SEGMENTATION OF CELL NUCLEI FROM HISTOLOGICAL IMAGES BY ELLIPSE FITTING

J. Hukkanen [1], *A. Hategan* [1], *E. Sabo* [2], *I. Tabus* [1]

[1]Department of Signal Processing
Tampere University of Technology
Tampere, Finland

[2]Department of Pathology
Rappaport Faculty of Medicine, Technion
Haifa, Israel

Email: jenni.hukkanen@tut.fi, andrea.hategan@tut.fi, e_sabo@rambam.health.gov.il, ioan.tabus@tut.fi

ABSTRACT

We propose a new algorithm for non-assisted segmentation of possibly clustered nuclei from histological images. We use elliptic shapes as parametric models to represent the nuclei contours and fit the parameters using the information present in the gray level intensity image and in the derived gradient image. Multiple seeds for each closed contour are found by ultimate erosion of an estimated edge image, resulting in an number of seeds generally larger than the number of nuclei. Our algorithm, called segmentation of nuclei by ellipse fitting (SNEF), constructs several candidate contours for each seed by fitting ellipses to selected subsets of edge pixels. In the end the algorithm selects the contours to be declared nuclei by comparing the values of a suitably chosen goodness of fit criterion. The proposed algorithm produces segmentations in agreement with an expert pathologist.

1. INTRODUCTION

Diagnosis for diseases that involve phenotypic changes in tissue pathology is typically made by the pathologist, who visually inspects histological images. When the consequences of the diagnosis are profound, diagnostic accuracy is understandably critical. However, occasionally, an interobserver variability may exist among the pathologists.

Manual segmentation of diagnostically important patterns from hematoxylin and eosin-stained (H&E) histological images is laborious, time-consuming, and inaccurate. Although many segmentation algorithms have been developed and used in various applications, the segmentation of histological images raises its specific problems, not completely solved yet. The two most difficult issues needed to be solved in view of getting an automatic segmentation of H&E images are the following: first, intensity variations within the borders of a nucleus will lead to the erroneous decision to split that nucleus into more than one object and thus will cause oversegmentation; a second demanding issue is that a number of nuclei appear clustered into a single compound object, which needs to be split into several components.

Thresholding is the simplest method for image segmentation. Typically, the threshold value is chosen based on histogram characteristics of the pixel intensities of the image [5]. Thresholding alone does not solve the problem of clustered nuclei and thus more complex methods are necessary.

During the last decade a number of refined segmentation methods have been introduced, e.g., [1, 3, 4, 7, 8], with specific algorithms designed for the segmentation of clustered objects, including morphological operations, watershed techniques, and model-based approaches where ellipses are fit to the contour of the cluster region. The nuclei images are usually convex and this shape prior can be used in the segmentation process.

The segmentation obtained from morphological operations alone, such as watersheds, is sensitive to noise and is biased towards over-segmentation, thus requiring postprocessing to eliminate the spurious contours. Level set methods are based on minimization of criteria involving region based or contour based functionals and produce in an iterative process accurate segmentations, but often require too high computational efforts. A special class of methods was intended for solving the separation of clustered shapes, such as touching or overlapping grains, based solely on the contour of the region containing the overlapping objects, by fitting parametric models, straight lines [4] or ellipses [1, 7, 8]. In order to get a good separation of the nuclei clusters, the initial contour of the region needed to be extracted with precision and needed to be smoothed, which was ensured by a complex preprocessing stage.

Our goal is to find a fast and reliable method to segment the nuclei in histological images, and provide reconstructions for the nuclei forming overlapping clusters. Since the subsequent utilization of these results for disease diagnosis requires such features as nuclei size, axes alignment, and eccentricity of the shapes, we utilize a parametric representation of the shapes by ellipses. The elliptic shape is well suited approximation for providing all the features of interest for the subsequent processing as it can be observed that most of the time the real nuclei shapes do not deviate very much from ellipses. Unlike other existing approaches, we utilize gradient information, not only outer contour information, since the edges obtained from gradients can convey important cues of the separation lines of the nuclei inside a clustered region, impossible to be guessed solely from the outer contour of the cluster region. We found that a simple criterion expressing the goodness of fit of each ellipse to its corresponding set of edge pixels can select the final contours very reliably and in a more principled way than in the previous approaches.

2. THE SEGMENTATION ALGORITHM

The description of the overall segmentation algorithm is compactly presented in Figure 1. In the following we present more details and some rationales behind its main steps.

