NeuroImage 51 (2010) 221–227

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation ☆

Rolf A. Heckemann [a,b,*], Shiva Keihaninejad [a], Paul Aljabar [c], Daniel Rueckert [c], Joseph V. Hajnal [d], Alexander Hammers [a,b]
The Alzheimer's Disease Neuroimaging Initiative [1]

[a] Division of Neuroscience and Mental Health, Faculty of Medicine, Imperial College, London, UK
[b] The Neurodis Foundation (Fondation Neurodis), Lyon, France
[c] Department of Computing, Imperial College, London, UK
[d] Imaging Sciences Department, MRC Clinical Science Centre, Imperial College, London, UK

ARTICLE INFO

Article history:
Received 3 November 2009
Revised 15 January 2010
Accepted 21 January 2010
Available online 28 January 2010

Keywords:
Image segmentation
Human brain
Brain anatomy
Brain atlas

ABSTRACT

Automatic anatomical segmentation of magnetic resonance human brain images has been shown to be accurate and robust when based on multiple atlases that encompass the anatomical variability of the cohort of subjects. We observed that the method tends to fail when the segmentation target shows ventricular enlargement that is not captured by the atlas database. By incorporating tissue classification information into the image registration process, we aimed to increase the robustness of the method. For testing, subjects who participated in the Oxford Project to Investigate Memory and Aging (OPTIMA) and the Alzheimer's Disease Neuroimaging Initiative (ADNI) were selected for ventriculomegaly. Segmentation quality was substantially improved in the ventricles and surrounding structures (9/9 successes on visual rating versus 4/9 successes using the baseline method). In addition, the modification resulted in a significant increase of segmentation accuracy in healthy subjects' brain images. Hippocampal segmentation results in a group of patients with temporal lobe epilepsy were near identical with both approaches. The modified approach (MAPER, multi-atlas propagation with enhanced registration) extends the applicability of multi-atlas based automatic whole-brain segmentation to subjects with ventriculomegaly, as seen in normal aging as well as in numerous neurodegenerative diseases.

Introduction

Anatomical segmentation plays an increasingly important role in information extraction from medical images. In the human brain, for example, a growing number of studies show correlations between clinical parameters and morphometric descriptors derived through segmentation of magnetic resonance (MR) images (Colliot et al., 2008; Duchesne et al., 2008; Heckemann et al., 2008; Kloppel et al., 2008), demonstrating the potential of such descriptors as biomarkers for diagnosis, disease monitoring and drug discovery.

To be practicable for studies on large cohorts and in routine clinical applications, anatomical segmentation needs to be automated. Current literature reports that multi-atlas based methods are among the best, both when dividing the whole brain into multiple segments (Heckemann et al., 2006) and when targeting individual structures of interest [e.g. the hippocampus (van der Lijn et al., 2008; Chupin et al., 2009)]. Initial approaches to automatic segmentation were based on manual atlases of single subjects, the labels of which were transformed to fit a target subject. The required transformation was estimated using image registration. Idiosyncrasies of the atlas individual, its manual segmentation, the image registration process and the target subject's anatomy led to uncertainties in the quality of the result. Subsequent work showed how automatic atlas-based segmentation could be made more accurate and consistent by combining multiple atlases (Rohlfing et al., 2004; Klein et al., 2005; Heckemann et al., 2006). Comparing various registration algorithms, Klein et al. (2009) showed that segmentation

☆ The Foundation for the National Institutes of Health (www.fnih.org) coordinates the private sector participation of the $60 million ADNI public-private partnership that was begun by the National Institute on Aging (NIA) and supported by the National Institutes of Health. To date, more than $27 million has been provided to the Foundation for NIH by Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson & Johnson, Eli Lilly and Co., Merck & Co., Inc., Novartis AG, Pfizer Inc., F. Hoffmann-La Roche, Schering-Plough, Synarc Inc., and Wyeth, as well as non-profit partners the Alzheimer's Association and the Institute for the Study of Aging.

* Corresponding author. Division of Neuroscience and Mental Health, Faculty of Medicine, Imperial College, London, UK.
E-mail address: soundray@imperial.ac.uk (R.A. Heckemann).

[1] Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A listing of ADNI investigators is available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI Manuscript Citations.pdf.

