity indices (SI) compared with EM-refined multi-atlas propagation from 75.6% to 76.2% on average. We apply our novel correction approach to segmentations obtained through either a locally weighted fusion strategy or an EM-based method and show significantly increased similarity indices.

1. INTRODUCTION

Whole brain anatomical segmentation of magnetic resonance (MR) images is an important task with the potential to assist clinical decision making in neuroimaging.[1–3] The labeling of an individual brain region enables automatic measurement of its volume, extents and shape. Manual delineation by trained experts is still considered the gold-standard for anatomical segmentation. However, manual methods do not scale well to the size of cohorts used in recent multicentre imaging studies, such as the Alzheimer's Disease Neuroimaging Inititative (ADNI, adni.loni.ucla.edu)[4] or TBIcare, a project that focuses on improving the diagnosis and treatment of traumatic brain injury (TBI) (www.tbicare.eu).

Widely used, fully automatic approaches to whole brain segmentation are based on propagating a set of manually or semi-automatically generated atlases to the target image. The propagated atlas label sets are then merged into a consensus segmentation using one of a choice of available fusion strategies.[5,6] Published methods range from simple, unweighted approaches[7–9] to more sophisticated strategies that weight the individual contributions locally, based on intensity information from the atlas and target images,[6] thereby achieving better accuracy. A popular representative of the multi-atlas segmentation category is MAPER,[10] which uses an unweighted strategy. A particularly promising approach is to use image intensity information to weight the contributing segmentations using the expectation-maximization (EM) algorithm.[11,12] Methods based on graph-cuts[13] are also based on image intensity information and are an alternative approach to EM refinement. In addition, fusion strategies based on statistical optimization have been proposed. The most popular representatives of this category are STAPLE[14] and its modifications.[15]

In this work, we employ intensity characteristics of various brain structures to improve labelings produced by two different multi-class, whole brain segmentation methods, namely MAPER-based[10] locally weighted fusion and EM refinement.[12] We calculate the intra-label intensity variance for each individual label of a segmentation produced by either of these automatic segmentation methods. By also calculating the intra-label intensity variance based on the gold-standard

* This work is partially funded under the 7th Framework Programme by the European Commision (http://cordis.europa.eu/ist/)

delineations we are able to estimate a correction value by which this intra-label intensity variance should be corrected. Based on this correction value we then modify single labels to better match the expected intra-label intensity variance. Furthermore, we show that merging segmentation results obtained with different techniques can improve the result of regional segmentations across the whole brain. Specifically we choose a different method depending on whether the region in question is cortical or subcortical. We used this approach in the recent "MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling". Our entry entitled "MALP-EM"[16] was evaluated as one of the top three performing methods in this challenge.

To evaluate the proposed approach, we generated label sets for 20 datasets of the OASIS database[17] and compared them with manual gold-standard segmentations. A distinct set of manually segmented images from the same cohort was used as training data. Using the proposed statistical correction step yields significantly improved similarity indices (SI) compared to our reference method MAPER.[10] The improvement is particularly distinct for cortical regions.

2. METHOD

2.1 Data

We used the data set provided as part of the "MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling". It consists of 35 T_1-weighted images with isotropic voxel size of $1 \times 1 \times 1$ mm from the OASIS database.[17] Corresponding labels had been created manually by experts† according to the protocol available at `www.braincolor.org`.[16] We split the data set, as done in the Grand Challenge, into training and testing subjects. The training data ($n = 15$) had been acquired from 15 subjects (10 female, 5 male, mean age: 23 ± 4.3), and testing data ($n = 20$) from 15 subjects, of which 5 subjects were scanned twice (12 female, 8 male, mean age: 40 ± 23).

2.2 Multi-Atlas Label Propagation with EM refinement (MALP-EM)

We calculate a subject-specific probabilistic brain atlas for an unsegmented T_1-weighted MR image $\mathbf{y_u}$ that is to be automatically labeled. The n voxels of $\mathbf{y_u}$ are indexed by $i = 1,\ldots,n$, so that the unseen image $\mathbf{y_u}$ is defined as $\mathbf{y_u} = \{y_1, y_2, \ldots, y_n\}$, where $y_i \in \mathbb{R}$ denotes the intensity value of the i-th voxel. We obtain the probabilistic priors by registering M manually generated atlases to the coordinate space of $\mathbf{y_u}$ using an approach that relies on a non-rigid registration based on free-form deformations (FFD).[18,19] Specifically we use an enhanced registration approach which incorporates tissue probability maps into the registration process (MAPER).[10]

