Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2017 February 11; 10133: . doi:10.1117/12.2254425.

Multi-Scale Hippocampal Parcellation Improves Atlas-Based Segmentation Accuracy

Andrew J. Plassard^a, Maureen McHugo^b, Stephan Heckers^b, and Bennett A. Landman^{a,c}

^aComputer Science, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN USA 37235

^bPsychiatry and Behavioral Sciences, Vanderbilt University Medical Center, 1601 23rd Avenue South, Nashville, TN USA 37212

^cElectrical Engineering, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN USA 37235

Abstract

Known for its distinct role in memory, the hippocampus is one of the most studied regions of the brain. Recent advances in magnetic resonance imaging have allowed for high-contrast, reproducible imaging of the hippocampus. Typically, a trained rater takes 45 minutes to manually trace the hippocampus and delineate the anterior from the posterior segment at millimeter resolution. As a result, there has been a significant desire for automated and robust segmentation of the hippocampus. In this work we use a population of 195 atlases based on T1-weighted MR images with the left and right hippocampus delineated into the head and body. We initialize the multi-atlas segmentation to a region directly around each lateralized hippocampus to both speed up and improve the accuracy of registration. This initialization allows for incorporation of nearly 200 atlases, an accomplishment which would typically involve hundreds of hours of computation per target image. The proposed segmentation results in a Dice similarity coefficient over 0.9 for the full hippocampus. This result outperforms a multi-atlas segmentation using the BrainCOLOR atlases (Dice 0.85) and FreeSurfer (Dice 0.75). Furthermore, the head and body delineation resulted in a Dice coefficient over 0.87 for both structures. The head and body volume measurements also show high reproducibility on the Kirby 21 reproducibility population (R^2 greater than 0.95, $p < 0.05$ for all structures). This work signifies the first result in an ongoing work to develop a robust tool for measurement of the hippocampus and other temporal lobe structures.

Keywords

Multi-Atlas Segmentation; Hippocampus

Introduction

Seated in the temporal lobe, the hippocampus is one of the most studied structures in the human brain. In particular, the hippocampus is known to be involved in aging [1], memory

[2], and spatial navigation [3]. The hippocampus has also been identified as a key structure in the pathophysiology of Alzheimer's disease [4], schizophrenia, and epilepsy [5].

Recent advances in magnetic resonance imaging (MRI) have provided researchers with new ways to visualize and study structural and functional changes in the hippocampus. Researchers will commonly acquire a structural T1-weighted image which provides contrast for delineation of the hippocampus from its neighboring structures [6, 7]. Several automated segmentation techniques have been proposed which utilized T1-weighted structural image to automatically segment the hippocampus. One such example of this is multi-atlas segmentation with the BrainCOLOR protocol [8-11]. Whole-brain multi-atlas segmentations typically are limited to 10-20 atlases due to the lengthy non-rigid registrations necessary. A secondary approach is the FreeSurfer cortical reconstruction which uses an energy model to identify and segment the cortical surface [12]. FreeSurfer has been shown to require significant manual intervention in populations other than healthy adults, e.g., the elderly [13], and thus is not ideal for large-scale studies.

Multi-atlas segmentation has provided a robust and accurate technique for segmenting the brain [9-11], abdomen [14], optic nerve [11, 15, 16], and several other structures [8]. A typical multi-atlas segmentation involves non-rigidly registering 10 or more structural and label image volume pairs, herein atlases, to a target image and aligning the labels with the learned transform. These registered label volumes are statistically fused together to produce a segmentation more consistent with the target's anatomy than any individual registered atlas. This process is limited by the number of atlases that can be reasonably non-rigidly registered to a target image (as non-rigid registration can often take hours per image pair). Several works have considered automatic identification of structures of interest without non-rigid registration [14, 17]. Localization of structures can be used to initialize registration and allow for incorporation of more atlases into multi-atlas segmentation.

In this work, we present a novel approach to fully automated multi-atlas hippocampus segmentation. In a two-stage approach, the hippocampus is first localized with a traditional whole brain multi-atlas segmentation. Then, a localized multi-atlas approach is applied with nearly 200 manually labeled hippocampus. This is the first work considering a quantity of atlases of this scale.

Methods

In a typical multi-atlas segmentation, 10-15 atlases of a similar field of view are registered to a target image; the registered images are then fused with a label fusion algorithm [8]. We propose that segmentation of the hippocampus does not require registration of the full brain. Rather, a full brain segmentation is used to localize the hippocampus, and then each lateralized hippocampus was segmented with 195 atlases within a region of interest isolated to the temporal pole and hippocampus. The resultant lateral segmentation was then inserted into the full brain space.