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.01.072

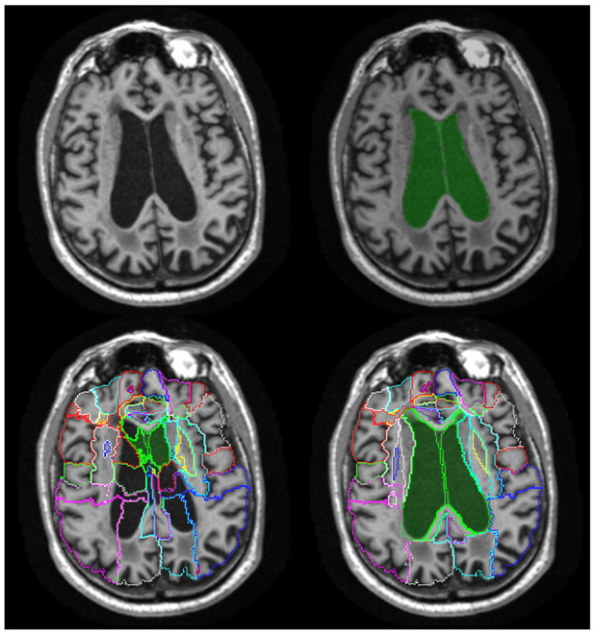

Fig. 1. Top left: transverse section through T1-weighted MR image volume in a subject with AD and ventriculomegaly (ADNI Image ID: I123024). Top right: reference segmentation of the ventricles based on tissue classification, superimposed as green label. Bottom: automatic segmentations superimposed as outlines, ventricle labels highlighted additionally in green. Bottom left: the purely intensity-based segmentation assigns incorrect labels to large parts of the lateral ventricles, while the cortical segmentations are acceptable. Bottom right: the MAPER procedure labels the lateral ventricles in their entirety with mild oversegmentation, while leaving cortical segmentations unaffected.

accuracy is correlated with the number of degrees of freedom in the registration. The same study also showed that free-form deformation-based registration using B-splines and normalized mutual information (NMI) as implemented in the Image Registration Toolkit [IRTK[2], Rueckert et al. (1999)] was among the best-performing registration methods for this purpose. In spite of the proven strength of this registration method for multi-atlas segmentation, we occasionally observed gross segmentation failures in target subjects with distended lateral ventricles (Fig. 1, bottom left panel) when using atlases based on images of young adult subjects. Commonly seen in normal aging as well as in subjects with neurodegenerative disease, ventricle distension or ventriculomegaly can cause the registration algorithm to align white matter in the atlases with the ventricular cerebrospinal fluid (CSF) in the target. This systematic failure limits the applicability of the method to sets of target images that are, within a certain range, anatomically similar to the set of atlases used, requiring manual intervention to identify target subjects that would violate this criterion. We set out to eliminate this problem by broadening the range of applicability of the method and enhancing the robustness of the method in the presence of abnormally sized ventricle spaces.

To achieve this, we modified the hierarchical registration approach underlying the segmentation method presented in Heckemann et al. (2006). This original approach uses the native MR intensity data and optimizes NMI at all six level of the hierarchy. The modification uses derived data sets instead for the first three levels. These are obtained using a tissue classification algorithm. We used cross-correlation to align the tissue maps, as it is a stronger driver than NMI for tissue probability maps. The resulting transformations were used to initialize registration of grey-scale intensity images at the remaining, higher levels of detail. The MAPER approach (multi-atlas propagation with enhanced registration) was tested on a diverse set of targets. Compared with the purely intensity-based method, the resulting segmentations agreed more strongly with reference segmentations, both in normal brains and in brains with pathology.

[2] http://www.doc.ic.ac.uk/dr/software.

Methods

Materials

T1-weighted 3D MR images from four different sources were used in this work. The first set (atlas set) consisted of 30 normal young adult subjects imaged at 1.5 Tesla. Segmentations of 83 structures, hand-drawn following protocols described by Hammers et al. (2003) and Gousias et al. (2008), available on http://www.brain-development.org, were used as label sources and as gold-standard references for assessing the quality of automatic segmentations.

The second set was selected from data acquired by the Oxford Project to Investigate Memory and Ageing [OPTIMA, Clarke et al. (1998)]. The total number of MR scans available was 88, acquired from 37 subjects with a median age of 75 years (range: 56–86 years). Our first observation of ventricular segmentation failure was made when processing this cohort, using the approach described in Heckemann et al. (2006). The phenomenon was seen in two images of different subjects. In the remaining 86 images, the whole-brain segmentation had been successful as judged by visual assessment. We therefore selected the two problematic images for this study. One subject fulfilled NINCDS criteria of probable Alzheimer's disease. The other subject was clinically normal, but among the oldest subjects in the cohort (85 years at the time of scanning).

The third source was the Alzheimer's Disease Neuroimaging Initiative (www.loni.ucla.edu/ADNI)[3]. From a database of 840 subjects with 4400 MR scans, we prospectively selected seven images based on criteria suggesting late-stage Alzheimer's disease (AD): (1) a recorded diagnosis of AD, (2) a follow-up examination at 24 months, (3) a recorded mini-mental state exam score ranging from 1 to 8 at 24 months (ADNI IDs of selected images: I91664, I103333, I123024, I123220, I123926, I132359, I132424).

