33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A novel approach to automated cell counting for studying human corneal epithelial cells

Namrata Bandekar, Alexander Wong, David Clausi and Maud Gorbet

Abstract— **A novel automated cell counting technique for cell sample images used to study the side-effects of lens cleaning solutions on human corneal epithelial cells is developed. The proposed multi-step approach integrates non-maximum suppression, seeded region growing, connected component analysis, and adaptive thresholding to produce segmentation and classification results that are robust to background illumination variation and clustering of cells. The proposed algorithm is computationally efficient, and experimental results show that the average detection rate of nucleated cells is greater than 90% with the proposed technique as opposed to the state-of-the-art level set method which gives an accuracy of less than 65%.**

I. INTRODUCTION

Cell counting has many applications in the field of biological and pathological imaging. Traditionally, cell counting is done manually, which is a time consuming, tedious and subjective process. Manual cell counting can lead to inconsistent and imprecise results. Hence there is a need to automate the process of counting cells. An automatic process will ensure consistency of results. There is also a need to classify the cells if there are different types of cells in the samples.

The purpose of this research is to develop a new automated cell counting technique for images of cell samples acquired for the purpose of studying the side-effects of lens wear and lens cleaning solutions on human corneal epithelial cells (HCECs) using an ex vivo cell collection system [1]. These samples have been used to measure cell viability and activation by counting the number of living cells, and contains three different types of cells: HCECs, white blood cells and ghost cells. Fig. 1 illustrates one such cell sample. AxioVision LE which can be purchased from Carl Zeiss Canada Ltd, Toronto has been used previously for this research study, however it did not yield the desired results. The software does not support a direct automated way to get a differential cell count. The dataset is fairly unique in terms of size, population and staining. The cells are exfoliated from the eye, washed and the images are acquired using fluorescence imaging after they settle at the bottom of a culture well. Due to the variation in cell structure and mode of imaging, a reasonable approach is to use segmentation and classification techniques.

II. BACKGROUND

Other methods for automatic cell counting have been proposed. The technique used depends on the type of image available and the nature of the dataset. Due to brightness variation of nuclei in the acquired cell sample images, the use of a global thresholding technique [2] is not suitable. Morphology [3] is not a feasible solution as the cell shape and size varies from one class of cells to another and also within a class. Zalewski *et al.* [4] proposed an algorithm for cell detection in greyscale images of yeast cells which uses local maxima detection and object recognition. The idea of detecting local maxima has been adopted to identify the location of the nuclei using non-maximum suppression [5].

Young *et al.* observed that template matching can be applied to detect the location of cells in images of cell samples [6]. This method however, requires the cells to be in focus and non-overlapping. Template matching is a computationally expensive technique. Moreover, the template for each cell type needs to be constructed using prior knowledge about the size, shape and the structure of cells. Since the shape and size of the cells varies, template matching will fail to detect cells which do not have the same shape, size and grey level variation as the template.

Many other automatic segmentation methods have been proposed such as active contours [7] and level set techniques [8]. Active contours require initialisation of a contour for each cell which makes the method semi-automatic and cumbersome. These techniques are not suitable for the dataset being used because they are not robust towards many challenges presented by the dataset such as background illumination variation, variation in cell structure and clustering of cells.

The novelty of the proposed approach lies in developing an effective algorithm pipeline to obtain a better result. A series of techniques have been integrated, which are described in Section III, to yield a segmentation and classification for the given images. Since the dataset used for this research has

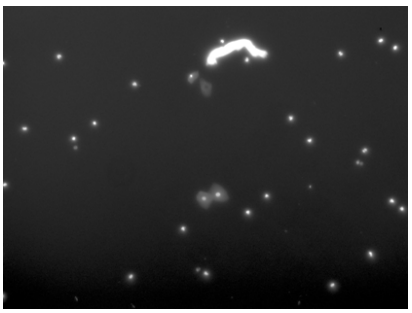


Fig. 1: An image of a cell sample showing the three different types of cells.

