

Splitting of overlapping nuclei guided by robust combinations of concavity points

Marina E. Plissiti^a, Eleni Louka^a and Christophoros Nikou^a

^aDepartment of Computer Science and Engineering,
University of Ioannina, 45110 Ioannina, Greece;

ABSTRACT

In this work, we propose a novel and robust method for the accurate separation of elliptical overlapped nuclei in microscopic images. The method is based on both the information provided by the global boundary of the nuclei cluster and the detection of concavity points along this boundary. The number of the nuclei and the area of each nucleus included in the cluster are estimated automatically by exploiting the different parts of the cluster boundary demarcated by the concavity points. More specifically, based on the set of concavity points detected in the image of the clustered nuclei, all the possible configurations of candidate ellipses that fit to them are estimated by least squares fitting. For each configuration, an index measuring the fitting residual is computed and the configuration providing the minimum error is selected. The method may successfully separate multiple (more than two) clustered nuclei as the fitting residual is a robust indicator of the number of overlapping elliptical structures even if many erroneous concavity points are present due to noise. Moreover, the algorithm has been evaluated on cytological images of conventional Pap smears and compares favorably with state of the art methods both in terms of accuracy and execution time.

Keywords: Overlapping nuclei segmentation, concavity points, ellipse fitting, Pap smear images;

1. INTRODUCTION

Cell overlapping is a widespread phenomenon in the analysis of cytological images obtained from microscopical slides, especially in the case of the well-known Pap smear. This is an effect that arises from the existence of different layers of the cervical specimen in a slide, which results in the partial obscurity of the cells lying in the lower layers by those lying in the upper layers of the slide. The interpretation of these images relies basically on the identification of the nuclei of the cells, which exhibit significant morphological changes in the presence of a disease, and the nuclei features have shown high discriminative ability in recognizing pathological cases.¹ For these reasons, the delineation of each nuclei in overlapping areas is a challenging issue, which attracts the scientific interest for the development of automated methods for the analysis of the specific images.

The first work addressing the problem of recognition of overlapping cells nuclei in Pap smear images was proposed in.² The overlap detection algorithm was based on information both from the nucleus contour and from the density profile. In recent studies, a spatially adaptive active physical model is used for the discrimination of each nucleus in nuclei clusters, which provides a compact description of the shape model and exploits local characteristics of the nuclei boundary and *a priori* knowledge about the expected shape of the nucleus.³ Although this method can be applied to nuclei clusters containing more than two overlapping nuclei, the total number of the nuclei in each cluster is not determined automatically and must be predefined, in order to provide acceptable results. Other methods that provide the automated calculation of the number of the overlapped nuclei have been proposed. The distance transform in a binary image containing the regions of the nuclei is calculated in Jung et al.,⁴ and the topographic surface generated by the distance transform is considered as a Gaussian mixture. The EM algorithm is then applied for the determination of the parameters of each nucleus cluster. In Jung et al.,⁵ the segmentation of clustered nuclei is treated as an optimization problem and a marker extraction scheme based on the H-minima transform is introduced to obtain the optimal segmentation result from the distance map. Finally, in Bai et al.,⁶ the boundary of the nuclei cluster is divided into several segments controlled by

Further author information: (Send correspondence to C.N.)
C.N. : E-mail: cnikou@cs.uoi.gr, Telephone: +30 265 100 8802

the concave points and then the best fitting of ellipses with regard to the boundary segments is estimated. This method can be applied to images containing an unknown number of overlapped nuclei. However, the use of multi-measures results in many thresholds that must be pre-defined empirically.

In this work, we present a novel method for the definition of each individual nucleus in nuclei clusters. The method exploits the geometrical characteristics of the boundary of the nuclei cluster and the expected nucleus shape. It must be noted that most of the nuclei usually have ellipse-like boundaries with the intensity of the pixels inside these boundaries being lower than those lying outside. Besides, this is the reason why many methods are based on ellipse fitting techniques⁴⁻⁶ or ellipse-template matching⁷ for the detection of the nuclei in cytological images.