The fourth set consisted of nine images of subjects affected by temporal lobe epilepsy (TLE). T1-weighted MRIs had been acquired at the National Society for Epilepsy in Chalfont St Peter, United Kingdom. Manual segmentations of both hippocampi in each image had been prepared following the protocol in Niemann et al. (2000). Acquisition and demographical details have been previously published (Hammers et al., 2007).

Image preprocessing

MR images were preprocessed using tools from the FSL suite (Version 4.1, Smith et al. (2004)). Preprocessing of the atlas, OPTIMA and TLE sets consisted of bias correction and brain extraction; ADNI images were already bias-corrected and thus subjected to brain extraction only. A coarse brain extraction was performed using "standard space roi" with the "−b" option for transforming a dilated brain mask from MNI standard space to the subject image. Bias correction was performed on the result using "fast" without partial volume estimation. The coarsely extracted, bias-corrected brain image was subjected to "bet" processing with the "−R" parameter for robust centre estimation, the "−m" option for obtaining a binary mask and, in the case of 3-Tesla images, a fractional intensity threshold of 0.4 to

[3] The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The Principal Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada.

Table 1
Evaluation of segmentations using the intensity-based method ("Intens"), ART, and the integrated method ("APER" for individual segmentations, "MAPER" for multi atlas based segmentations). Row labels show image identifiers for ADNI and OPTIMA subjects (O1 was an elderly subject with no clinical evidence of AD, O2 was a subject with probable AD). Column "Ref." shows the volume of the reference label generated for central CSF based on tissue classification [mean for young adults: 18.8 ± 8 ml, Hammers et al. (2003)]. After the visual review of the whole-brain segmentation, the ventricle label was superimposed on the MR grey scale display. A match is indicated by ■, a mismatch by □. JC = Jaccard coefficient, JC_{mean} = mean Jaccard coefficient for n = 30. The column "Volume error" shows the number of individual segmentations that underestimated the ventricular CSF volume by 50% or more (out of 30).

	Ref.	Fused					Individual			
		Visual review			JC		Volume error		JC_{mean}	
	ml	Intens	ART	MAPER	Intens	MAPER	Intens	APER	Intens	APER
(1) I91664	90	□	□	■	0.23	0.86	23	0	0.28	0.81
(2) I132359	74	■	□	■	0.69	0.69	2	0	0.63	0.66
(3) I103333	45	■	□	■	0.85	0.83	11	0	0.5	0.77
(4) I123024	137	□	□	■	0.10	0.84	28	0	0.15	0.80
(5) I123220	67	■	■	■	0.76	0.76	3	0	0.65	0.71
(6) I132424	52	■	■	■	0.59	0.59	0	0	0.55	0.55
(7) I123926	119	□	□	■	0.08	0.82	28	0	0.12	0.78
(8) O1	54	□	□	■	0.43	0.64	15	0	0.34	0.62
(9) O2	101	□	□	■	0.12	0.84	24	0	0.23	0.81

reduce the incidence of false-negative voxels compared to the default of 0.5. The binary mask was dilated three times with a $3 \times 3 \times 3$ kernel. The dilated brain mask was added to the "bet"-extracted image (with voxel values ranging typically from 0 to 300) in order to obtain a volume that contained a layer of voxels with a value of 1 in the close vicinity of the brain and values of zero beyond. Registration times were reduced by ignoring zero voxels during optimization and the presence of the layer of value 1 voxels prevented boundary artifacts at the brain surfaces, allowing their correct matching.

Tissue classification was performed on all data sets using the FAST tool from the FSL suite (Zhang et al., 2001). For the OPTIMA and ADNI data sets, we created reference labels of ventricular CSF semi-automatically. This was achieved by taking the binary CSF label generated by FAST as a starting point and separating ventricular from cisternal and sulcal CSF: a binary brain mask was eroded until it visually matched mainly central CSF; the CSF label was masked with the eroded brain mask; finally the ventricular label was "cleaned" by retaining only the largest connected component.

Image registration

Registrations were carried out on Intel dual processor servers with quadruple core Xeon processors ("Harpertown", Intel Corp., Santa Clara, CA, USA) with 16 GBytes of RAM per server, which are part of a cluster installation provided by the Imperial College High Performance Computing Service[4].

IRTK registration based on signal intensity

This baseline approach builds on Heckemann et al. (2006), but contains methodological improvements and is described in detail to highlight differences with the registration procedure underlying the new MAPER approach.

