Patch-Based Segmentation from MP2RAGE Images: Comparison to Conventional Techniques

Erhard T. Næss-Schmidt[1,2], Anna Tietze[1,3], Irene K. Mikkelsen[1], Mikkel Petersen[1], Jakob U. Blicher[1,4], Pierrick Coupé[5], José V. Manjón[6], and Simon F. Eskildsen[1(✉)]

[1] Center of Functionally Integrative Neuroscience,
Aarhus University, Aarhus, Denmark
`seskildsen@cfin.au.dk`

[2] Hammel Neurorehabilitation Centre, Aarhus University, Aarhus, Denmark

[3] Department of Neuroradiology, Aarhus University Hospital, Aarhus, Denmark

[4] Department of Neurology, Aarhus University Hospital, Aarhus, Denmark

[5] Laboratoire Bordelais de Recherche en Informatique, PICTURA Research Group, Unité Mixte de Recherche CNRS (UMR 5800), Talencecedex, France

[6] Instituto de Aplicaciones de las Tecnologías de la Información y de las ComunicacionesAvanzadas (ITACA), Universitat Politècnica de València, Valencia, Spain

Abstract. In structural and functional MRI studies there is a need for robust and accurate automatic segmentation of various brain structures. We present a comparison study of three automatic segmentation methods based on the new T1-weighted MR sequence called MP2RAGE, which has superior soft tissue contrast. Automatic segmentations of the thalamus and hippocampus are compared to manual segmentations. In addition, we qualitatively evaluate the segmentations when warped to co-registered maps of the fractional anisotropy (FA) of water diffusion. Compared to manual segmentation, the best results were obtained with a patch-based segmentation method (volBrain) using a library of images from the same scanner (local), followed by volBrain using an external library (external), FSL and Freesurfer. The qualitative evaluation showed that volBrain local and volBrain external produced almost no segmentation errors when overlaid on FA maps, while both FSL and Freesurfer segmentations were found to overlap with white matter tracts. These results underline the importance of applying accurate and robust segmentation methods and demonstrate the superiority of patch-based methods over more conventional methods.

Keywords: Patch-based segmentation · MRI · volBrain · Freesurfer · FSL · MP2RAGE

1 Introduction

The extensive use of imaging techniques to investigate brain diseases and the need to outline specific region of interests (ROIs) for quantitative analysis emphasize the importance of accurate and robust segmentation methods. Accurate tracing of deep

brain structures, such as the thalamus and hippocampus, requires a high degree of expertise and preferably standardized outlining protocols. Even though an acceptable intra- and inter-rater reliability can be achieved using standardized protocols [1], manual segmentation is very time consuming. In large datasets, segmentation can become a bottleneck in post-processing and data analysis. Moreover, manual region outlining is prone to inconsistencies. Automatic or semi-automatic segmentation methods have the potential to solve these issues.

Several software solutions for automatic segmentation are publicly available. Functional MRI of the Brain (FMRIB) Software Library (FSL) and Freesurfer are tools frequently used for segmentation and appear to be reasonably reliable [2, 3]. However, there are still a potential to improve the automatic segmentation methods, especially in longitudinal studies [4] and for diseases that cause small structural changes.

Novel segmentation methods utilize redundancy in images to exploit a representative image library with corresponding validated structure labels [5–7]. These methods are called non-local means patch-based segmentation (NLM-PBS), since similar image patches are searched for in a non-local fashion, i.e. spatially located in a neighborhood around the target structure. NLM-PBS has been shown to be superior to conventional atlas-based techniques and even to other library-based methods [5, 6]. State-of-the-art segmentation methods like NLM-PBS have been shown to perform well even in a longitudinal setting [8].

An MRI sequence that has become widely used to obtain T1-weighted (T1w) anatomical images with good grey matter (GM)/white matter (WM) contrast is the magnetization-prepared rapid gradient-echo sequence (MPRAGE). However, at high static field strengths, increasing B1 field inhomogeneity leads to high intensity variations across the image. To mitigate this bias field, an improved MPRAGE sequence was recently proposed. By acquiring two MPRAGE images at different inversion times, this so-called MP2RAGE sequence is less influenced by B1 as well as M0 and T2* [9]. The resulting T1w image contrast is improved, but is also different from conventional MPRAGE images. Thus, current segmentation methods are not performing well on this new sequence [10].