2.1 Preprocessing

The hematoxylin and eosin (H&E) stained histological images are originally represented in the red-green-blue (RGB) colorspace. We convert the RGB image to CIE L*a*b co-

 ISSN 2076-1465

lorspace and the luminosity component L is denoted in the sequel as the gray-level image I, which is further processed to get the segmentation.

2.2 Combining intensity and gradient information

We present next the Step 1 of SNEF algorithm in more detail and refer to Figure 2 for illustrations of the intermediate images introduced here. The image I is thresholded to a binary image B by dual thresholding [3]. The set of background pixels is $\mathscr{B}_0$ and the set of border pixels of the object is $\mathscr{F}_1$. We construct a gradient magnitude image G from the gray-level image I using Sobel operator [6].

The gradient magnitude image G is thresholded by dual thresholding in order to eliminate weak contours. We further remove from the resulted binary image those pixels which belong to the background pixel set $\mathscr{B}_0$, resulting in the binary image E. In order to guarantee closed borders in the image E, we additionally set ON all the pixels of the border set $\mathscr{F}_1$. From this image we remove the isolated regions of less than 8 pixels and denote H the resulting edge image. The set of edge pixels is denoted $\mathscr{H}_1$.

2.3 Finding seeds and fitting ellipses

A set of seeds $\mathscr{S}$ for closed contours are found by ultimate erosion applied to the image H. We prefer to overestimate the number of seeds since the segmentation algorithm tolerates to have more seeds than the true number of contours. The ultimate erosion performs the iterative erosion of an object within the image until a last stage, when the object disappears; those objects existing in the image at the stage immediately before the last stage are considered as seeds.

Now we shall describe the process of determining the ellipse that surrounds an arbitrary seed and fits best the intensity and gradient information in image H. From each seed $S_i = (x_{oi}, y_{oi}) \in \mathscr{S}$ a ray is rotated at all angles $\alpha \in \{1^o, \ldots, 360^o\}$ and at a generic angle α the radius r is incremented to generate the points $x_i = x_{oi} + r\cos\alpha$, $y_i = y_{oi} + r\sin\alpha$ on a line, until a pixel of the edge set $\mathscr{H}_1$ is met; we denote this pixel $C_\alpha(x_{oi}, y_{oi})$. After a complete rotation of the ray the obtained pixels $C_\alpha(x_{oi}, y_{oi})$ with $\alpha \in \{1^o, \ldots, 360^o\}$ are grouped into connected components, denoted $\mathscr{C}_1, \ldots, \mathscr{C}_{n_c}$. The connected components are arranged in a list, into a preference order given by the smallest distance from each connected component to the seed. The list is then sequentially processed and the connected components are incrementally appended into a set $\mathscr{D}$. At stage ℓ, after appending a new connected component, an ellipse is fitted to the pixel coordinates in $\mathscr{D}$ by direct least squares fitting of ellipses [2], resulting in the parameter set $\Theta(x_{oi}, y_{oi}, \ell)$. The ellipse pixel set, $\mathscr{E}(x_{oi}, y_{oi}, \ell)$, is generated by using the parameters $\Theta(x_{oi}, y_{oi}, \ell)$ in the equation of the ellipse, and rounding the obtained coordinates at the image grid resolution. Thus, considering all stages $\ell = 1, 2, \ldots$ we obtain a number of candidate ellipses, out of which we need to keep a single winning ellipse, which we will associate to seed S_i. The value of goodness of fit used for ranking the ellipses is described in the next subsection.

2.4 Goodness of fit criterion

The goodness of fit of an ellipse to a potential contour of a nucleus is defined to take into account two important features: the first tells the percentage of ellipse points which are

0 *Preprocessing step*
Consider as gray-level intensity image I the component L obtained in the conversion of the (H&E) stained histological image from RGB colorspace to CIE L*a*b;

1 *Combine intensity and gradient information*

1.1 Construct a binary image B by thresholding the gray-level image I. Denote the background pixel set $\mathscr{B}_0$ and construct the border image F by extracting the set $\mathscr{F}_1$ of border pixels from the binary image B.

1.2 Construct a gradient magnitude image G from the gray-level image I using the Sobel operator.

1.3 Threshold the gradient magnitude image G and perform AND operation with image B to obtain the intermediate edge image E.

1.4 Combine the border image F and the intermediate edge image E to obtain H, where the pixels set ON are called edge pixels, forming the set $\mathscr{H}_1$.