Using the Image Registration Toolkit [IRTK, Rueckert et al. (1999)], each atlas was aligned with the target. The corresponding pair of grey scale MR images were registered using rigid followed by affine registration, maximizing normalized mutual information (NMI). Following this global alignment, matching of anatomical detail was achieved by subsequent nonrigid registration. Displacements were applied to the atlas image via a lattice of control points and blended using B-spline basis functions, again maximizing NMI (Rueckert et al., 1999). A hierarchical coarse-to-fine approach was used with control point spacings of 20 mm, 10 mm, 5 mm and 2.5 mm. At each stage, the output transformation of the previous step was used as a starting point.

The stopping condition for the optimization was either no further improvement in NMI or the reaching of a maximum number of iterations (defined separately for each level). A third stopping criterion, δ, was based on the maximum amount of displacement across all control points in the most recent iteration. This was set to 1/10th of the control point spacing and was used to stop the optimization when iterations made only small updates to the transformation.

The registration required approximately 6 hours of processing time on a single CPU core per registration pair.

Integrated IRTK registration based on tissue classification and signal intensity (MAPER-type registration)

FAST tissue probability maps were combined into a multi-spectral image volume using the "fslmerge" tool, with each of the three channels of the image representing partial volume estimates for one tissue class. The atlas and target images were then aligned using rigid, affine and coarse nonrigid (20 mm) registration, maximizing the summed cross-correlation across all channels of the multi-spectral image volume. These registration steps were carried out with versions of the "rreg," "areg," and "hreg" tools from IRTK which had been modified to deal with multi-spectral volumes (these modifications have since become part of the published version of IRTK). Up to this point, the procedure considers exclusively the tissue classification output from FAST. The resulting transformation was then used as a starting point for detailed (10, 5 and 2.5 mm) registration using "hreg," which was run with parameters identical to those of the reference approach (IRTK registration based on signal intensity), using signal intensity image pairs and NMI as the similarity measure.

The MAPER type registration required approximately 4 hours of processing time on a single CPU core per registration pair.

Registration using ART and ANTS

The comparison by Klein et al. (2009) showed ART[5] and SyN (now a part of the ANTS toolkit[6]) to perform consistently well across the experiments in that study. We attempted to apply both ART and SyN to assess the robustness of these methods in comparison to IRTK.

We applied the 3dwarper program from ART (version released 12 March 2009) to register all 30 images in the atlas set to the ADNI, OPTIMA and TLE data sets. We attempted to use ANTS/SyN Version

[4] URL: http://www3.imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing.

[5] http://www.nitrc.org/projects/art/.

[6] http://www.picsl.upenn.edu/ANTS/.

1.6 in the same manner, but found unsatisfactory results due to the varied nature of the target data sets. ANTS software has been developed with a view to registering image pairs acquired using identical scanning parameters. The creators of ANTS are currently "working on a principled solution to [the problem of combining] data from different scanners" (B. Avants, personal communication).

Measures of segmentation success

To assess the accuracy of automatic segmentation in normal brain images, we used the first data set (30 normal subjects) and performed leave-one-out cross-comparisons. For each test subject, 29 individual segmentations resulting from propagating the other subjects' labels into the test subject's anatomical space were combined into one using vote–rule decision fusion (Heckemann et al., 2006). The overlap of the output label set with the manual label set of the respective target is used to express registration accuracy (i.e., we used manual segmentation as the gold standard).

Two commonly used ratios for measuring overlap are the similarity index (SI) or Dice coefficient [intersection divided by mean volume, Dice (1945)] and the overlap ratio or Jaccard coefficient (JC) [intersection divided by union, Jaccard (1901)]. Both JC and SI vary in the range zero to one but JC is a convex monotonically increasing function of SI: $JC = SI/(2-SI)$. This gives distributions of overlaps at the top end of the scale a greater variance when measured with JC (Shattuck, 2001). Mean JC values from all 83 segmented regions were used as summary measures (JC_m).

Automatic segmentations for the OPTIMA, ADNI and TLE sets were generated by propagating labels from all 30 atlases using three registration approaches: baseline (intensity-based) IRTK, MAPER, and ART. Vote-rule decision fusion was applied to consolidate all 30 segmentations in target space.

The accuracy of fused segmentations on the OPTIMA and ADNI sets was assessed visually. A given segmentation was rated a "mismatch" when the ventricle label outlines failed to align with the grey scale boundary between lateral ventricles and cerebral white matter. In addition, we assessed the individual segmentations quantitatively by comparing with the ventricular CSF reference label. The automatically generated ventricle labels (lateral ventricles, temporal horns and third ventricle) were merged and assessed against the reference label, using JC and relative volume error.

For the TLE set, the quality of the fused hippocampal segmentations was assessed by measuring the overlap using reference labels created by expert manual outlining Hammers et al. (2007).