To the best of our knowledge, the accuracy of different automated segmentation methods has not been compared using MP2RAGE images. Furthermore, NLM-PBS has not yet been directly compared to more conventional methods. In this study, we compared the performance of NLM-PBS (with two different libraries) to two widely used methods (Freesurfer and FSL) using manual segmentation as the gold standard. We measured the segmentation accuracy on two deep brain structures, thalamus and hippocampus, imaged with MP2RAGE.

2 Methods

2.1 Participants, MRI Acquisition and Pre-processing

For this study we collected 22 healthy subjects (age range 19–40 years, 12 females) from another internal research project. MP2RAGE images were obtained as part of the study protocol in all subjects, and 10 subjects were additionally examined with diffusion weighted imaging (DWI) as approved by the Regional Ethics Committee.

All subjects were scanned on a Siemens Magnetom Skyra 3T MRI system with a 32 channel head coil. MP2RAGE parameters were TR = 5 s, TI1 = 0.7 s, TI2 = 2.5 s, $\alpha1 = 4°$, $\alpha2 = 5°$ reconstructed at isotropic 1 mm^3 resolution (acquisition matrix: 240 × 256, 176 sagittal slices). The final MP2RAGE images were reconstructed by combining the two inversion times as described in [9]. DWI was acquired with 32 directions and 5 B0 maps. Parameters were TR = 10.9 s, TI = 2.1 s, reconstructed at isotropic 2.3 mm^3 resolution (acquisition matrix: 96 × 96, 38 axial slices).

MP2RAGE images have amplified background noise due to the reconstruction process. In our experience, Freesurfer and FSL perform poorly with this artificially amplified background noise, thus we masked out the background noise prior to applying the segmentation methods. Diffusion images were preprocessed using ExploreDTI [11]. We applied eddy current correction, motion correction and distortion correction before calculation of fractional anisotropy (FA) maps and co-registration to the MP2RAGE images. Using the inverse transformation, manual and automatic segmentation masks were then warped to DWI space and overlaid the FA maps.

2.2   Manual Segmentation

The thalamus and hippocampus from the 22 MP2RAGE images were manually segmented by an experienced neuroradiologist (EN) and a trained assistant (TA) using ITK-SNAP (www.itk-snap.org) [12]. First, EN manually traced the thalami in the axial plane using anatomical landmarks. Then, both EN and TA adjusted the thalami in all three principal planes using the protocol outlined by Power et al. [13]. The hippocampi were outlined according to the EADC-ADNI segmentation protocol [1] by TA supervised by EN. All segmentations were performed in MNI space to have similar orientation and make consistent decisions according to the protocols. The final segmentations were transformed back to scanner native space for comparison.

2.3   Automatic Segmentation Methods

We used a publicly available implementation (volBrain) of NLM-PBS [5]. For comparison we selected the publicly available and widely used segmentation tools FSL and Freesurfer. Default settings were used for all pipelines except for the added noise removal as described above. The following provides a brief overview of the three segmentation methods along with the applied settings.

FSL: Images were processed using FMRIB's Integrated Registration & Segmentation Tool (FIRST) from FSL v5.0, a tool to segment subcortical structures [14]. FIRST is a model-based segmentation tool, which uses training data from 317 manually segmented images. The manual labels are parameterized as surface meshes and modelled as a point distribution model. The deformable surfaces are then used to automatically parameterize the volumetric labels in terms of meshes and are constrained to preserve vertex correspondence across the training data. In addition, normalized intensities along the surface normals are sampled and modeled. We omitted the bias field correction step as MP2RAGE images are minimally affected by B1 field inhomogeneity. We used the

default settings of FIRST, as they have been empirically optimized and include shape and boundary correction.

Freesurfer: Images were processed with Freesurfer version 5.3 [15]. Briefly, the processing includes removal of non-brain tissue, spatial normalization, segmentation of the subcortical WM and deep GM structures, and intensity normalization. The segmentation maps are created using spatial intensity gradients across tissue classes and are therefore not simply reliant on absolute signal intensity. Therefore, both intensity and continuity information are being carried out in this segmentation method.

volBrain: The volBrain system (http://volbrain.upv.es) is based on an advanced pipeline providing automatic segmentations of several brain structures from T1w MRI. Images are denoised using an adaptive non-local means filter [16], registered to MNI space using ANTS [17], inhomogeneity corrected using SPM8 routines [18], and intensity normalized. Then, thalamus, hippocampus and six other subcortical structures are segmented using and updated version of NLM-PBS [5]. We tested the segmentation method using two different libraries: 1) the default volBrain library (external) of 50 conventional T1w images (MPRAGE and SPGR), and 2) our own manually segmented library of 22 MP2RAGE images in a leave-one-out fashion (local). In both cases, the images were flipped across the mid-sagittal plane to artificially increase the library size as done in related work [6].