Thus, given the ellipse-like nucleus shape, the concavity points of the global boundary are estimated, which separate the boundary of the clustered nuclei into two or more independent parts, depending on the number of the nuclei in the cluster. These parts are then used for the calculation of the best ellipse fitting criterion, which results in the detection of the individual nuclei. The main idea of the method is that we expect each concavity point to be a common point for two adjacent nuclei boundaries. The proposed method does not depend on prespecified thresholds, it is fast, robust and it provides reliable detection of each nucleus boundary. Furthermore, in comparison with other methods, it exhibits more accurate boundaries of the overlapped nuclei in terms of the Euclidean distance and in most cases, it requires shorter processing time.

2. METHOD

The proposed method consists of two main steps: the first step is the determination of the global boundary of the nuclei cluster and the calculation of strong concavity points along the boundary. The second step is the determination of the overlapping nuclei using the concavity points along with an ellipse fitting criterion.

2.1 Localization of strong concavity points on the cluster boundary

A preprocessing procedure is first performed in the cytological image, for the reduction of the noise effects. Thus, a Gaussian filter and a contrast enhancement filter are sequentially applied to the gray scale counterpart of the initial image. Then, a standard global thresholding⁸ results in the detection of the global boundary of the nuclei cluster (Fig. 1(b)). This is a simple and effective technique, especially for real time applications, since the nuclei are usually darker than the background, and a global threshold can be used in order to define these distinct parts of the image. Finally, opening by reconstruction and closing by reconstruction are sequentially performed, in order to provide a smoother nuclei cluster boundary.

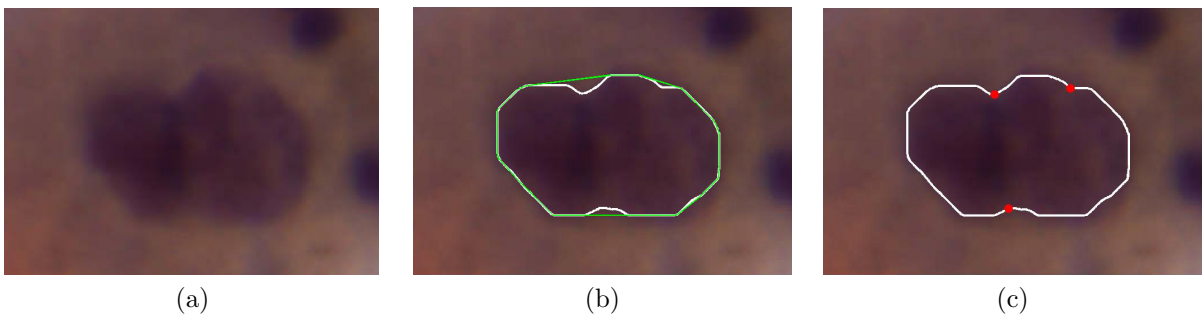


Figure 1. (a) Initial image, (b) the nuclei cluster boundary (in white) with its convex hull (in green) and (c) the concavity points. The figure is better seen in color.

Ideally, the global boundary of the nuclei cluster can be considered as the concatenation of parts of ellipses, as the shape of each nucleus is ellipse-like. Thus, each concavity point would indicate the position of overlapping of two single nuclei. However, the existence of noise and the artifacts that are introduced by uneven staining of the cells, result in the detection of more concavity points along the boundary of the nuclei cluster. These concavity points belong to individual nuclei and they are due to the detection of jugged parts in the nucleus

boundary. For the calculation of all the concavity points of the boundary, we first calculate the convex hull of the global boundary (Fig. 1(b)). Then, for each line segment of the convex hull that does not belong to the global boundary, we identify the points of the boundary that have the largest distance from the convex hull (Fig. 1(c)). In order to identify the strong concavity points, we consider a specific point on the boundary to be a concavity point if the distance from the convex hull is larger than some pixels (e.g. 3 pixels for the experiments presented here on Pap smear images), otherwise we reject this point. In this way, we avoid to process points near the convex hull, as they are probably produced by the inhomogeneity of the image.