N. Bandekar, A. Wong, D. Clausi and M. Gorbet are with the Department of Systems Design Engineering, University of Waterloo, Ontario, Canada, N2L 3G1. nbandekar,a28wong,dclausi,mgorbet@uwaterloo.ca

various challenges such as cell clustering, variation in cell size and shape, brightness variation across cells and debris, the objective of this paper is to build a robust segmentation algorithm.

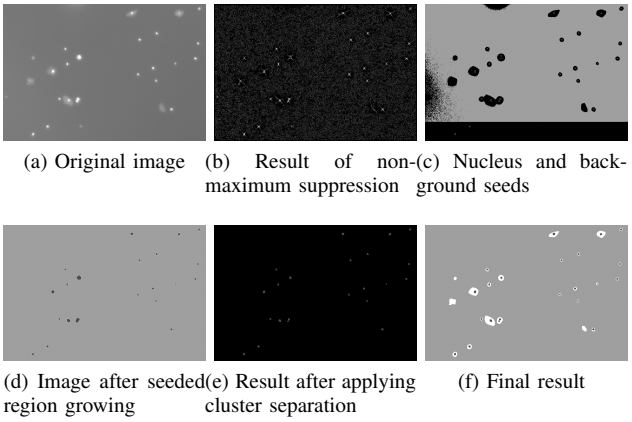
The performance of the proposed technique is investigated in terms of the rate of detection of nucleated cells and the speed of the algorithm to process the images. The performance of the algorithm is also compared to the state-of-the-art level set technique [8] and these results are discussed in Section IV. Finally, conclusions and future work are discussed in Section V.

III. MATERIALS AND METHODS

The algorithm has been built using a series of techniques as shown in Fig. 3. Since the nuclei are the brightest regions in the image, non-maximum suppression is used to detect them as described in Section III-A. Once seed pixels are obtained for nuclei, the other pixels belonging to these nuclei need to be obtained. A seeded region growing method, which is discussed in Section III-B, has been used for this purpose. Background seeds are also needed as an input to the seeded region growing technique. These have been obtained using global thresholding because the background is the darkest region in all images. Since the regions are overgrown in the previous step, this is followed by a cluster segmentation algorithm which is presented in Section III-C. Cell body seeds need to be obtained and they are acquired using a classification criterion which is based on spatial proximity to nucleus pixels and grey level. Seeded region growing is applied to these pixels and a classification of the image pixels is obtained based on three classes: nucleus, cell body and background. An example of these steps is shown in Fig. 2.

A. Non-maximum suppression

The nucleus is observed to be brighter than the rest of the cell body in all images. Also, there is no significant noise or artefacts in the images. Hence, non-maximum suppression (NMS) [5] is adopted to detect the nuclei represented by local maxima. NMS is a preferred starting point to obtain seed pixels belonging to nuclei as compared to edge detection or global thresholding algorithms. NMS is a local thresholding technique. Global thresholding such as Otsu's algorithm are not effective because brightness of nuclei is not constant. Other techniques such as edge detection are very sensitive to noise and will pick up erroneous edges.



(a) Original image (b) Result of non-maximum suppression (c) Nucleus and background seeds

(d) Image after seeded region growing (e) Result after applying cluster separation (f) Final result

Fig. 2: Different stages involved in the proposed technique.

B. Region growing

A few pixels are identified as the local maximum and belonging to the nucleus of some cell using NMS. Next, a seeded region growing (SRG) approach [9] is applied to find the rest of the pixels belonging to the nucleus. A region growing solution is preferred over a clustering technique. Both the grey level and spatial information are exploited in the region growing technique whereas clustering methods such as K-means [10] do not use contextual information. Seed pixels for background are obtained by global thresholding using the minimum grey level in the image as the threshold. SRG uses both spatial proximity and intensity of the pixels when determining the class to which the pixel belongs. The success of this algorithm depends largely on the seeds chosen at the beginning. The algorithm can be sensitive to noise in the image if the seed falls on a noisy pixel. However, this is avoided by selecting a group of seeds for each region which averages out the noise.