Results

As noted in the introduction, the simple intensity-based procedure was unsuccessful at segmenting the lateral ventricles in a subset of patients with ventriculomegaly (example in Fig. 1). Visual review of multi-atlas segmentations generated with the purely signal-based procedure showed such failures in both OPTIMA subjects and in 3/7 ADNI subjects (I91664, I123024 and I123926). Multi-atlas segmentations generated after registration with ART failed in a similar fashion in 5 out of 7 ADNI subjects (I91664, I132359, I103333, I123024, I123926) and both OPTIMA subjects. Segmentations generated with MAPER showed good results for the whole brain in the two OPTIMA and seven ADNI subjects. Individual segmentations of these images also showed increased robustness of MAPER. Detailed results for OPTIMA and ADNI sets are shown in Table 1 and Fig. 2.

Leave-one-out cross-comparisons in a cohort of normal adults using both approaches yielded the results summarized in Table 2. The intensity-based approach resulted in a mean JC_m of 0.689. MAPER achieved a mean JC_m of 0.691, a significant improvement ($p<0.05$ on Welch's two-tailed paired t-test comparing 30 pairs of JC_m values). For most of the individual structures, no significant difference could be observed. Welch t-tests on 11/83 structures showed that MAPER was "significantly" better, while in 5/83 the MAPER overlap was worse, but none of these individual structure differences are still significant after correction for multiple comparisons.

Table 2
Leave-one-out cross-comparison results for two implementations of multi-atlas segmentation in healthy young adults. JC_m is shown for the aggregate measure, mean JC for individual structures ($n=30$). Where one of the methods yielded a significantly higher JC (Welch's two-tailed paired *t*-test before adjustment for multiple comparisons), the relevant JC value is shown in bold face. Intens: intensity-based method, np: not paired, ant.: anterior, post.: posterior, inf.: inferior, lat.: lateral, temp.: temporal, par.: parietal.

Structure name Aggregate (np)	Right hemisphere		Left hemisphere	
	Intens 0.689	MAPER 0.691	Intens	MAPER
Temporal lobe				
Hippocampus	0.716	0.716	0.700	0.698
Amygdala	0.656	0.658	0.646	0.645
Ant. temp. lobe, medial part	0.746	**0.756**	0.752	0.755
Ant. temp. lobe, lat. part	0.625	**0.642**	0.626	0.632
Parahippocampal gyri	0.703	0.704	0.692	0.692
Superior temp. gyrus, post. part	0.786	0.791	0.787	**0.792**
Middle and inf. temp. gyrus	0.787	**0.793**	0.777	0.782
Fusiform gyrus	0.635	0.639	0.596	0.599
Post. temp. lobe	0.762	0.761	0.748	0.752
Structure name	Intens	MAPER	Intens	MAPER
Superior temp. gyrus, ant. part	0.709	**0.714**	0.702	0.710
Posterior fossa				
Cerebellum	0.935	0.935	0.938	0.938
Brainstem (np)	0.883	0.883		
Insula and cingulate gyri				
Insula	0.767	0.767	0.766	0.766
Cingulate gyrus, ant. part	0.660	**0.669**	0.714	**0.721**
Cingulate gyrus, post. part	0.736	0.738	0.698	**0.703**
Frontal lobe				
Middle frontal gyrus	0.749	0.748	0.752	0.752
Precentral gyrus	0.733	0.732	0.742	0.743
Ant. orbital gyrus	**0.618**	0.607	0.658	0.651
Inf. frontal gyrus	0.736	0.736	0.732	0.732
Superior frontal gyrus	0.758	0.762	0.765	0.767
Medial orbital gyrus	**0.614**	0.609	0.617	0.616
Lat. orbital gyrus	0.419	0.420	0.468	0.462
Post. orbital gyrus	0.627	0.629	0.645	0.649
Subgenual frontal cortex	0.607	0.608	0.573	0.574
Subcallosal area	0.338	0.338	0.336	0.336
Pre-subgenual frontal cortex	0.510	**0.531**	0.474	**0.494**
Occipital lobe				
Lingual gyrus	0.733	0.734	**0.717**	0.711
Structure name	Intens	MAPER	Intens	MAPER
Cuneus	0.702	0.700	0.705	0.700
Lat. remainder of occipital lobe	0.729	0.729	0.717	0.715
Parietal lobe				
Straight gyrus	0.688	0.686	**0.702**	0.693
Postcentral gyrus	0.713	0.712	0.705	0.706
Superior par. gyrus	0.769	0.769	0.774	0.776
Inf. lat. Remainder of par. lobe	0.763	0.761	0.746	0.748
Central structures				
Caudate nucleus	0.806	0.806	0.811	0.811
Nucleus accumbens	**0.530**	0.524	0.512	0.510
Putamen	0.804	0.805	0.812	0.812
Thalamus	0.800	0.799	0.794	0.795
Pallidum	0.628	0.628	0.626	0.628
Corpus callosum (np)	0.766	0.766		
Substantia nigra	0.548	**0.555**	0.548	0.549
Ventricles				
Lat. ventricle (excl. temp. horn)	0.817	0.817	0.833	0.833
Lat. ventricle, temp. horn	0.517	0.516	0.482	0.481
Third ventricle (np)	0.710	0.710		

