Improving the Robustness of Cell Nucleus Segmentation

Pascal Bamford and Brian Lovell
Cooperative Research Centre For Sensor Signal and
Information Processing,
University of Queensland, QLD 4072, Australia
[bamford—lovell]@csee.uq.edu.au

Abstract

A highly successful active contour implementation, for the automatic segmentation of cervical cell nuclei, is shown to lend itself well to a framework that further increases its success rate. The method is based upon measuring changes in the final contour as the one parameter that governs its behaviour is varied. Only one object of interest is contained in the image, but as artefacts often appear as well, the contour varies with the parameter as it finds different solutions. In contrast, simple images with no artefacts are very stable. Therefore a stability measure is calculated and each image is classified according to its degree of 'difficulty'. A system can then choose at which level to operate to ensure only high quality examples are processed after segmentation.

1 Introduction

The completely automatic segmentation of Papanicolaou stained cervical cell nuclei has been the subject of active research for many years. The main frustration that research groups had encountered was due to the fact that it is a deceptively difficult problem. The images often appear trivial to segment; however no one technique had been found to perform suitably well over a large data set. This was mainly due to the fact that the cytoplasm is also stained which greatly reduces the nucleus-cytoplasm contrast as well as producing many artefacts in the images. An active contour method has recently been found to be highly suited to this application with a success rate of 99.64% correct segmentations on a data-set of 20,130 images [1]. This is a far superior result to those that have been reported in the past [8][9][5] and appears to finally demonstrate the feasibility of a completely automatic segmentation procedure for this task. However, there remains the possibility of sample contamination from the few remaining failures that the algorithm produces. In order to prevent this, the need would still exist for a human to view the output of this stage, undermining its utility in a practical system. Fortunately, the segmentation algorithm lends itself well to the implementation of an error checking and 'segmentation stability' stage which enables a further increment upon the success rate, effectively removing this requirement for human supervision.

BMVC 1998 doi:10.5244/C.12.52

2 The Segmentation System

The use of active contours in bio-medical applications is well established and global minimum searching methods have been found to be particularly useful in the presence of the many artefacts usually associated with these images [3][4][6]. Here, a dynamically programmed search method was implemented which was based upon a suggestion in [7]. A search space is first set up within the image, bounded by two concentric circles centralised upon a point found by an initial rough segmentation. This search space is then sampled by discretising both circles and a number of radii joining them. Every possible contour that lies upon the points of the search space is then considered and an associated cost function is calculated. This cost represents the traditional snake property of a balance between the contour's smoothness and how much it lies upon pixels of high image gradient. This balance is controlled by a single regularisation parameter, $\lambda \in [0, 1]$. By choosing a high value of λ, the smoothness term dominates which may lead the contour to ignore important image edges. A low value of λ will allow sharp corners to develop in the contour as it attempts to lie upon high gradient areas, which may not necessarily be on the desired objects edge. Once every contour has been evaluated, the single contour with least cost is chosen as the solution. As the focus of this contribution is on the development a subsequent error checking stage, the reader is referred to [1] for a fully detailed explaination of the above techniques.

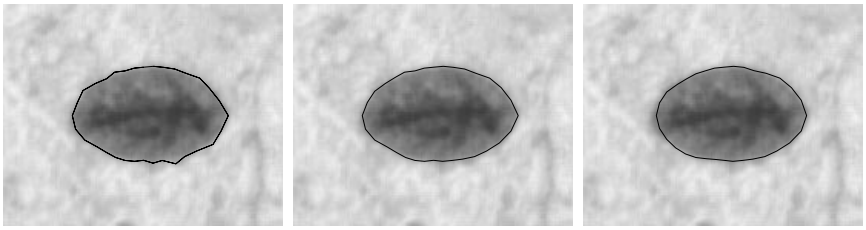
3 Segmentation Accuracy

The single parameter which affects the behaviour of the algorithm, λ, was chosen to be 0.8 after viewing its effect upon a small sub-set of the images. All of the images were then segmented and subsequently verified by eye. Of the 20,130 images, 99.64% were found to be correctly segmented. Two main classes of failure were identified. Forty four of the failures were due to the nuclei lying close to the cytoplasm boundary. As the background/cytoplasm contrast was much greater than that of the nucleus/cytoplasm, the contour tended to be attracted to the background/cytoplasm boundary, causing the segmentation errors. Twenty-six of the failures were caused by the inappropriate choice of λ for that individual image (they all subsequently produced correct segmentations with different values of λ.) The remaining three images were found to fail at all attempts — caused by the incorrect setting of microscope focus and illumination. The failures due to the presence of the background in the nucleus images were deemed to be preventable at a previous cell-finding stage. Here, the location of the cytoplasm/background boundary is known and can therefore be prevented from appearing in the images presented to the active contour algorithm. The problem of the choice of λ is therefore the major issue.