2 *Find a set of seeds $\mathscr{S}$*
Apply ultimate erosion to H.

3 *For each seed $S_i = (x_{oi}, y_{oi}) \in \mathscr{S}$ find several candidate ellipses and choose the best*

3.1 *Rotate a ray centered at the seed:*
For each angle α consider the line $x_i = x_{oi} + r\cos\alpha$, $y_i = y_{oi} + r\sin\alpha$ where r is incremented until a pixel, denoted $C_\alpha(x_{oi}, y_{oi})$, on the edge set $\mathscr{H}_1$ is reached.

3.2 Group all obtained points $C_\alpha(x_{oi}, y_{oi})$ into connected components, denoted $\mathscr{C}_1, \ldots, \mathscr{C}_{n_c}$. Arrange the connected components into increasing order based on the smallest distance from the connected component to the seed.

3.3 *Loop incrementally appending more connected components into a set $\mathscr{D}$:*
For $\ell = 1$ to n_c

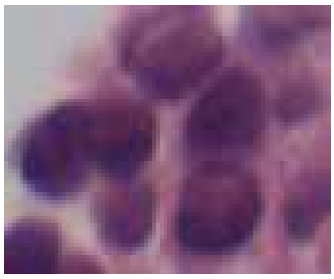
3.3.1
$$\mathscr{D} \leftarrow \mathscr{D} \cup \mathscr{C}_\ell$$

3.3.2 Fit an ellipse to the pixels coordinates in $\mathscr{D}$, resulting in the parameter set $\Theta(x_{oi}, y_{oi}, \ell)$.

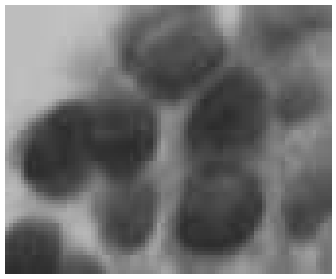
3.4 Choose out of the ellipses $\Theta(x_{oi}, y_{oi}, 1), \ldots, \Theta(x_{oi}, y_{oi}, n_c)$ the one which maximizes the criterion (1).

4 *Selecting the final segmentation*
Order the seeds in decreasing order of criterion (1). For every seed check if its ellipse has an overlap larger than 60% with any of the previously chosen ellipses, and if yes remove it.

Figure 1: The algorithm for segmentation of nuclei by ellipse fitting (SNEF)

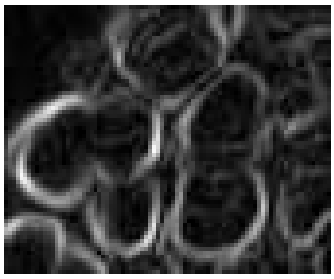


(a) Selected part of the original image



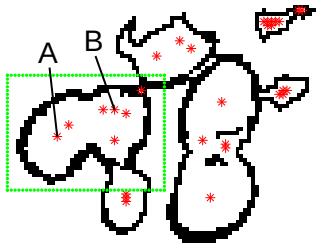
(b) Gray-level image I

(c) Binary image B obtained by thresholding I



(d) Gradient magnitude G

(e) Intermediate edge image E obtained by thresholding the gradient image G and performing AND operation with B.



(f) The final edge image H and the seeds (red stars). The segmentation of the selected part of the image (green box) is illustrated in Fig. 3 for the two seeds marked as A and B.

Figure 2: Illustration of the Steps 0-2 of the SNEF algorithm: preprocessing, combining intensity and gradient information, and the set of seeds found.

in the immediate vicinity of any edge pixel belonging to $\mathscr{H}_1$; while the second tells the percentage of connected component pixels in the set $\mathscr{D}$ that are in the vicinity of an ellipse pixel. The first term is high when the ellipse corresponds to almost a close contour made out of edge pixels. But for overlapping nuclei we may have incomplete contours in the edge image, and thus we may want to reward also the situations with incomplete edge contours, when the edge pixels in $\mathscr{D}$ forming just a partial contour are fitting very well with the ellipse.