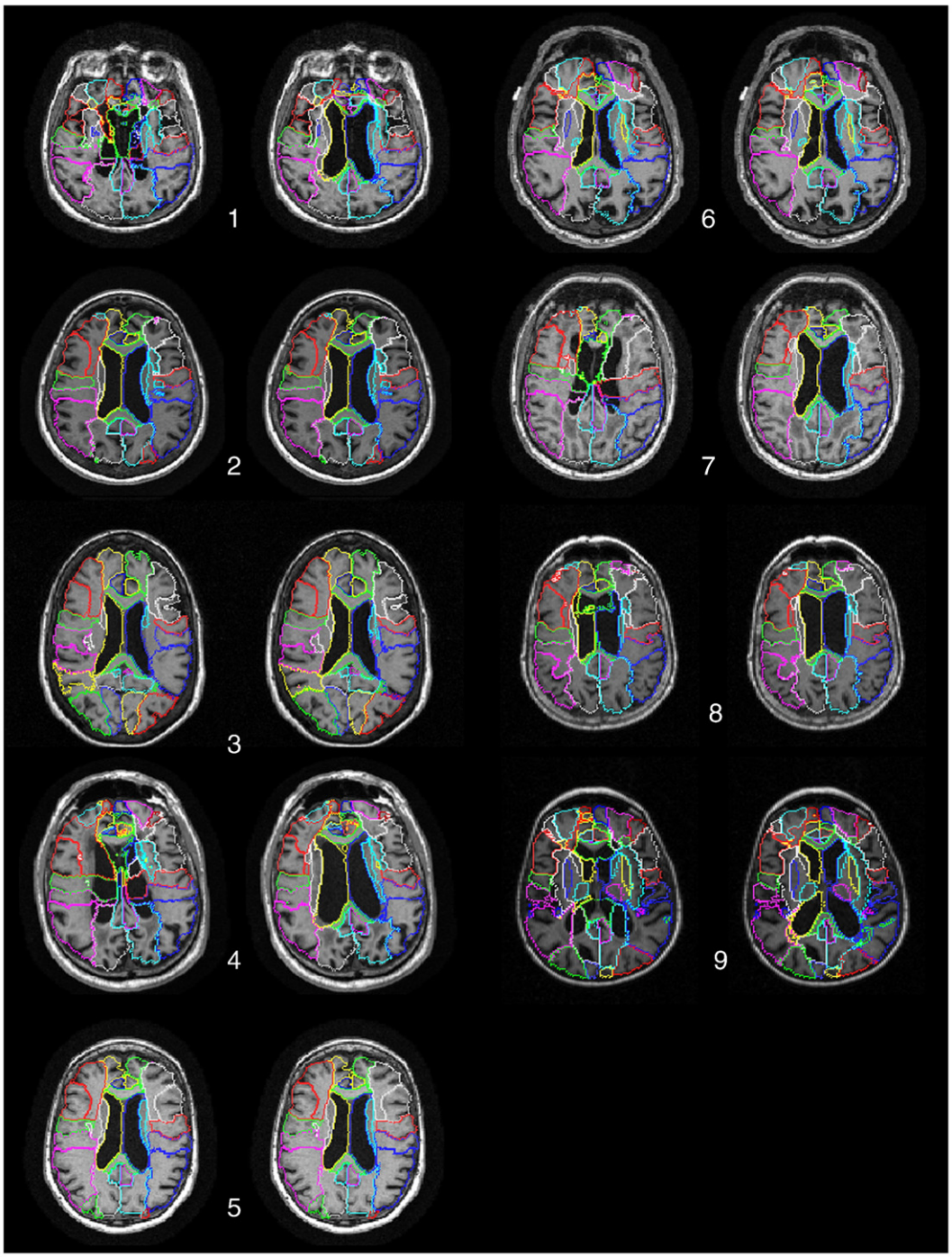

Fig. 2. Intensity-based (left) and MAPER (right) segmentations paired for comparison. Label outlines are superimposed on transverse sections selected to show the lateral ventricles. Subjects are shown in the same order as in Table 1.

In the temporal-lobe epilepsy cohort, the intensity-based and MAPER approaches yielded near identical overlap results for hippocampi (mean JC for right and left hippocampi: intensity-based 0.644, MAPER 0.645, ART 0.638, $n = 14$).

Discussion

The novel result of this study is that integrating tissue class information into intersubject brain image registration processes can extend the range of applicability of automatic segmentation based on such registrations. We found that the modification maintains the high level of accuracy of multi atlas label propagation with decision fusion, while increasing robustness in applications where segmentation targets with large aberrations of ventricular volume are encountered.