4 Lambda Sensitivity

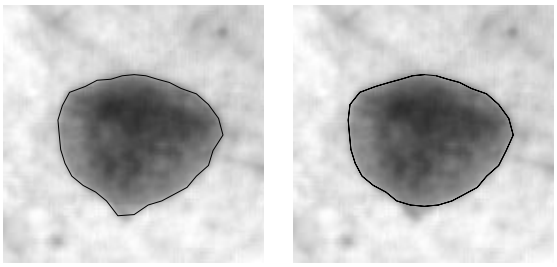
For the majority of relatively simple images with little ambiguity in the true location of the nuclear boundary, the final segmentation can be fairly insensitive to λ over a wide range of values (figure 1).

By contrast, 'difficult' images (even for humans) produce very different contours depending upon the choice of λ (figure 2). These images usually contain artefacts near or



(a) (b) (c)

Figure 1: Segmentation results showing insensitivity to different values of λ. (a) $\lambda = 0.1$ (b) $\lambda = 0.5$ (c) $\lambda = 0.9$

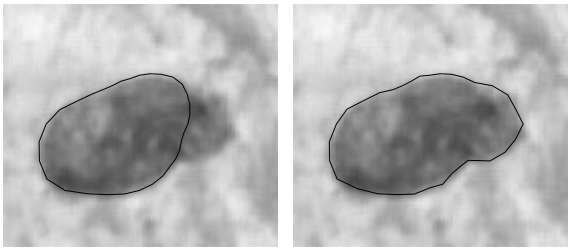


(a) (b)

Figure 2: Example of an image which is not stable over a range of λ (a) $\lambda = 0.5$ (b) $\lambda = 0.8$

on the nuclear boundaries that make the 'true' border hard to find. Another example of λ sensitivity is shown in figure 3. In this case, the nucleus has been folded or has been distorted in some way and contains a sharp corner. The examples in figures 1, 2 and 3 show that no single value of λ is capable of accurately segmenting all of the images. They do, however, indicate that different examples have a varying 'sensitivity' to λ. Simple images result in a contour which remains fairly stable over wide ranges of λ (figure 1) whereas harder images cause the contour to move or switch with λ (figures 2&3). Therefore, if a measure of an image's stability with varying λ could be gleaned, then a method exists for subsequent stages to only choose 'easy' images to segment. In this application, the pruning of the data in order to enhance the sample has no effect upon the final classification. This idea naturally has similarities with previous methods that have used a number of different segmentation algorithms upon one image. A choice is then made whether to accept or reject the segmentation depending on how well they all agree on the solution [9][2]. However, the advantage here is in the use of the exact same framework and initialisation for each run, with simple comparison of results. In order to pursue this method, a sub-set of 772 images was segmented at values of λ from 0.0 to 1.0 at 0.1 increments. The results are plotted in figure 4.

With λ set at 0.0, the smoothness constraint is completely ignored and the point of greatest gradient is chosen along each search space radius. Figure 4 shows that for 65.65% of images, all the points of greatest gradient actually lie upon the nucleus/cytoplasm border (figure 5(a)). For the remaining 34.35% of images, a large gradient due to an artefact



(a) (b)

Figure 3: Folded or distorted nucleus (a) $\lambda = 0.8$ does not allow the contour to bend enough to capture sharp corner (b) $\lambda = 0.5$