2.2 Determination of the overlapping nuclei

The detected concavity points are used for the separation of the global cluster boundary into its parts. Each possible combination of concavity points separates the global boundary into independent parts. These parts are candidate boundaries for each nucleus in the nuclei cluster. As we can see in Fig. 2(b), only one combination of concavity points provides the correct identification of each nucleus boundary. For the determination of the specific boundaries of the individual nuclei, we perform ellipse fitting to each of the parts of the global boundary.

The boundary of the nuclei cluster contains parts of the boundary of each individual nucleus. We expect these boundaries to be ellipse-like, as the true shape of a nucleus.⁴⁻⁷ For this reason, we proceed with the approximation of each part of the boundary by an ellipse whose parameters are obtained using the direct least squares ellipses algorithm.⁹

The number of combinations of different candidate ellipses that fit to the boundary depends on the number of the detected concavity points. More specifically, for N candidate concavity points, considering that only m out of them are true concavity points, that is, they are not generated due to noise, the number of candidate ellipses along the boundary is given by:

$$C_m^N = \binom{N}{m} = \frac{N!}{m!(N-m)!}.$$

As a consequence, in images containing noise or artifacts and the nuclei cluster boundary is not smooth, the detection of many concavity points will result in the estimation of a large number of ellipses as the total number of candidate ellipses is $\sum_{m=1}^N C_m^N$. Nevertheless, only one combination of these ellipses corresponds to the actual nuclei boundaries. These ellipses are selected using the value of a distortion evaluation function which measures the averaged fitting residual (AFR)⁵ of each boundary point to the respective ellipse.

Let $\mathbf{B}_{(m,N)}^j = \{\mathbf{b}_{(m,N)}^{j1}, \mathbf{b}_{(m,N)}^{j2}, \dots\}$ be the splitting of the boundary corresponding to the j^{th} out of C_m^N combinations of boundary points. Let also $\mathbf{E}_{(m,N)}^j = \{\mathbf{e}_{(m,N)}^{j1}, \mathbf{e}_{(m,N)}^{j2}, \dots\}$ be the corresponding set of ellipses that fit to the parts of the boundary. Then, the following measure between a given splitting of the boundary and the corresponding ellipses, which integrates the residuals from the ellipses, is calculated:

$$D\left(\mathbf{B}_{(m,N)}^j, \mathbf{E}_{(m,N)}^j\right) = \sum_i r\left(\mathbf{b}_{(m,N)}^{ji}, \mathbf{e}_{(m,N)}^{ji}\right), \quad (1)$$

where $r(\cdot, \cdot)$ represents the residuals of the fitted ellipses to the boundary points, for the underlined combination of concavity points. The index i in the summation depends on the number of ellipses of the involved combination. The optimal segmentation is achieved by the combination of concavity points minimizing (1), that is

$$\{m^*, j^*\} = \arg \min_{m,j} D\left(\mathbf{B}_{(m,N)}^j, \mathbf{E}_{(m,N)}^j\right),$$

which indicates the number of concavity points (m) and their exact combination (j) that provide the minimum distortion. The overall procedure may be summarized in Algorithm 1.

In Fig. 2, we may observe the results of the method in a conventional image containing a nuclei cluster. For the specific image, three concavity points were detected on the global boundary (Fig. 1(c)). The combination of these concavity points by two, would result in the division of the boundary into three possible ways (Fig. 2(a,b,c)). In each case, the global boundary is divided into two parts. After ellipse fitting, the resulting ellipses

Algorithm 1 Overlapping nuclei separation

Input: Image of clustered nuclei

Output: The elliptical boundary of each nucleus

- Extract the boundary of the nuclei cluster.
- Detect the concavity points of the boundary.
- For each combination of concavity points
 - Fit the ellipses to the parts of the boundary determined by the concavity points and compute the average distortion (1).
- Select the combination of the set of ellipses with the minimum distortion.



Figure 2. Three concavity points were detected on the global boundary in this example. (a)-(c) Their combinations by two, result in the division of the boundary into three possible ways with two nuclei. (d) The last case considers that there are three nuclei in the image. The optimal results are depicted in (b). The figure is better seen in color.

are depicted and the calculated distortion evaluation function (1) for the first, second and third case are 0.0155, 0.0133 and 0.0337 respectively. Furthermore, in Fig. 2(d) the result of the combination of three concavity points by three is depicted, and the corresponding value for the distortion evaluation function is 1.229. Thus, the second case in Fig. 2(b) is the optimal segmentation, as it was expected by visual inspection.

