Fully automated grey and white matter segmentation of the cervical cord in vivo

Ferran Prados[1,2], Manuel Jorge Cardoso[1], Marios C Yiannakas[2], Luke R Hoy[2], Elisa Tebaldi[2], Hugh Kearney[2], Martina D Liechti[2], David H Miller[2], Olga Ciccarelli[2], Claudia Angela Michela Gandini Wheeler-Kingshott[2,3], and Sebastien Ourselin[1]

[1]Translational Imaging Group, Medical Physics and Biomedical Engineering, University College London, London, United Kingdom, [2]NMR Research Unit, Queen Square MS Centre, Department of Neuroinflammation, UCL Institute of Neurology, University College London, London, United Kingdom, [3]Brain Connectivity Center, C. Mondino National Neurological Institute, Pavia, Italy

Synopsis

We propose and validate a new fully automated spinal cord (SC) segmentation technique that incorporates two different multi-atlas segmentation propagation and fusion techniques: Optimized PatchMatch Label fusion (OPAL) and Similarity and Truth Estimation for Propagated Segmentations (STEPS). We collaboratively join the advantages of each method to obtain the most accurate SC segmentation. The new method reaches the inter-rater variability, providing automatic segmentations equivalents to inter-rater segmentations in terms of DSC 0.97 for whole cord for any subject.

Introduction

Axonal loss in the spinal cord is a major cause of irreversible clinical disability in multiple sclerosis (MS). In vivo, axonal loss can be inferred indirectly, by estimating the reduction of the cord cross-sectional area (CSA) over time (i.e. a measure of atrophy), which may be obtained by means of image segmentation using magnetic resonance imaging (MRI); such a measure has been shown to correlate with clinical scores of disability and has been suggested as a plausible endpoint to clinical trials of neuroprotection[1]. However, the measure of cord CSA cannot elucidate between the individual rates of GM and WM atrophy, which could have disparate prognostic implications, and may be best studied independently[2].

We propose and validate a new fully automated spinal cord (SC) segmentation technique that incorporates two different multi-atlas segmentation propagation and fusion techniques: Optimized PatchMatch Label[3] fusion (OPAL) and Similarity and Truth Estimation for Propagated Segmentations[4] (STEPS). We collaboratively join the advantages of each method to obtain the most accurate SC segmentation.

Methods

The presented method is based on a collaborative effort that takes advantage of the key characteristics of two different label fusion algorithms in order to build a fully automated and robust slice-wise pipeline.

Firstly, OPAL is quick and precise, however as it is based on computing distance between patch intensities, detecting small structures such as GM within the SC can be difficult. Due to partial volume effects and the presence of pathology, such as MS lesions, it is difficult to algorithmically distinguish the boundaries between GM and lesions.

Secondly, STEPS is a method that requires a good match between all the library's templates and the image that needs to be segmented. This good matching can be obtained through linear registration, to first align images, and then through non-linear registration to deform the template model to fit input image. In order to increase the performance of the method and ensure a good registration, we need a rough segmentation of the input image to ensure that we are registering the center of each template to the center of the SC image being segmented. While this rough SC segmentation could be drawn manually, we used the result of OPAL as an initialization of the registration required for the STEPS algorithm, avoiding user intervention.

Data comprised 25 healthy subjects scanned in a 3T Philips Achieva with the center of the imaging volume positioned at the level of C2-3 intervertebral disc. All underwent a FFE sequence of 0.5x0.5x5 mm^3. Three raters manually outlined GM and semi-automatically outlined the cord in all participants using JIM.

The template library used for label fusion relies on labeled images from 25 healthy subjects. For each subject there were three slices, with the middle slice centered at C2-3 level. In order to maximise the size of the library, all the scans were left-right flipped, resulting in a final template library of 150 2D slices (25 datasets * 3 slices * 2 L/R flip). For each image in the template library, associated consensus segmentation of the three raters for whole cord and GM was also available. Table 1 presents used parameters.

Results

The proposed method was compared to the consensus segmentation of 3 raters (see Figure 1). Such comparison demonstrates whether or not the proposed method can perform similarly to a single human rater. To assess the performance, the Dice Score Coefficient (DSC), Hausdorff Distance (HD) and Mean Surface Distance (MSD) between the masks, are provided (see Table 2). In order to remove possible bias, a leave one out strategy was used, i.e. a segmented image as well as its left-right flipped version were removed from the template library. Results show that the proposed method is able to segment the whole cord with an accuracy similar to rater 1 (p>0.001) and it achieves good results for GM segmentation, though less consistent when compared to all raters.

Conclusions

This work has introduced a fully-automated GM and whole SC segmentation technique based on a collaborative effort between two cutting-edge multi-label fusion techniques. The new method reaches the inter-rater variability, providing automatic segmentations equivalents to inter-rater segmentations in terms of DSC 0.97 for whole cord for any subject. Regarding the GM, the results are close to inter-rater segmentations DSC 0.84. Future work will explore to support multi-modality images and compare to other published methods.

Acknowledgements

References

1) Lossef, Brain, 1996   2) Yiannakas, NeuroImage, 2012   3) Ta, MICCAI, 2014   4) Cardoso, MedIA, 2013

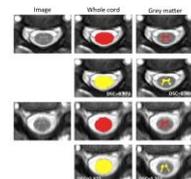Abstract ID: 214

Figures

Figure 1: Row 1 and 2 show: image with no motion artifact, three raters consensus segmentation masks and best GM automatic segmentation results with corresponding whole cord segmentation masks. Row 3 and 4 worst GM segmentations results and corresponding whole cord segmentation masks, image has motion artifact.

OPAL	STEPS
Patch size: 5x5x5 voxels	σ=1.5
Iterations: 5	Spatially consistency 0.35
Number best matches: 10	Number best templates: 13
Number of threads: 10	Mask threshold 0.5

Table 1: Selected parameters for OPAL and STEPS

R=23		Rater 1	Rater 2	Rater 3	Automatic
Whole cord	DSC	0.96 (0.01) p<0.001	1.00 (0.00) p<0.001	1.00 (0.00) p<0.001	0.97 (0.00)
	HD	1.70 (0.54) p<0.001	1.49 (0.39) p<0.001	1.40 (0.15) p<0.001	1.58 (0.22)
	MSD	0.13 (0.06) p<0.001	0.10 (0.04) p<0.001	0.10 (0.01) p<0.001	0.14 (0.04)
Grey matter	DSC	0.89 (0.04) p<0.001	0.93 (0.02) p<0.001	0.90 (0.01) p<0.001	0.84 (0.04)
	HD	1.44 (0.27) p<0.001	1.28 (0.13) p<0.001	1.31 (0.13) p<0.001	1.43 (0.26)
	MSD	0.13 (0.03) p<0.001	0.08 (0.03) p<0.001	0.07 (0.03) p<0.001	0.08 (0.03)

Table 2: Mean and standard deviation of Dice Similarity Coefficient (DSC), Hausdorff Distance (HD) and Mean Surface Distance (MSD) for each segmented mask to the three human raters consensus mask. Pair t-test against automated method. Higher DSC values, lower HD and lower MSD mean better result.