Manual Labeling Protocol

195 subjects (90 healthy adults, 105 adults with schizophrenia) were scanned with a T1-weighted MPRAGE scan (TI/TR/TE=860/8.0/3.7ms). These 195 subjects were then manually traced with the left and right anterior and posterior hippocampus following the protocol in [18]. Briefly, for each patient their scan was first loaded into 3D Slicer for simultaneous visualization in the axial, coronal, and sagittal planes. For the right hippocampus, the scan was traced from the lateral section to medial in the sagittal plane. The tracing was then refined in the coronal plane. The anterior and posterior hippocampus were then divided coronally at the slice where only one cut through the hippocampus remained. This delineation was then verified sagittally. This procedure was repeated on the left hippocampus.

A series of 10 subjects were imaged with the same T1-weighted MPRAGE sequence as the atlases. These scans, herein the reproducibility atlases, were labeled with the previously described protocol by two unique raters, herein “Rater 1” and “Rater 2”. These 10 reproducibility atlases were distinct from the 195 atlas images and were used for validation.

Hippocampus Multi-Atlas Segmentation

The 195 manually traced subjects were used as atlases. To reduce computation time and allow for simultaneous use of all 195 atlases, a whole-brain multi-atlas segmentation (WBS) with the BrainCOLOR protocol (www.neuromorphometrics.com) was performed on each subject. For each scan, 15 atlases labeled with the BrainCOLOR protocol were registered to the scan using NiftyReg [19] for affine registration and the Advanced Normalization Tools (ANTs) and then Symmetric Normalization (SyN) algorithm [20] for non-rigid registration using the parameters described in [21]. The registered atlases and labels were then deformed to the target image space and fused with Hierarchical Non-Local Spatial STAPLE (H-NLSS) [9-11]. The bounding box for the left and right hippocampus were then determined from the WBS and a 5mm padding was added in each direction. The bounding box and padding region was extracted for the left and right hippocampus, resulting in 195 lateralized atlases.

For a given target scan, the WBS was performed following the same procedure as the atlases and the left and right hippocampi were extracted. For the right hippocampus, the 195 right hippocampus atlases were registered to the target image scan using NiftyReg [19] for affine registration and the Advanced Normalization Tools (ANTs) and then Symmetric Normalization (SyN) algorithm [20] for non-rigid registration using the parameters described in [21]. The registered atlases and labels were then deformed to the target image space and fused with Joint Label Fusion (JLF) [22]. This process was then repeated for the left hippocampus using the left hippocampus atlases. After both lateral hippocampus segmentations were completed, the segmentations were returned to their corresponding target image space.

Results

To test the accuracy and robustness of the proposed algorithm, herein denoted VU Seg, two experiments were considered. First, a population of 10 atlases, separate from the 195, were

used for validation. These atlases were labeled independently by two unique raters. Second, the Kirby21 multi-modal reproducibility dataset was segmented following the algorithm proposed above to show that the volumes are consistent between the scans.

Reproducibility Atlases

The reproducibility atlases were then segmented with the BrainCOLOR protocol, FreeSurfer [12], and the VU Seg. First, Dice similarity coefficient was calculated between the BrainCOLOR, FreeSurfer, and whole left and right hippocampus from the VU Seg against the whole left and right hippocampus from Rater 1 and Rater 2. VU Seg showed a significant improvement ($p < 0.05$; Wilcoxon sign-rank test) in Dice for the left and right hippocampuses compared with BrainCOLOR, FreeSurfer, and the reproducibility between Rater 1 and 2 (Figure 1).

Second, Dice was calculated between the left and right anterior and posterior hippocampus segments in VU Seg compared with each Rater 1 and Rater 2. For the right anterior and posterior hippocampuses, VU Seg showed a significantly higher Dice compared with either Rater 1 or 2 than the reproducibility between Rater 1 and 2. For the left anterior hippocampus VU Seg showed no change between the reproducibility in Dice calculated against Rater 2 but a slight decrease in Dice calculated against Rater 1 (median 0.905 for reproducibility and 0.891 for VU vs Rater 1). For the right anterior hippocampus VU Seg showed no change between the reproducibility in Dice calculated against Rater 2 but a slight decrease in Dice calculated against Rater 1 (median 0.901 for reproducibility and 0.877 for VU vs Rater 1; Figure 2).

