Evaluation of a Framework for On-line Interactive Segmentation of Similar 3-D Images based on a Single Example

Emmanouil Moschidis
emmanouil.moschidis@postgrad.manchester.ac.uk
Jim Graham
jim.graham@manchester.ac.uk

Imaging Science and Biomedical Engineering
School of Cancer and Enabling Sciences
The University of Manchester

Abstract

We present the results of the performance evaluation of an on-line interactive segmentation framework that addresses the problem of propagation of a single three dimensional (3-D) example to similar images. The aim is to minimise user intervention during interactive segmentation when an example of the desired outcome exists. The framework tackles the segmentation problem as a process consisting of two distinct tasks: the localisation of the region of interest and its boundary delineation. We have previously applied and evaluated this approach on the segmentation of the prostate with encouraging results. Here, we extend the evaluation study on additional datasets and application domains. For each stage of the framework, different candidates are evaluated against delineations provided by experts. The results suggest that a two stage strategy increases the framework's repeatability and that our approach can provide automated results close to the ground truth, allowing the user to intervene only if necessary at a later refinement stage or in case of provision of unsatisfactory segmentation outcomes.

1 Introduction

Interactive segmentation has attracted a lot of attention during the past decade. This is due to the advent of computationally efficient methods, which allowed human experts to steer the entire process and obtain arbitrary results that match their perception of ground truth in short time with limited interaction. In the context of this segmentation paradigm, images are treated on an individual basis and segmented one-by-one, involving a certain amount of interaction and consequent cognitive load.

We address the problem of miminimising this interaction when similar images are processed, given a single segmented image as an example of the desired outcome. We tackle this by propagating the segmentation example onto the subsequently processed images. This way the user is freed from the entire process and may intervene only if necessary at a later refinement stage or in case the framework fails to provide satisfactory outcomes.

We consider two images as similar when they depict the same anatomy of interest and they are acquired by the same imaging modality using the same protocol. Our main assumption is that the processed images do not exist as a dataset, but they rather appear one at a time

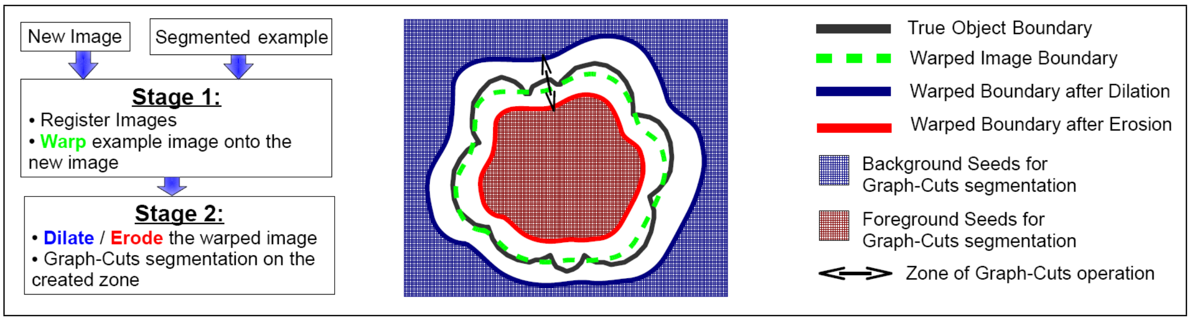

Figure 1: Overview of the suggested framework

(on-line), as it often happens in real life. Therefore, group-wise approaches are not included in the evaluation. Also, since one single image cannot capture the variation of a population, we exclude model-based methods and we restrict our study to data-driven ones.

We consider segmentation as a process that consists of two distinct tasks: the localisation/recognition of the anatomy of interest and its boundary delineation, as suggested in [15]. Consequently, we suggest a two staged framework that handles these two tasks separately. For each stage we identify and evaluate potential candidate methods against ground truth provided by a radiologist. We also seek methods able to perform in real-time as part of an interactive segmentation system of three dimensional (3-D) medical images, which generates ground truth for subsequent building of statistical models based on the segmented images.

We have previously applied and evaluated this approach on the segmentation of the total prostate with encouraging results [11]. Here, we extend the evaluation study on additional datasets and application domains. The results show that the suggested framework can provide results close to the ground truth without any user interaction and that the two stage strategy improves its repeatability.

2 Methods

A set of 22 T2 weighted fat suppressed MR images of the prostate, a set of 10 CT images of the wrist and a set of 5 MR images of the brain were used in this study. The segmentation tasks of the total prostate and the central gland were associated with the first set, whereas the segmentation task of the carpal bones and brain ventricles were associated with the second and third set respectively.

The suggested framework (fig. 1) operates in two stages. The methods employed in the first stage are intended to provide an approximate position of the anatomy of interest in the new processed image, given one single example. In this context, we evaluate four established non-linear registration methods (Demons [14], [9], B-Splines [12], [9], DROP [5], [7] and a group-wise method (GWR) [4], used in a pair-wise fashion) and Active Graph Cuts (AGC) [6], [2], as means of segmentation propagation.