C. Cluster segmentation

If there is a cluster of nuclei in the image, the region growing step overcompensates by forming one region for all the nuclei in the cluster. This can be corrected by using the proposed recursive cluster segmentation technique. Other techniques for separating individual cells in clusters such as ellipse fitting and template matching are tedious and computationally expensive. Connected component analysis is applied to determine the oversized clusters. Then the cluster separation algorithm, which is described below, examines each of the connected regions individually.

1) Determine if the region is "valid". A region is valid if it has a size which falls in a range of values (42-172 μm^2) representing the average size of a nucleus. The valid region is then added to a list of pixels which belong to nuclei. It is removed from the image to process the remaining invalid regions.
2) If the invalid region is smaller than the range of acceptable sizes, it is not processed further. If the invalid region is bigger than the allowed nucleus size, the algorithm searches for a sub-region within this

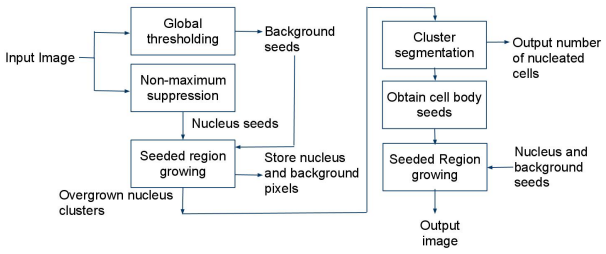
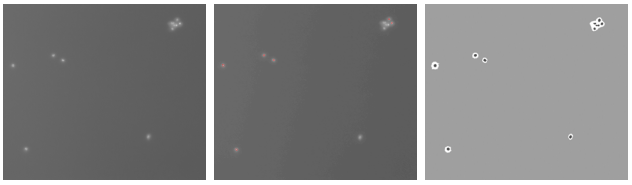
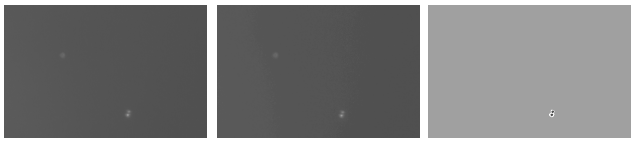


Fig. 3: Flowchart for the proposed technique.



(a) Test image (b) DRLSE result (c) Result of proposed algorithm

Fig. 4: Example where the level set algorithm fails to detect cell clusters whereas the proposed algorithm can detect the individual cells reasonably well.



(a) Test image (b) DRLSE result (c) Result of proposed algorithm

Fig. 5: Example where the level set method fails to detect cells with low contrast but the proposed technique is more effective.

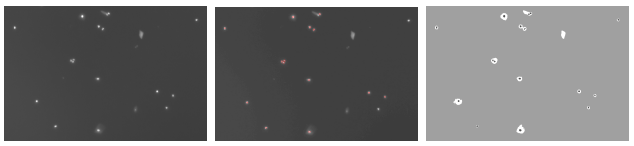
region which has an acceptable size using adaptive thresholding. The threshold is increased iteratively starting from the minimum grey level in the cluster until a valid region is found. If the threshold becomes greater than the maximum grey value of 255, the region is not processed further.

After segmentation of the nuclei is complete, the rest of the cell body is detected. A few cell body seeds are obtained by considering the neighbours of the nucleus pixels. The criterion for determining cell body seeds combines both spatial location of the pixels and their grey level. If the grey level of each of these pixels falls within the range of the mean background and nucleus grey levels, then they are accepted as valid cell body seeds. The seeded region growing algorithm described in Section III-B is used to classify all the pixels as belonging to either of the three classes: nucleus, cell body or background.