After mapping the M label maps to the individual target space of $\mathbf{y_u}$, we create the probabilistic atlas using a locally weighted multi-atlas fusion strategy[5] based on a Gaussian weighted sum of squared differences on rescaled, intensity-normalized images. The transformed atlases are denoted by $\mathbf{A_m}$ with $m = 1,\ldots,M$. For K structures and M propagated atlases the probabilistic atlas at voxel i for class k is calculated as:

$$\pi_{ik} = \frac{\sum_{m=1}^{M} w_{im}\gamma_{ik}^{m}}{\sum_{j=1}^{K}\sum_{m=1}^{M} w_{im}\gamma_{ij}^{m}} \tag{1}$$

With the difference image $\mathbf{D_m}$, for the image $\mathbf{y_u}$ and atlas $\mathbf{A_m}$ defined as $\mathbf{D_m} = (\mathbf{y_u} - \mathbf{A_m})^2$, we calculate w_{im} and γ_{ik}^{m} as:

$$w_{im} = [G_\sigma \star \mathbf{D_m}]_i^{-1}\,, \qquad \gamma_{ik}^{m} = \begin{cases} 1, & \text{if atlas } \mathbf{A_m} \text{ assigns label } k \text{ to voxel } i \\ 0, & \text{else} \end{cases} \tag{2}$$

Here G_σ is a zero centered Gaussian with parameter σ and $\star$ denotes a convolution. We thus take the spatial distance of voxels from the voxel of interest into account when comparing neighborhoods. For our experiments we set $\sigma = 2.5$.

We then estimate the hidden segmentation of $\mathbf{y_u}$, denoted by $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n\}$, where $\mathbf{z}_i$ is a vector of size K and the k^{th} component represents the probability that a voxel i belongs to a region k, by means of the observed intensities employing the widely used EM optimization presented by van Leemput et al. (1999).[20] In this EM optimization it is assumed that the log-transformed intensities of voxels belonging to a certain label k are normally distributed with mean μ_k and standard deviation σ_k. The model parameters are thus $\phi = \{(\mu_1,\sigma_1),(\mu_2,\sigma_2),\ldots,(\mu_K,\sigma_K)\}$. Smoothness of the final segmentation is enforced through global and stationary Markov Random Fields (MRF).[21]

In this application, the EM algorithm consists in interleaving the estimation of the class probabilities z_{ik}, while assuming that the model parameters ϕ are fixed, with the maximization of the overall probability of observing an image $\mathbf{y_u}$.

†provided by Neuromorphometrics, Inc. (http://Neuromorphometrics.com/) under academic subscription.

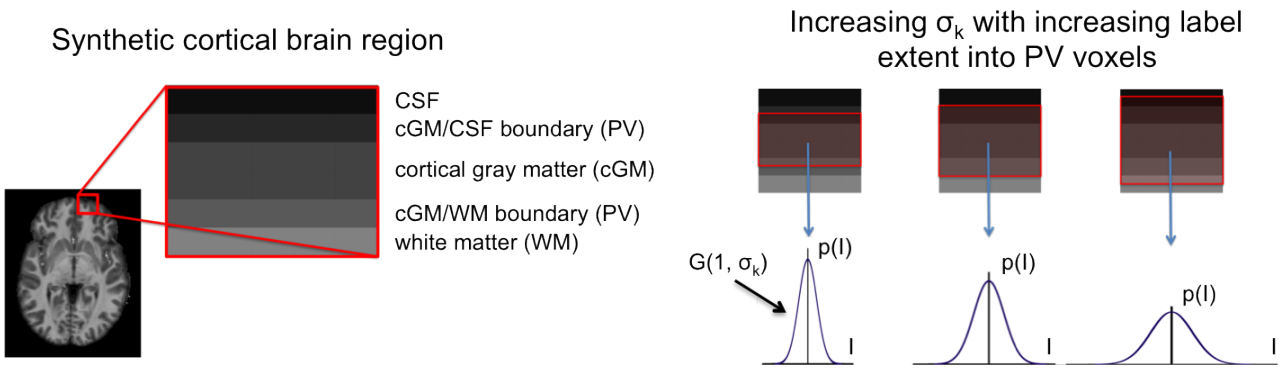

Figure 1. The intra-label intensity variance σ_k^2 is a good indicator for the number of partial volume voxels included at the boundary of a certain label k. This figure illustrates qualitatively the increase of the intra-class variance within a cortical label with increasing extent of the labeled region. We calculate the intra-class variance σ_k^2 by normalising each class to intensity mean $\mu_k = 1$ and estimate the according Gaussian distribution G(1,σ_k). We assume that the intensities within a certain label k are normally distributed.