The goodness of fit for the ellipse $\Theta(x_{oi},y_{oi},\ell)$ is thus evaluated as follows:

$$V(x_{oi},y_{oi},\ell) = \frac{|\mathscr{E}(x_{oi},y_{oi},\ell)\cap\mathscr{H}_1'|}{|\mathscr{E}(x_{oi},y_{oi},\ell)|} + \\ + \frac{|\mathscr{D}(x_{oi},y_{oi},\ell)\cap\mathscr{E}'(x_{oi},y_{oi},\ell)|}{|\mathscr{D}(x_{oi},y_{oi},\ell)|}, \qquad (1)$$

where $\mathscr{E}'$ denotes the set of pixels $\mathscr{E}$ dilated by the cross structural element, and similarly $\mathscr{H}_1'$ is the dilation of the set $\mathscr{H}_1$. We use the dilated sets so that we count in the intersection of sets not only exact matching of pixels of the ellipse and the edge set involved, but also we count the almost matching when the pixel of the ellipse is in the four neighbour vicinity of an edge pixel.

2.5 Selecting the ellipses for the final segmentation

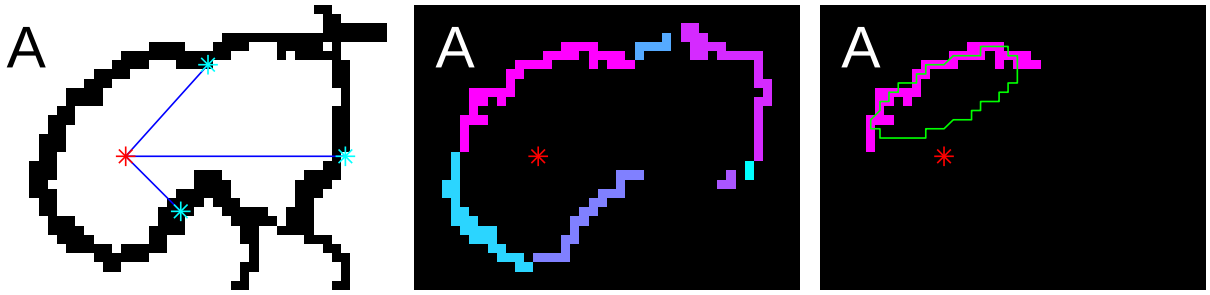
The best fitting ellipse for a seed $S_i = (x_{oi},y_{oi}) \in \mathscr{S}$ is chosen out of the ellipses $\Theta(x_{oi},y_{oi},1),\ldots,\Theta(x_{oi},y_{oi},n_c)$ so that the chosen ellipse has the highest value of criterion (1).

Since we have anticipated a higher number of seeds than nuclei (this usually happens with the ultimate erosion operator), the winning ellipses obtained for different seeds can overlap, as some of the seeds can be near each other and represent same nuclei. Therefore, we need to decide which one of several competing ellipses is really representing the nucleus. For this, we first order the seeds in decreasing order of criterion (1) and for each seed we check if its ellipse has an overlap larger than 60% with any of the previously chosen ellipses. If such an overlap exists, the seed and its ellipse are removed from the list. The list of ellipses obtained in the end represents the final segmentation.

3. ILLUSTRATION OF THE ALGORITHM

We illustrate the algorithm using the image in Figure 2(a), which is one of the difficult parts of the original image in Figure 4(a). The Steps 0 and 1 of the SNEF algorithm in which we obtain the edge image H are illustrated in the Figures 2(b) to 2(f). In order to illustrate the segmentation Steps 2 to 4 of the algorithm we continue only with the small rectangle shown in green in Figure 2(f). The results of the operations performed in Step 3 for the seed A are shown in images 3(a) to 3(f) and similarly the results for the seed B are shown in images 3(g) to 3(h). In Figure 3(i) the overall segmentation results for Figure 2(a) are presented.

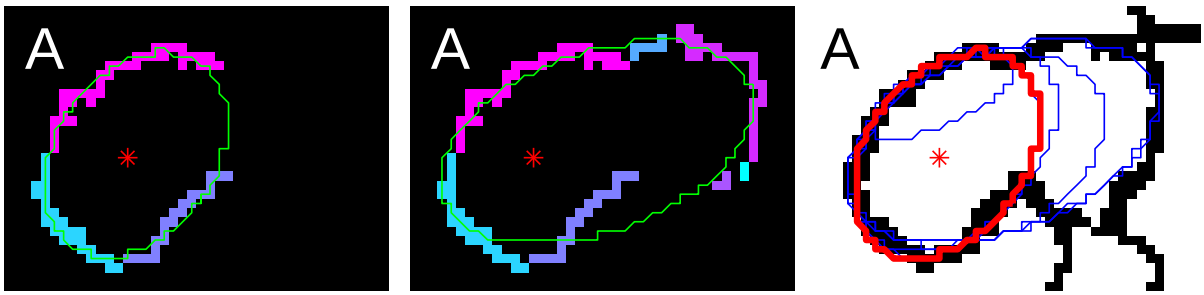
In Figure 3(a) is illustrated the Step 3.1 of the algorithm in which the edge pixels reachable from a seed are obtained. In Figure 3(b) the obtained pixels are grouped into connected components (Step 3.2 of the algorithm). From different seeds one can reach different edge pixels and therefore each seed will have its different connected components. This can be seen by comparing the connected components obtained for the seed A (Figure 3(b)) and the seed B (Figure 3(g)).