IV. Evaluation

The images in the database have a maximum size of 1388x1040 pixels with a 100x magnification. The spatial resolution of each image is approximately 0.7 μm per pixel. The white blood cells can be distinguished from other cells based on their size. The white blood cells have an average diameter of 10 μm, whereas the corneal cells and ghost cells have an average diameter of 30-70 μm. Ghost cells are the only cells without a nucleus, which is seen as a bright spot inside cells.

The results for the algorithm were averaged over 27 images from the dataset. The proposed algorithm was compared to the distance regularized level set technique (DRLSE) [8] and both were run using MATLAB. The initial contour for DRLSE was set to wrap around the entire image. The



(a) Test image (b) DRLSE result (c) Result of proposed algorithm

Fig. 6: Example where the proposed method detects ghost cells but the level set algorithm fails.

TABLE I: Accuracy and F1-measure for the proposed algorithm

Metric	Nucleated cells	Ghost cells	Other
Accuracy	0.957 $\pm$0.067	0.879 $\pm$0.039	0.860 $\pm$0.075
F1-measure	0.973 $\pm$0.046	0.832 $\pm$0.081	0.910 $\pm$0.067

algorithm was run for 4500 iterations for all images for the method to converge. From the results, it was observed that DRLSE could only detect the nuclei and hence failed to identify ghost cells. The performance of both methods was evaluated based on the accuracy of detection of nucleated cells.

The F1-measure [11] is a popular metric for evaluating the correctness of a classification algorithm. It is defined as the harmonic mean of precision and recall.

$$F_1 \text{ measure} = \left(\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right) \quad (1)$$

where, precision $= \frac{TP}{TP+FP}$ and recall $= \frac{TP}{TP+FN}$, TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives in the segmentation and classification result. The accuracy of the algorithm for a class of cells is defined as

$$\text{Accuracy} = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \quad (2)$$

where, TN is the number of true negatives in the classification result.

Table I shows the average accuracy and F1-measure for different classes: nucleated cells and anucleated cells obtained by using the proposed algorithm. Debris, background and other anomalies are included in the "other" class. The nucleated cells are detected with an accuracy of greater than 90% however the method is not reliable to detect ghost cells. Ghost cells have much greater variation in brightness and are sometimes misclassified as background or nucleated cells. Although debris are difficult to identify due to similarity to cells in structure and greylevel, the algorithm is effective at detecting them.

Both DRLSE and the proposed algorithm were run on a 12 GB RAM, 3.2 GHz, dual core machine over all images and the average computation time was recorded. For a fair comparison the proposed algorithm was implemented in MATLAB. Table II compares the average F1-measure and

TABLE II: Comparison between level set and the proposed technique for nucleated cells

	Level Set	Proposed algorithm
Accuracy	0.613 $\pm$0.181	0.957 $\pm$0.067
F1-measure	0.773 $\pm$0.105	0.973 $\pm$0.046
Computational speed (seconds)	1200	36

(a) Test image (b) Result of proposed algorithm

Fig. 7: Example where the proposed algorithm fails to segment the cells due to background illumination gradient.

accuracy for both algorithms for nucleated cells. It is evident from the results that the proposed algorithm outperforms DRLSE in terms of accuracy in detecting nucleated cells and computational speed.

In some cases DRLSE failed to detect nucleated cells but the proposed algorithm yielded good results. Fig. 4 shows an example test image which has cell clusters. The DRLSE algorithm fails to detect the individual cells in the cluster.

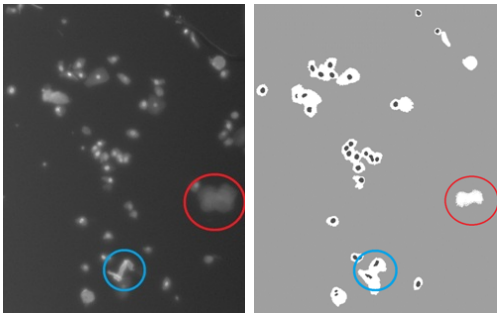
The DRLSE method is sensitive to variation in contrast between nucleus and the cell body. As shown in Fig 5, the proposed algorithm is successful in detecting the nucleated cells but the level set method fails to detect the two cells in the image. The DRLSE algorithm cannot detect ghost cells as illustrated in Fig. 6.