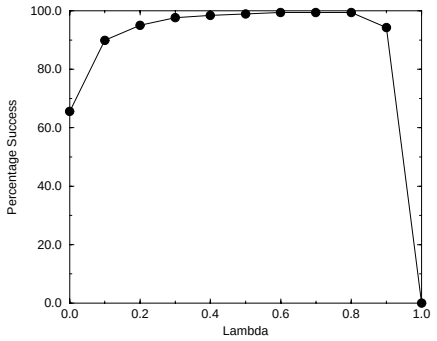
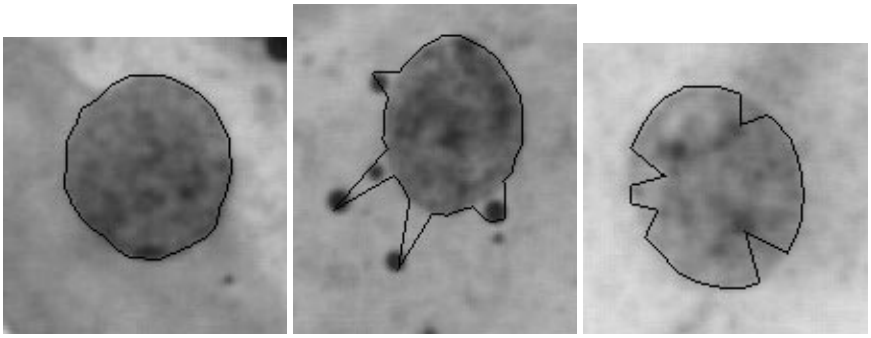


Figure 4: Percentage of successful segmentations versus λ

or darkly stained chromatin will draw the contour away from the true border (figures 5(b),(c)). As λ is increased, the large curvatures present in figure 5(b) and (c) increase the energy of these contour configurations and they therefore become less probable. Figure 6(a), (b) and (c) show how the contour is affected by increasing λ. Between $\lambda = 0.3$ and 0.8, the curve in figure 4 flattens out with high success rates. The plateau is remarkable broad, demonstrating that the method is well-suited to this application. As λ increases beyond 0.8, some sharper nuclear borders become discriminated against (figure 7). However, the rare occurance of this type of nuclear feature is reflected in the small drop in success rate to 94.3% at $\lambda = 0.9$.

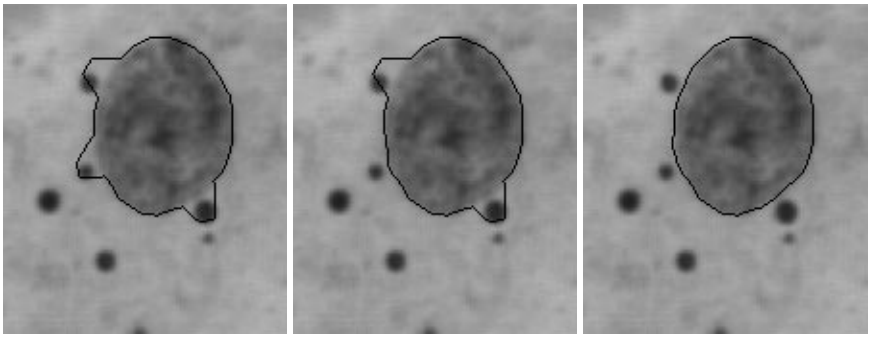
5 Image grading

It was observed from the data that the performance was monotonically increasing for $0 < \lambda < 0.8$ with the solutions for λ_1 being a strict subset of solutions for λ_2, where $\lambda_1 < \lambda_2$. Although this was true for the 772 images tested, it could not be assumed true for every image. However, it was proposed that any image not satisfying this condition would be a sufficiently rare event. Therefore, by segmenting any image at a highly probable value of λ for success, say 0.7 (middle of the plateau) and again at 0.0, a similarity between



(a) (b) (c)

Figure 5: Examples for $\lambda = 0.0$ (a) largest gradients occur on the nucleus/cytoplasm border (b) dark artefacts generate larger gradients than the border (c) darkly stained chromatin generates larger gradients than the border.



(a) (b) (c)

Figure 6: The effect of increasing λ (a) $\lambda = 0.1$ (b) $\lambda = 0.2$ (c) $\lambda = 0.3$

the resulting contours would indicate a high level of stability (figure 1). This image could then be classified a level 0 image, or a 'very easy' image to segment. Lack of similarity could lead to a second run at $\lambda = 0.1$ and another comparison made with the first contour at $\lambda = 0.7$. Similarity would lead to a classification of level 1 and so on until either the risk of including an image which fails at all attempts is too high, or $\lambda = 0.6$ is reached.