2.3 Estimation of statistical correction rates

Statistical information based on the intensity distribution within a certain label is an important characteristic. We incorporate this statistical information into the segmentation process to enhance label overlaps. In particular, the standard deviation σ_k of intensities within a certain label k is a good indicator of the size of the label. For example, in cortical grey matter structures the label-specific standard deviation (or intra-class variance) will increase if voxels, often partial volume voxels, at the grey matter/white matter or grey matter/cerebrospinal fluid (CSF) boundary are added to the label. This assumption is illustrated in Figure 1. Ideally, automatically obtained labels should therefore have the same intra-class variance as the labels of the gold-standard.

Based on the available gold-standard segmentations and corresponding automatically computed segmentations, we calculate the normalized intra-class variance for each region k and each atlas $\mathbf{A_m}$, denoted by $(\overline{\sigma^m_{\text{gold},k}})^2$ in the gold-standard and $(\overline{\sigma^m_{\text{auto},k}})^2$ for the automatic method for region k . We normalize the standard deviation of each region k and atlas $\mathbf{A_m}$ by the intensity mean μ_k^m of this label as: $\overline{\sigma_k^m} = \frac{\sigma_k^m}{\mu_k^m}$. We then estimate a value Δ_k by which $\overline{\sigma_{\text{auto},k}}$ should be corrected in order to better match $\overline{\sigma_{\text{gold},k}}$. In particular we minimize:

$$\Delta_k = \min_\Delta \sum_{m=1}^{M} (\overline{\sigma^m_{\text{gold},k}} - (\overline{\sigma^m_{\text{auto},k}} + \Delta))^2 \tag{3}$$

2.4 Refinement based on statistical correction rates

With these class-specific learned statistical correction rates Δ_k we correct the intra-class standard deviation $\overline{\sigma_{\text{auto},k}}$ of an automatically obtained segmentation for a subject for which no gold-standard segmentation is available. We implement this in a refinement step, where we correct the intra-class statistics of each class by subsequently adding candidate voxels to the label. Voxels i are candidates for addition to a label k if they fulfil the requirements:

1) Their label probability z_{ik} is greater than α, with $\alpha = 0.1$ set heuristically.

2) They are labeled as CSF (background) or cerebral white matter.

Condition 1) ensures that only voxels with a minimum baseline probability are added to the label. Condition 2) ensures that labels are extended into the adjacent white matter or CSF, but not into neighboring cortical labels. We then sort all candidate voxels in descending order by their likelihood z_{ik} of being labeled as class k. From this sorted pool of voxels we successively add voxels to the label k until the intra-class standard deviation $\overline{\sigma_{\text{auto},k}}$ increases by Δ_k.

3. EXPERIMENTAL RESULTS

For evaluation, we used the 35 T_1-weighted MR scans with corresponding manually annotated label sets provided through the "MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling". We used a fixed subset of 15 scans and corresponding label sets as atlases ($M = 15$) and for the estimation of the correction rates σ_k. These 15 atlases were used to segment the remaining 20 images forming the testing set.

In total we compared six different approaches:

- MAPER: Standard MAPER as published by Heckemann et al. (2010);[10]
- MALP: Segmentations obtained according to Equation 1;
- MALP SC: MALP with the proposed statistical correction (SC);
- MALP-EM: MALP segmentations with EM-based refinement;
- MALP-EM SC: MALP-EM with the proposed SC;
- Merged: obtained by merging subcortical MALP SC with cortical MALP-EM SC labels.

In a first step we segmented the 15 training subjects using MALP (compare Equation 1) and MALP-EM in order to learn the statistical correction rates Δ_k. We determined all correction rates by comparing the intra-class variance based on segmentations obtained by locally weighted fusion (MALP) or EM-refinement (MALP-EM) to the statistics based on the gold-standard label sets as presented in Section 2.3. We then used these learned correction rates to correct both results obtained using MALP and MALP-EM on the testing set.