(a) A ray centered at the seed A (red star) rotates picking at each angle one pixel (cyan star) from the edge pixel set $\mathscr{H}_1$ (only three ray positions are shown).

(b) Edge pixels picked by the ray are grouped into seven connected components (each separate connected component has its own color).

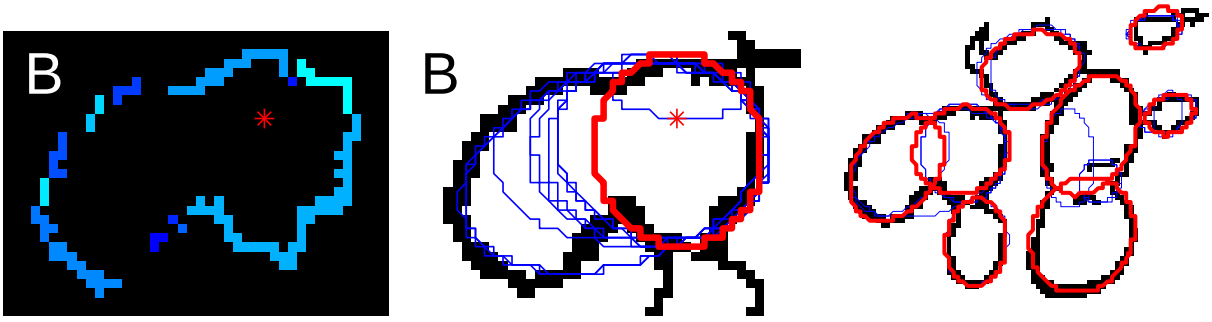
(c) The connected component closest to seed A and the fitted ellipse.



(d) The three connected components closest to seed A and the fitted ellipse.

(e) All seven connected components and the fitted ellipse.

(f) In blue: all the fitted ellipses for seed A. In red: the winning ellipse, having the largest value of criterion (1).



(g) The 12 connected components resulted for seed B.

(h) In blue: all the fitted ellipses for the seed B. In red: the winning ellipse, having the largest value of criterion (1).

(i) In blue: the best fitted ellipses for all different seeds. In red: the final segmentation obtained after removing the overlapping ellipses in the Step 4 of our algorithm.

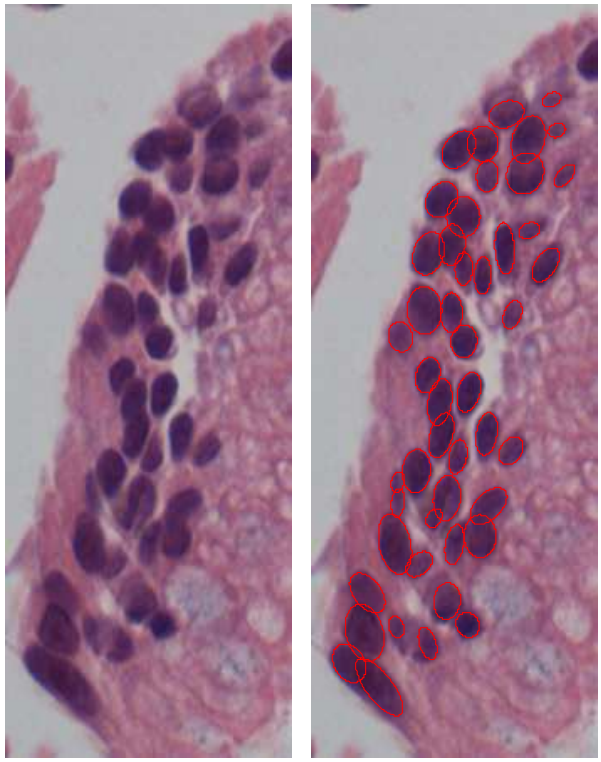
Figure 3: Illustration of Steps 3 and 4 of SNEF algorithm for fitting ellipses and selecting the final segmentation. The case of seed A is considered in (a) - (f) while the case of seed B is considered in (g) - (h). In (i) the overall segmentation results for the image in Figure 2(a) are presented superposed over the edge image H from Figure 2(f).