For all segmentation methods, error logs were recorded, and quality was visually inspected with ITK-SNAP, overlaying the segmentations onto the T1w image.

2.4 Comparison Metrics

The segmentations obtained from the four automatic methods were compared to the manual segmentations using Dice similarity index (DSI) given by $\frac{2|A \cap B|}{|A|+|B|}$, where A is the set of voxels in the proposed segmentation and B is the set of voxels in the reference (manual) segmentation and $|\cdot|$ is the cardinality. DSI ranges from zero to one where one indicates a perfect match. Furthermore, the false positive and false negative rate (FPR, FNR) of the automatic segmentations were calculated.

3 Results

Figure 1 shows examples of manual segmentations and the corresponding automatic segmentations of the thalamus and hippocampus generated by the four evaluated methods overlaid on the T1w image and the FA map. As the examples illustrate, the thalamus is over-segmented by Freesurfer and to a lesser extent by FSL. As can be seen from the FA map, the internal capsule is partly included in the segmentation. volBrain local does not include any WM tracts, while volBrain external slightly over-segments the thalamus. This observation is reflected in the significantly larger FPRs of FSL and Freesurfer compared to volBrain using both libraries (Fig. 2). The consistent over-segmentation of FSL results in relatively few false negatives, while Freesurfer

Best volBrain local (thalamus DSI=0.929, hippocampus DSI=0.911):
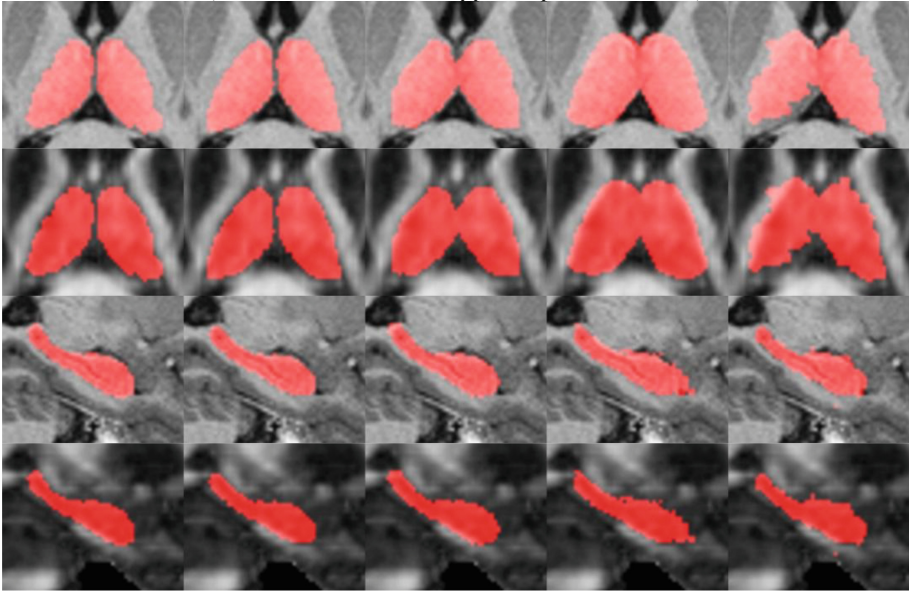
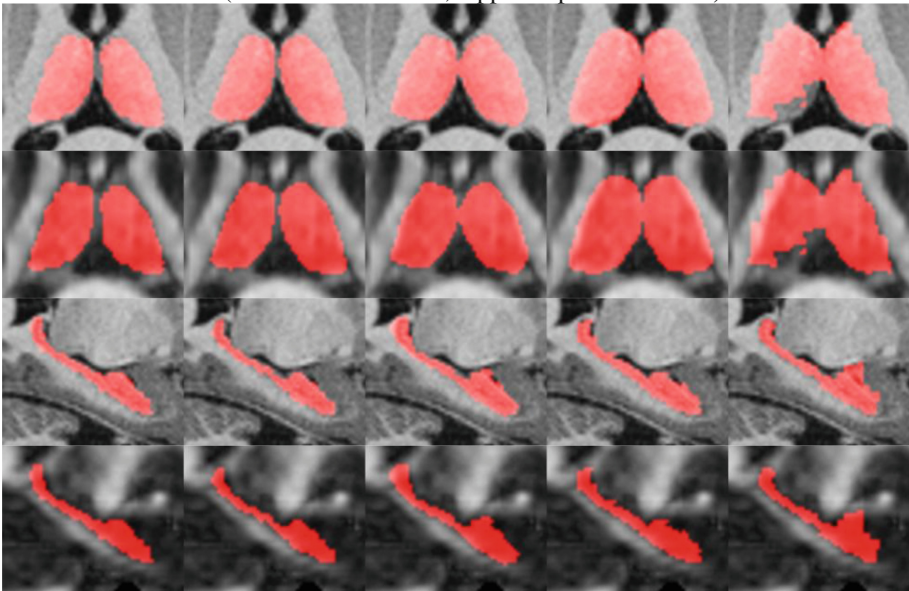Worst volBrain local (thalamus DSI=0.882, hippocampus DSI=0.885):