For the second framework stage, we seek computationally efficient methods that can delineate accurately the anatomy of interest, once initialised by the first stage's output. Such methods exist in the interactive segmentation literature. In a recent evaluation study comparing several such methods [10], it was shown that Graph-Cuts [1] (GC) provides accurate and repeatable results, even with limited user input. Here, we assess the performance of GC against GC coupled with a Canny [3] (GC+C), a Phase Congruency [8] (GC+PC) and a SUSAN [13] (GC+S) boundary detector, as means of accurate boundary delineation.

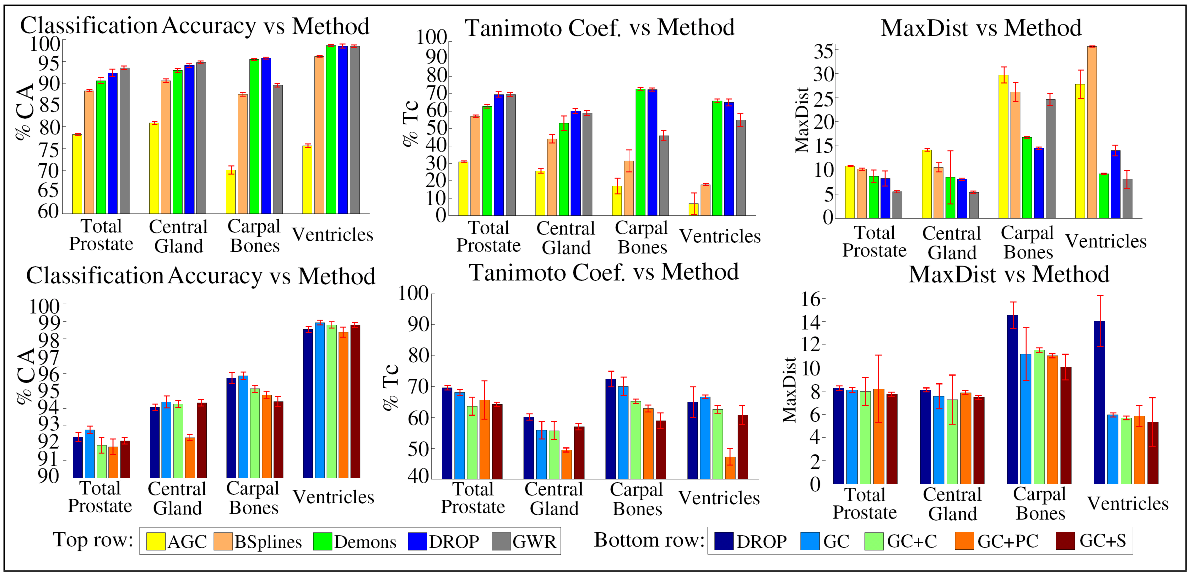

Figure 2: Summary of the performance of the first (top row) and second (bottom row) stage's candidate methods with respect to the accuracy metrics. The error bars represent the $\pm 1.96\times$ standard error of the mean.

We quantify the performance of methods with a score of classification accuracy (*CA*), calculated as in eq. 1, the Tanimoto coefficient (Tc), calculated as in eq. 2 and a maximum point to surface distance between the segmentation and the ground truth surface (*MaxDist*), calculated via a 3-D distance transform. Each metric reveals a different aspect of the segmentation accuracy. In the case of *CA* and Tc, larger values represent improved segmentation, whereas with *MaxDist*, this corresponds to smaller values.

$$CA = 100 \times \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}\% \qquad (1)$$

$$Tc = 100 \times \frac{|TP|}{|TP| + |FP| + |FN|}\% \qquad (2)$$

3 Experiments and Results

The top row of fig. 2 summarises the results of the performance evaluation of the four registration methods and AGC. During the experiments each image in every dataset was selected once as a template image and its ground truth surface was propagated to the remaining images of the same dataset with each of the assessed methods. GWR provided the most accurate results, except for the segmentation task of the Carpal Bones. DROP performed noticeably better in this case, whereas in the other three segmentation tasks it produced results comparable to GWR. Demons, BSplines and AGC provided less accurate results. In terms of computational efficiency, DROP was by far the most computationally efficient method and GWR the most expensive. Given our requirement for real-time processing as well as accuracy, DROP was selected as the appropriate method for our framework's first stage.

In order to initialise GC from the DROP output, a zone around the boundary suggested by the latter was created by successive erosions and dilations of the warped volume and foreground/background seeds were configured as shown in fig. 1. The zone width is a user