This work was motivated by the observation that an automatic segmentation procedure using multi-atlas label propagation and decision fusion, previously shown to work very well in normal human brain images (Heckemann et al., 2006), failed in some subjects with grossly distended lateral ventricles. The registration procedure aligned white matter in the atlas with ventricular CSF in these cases. The failure was evident in the result of intermediate nonrigid registration steps with hundreds to thousands of degrees of freedom. Subsequent hierarchical registration steps (high dimensional with millions of degrees of freedom, but limited lengths of control point

displacement) did not compensate the mismatch. In line with expectations, similar underestimations of enlarged ventricles were seen when basing the segmentation on ART. This registration algorithm is also based on high-dimensional free-form deformations, but uses normalized cross-correlation as the similarity measure.

To address the problem, we modified the segmentation procedure in such a way that an improved starting estimate for high-dimensional registration would be achieved by using tissue class information to coarsely pre-align the atlas image with the target image. As a baseline for comparison, we used a purely intensity-based procedure, where all the registration steps were carried out with the T1 signal intensity images. All other registration parameters were identical. We found that the modified procedure (MAPER) increases the robustness of the method when applied to images of subjects with ventriculomegaly, without sacrificing accuracy. This absence of a disadvantage is shown in a different application (hippocampal segmentation in subjects with temporal lobe epilepsy). It is also shown for regions beyond the ventricles in the leave-one-out cross-comparison based on atlases of normal subjects: on an aggregate measure based on 83 regions, the modification even yields a small improvement.

MAPER also led to a reduction in processing time by one third—this is merely an unexpected side effect, as the focus has been on increasing robustness before optimizing the speed of the segmentation process.

In related work, Khan et al. (2008) have found that deformable registrations can be improved through initialization using Freesurfer (Fischl et al., 2004). Their approach requires individual setup for each region of interest. Accordingly, the authors report results only for a small number of regions, albeit on a spectrum of pathologies. For patients with Alzheimer's disease, only hippocampal segmentations are reported. Comparisons were made with manual segmentations and results provided as SI (MAPER SI results for healthy subjects are shown in brackets): right caudate nucleus 0.81 (0.90), putamen 0.83 (0.90), pallidum 0.74 (0.77), thalamus 0.86 (0.89), hippocampus 0.75 (0.83; for TLE subjects: 0.78). We conclude that MAPER should be preferred over the approach of Khan et al. if simultaneous segmentation of multiple regions at high levels of accuracy are required.

Another anatomical segmentation method that considers tissue class information is HAMMER (Shen and Davatzikos, 2002). To obtain adequate segmentation results with HAMMER, an input volume is needed that distinguishes between ventricles and other CSF spaces. As this spatial classification would need to be prepared either manually or with other software, a direct comparison between HAMMER and MAPER will not be meaningful.

We have shown that MAPER generates detailed anatomical segmentations covering the entire brain even in the presence of gross ventricle distension. Nevertheless, in applications that require delineation of a single region of interest, specialized approaches may be superior. For the lateral ventricles, Chou et al. (2008) present a multi-atlas label propagation approach where the transformed labels are converted into mesh-based surface models and averaged. The resulting segmentations show a high level of agreement with reference segmentations and reveal ventricular shape differences related to the subject's genetic makeup.

The problem addressed by our proposed enhancement results from anatomical discrepancies between the atlas population and the target subjects. An alternative solution would be to manually generate atlases specifically for an elderly or demented population, but this would involve a large amount of expert effort. Another option would be to propagate atlases via intermediate subjects that bridge the anatomical dissimilarity between a specific subject and the atlas population. This approach has been shown to work well for hippocampal segmentation when intermediate subjects are chosen based on intersubject similarity of the hippocampal region ["LEAP", Wolz et al. (2009)]. LEAP is designed with a view to improving the segmentation of a single specified structure, while MAPER improves the quality of whole-brain segmentations. Still, the potential of combining LEAP and MAPER will be investigated in future work.

Conclusion

Integrating tissue class information into intersubject brain image registration processes can improve the resulting anatomical correspondence estimate. The MAPER method is suitable for creating accurate atlas-based segmentations, while being more robust in the presence of pathology than previous approaches.

Acknowledgments

The authors thank Professor David Smith, Oxford, for the provision of data from the OPTIMA study.

RAH was supported by a research grant from the Dunhill Medical Trust, UK.

References

Chou, Y.-Y., Lepore, N., de Zubicaray, G.I., Carmichael, O.T., Becker, J.T., Toga, A.W., Thompson, P.M., 2008. Automated ventricular mapping with multi atlas fluid image alignment reveals genetic effects in Alzheimer's disease. NeuroImage 40 (2), 615–630 April URL http://dx.doi.org/10.1016/j.neuroimage.2007.11.047.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, 2009. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19 (6), 579–587 June URL http://dx.doi.org/10.1002/hipo.20626.