Third, the mean volume of the left and right anterior and posterior hippocampus segments for each of the 10 reproducibility atlases was calculated for Rater 1, Rater 2, and VU Seg. The variance between the mean volume of the structure and each of the segmentations was calculated. VU Seg showed a significant decrease in variance for each structure compared with Rater 1 and Rater 2 ($p < 0.05$, Wilcoxon sign-rank test; Figure 3).

Kirby21 Multi-Modal Reproducibility

Collected in 2009, the Kirby21 Multi-Modal Reproducibility data is a population of 21 subjects scanned twice with a series of structural and quantitative 3T imaging sequences [23]. The purpose of this dataset is to test the reproducibility of algorithms on standardized imaging data. A T1-weighted MP-RAGE image volume was collected (TI/TR/TE=842/6.7/3.1) with a $1.0 \times 1.0 \times 1.2 \text{ mm}^3$ resolution. The VU Seg algorithm was performed on each subject in this population. The volumes between the first and second scanning sessions showed a significant correlation ($R^2 > 0.95$, $p < 0.05$; Figure 4A). The highest percent volume difference for any structure between any two scans was 3% (Figure 4B).

Computing Resources and Processing Time

All processing was done on Vanderbilt's Advanced Computing Center for Research & Education computational cluster (ACCRES). ACCRES consists of a diverse collection of compute nodes, consisting of processor speeds ranging from 1.9 to 3.0 GHz. The average

processing time for the WBS was 24 hours with the entire processing running serially. The average time for the hippocampus segmentation was 7 hours. On average, each registration took 72 seconds for the reduced-field-of-view hippocampus segmentation.

Discussion

There are two major contributions of this work. First, we proposed an algorithm for segmentation of the anterior and posterior hippocampus. This algorithm, VU Seg, was shown to be at least as good as the inter-rater variability for segmentation of the full hippocampus and the anterior/posterior hippocampus division. Furthermore, the VU Seg algorithm was shown to be significantly more accurate than both a modern multi-atlas segmentation approach with BrainCOLOR atlases and FreeSurfer. Visual inspection of the data (Figure 5) reinforces the quantitative improvement via qualitative appearance of a smoother, more accurate, and more consistent segmentation than the other algorithms considered. VU Seg was also shown to be reproducible on the Kirby21 dataset, a dataset which was acquired on a different physical scanner and at a lower resolution than the atlas population.

Second, this work presents a technique for incorporation of hundreds of atlases into a segmentation algorithm with practical resource constraints. Typical multi-atlas segmentations in the brain are limited to tens of atlases due to high computational requirements of non-rigid registration. This incorporation of hundreds of atlases can allow for more accurate segmentation of challenging structures, particularly when variable anatomies are present in the target population.

Acknowledgments

This project was supported by NSF CAREER 1452485, the National Center for Research Resources, Grant UL1 RR024975-01 (now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06), R01-EB006136, NS095291, T32LM012412, and the Michael J. Fox Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

References

1. Resnick SM, Pham DL, Kraut MA, et al. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *The Journal of Neuroscience*. 2003; 23(8):3295–3301. [PubMed: 12716936]
2. Bannerman D, Rawlins J, McHugh S, et al. Regional dissociations within the hippocampus—memory and anxiety. *Neuroscience & Biobehavioral Reviews*. 2004; 28(3):273–283. [PubMed: 15225971]
3. Eichenbaum H, Dudchenko P, Wood E, et al. The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron*. 1999; 23(2):209–226. [PubMed: 10399928]
4. Hyman BT, Van Hoesen GW, Damasio AR, et al. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science*. 1984; 225(4667):1168–1170. [PubMed: 6474172]
5. Yonekawa WD, Kapetanovic IM, Kupferberg HJ. The effects of anticonvulsant agents on 4-aminopyridine induced epileptiform activity in rat hippocampus in vitro. *Epilepsy research*. 1995; 20(2):137–150. [PubMed: 7750510]
6. Plassard AJ, Harrigan RL, Newton AT, et al. On the fallacy of quantitative segmentation for T1-weighted MRI. :978416–978416-7.