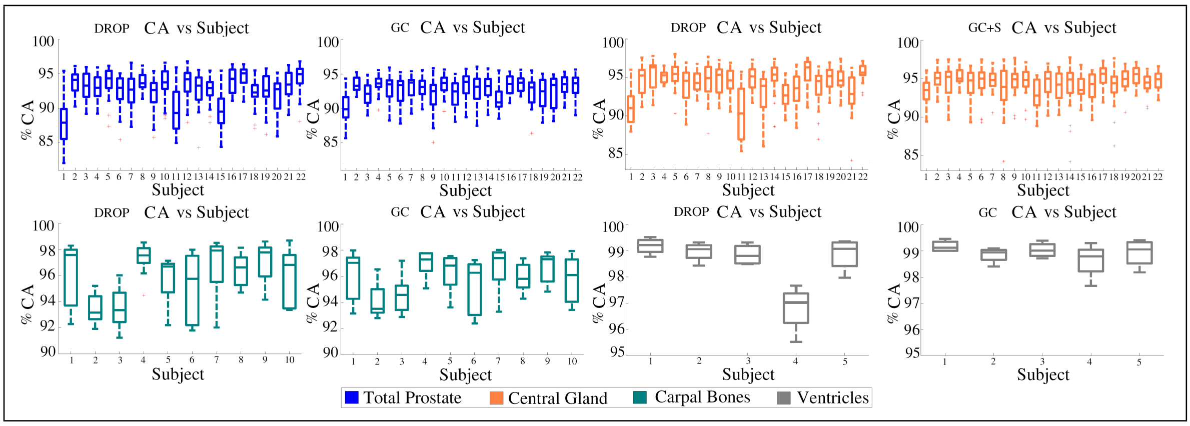

Figure 3: Pairs of box and whisker plots depicting the increase of the framework's repeatability from stage 1 (left image) to stage 2 (right image) for each dataset. The whiskers are $1.5\times$ the interquartile range. Values outside them are considered outliers (red crosses).

defined parameter that was kept constant for every dataset, to allow for unbiased experimental results. The bottom row of fig. 2 summarises the results of the performance evaluation of the candidate methods for the second framework stage. The results are also depicted against DROP, to allow for direct observation of the effect of the additional processing on the DROP outcome. Overall, the changes in segmentation performance with respect to accuracy due to it are small. The main effect is the reduction of the *MaxDist* error, which is noticeable on the wrist CT and brain MR image datasets. Also, apart from the segmentation of the central gland, where GC+S gave less variable results than GC, the use of edge detectors did not seem to provide any advantage over the original GC, possibly due to the already good object/background contrast. However, the major advantage of the additional processing step is the increase of the framework's repeatability. Fig. 3 depicts the framework's performance with respect to the *CA* metric for each subject selected as the template image. In all datasets the second stage offers an improvement of the framework's dependency on the selected template.

4 Concluding Remarks

We presented the results of a performance evaluation study of candidate methods for an interactive segmentation framework, which leverages prior knowledge from one single image example, in order to minimise the user's intervention when similar images are processed. The suggested framework operates in two stages. The results show that it can provide results close to the ground truth without any user intervention, when a deformation-based registration is followed by a Graph-Cuts segmentation. Moreover, the two stage approach improves the overall repeatability. While we have addressed this in the context of interactive segmentation, the results of this study can be applied in "automatic" Atlas-Based segmentation as well, where a single image is often used as a template.

In this study, we addressed the problem of using the prior knowledge of a single image. However, in an on-line set-up more image examples are accumulated as the segmentation proceeds. These can serve as a rich source of information about subsequent images in the set. Therefore, future work will concentrate on leveraging the additional amount of prior knowledge produced as the segmentation process goes on.

5 Acknowledgments

The project is funded by the Biotechnology and Biological Sciences Research Council.

References

[1] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Proc. ICCV*, volume 1, pages 105–112, 2001.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE T-PAMI*, 26(9):1124–1137, 2004.

[3] J. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8(6):679–698, 1986.

[4] T. F. Cootes, C. J. Twining, V. S. Petrović, K. O. Babalola, and C. J. Taylor. Computing accurate correspondences across groups of images. *IEEE T-PAMI*, 32(11):1994–2005, 2010.

[5] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through MRFs and efficient linear programming. *MedIA*, 12(6):731–741, 2008.

[6] O. Juan and Y. Boykov. Active graph cuts. In *Proc. CVPR*, pages 1023–1029, 2006.

[7] N. Komodakis, G. Tziritas, and N. Paragios. Fast approximately optimal solutions for single and dynamic MRFs. In *Proc. CVPR*, pages 1–8, 2007.

[8] P. Kovesi. Image features from phase congruency. *Videre*, 1(3):1–27, 1999.

[9] D. J. Kroon and C. H. Slump. MRI modality transformation in demon registration. In *Proc. ISBI*, pages 963–966, 2009.

[10] E. Moschidis and J. Graham. A systematic performance evaluation of interactive image segmentation methods based on simulated user interaction. In *Proc. ISBI*, pages 928–931, 2010.

[11] E. Moschidis and J. Graham. Propagating interactive segmentation of a single 3-D example to similar images: An evaluation study using MR images of the prostate. In *Proc. ISBI*, pages 1472–1475, 2011.

[12] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations:application to breast images. *IEEE T-MI*, 18(8):712–721, 1999.

[13] S. M. Smith and M. J.Brady. SUSAN-A New approach to low level image processing. *IJCV*, 23(1):45–78, 1997.

[14] J. P. Thirion. Image matching as a diffusion process. *MedIA*, 2(3):243–260, 1998.

[15] J. K. Udupa, V. R. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn. A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics*, 30(2):75–87, 2006.