We segmented each of the images into the 138 regions provided. We calculated Dice overlaps (similarity indices, SI) with the available gold-standard segmentations on 134 regions, excluding, as in the Grand Challenge, the regions "vessel" and "cerebral exterior" on both right and left. SI values for the different segmentation methods are shown in Table 1 and visually in Figure 2. The improvements in cortical SI obtained through statistical correction are significant (Student's two-sided paired t-test, $p < 0.001$) for both the MALP and MALP-EM evaluations. We observed that the proposed statistical correction is significantly beneficial for cortical structures, while it did not lead to significant improvements in subcortical structures. The boxplot in Figure 3 illustrates the "success rate" (fraction of relabeled voxels for which the label change constituted an improvement) of the statistical correction step. For cortical voxels, the mean success rate was 64%. For subcortical structures, it was 54%, confirming the observations made when comparing similarity indices.

In order to take advantage of the strengths of the respective approaches – locally weighted fusion for subcortical regions and EM-based refinement for cortical regions – we also provide results for a merged label set "Merged", which we obtained by combining the subcortical MALP SC with the cortical MALP-EM SC labels after statistical correction. The main reason for an improved average SI in subcortical regions, comparing "Merged" to MALP SC, is due to an improved cerebral white matter segmentation. The cerebral white matter MALP SC labels benefit from the combination with more accurate cortical MALP-EM SC labels. The effect of statistical correction and an exemplary segmentation result is illustrated in Figure 3.

We also compared our results "Merged" to the best performing method in the Grand Challenge called "PICSL_BC" with an average SI of 76.5%. "PICSL_BC" employs a learning-based wrapper method presented in Wang et al. (2011)[22] where segmentation bias with respect to the gold-standard segmentations is learned. The comparison with this result is reasonable since the same images were used for training and testing and overlaps are calculated based on the same 134 regions. No significant difference could be found for the average similarity index over all 134 regions.

4. DISCUSSION

To our knowledge this is the first method using learned intra-label intensity variance errors as intensity-based statistical characteristic to improve multi-atlas whole brain segmentations. We used a similar approach incorporating statistical characteristics, also called MALP-EM, in the segmentation challenge "MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling".

Our results indicate that incorporating learned statistical characteristics in the segmentation approach has the potential to improve label overlaps. In this work we have shown that the consideration of the intra-label intensity variance is

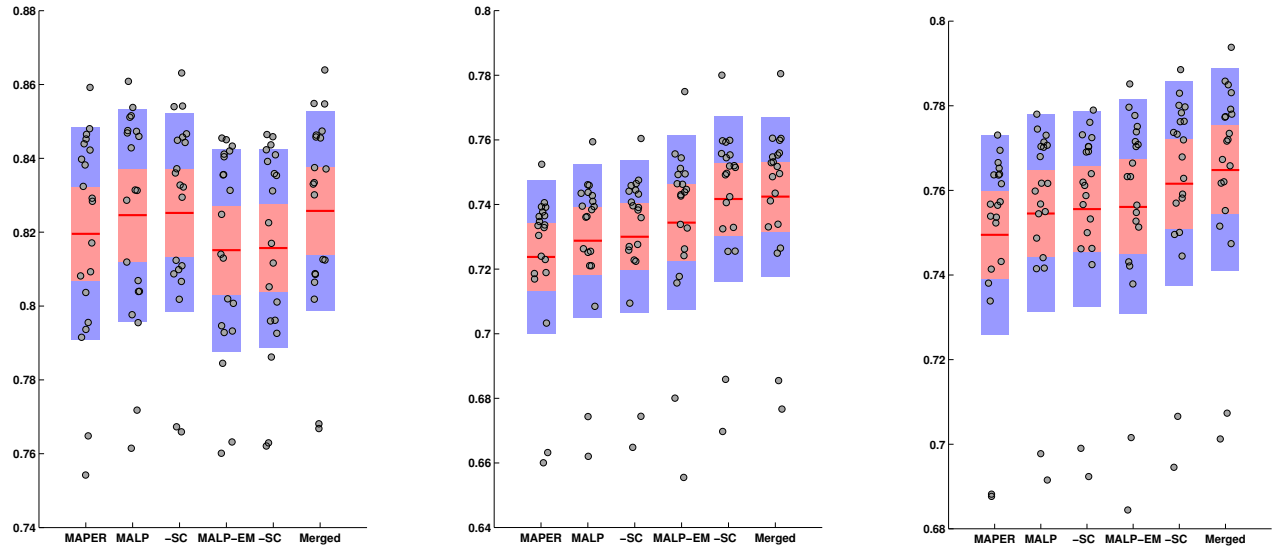

Figure 2. Boxplot of all compared methods applied to the 20 testing subjects. Average similarity indices (SI) of 36 subcortical structures (left), 98 cortical structures (middle) and all 134 regions (right).