7. Deichmann R, Good C, Josephs O, et al. Optimization of 3-D MP-RAGE sequences for structural brain imaging. *Neuroimage*. 2000; 12(1):112–127. [PubMed: 10875908]
8. Iglesias JE, Sabuncu MR. Multi-Atlas Segmentation of Biomedical Images: A Survey. *arXiv preprint arXiv:1412.3421*. 2014
9. Asman AJ, Landman BA. Characterizing spatially varying performance to improve multi-atlas multi-label segmentation. *Inf Process Med Imaging*. 2011; 22:85–96. [PubMed: 21761648]
10. Asman AJ, Landman BA. Non-local STAPLE: an intensity-driven multi-atlas rater model. *Med Image Comput Comput Assist Interv*. 2012; 15(Pt 3):426–34. [PubMed: 23286159]
11. Asman AJ, Landman BA. Hierarchical performance estimation in the statistical label fusion framework. *Med Image Anal*. 2014; 18(7):1070–81. [PubMed: 25033470]
12. Fischl B. FreeSurfer. *Neuroimage*. 2012; 62(2):774–781. [PubMed: 22248573]
13. Huo Y, Plassard AJ, Carass A, et al. Consistent cortical reconstruction and multi-atlas brain segmentation. *NeuroImage*. 2016
14. Xu Z, Burke RP, Lee CP, et al. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Medical image analysis*. 2015; 24(1):18–27. [PubMed: 26046403]
15. Harrigan RL, Panda S, Asman AJ, et al. Robust optic nerve segmentation on clinically acquired computed tomography. *Journal of Medical Imaging*. 2014; 1(3):034006–034006. [PubMed: 26158064]
16. Panda S, Asman AJ, Khare SP, et al. Evaluation of multitlas label fusion for in vivo magnetic resonance imaging orbital segmentation. *Journal of Medical Imaging*. 2014; 1(2):024002–024002. [PubMed: 25558466]
17. Criminisi A, Robertson D, Konukoglu E, et al. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical image analysis*. 2013; 17(8):1293–1303. [PubMed: 23410511]
18. Woolard AA, Heckers S. Anatomical and functional correlates of human hippocampal volume asymmetry. *Psychiatry Research: Neuroimaging*. 2012; 201(1):48–53. [PubMed: 22285719]
19. Ourselin S, Roche A, Prima S, et al. Block matching: A general framework to improve robustness of rigid registration of medical images. :557–566.
20. Avants BB, Tustison NJ, Song G, et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011; 54(3):2033–44. [PubMed: 20851191]
21. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*. 2009; 46(3):786–802. [PubMed: 19195496]
22. Wang H, Suh JW, Das SR, et al. Multi-atlas segmentation with joint label fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2013; 35(3):611–623.
23. Landman BA, Huang AJ, Gifford A, et al. Multi-parametric neuroimaging reproducibility: a 3-T resource study. *Neuroimage*. 2011; 54(4):2854–2866. [PubMed: 21094686]

Figure 1.

Segmentation results comparing BrainCOLOR, FreeSurfer and VU Seg to the inter-rater reproducibility results. VU Seg shows a significant improvement ($p < 0.05$, Wilcoxon sign-rank test) over all techniques including reproducibility in all structures except for the left hippocampus for Rater 2 where it was statistically indistinguishable.

Figure 2.

Segmentation results comparing VU Seg to the inter-rater reproducibility. In the right anterior and posterior hippocampus, VU Seg out-performed inter-rater reproducibility ($p < 0.05$, Wilcoxon sign-rank test). In the left anterior hippocampus, VU Seg was not significantly different than inter-rater reproducibility. In the left posterior hippocampus VU Seg slightly underperformed inter-rater reproducibility.

Figure 3.

Segmentation results comparing VU Seg to the inter-rater reproducibility. VU Seg showed a significant decrease in volumetric variance with respect to the mean hippocampus volume across all three segmentations ($p < 0.05$, Wilcoxon sign-rank test).

Figure 4.

Volumetric results from the segmentation of the Kirby21 reproducibility scans. The volumes between the first scan (Volume Estimate 1) and the second scan (Volume Estimate 2) shows a high correlation across all structures ($r^2 > 0.95$, $p < 0.05$; A). Also, the variation between any two scans is at most 3% (B) and most cases are below 1% inter-scan variation.

Figure 5.

Qualitative segmentation results comparing the two manual tracings (Rater 1 and Rater 2) to three automated segmentation approaches for the subject with the median Dice. In general, VU Seg is comparable to the reproducibility results. On the other hand, BrainCOLOR is less consistent and tends to over-segment into the white matter. FreeSurfer is in general less consistent and over-segments several boundaries. Green arrows highlight areas of differences between one manual rater (shown in orange on all plots).