Fig. 1. Examples of manual and automatic segmentations of thalamus and hippocampus overlaid MP2RAGE and FA images. Examples are selected as respectively the best and worst volBrain local cases. From left to right: manual, volBrain local, volBrain external, FSL, Freesurfer.

also suffers from a relatively large FNR. In general, volBrain local performs best on thalamus segmentation with very high DSI (0.913 ± 0.014) followed by volBrain external (0.868 ± 0.024), FSL (0.806 ± 0.034) and Freesurfer (0.798 ± 0.049).

In terms of segmentation accuracy, the hippocampus follows a similar pattern with high DSI for volBrain local (0.892 ± 0.016), followed by volBrain external (0.859 ± 0.014), FSL (0.808 ± 0.017), and Freesurfer (0.771 ± 0.022) (Fig. 2). In terms of FPR and FNR, the pattern for hippocampus is slightly different from that of thalamus. FPR is reflecting the same order as DSI, with volBrain local performing best (8.9 % ± 2.7 %) and Freesurfer performing worst (41.2 % ± 7.2 %). However, in terms of FNR the methods are very similar with a relatively short range (mean FNR: 5.2 %– 12.4 %). The consistent over-segmentation of FSL and Freesurfer naturally leads to relatively low FNRs. volBrain local is the only method with well-balanced FPR and FNR for hippocampus, while volBrain using both libraries demonstrate balanced over- and under-segmentations on thalamus.
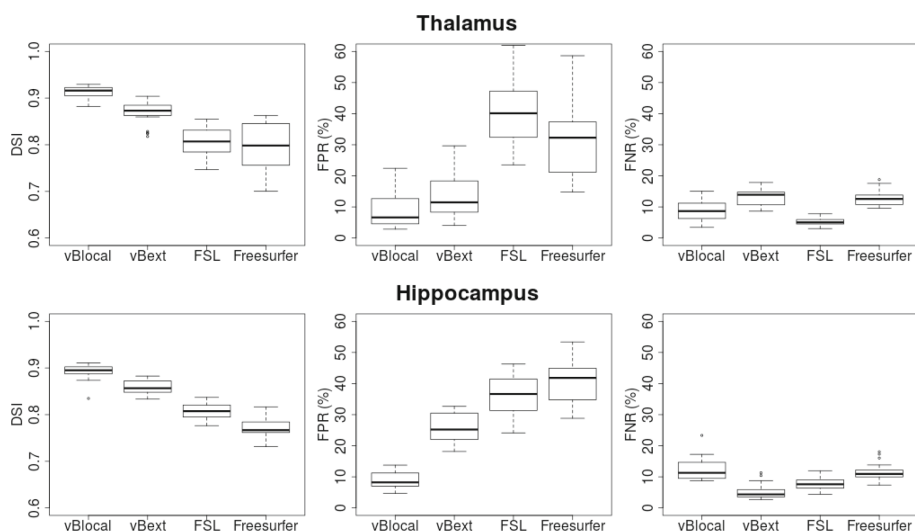

Fig. 2. Dice overlap, false positive rate and false negative rate for segmentations of the thalamus and the hippocampus using the four methods under evaluation. Box lines indicate 1st quartile, median, and 3rd quartile. Whiskers indicate extreme values, which are within the range of two times the length of the box. Dots are values outside this range.