6 Contour comparison

As the contours to be compared were the result of the same algorithm and indeed the same search space in the image, the comparison between the two contours was trivial. The distance between each chosen point on each of the search space radii for each contour was calculated and the maximum absolute deviation (MAD) returned. For the subset of 772 images, the minimum MAD was evaluated for the transition between a good segmentation and a bad segmentation as λ was varied. This resulted in the threshold of 2.3 pixels to be set. Any two contours whose MAD was greater than this threshold was deemed as 'not

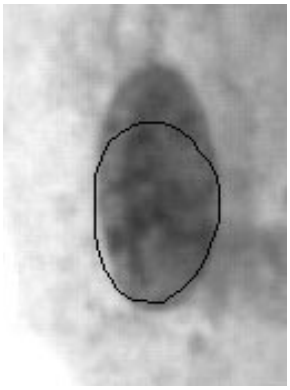


Figure 7: Bend at the top of the nucleus is deemed to be too sharp with $\lambda = 0.9$

similar'.

7 Grading Results

From the remaining database, a test set of 18,718 images was graded between a level 0 and 6 using the method described above. The percentage of the total number of images included in each level was calculated and the number of the known failures in each level noted. The results are summarised in table 2. From the table, it can be seen that by

Level	0	1	2	3	4	5	6
% Images at level	54.65	73.12	80.53	85.93	90.07	92.98	95.83
No. of failures	0	0	1	3	9	14	19

Table 1: Level of segmentation ease with percentage of images at each level and the number of failures included in that level.

grading the images it is possible to achieve a 100% accuracy by choosing level 0. In this case approximately half of the images would, however, be discarded. By increasing the acceptance level to one, 73.12% of the images would be accepted still with 100% accuracy.

8 Conclusions and Future Work

A dual-search based active contour was implemented and tested upon a large dataset of cell nuclei images. The remarkable success rate of this algorithm has been shown to have the potential of further improvement by a method of grading the images under 'ease of segmentation'. Naturally the reported 100% accuracy rates cannot be expected to be maintained in practice, however it is expected that only a very few incorrect segmentations would be missed. This is because an image is required to be very stable over a large range of λ values. As problem segmentations are mainly caused by artefacts distorting the contour away from the nucleus border, they often lie in areas of high image energy which

makes them unstable. The threshold for the comparison of two contours was calculated by evaluating the maximum absolute deviation (MAD) between the good/bad segmentations of the subset. As this occurred between two consecutive values of λ, it may not be the best choice for a practical system. This is because as λ is increased from a bad segmentation at, say, 0.1 to a good segmentation at 0.7 then the contour may creep towards the point where it was declared as 'good', with all previous attempts having been declared as 'bad'. Therefore, if it was decided to run the system at level 1, then the threshold should be re-calculated as the minimum MAD between good/bad segmentations at 0.1 and 0.7. This would result in a less severe threshold and reduce the number of correct segmentations being discarded.

References

[1] Pascal Bamford and Brian Lovell. Unsupervised cell nucleus segmentation with active contours. *Signal Processing Special Issue: Deformable models and techniques for image and signal processing*, 1998. To Appear.

[2] Chen-Chau Chu and J. K. Aggarwal. The integration of region and edge-based segmentation. In *Proceedings of the Third International Conference on Computer Vision*, pages 117–20, Los Alamitos, CA, USA, 1990. IEEE Comput. Soc. Press.

[3] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Trans. Patt. Anal. Mach. Int.*, 15(11):1131–1147, 1993.

[4] Chris A. Davatzikos and Jerry L. Prince. An active contour model for mapping the cortex. *IEEE Transactions on Medical Imaging*, 14(1):65–80, 1995.

[5] Gerardo Luis Garcia. *Feasibility of contextual analysis in an automated cervical prescreening system*. PhD thesis, Worcester Polytechnic Institute, 1986.

[6] D. Geiger, A. Gupta, L. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking, and matching deformatble contours. *IEEE Trans. Patt. Anal. Mach. Int.*, 17(3):294–302, 1995.

[7] Steve R. Gunn. *Dual Active Contour Models for Image Feature Extraction.* PhD thesis, University of Southampton, May 1996.

[8] Calum Eric MacAulay. *Development, implementation and evaluation of segmentation algorithms for the automatic classification of cervical cells.* PhD thesis, University of British Columbia, August 1989.

[9] Bo Nordin. *The development of an automatic prescreener for the early detection of cervical cancer: Algorithms and implementation.* PhD thesis, Uppsala University, 1989.