Clarke, R., Smith, A.D., Jobst, K.A., Refsum, H., Sutton, L., Ueland, P.M., 1998. Folate, vitamin B12, and serum total homocysteine levels in confirmed Alzheimer disease. Arch. Neurol. 55 (11), 1449–1455 November URL http://view.ncbi.nlm.nih.gov/pubmed/9823829.

Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., Dubois, B., Garnero, L., Eustache, F., Lehéricy, S., 2008. Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. Radiology 248 (1), 194–201 July URL http://dx.doi.org/10.1148/radiol.2481070876.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302 July.

Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G.B., Collins, D.L., 2008. MRI-based automated computer classification of probable AD versus normal controls. Med. Imag., IEEE Trans. on 27 (4), 509–520 URL http://dx.doi.org/10.1109/TMI.2007.908685.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14 (1), 11–22 January URL http://dx.doi.org/10.1093/cercor/bhg087.

Gousias, I.S., Rueckert, D., Heckemann, R.A., Dyet, L.E., Boardman, J.P., Edwards, D.A., Hammers, A., 2008. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. NeuroImage 40 (2), 672–684 April URL http://dx.doi.org/10.1016/j.neuroimage.2007.11.034.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19 (4), 224–247 August URL http://dx.doi.org/10.1002/hbm.10123.

Hammers, A., Heckemann, R.A., Koepp, M.J., Duncan, J.S., Hajnal, J.V., Rueckert, D., Aljabar, P., 2007. Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof of principle study. Neuroimage 36 (1), 38–47 May URL http://dx.doi.org/10.1016/j.neuroimage.2007.02.031.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage July URL http://dx.doi.org/10.1016/j.neuroimage.2006.05.061.

Heckemann, R.A., Hammers, A., Rueckert, D., Aviv, R.I., Harvey, C.J., Hajnal, J.V., 2008. Automatic volumetry on MR brain images can support diagnostic decision making. BMC Med. Imaging 8, 9+ May URL http://dx.doi.org/10.1186/1471-2342-8-9.

Jaccard, P., 1901. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 241–272.

Khan, A.R., Wang, L., Beg, M.F.F., 2008. FreeSurfer-initiated fully automated subcortical brain segmentation in MRI using Large Deformation Diffeomorphic Metric Mapping. NeuroImage 41 (3), 735–746 July URL http://dx.doi.org/10.1016/j.neuroimage.2008.03.024.

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46 (3), 786–802 July URL http://dx.doi.org/10.1016/j.neuroimage.2008.12.037.

Klein, A., Mensh, B., Ghosh, S., Tourville, J., Hirsch, J., 2005. Mindboggle: automated brain labeling with multiple atlases. BMC Med. Imaging 5 (1), 7+ October URL http://dx.doi.org/10.1186/1471-2342-5-7.

Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681–689 March URL http://dx.doi.org/10.1093/brain/awm319.

Niemann, K., Hammers, A., Coenen, V.A., Thron, A., Klosterkotter, J., 2000. Evidence of a smaller left hippocampus and left temporal horn in both patients with first episode schizophrenia and normal control subjects. Psychiatry Res. 99 (2), 93–110 August URL http://view.ncbi.nlm.nih.gov/pubmed/10963985.

Rohlfing, T., Russakoff, D.B., Maurer, C.R., 2004. Performance based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans. Med. Imaging 23 (8), 983–994 August URL http://view.ncbi.nlm.nih.gov/pubmed/15338732.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging 18 (8), 712–721 August URL http://dx.doi.org/10.1109/42.796284.

Shattuck, D., 2001. Magnetic Resonance Image Tissue Classification Using a Partial Volume Model. NeuroImage 13 (5), 856–876 May URL http://dx.doi.org/10.1006/nimg.2000.0730.

Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging 21 (11), 1421–1439 November URL http://view.ncbi.nlm.nih.gov/pubmed/12575879.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (Suppl. 1), S208–S219 URL http://dx.doi.org/10.1016/j.neuroimage.2004.07.051.

van der Lijn, F., den Heijer, T., Breteler, M.M., Niessen, W.J., 2008. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. NeuroImage 43 (4), 708–720 December URL http://dx.doi.org/10.1016/j.neuroimage.2008.07.058.

Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D., the Alzheimer's Disease Neuroimaging Initiative, 2009. LEAP: learning embeddings for atlas propagation. NeuroImage October URL http://dx.doi.org/10.1016/j.neuroimage.2009.09.069.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. IEEE Trans. Med. Imaging 20 (1), 45–57 January URL http://dx.doi.org/10.1109/42.906424.