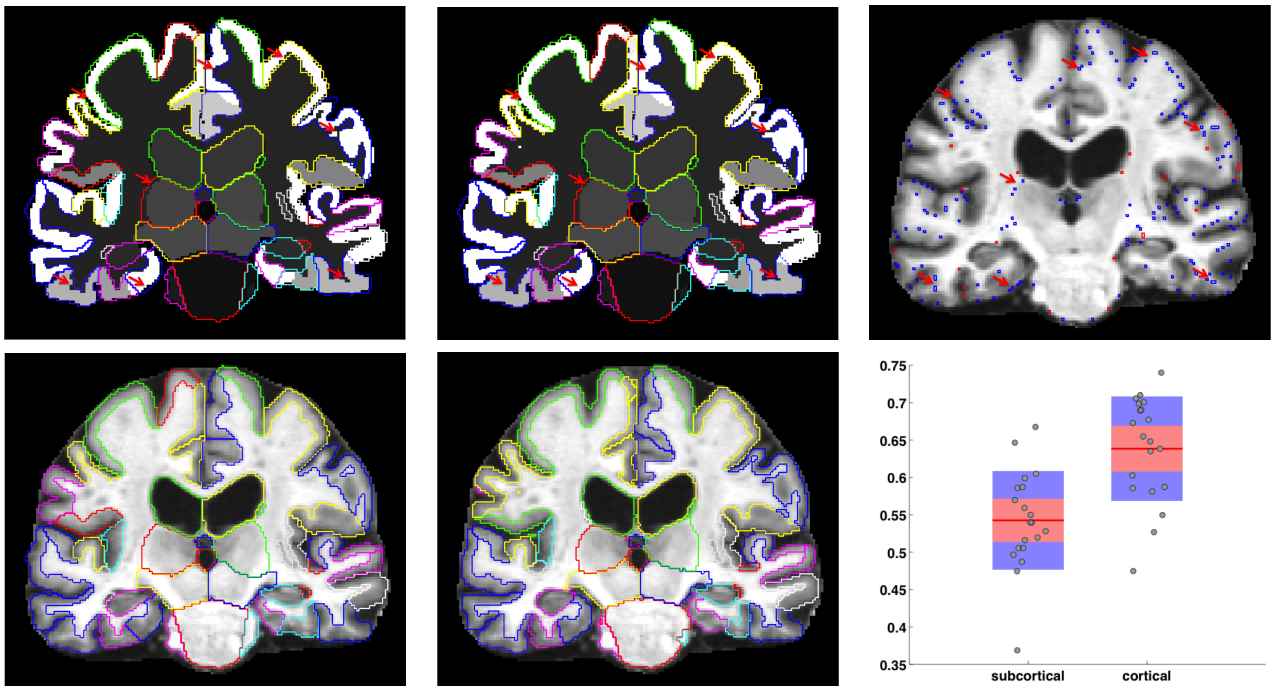

Figure 3. Label maps (greyscale) of MALP-EM (top, left) and MALP-EM with corrected statistics (top, middle) with overlaid gold-standard segmentation. Corresponding T_1-weighted MR image with voxels labeled correct (blue) and false (red) through statistical correction (top, right). Exemplary improved regions indicated through red arrows. An exemplary coronal slice segmented by a trained expert (bottom, left) and with our proposed method "Merged" (bottom, middle). The success rate for the 20 testing subjects with which a certain voxel is relabeled during the statistical correction is illustrated in the box plot (bottom, right).

