Evaluation of four hippocampal segmentation methods in healthy and pathological subjects

Abstract No:
999

Authors:
Courtney Bishop[1], Dorit Merhof[2], David Glahn[3], David Kennedy[4], Jerome Declerck[5], Mark Jenkinson[1]

Institutions:
[1]FMRIB Centre, University of Oxford, Oxford, United Kingdom, [2]University of Konstanz, Konstanz, Germany, [3]Yale, Hartford, United States, [4]University of Massachusetts Medical Center, Worcester, MA, [5]Siemens Molecular Imaging, Oxford, United Kingdom

Introduction:
With hippocampal atrophy proposed as a clinical biomarker for early Alzheimer's Disease (AD), there is much interest in the accurate, reproducible delineation of this region of interest (ROI) in structural MR images; a task complicated by its complex shape, large inter-subject variability and poor contrast hippocampus-amygdala border. Conventionally used manual segmentation requires expert knowledge and can be extremely time consuming, thus impractical for large-scale clinical studies, fuelling the development of semi-automated and automated segmentation methods for this purpose. Whilst these segmentation tools tend to perform well on healthy subject data, they often fail to capture the substantial and specific atrophy displayed in clinical data. Furthermore, with most studies focusing on development and evaluation of an individual segmentation method on a single dataset, reporting any number of performance measures, direct comparison of hippocampal segmentation methods is compromised.

Methods:
This work evaluates four MR hippocampal segmentation methods (FMRIB's Integrated Registration and Segmentation Tool (FIRST) [Patenaude 2007], Freesurfer's (FS) Aseg [Fischl et al. 2002], Classifier Fusion (CF) and an early development of a novel Fast Marching approach (denoted FMClose)), with segmentation performance on two clinical datasets assessed according to five common measures (Dice, sensitivity, specificity, false positive rate (FPR) and false negative rate (FNR)). We gain insight into the modes of success/failure of the different methods and potential means to make them more robust and accurate in the presence of specific and substantial pathology. The first clinical dataset contains 9 normal controls (NC) and 8 highly-atrophied AD patients, supplied by the Centre for Morphometric Analysis (CMA), whilst the second is a collection of 16 NC and 16 bipolar (BP) patients from a collaborator in San Antonio (denoted BPSA). The CMA images are of noticeably lower quality than the BPSA data, with voxel dimensions of 0.94 x 1.5 x 0.94mm and 0.8 x 0.8 x 0.8mm respectively.

Results:
Plots of Dice, FPR and FNR for the CMA and BPSA data are given in Figures 1 and 2 respectively. Decreased performance of FIRST, and to a lesser extent FS, on the BPSA data suggests that these model-based methods are biased towards the CMA

data used in their training. Whilst FIRST dominates on the CMA data, overall performance on the BPSA data is hindered by poor initial registration of a few subjects to the standard space. Likewise, CF performance on the CMA data is restricted by reduced image quality, poor brain extraction and subsequent registration error. However, it outperforms all other methods on the BPSA data, suggesting a potential advantage of using disease- and subject-specific selection methods like CF over pre-existing models trained from much more diverse data. In general, FS displays good Dice results and significantly reduced FNR compared to other methods, but has a tendency for 'greedy' labeling, with the highest FPR on both datasets. FMClose suffers from both spillover in anterior and inferior regions and under-estimation at medial hippocampal boundaries, contributing to both FPR and FNR. This method also excludes sub-hippocampal regions of contrasting intensity, such as the dentate gyrus, from the segmentation estimate, suggesting that intensity-based methods may be useful for segmentation of hippocampal sub-structures.

Conclusions:
With this work bringing to light several strengths and weaknesses of the evaluated hippocampal segmentation methods, future work should focus on development of these methods to make them more robust and accurate in the presence of specific and substantial pathology. For the BPSA data, FIRST registration errors could be addressed with a three-step affine registration process, similar to the CF method presented here. For CF, registration errors could be addressed with bias-field correction and removal of excess neck prior to registration, with current classifier selection strategies extended using higher-dimensional clustering algorithms. Finally, extension of the FM algorithm to include a spatial prior could prevent spillover at poor contrast hippocampal boundaries and act as a final stopping criteria for the propagating FM front. During development of these segmentation methods, it may prove optimal to combine aspects of each in a type of 'multi-approach' segmentation method.