In Figures 3(c) to 3(f) are illustrated Steps 3.3. and 3.4 of the algorithm in which connected components are incrementally appended into a set $\mathscr{D}$; after each appending an ellipse is fitted to pixels in the set $\mathscr{D}$; finally the value of criterion (1) is calculated for each fitted ellipse. The closest connected component to the seed A (in the sense of the minimum distance between the seed and the pixels of the connected component) is shown in Figure 3(c), together with the fitted ellipse. The value of criterion (1) for this ellipse is 1.65. The first term of the criterion is 0.68 and the second term 0.97 meaning that the ellipse fits well the connected components in the set $\mathscr{D}$ (second term) but the relatively low percentage of ellipse pixels on edge pixels $\mathscr{H}_1$ (first term) penalizes the criterion. The three closest connected components and their fitted ellipse are illustrated in Figure 3(d). The first term of the criterion (1) is 0.86 and the second term 0.99. Thus, the value of criterion (1) is 1.85, which is better than for the ellipse presented in Figure 3(c). The biggest change in the value of the criterion (1) is in the first term of the criterion, caused by ellipse pixels touching more edge pixels. In Figure 3(e) there are all seven connected components and their fitted ellipse. Now, the value of criterion is 1.74. consisting of 0.96 and 0.78 as a first and second terms, respectively. Although the ellipse fits better the edge pixels (first term), the fit to connected components (second term) is lower than in case of one or three connected components.

In Figure 3(f) are presented all the fitted ellipses of different combinations of connected components and emphasized in red is the best fitted ellipse based on the criterion (1) in the case of seed A. In case of seed B the similar results are presented in Figure 3(h).

In Figure 3(i) complete results are presented, for all seeds, after processing the image in Figure 2(a). From each seed of Figure 2(f) only the best fitting ellipse based on the criterion (1) is taken and presented in blue in Figure 3(i). However, the ellipses of different seeds can overlap. Thus, the seeds are arranged in the order of criterion and if there is more than 60% overlap between better fitting ellipse, the ellipse is removed (Step 4 of SNEF).

In Figure 4 the final results of the algorithm are presented for a large section of a histological image from the biopsy of Barrett's esophagus mucosa. The original image is presented for comparison. It can be seen that our segmentation algorithm gives accurate results for the segmentation of cell nuclei from histological images. The results have been checked by the expert pathologist who was in complete agreement with the segmentation.



(a) The original tissue image (b) The segmentation results

Figure 4: The segmentation results

4. CONCLUSIONS

We proposed a new algorithm for segmentation of possibly clustered nuclei from histological images using ellipse fitting. To separate the overlapping nuclei, the algorithm utilizes the information from both intensity and gradient images. We proposed also a criterion to select between several competing ellipses. We evaluated the algorithm on real histological images and the segmentation results were in agreement with the segmentation proposed by an expert pathologist.

REFERENCES

[1] X. Bai, C. Sun, and F. Zhou, "Splitting touching cells based on concave points and ellipse fitting," *Pattern Recognition*, vol. 42, pp. 2434–2446, 2009.

[2] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square of ellipse," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 476–480, 1999.

[3] M. Hu, X. Ping, and Y. Ding, "Automated cell nucleus segmentation using improved snake", *International Conference on Image Processing*, vol. 4, pp. 2737–2740, 2004.

[4] S. Kumar, S. H. Ong, S. Ranganath, T. C. Ong, and F. T. Chew, "A rule-based approach for robust clump splitting", *Pattern Recognition*, vol. 39, pp. 1088–1098, 2006.

[5] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", *Journal of Electronic Imaging*, vol. 13, pp. 146–165, 2004.

[6] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. Brooks/Cole Publishing Company, Pacific Grove, CA, 1999.

[7] H. Talbot and B. C. Appleton, "Elliptical distance transforms and object splitting", *Proc. of VIth International Symposium on Mathematical Morphology*, April 3-5. Sydney, Australia, 2002.

[8] G. Zhang, D. S. Jayas, and N. D. G. White, "Separation of touching grain kernels in an image by ellipse fitting algorithm", *Biosystems Engineering*, vol. 92, pp. 135–142, 2005.