	MAPER	MALP	+ SC	MALP-EM	+ SC	Merged
SI 134 classes	74.9	**75.5**	75.6	75.6	**76.2**	**76.5**
SI 36 subcort. classes	82.0	**82.5**	82.5	81.5	81.6	82.6
SI 98 cort. classes	72.4	**72.9**	**73.0**	73.4	**74.2**	74.2

Table 1. Similarity indices (SI) averaged over 20 subjects for all 134 classes, all 36 subcortical classes and all 98 cortical classes obtained with MAPER using majority vote, MALP using a locally weighted fusion strategy and optional statistical correction SC, MALP-EM refining MALP with the EM algorithm and optional statistical correction SC. "Merged" denotes a label set obtained by merging the cortical labels from MALP-EM SC and the subcortical labels from MALP SC. bold = significantly better than best method in columns to the left (Student's two-sided paired t-test, $p < 0.001$).

significantly beneficial for cortical regions. However, it did not significantly influence segmentation accuracy of subcortical structures. There are two potential reasons for this: First, the proposed statistical correction is a rather coarse procedure. If the baseline segmentation already has an accuracy of 80% SI or higher, it does not seem to yield a strong benefit. For such structures, which are prevalent in the subcortical set, more sophisticated correction models may need to be learned. Second, the manual segmentation protocol for subcortical labels defines some boundaries on geometrical criteria, rather than intensity gradients. Our statistical correction procedure does not incorporate this type of criterion and may thus falsely expand a given subcortical label. For cortical labels, on the other hand, the statistical correction step accounts for differences in the partial volume characteristics of segmentation methods, leading to significant benefits.

In future work, we plan to investigate more complex models to further improve segmentation accuracy for cortical as well as subcortical structures.

5. CONCLUSION

We presented a novel, computationally efficient technique to refine cortical labels of multi-class whole brain segmentations based on learned statistical characteristics. By comparison with EM-refined label maps, we showed that the correction of intensity characteristics is an effective approach to enhance the quality of a given segmentation significantly. Using a recently published atlas set, we confirmed that locally weighted fusion strategies outperform simple majority vote fusion, and that results obtained through multi-atlas fusion can be improved in the cortex using refinement based on EM optimization. We furthermore confirmed that segmentation accuracy can be increased in a straightforward way by combining segmentation results obtained with different labeling techniques. Our method is comparable to the leading methods presented at a recent segmentation challenge.

REFERENCES

[1] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.

[2] S. Klöppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak, "Accuracy of dementia diagnosis–a direct comparison between radiologists and a computerized method," *Brain*, vol. 131, no. 11, pp. 2969–2974, 2008.

[3] R. Heckemann, A. Hammers, D. Rueckert, R. Aviv, C. Harvey, and J. Hajnal, "Automatic volumetry on MR brain images can support diagnostic decision making," *BMC Medical Imaging*, vol. 8, no. 1, p. 9, 2008.

[4] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, 2005.

[5] X. Artaechevarria, A. Munoz Barrutia, and C. d. S. Ortiz, "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.

[6] M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1714–1729, 2010.

[7] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.

[8] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[9] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, 2009.

[10] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *NeuroImage*, vol. 51, no. 1, pp. 221–227, 2010.

[11] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage*, vol. 49, no. 3, pp. 2352–2365, 2010.

[12] C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen, R. A. Heckemann, A. Hammers, and D. Rueckert, "Multi-class brain segmentation using atlas propagation and EM-based refinement," *Proceedings of ISBI 2012*, pp. 896–899, 2012.

[13] R. Wolz, R. A. Heckemann, P. Aljabar, J. V. Hajnal, A. Hammers, J. Lötjönen, and D. Rueckert, "Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI," *NeuroImage*, vol. 52, no. 1, pp. 109–118, 2010.

[14] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, 2004.

[15] A. J. Asman and B. A. Landman, "Non-local staple: An intensity-driven multi-atlas rater model," vol. 7512, pp. 426–434, Lecture Notes in Computer Science, MICCAI 2012, 2012.

[16] C. Ledig, R. A. Heckemann, P. Aljabar, R. Wolz, J. V. Hajnal, A. Hammers, and D. Rueckert, "Segmentation of MRI brain scans using MALP-EM," *MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling*, pp. 79–82, 2012.

[17] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (oasis): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[18] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, and D. J. Leach, M. O.and Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[19] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 3, pp. 278–284, 2010.

[20] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897–908, 1999.

[21] M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin, "Load: A locally adaptive cortical segmentation algorithm," *NeuroImage*, vol. 56, no. 3, pp. 1386–1397, 2011.

[22] H. Wang, S. R. Das, J. W. Suh, M. Altinay, J. Pluta, C. Craige, B. B. Avants, and P. A. Yushkevich, "A Learning-Based Wrapper Method to Correct Systematic Errors in Automatic Image Segmentation: Consistently Improved Performance in Hippocampus, Cortex and Brain," *NeuroImage*, vol. 55, no. 3, pp. 968–985, 2011.