References:
Patenaude, B. (2007), 'Bayesian statistical models of shape and appearance for subcortical brain segmentation', *In: FMRIB Technical Report TR07BP1 Oxford*.
Fischl, B. (2002), 'Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain', *Neuron*, vol. 33, pp. 341-55.

Categories
- Alzheimer and Dementia (Disorders of the Nervous System)
- Anatomical MRI (Imaging Techniques and Contrast Mechanism)
- Flattening, Segmentation (Modeling and Analysis)
- Anatomical Studies (Neuroanatomy)

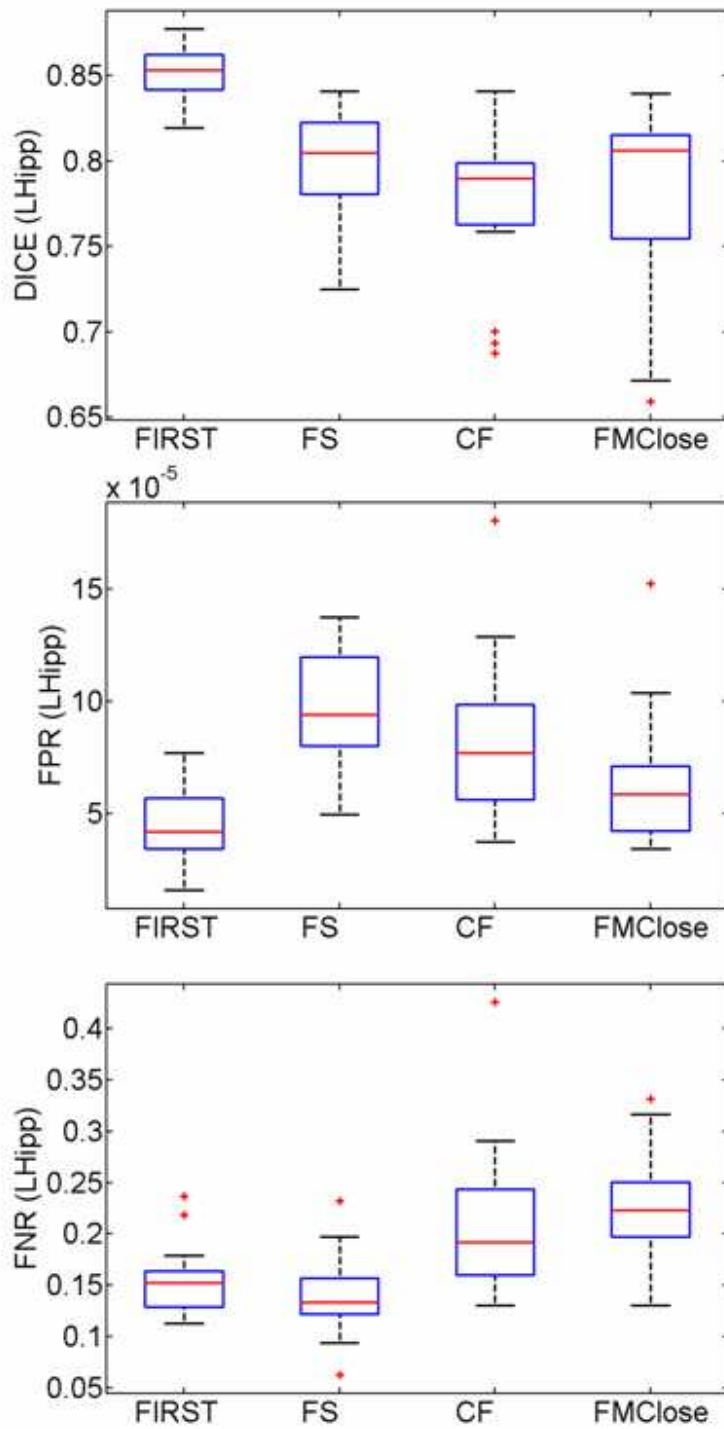## Figure 1: Results for the CMA data, left hippocampus (LHipp)

Figure 2: Results for the BPSA data, left hippocampus (LHipp)