4   Discussion

In this study we evaluated the performance of a recent patch-based segmentation method [5] and compared the results to those of FSL and Freesurfer, two widely applied methods in the neuroimaging community. Using MP2RAGE, a recently proposed T1w MRI sequence with superior soft tissue contrast, we tested the algorithms on two often investigated deep brain structures, the hippocampus and the thalamus.

The results demonstrated that the patch-based method outperforms both Freesurfer and FSL on these structures.

The accuracies we obtained on MP2RAGE images are similar to previously reported accuracies on the hippocampus using conventional MPRAGE [5, 7, 19]. For thalamus, average accuracies are in the same range as hippocampal accuracies for all four methods. However, for FSL and Freesurfer thalamic segmentation accuracies varied more than for hippocampus (Fig. 2). This may be caused by the fuzzy boundary of the thalamus where the image texture is important for making segmentation decisions, not just the image intensity and gradient. Patches can capture texture similarities, and this is perhaps why NLM-PBS attains consistently high accuracy on thalamus.

Using volBrain with a local library provided the best results. In this case the training data was matched perfectly to the test data, while the external library consisted of different imaging sequences from different scanners and manually labeled by different experts. The differences between local and external library reflects the importance of using a coherent labeling protocol and a similar image type within the template library. However, it is worth to note that even with these differences, volBrain external was able to provide good results highlighting the robustness of the method.

FSL and Freesurfer excessively over-segmented the structures with FPRs in the range 15 %–62 %. This resulted in consistent inclusion of WM in the segmentation of the two evaluated GM structures as qualitatively verified using FA maps. This is a major problem for morphometric as well as functional studies, where the over-segmentation leads to increased variance and impaired ability to detect differences and changes. Only volBrain external on hippocampus were found to over-segment. This may be due to differences in how the raters interpret the EADC-ADNI protocol.

The protocols for manual segmentation were based only on T1w images. As can be seen from the overlay on FA maps, it seems that WM voxels are occasionally included in the manual mask. This may be due to difficulty in determining the correct border when using T1w contrast only or simply due to co-registration errors between T1 and DWI. If the former, an improved manual segmentation may be obtained using multi-spectral data combining T1 and FA. Also, the automatic methods will most likely benefit from a multispectral approach. However, for a method to be versatile it is desired to work well on just T1w sequences as acquired in most MRI studies.

References

1. Boccardi, M., et al.: Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. Alzheimer's Dement. J. Alzheimer's Assoc. **11**(2), 126–138 (2015)
2. Nugent, A.C., et al.: Automated subcortical segmentation using FIRST: test-retest reliability, interscanner reliability, and comparison to manual segmentation. Hum. Brain Mapp. **34**(9), 2313–2329 (2013)
3. Han, X., et al.: Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage **32**(1), 180–194 (2006)

4. Mulder, E.R., et al.: Hippocampal volume change measurement: quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. NeuroImage **92**, 169–181 (2014)
5. Coupé, P., et al.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage **54**(2), 940–954 (2011)
6. Eskildsen, S.F., et al.: BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage **59**(3), 2362–2373 (2012)
7. Tong, T., et al.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage **76**, 11–23 (2013)
8. Coupé, P., et al.: Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. NeuroImage Clin. **1**(1), 141–152 (2012)
9. Marques, J.P., et al.: MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. NeuroImage **49**(2), 1271–1281 (2010)
10. Fujimoto, K., et al.: Quantitative comparison of cortical surface reconstructions from MP2RAGE and multi-echo MPRAGE data at 3 and 7 T. NeuroImage **90**, 60–73 (2014)
11. Leemans, A., et al.: ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In: 17th Annual Meeting of International Society Magnetic Resonance Medicine, Hawaii, USA (2009)
12. Yushkevich, P.A., et al.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage **31**(3), 1116–1128 (2006)
13. Power, B.D., et al.: Validation of a protocol for manual segmentation of the thalamus on magnetic resonance imaging scans. Psychiatry Res. **232**(1), 98–105 (2015)
14. Patenaude, B., et al.: A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage **56**(3), 907–922 (2011)
15. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage **9**(2), 179–194 (1999)
16. Manjon, J.V., et al.: Adaptive non-local means denoising of MR images with spatially varying noise levels. J. Magn. Reson. Imaging JMRI **31**(1), 192–203 (2010)
17. Avants, B.B., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011)
18. Weiskopf, N., et al.: Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). NeuroImage **54**(3), 2116–2124 (2011)
19. Morey, R.A., et al.: A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. NeuroImage **45**(3), 855–866 (2009)