The proposed algorithm was observed to fail in a few situations. It is less effective if there is gradual background illumination variation in the image as illustrated in Fig. 7. This is due to the nature of the seeded region growing algorithm. The background seeds are selected as pixels with low grey levels in the image. So the mean of the background class for seeded region growing algorithm tends to be biased due to the low intensity pixels on the right of the image. The background pixels on the left are closer to the mean grey level of the cell body class. Hence, they are misclassified.

Debris are misclassified as nucleated cells as shown in Fig. 8. Individual ghost cells are not detected if they are in a cluster as shown in Fig. 8. The algorithm only makes use of the greylevel and spatial information and to detect the individual ghost cells more information is needed such as concavities or shape of cells. Since the shape of the cells varies, a notch detection [12] approach would be more effective.

V. CONCLUSIONS AND FUTURE WORKS

A novel algorithm has been implemented for segmentation and classification of cell images for the purpose of automated cell counting. This algorithm is superior to the DRLSE with respect to accuracy, speed and complexity. The technique has contributed to a higher speed and accuracy in counting cells as compared to the AxioVision LE software.



(a) Test image (b) Result of proposed algorithm

Fig. 8: Example where the proposed algorithm fails to segment the debris (blue) and the individual ghost cells (red).

Future work should extend the algorithm to classify nucleated cells into one of two classes: corneal and white blood cells. Morphology and connected component analysis can be used together to obtain the number of different types of cells in the image. In cases where the background is not homogeneous, a background illumination subtraction algorithm can be used. Concavity or notch detection can be applied to detect ghost cell clusters.

REFERENCES

[1] R. Peterson, M. Gorbet, D. Cira, C. Amos, C. Woods and D. Fonn, "Non-Invasive Collection and Examination of Human Corneal Epithelial Cells", *Optometry and Vision Science 2011*, In press.

[2] M. Pons, H. Vivier, J. Remy and J. Dodds, "Morphological characterization of yeast by image analysis", *Biotechnology and Bioengineering*, vol. 42, no. 11, pp. 1352-1359, 1993.

[3] C. Caetano, L. Ventura, S. Sousa and R. Lotufo, "Identification and segmentation of cells in images of donated corneas using mathematical morphology", *Computer Graphics and Image Processing, 2000. Proceedings XIII Brazilian Symposium on*, pp. 344, 2000.

[4] K. Zalewski and R. Buchholz, "Morphological analysis of yeast cells using an automated image processing system", *Journal of Biotechnology*, vol. 48, no. 1-2, pp. 43-49, 1996.

[5] P. Kovesi, "Phase congruency: A low-level image invariant", *Psychological Research*, vol. 64, no. 2, pp. 136-148, 2000.

[6] D. Young, C. Glasbey, A. Gray and N. Martin, "Towards automatic cell identification in DIC microscopy", *Journal of Microscopy*, vol. 192, no. 2, pp. 186-193, 1998.

[7] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours",*Signal Processing*, vol. 71, no. 2, pp. 203-213, 1998.

[8] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance Regularized Level Set Evolution and its Application to Image Segmentation",*IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3243-3254, 2010.

[9] R. Adams and L. Bischof, "Seeded region growing", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 6, pp. 641-647, 1994.

[10] N. Pal and S. Pal, "A review on image segmentation techniques", *Pattern Recognition*, vol. 26, no. 9, pp. 1277-1294, 1993.

[11] C. van Rijsbergen, *Information Retrieval (2nd ed.)*, Butterworth-Heinemann, Newton, MA, USA, 1979.

[12] S. Kothari, Q. Chaudry and M. Wang, "Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques", *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp.795-798, 2009.