3. RESULTS AND DISCUSSION

Our method was evaluated on a data set of 30 images from conventional Pap smears containing clusters of overlapping nuclei. The number of the individual nuclei in these images varied from two to four. The images were acquired through a CCD camera (Olympus DP71) adapted to an optical microscope (Olympus BX51). We used a $40\times$ magnification lens and the acquired images were stored in JPEG format. As it was verified by the results, in all the images of our data set, the method is able to identify the distinct areas of the overlapping nuclei and to approximate their boundary with an ellipse. The calculation of the distortion evaluation function ensures that from all the ellipses that will be obtained for a specific image, only the set of ellipses with the best match will be selected for the final result.

In our data set, we have included complex images with uneven staining of the cells and inhomogeneity in the intensity of the background and the nuclei areas. Even in those cases (Fig. 3) where the cluster boundary presents jagged parts, the boundaries of each nucleus are accurately identified. As we can observe in Fig. 3(b), the detected boundary of the nuclei cluster is quite rugged and presents several concavity points (Fig. 3(c)). The method exhaustively process all of the possible combinations of concavity points in order to provide two ellipses, which fit in the specific parts of the boundary (Fig. 3(d-f)). From the entire set of combinations, only one pair is acceptable by the use of the distortion evaluation function. Thus, the method achieves to identify the individual nuclei in the cluster, as the actual concavity points that separate the different nuclei are identified in every case. This leads to the detection of the parts of the boundary that correspond to each nucleus.

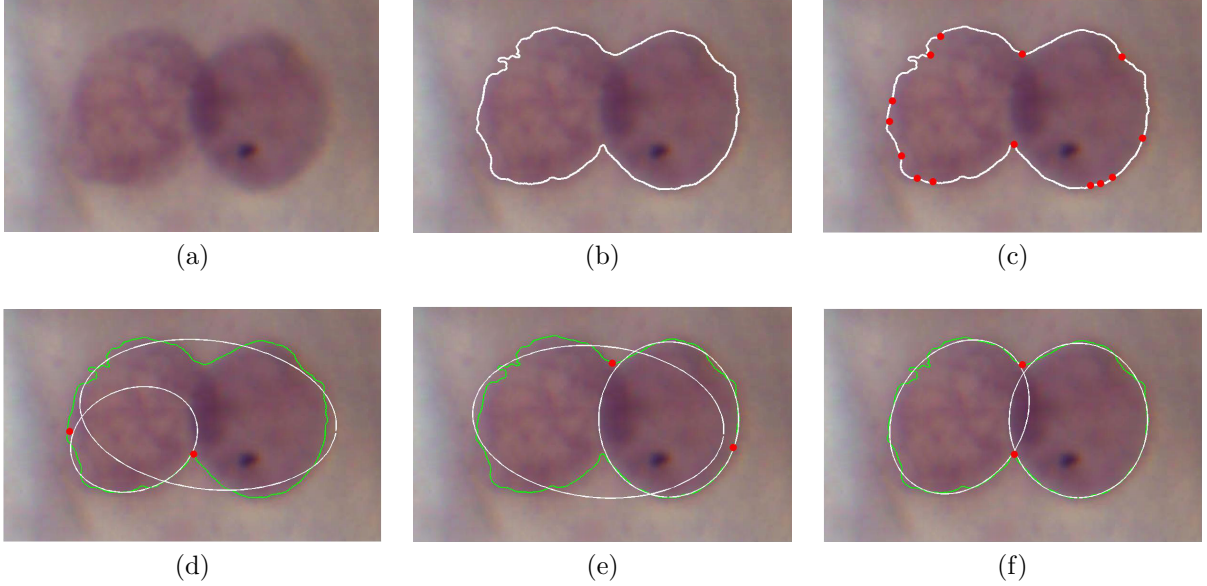


Figure 3. (a) Initial image, (b) the nuclei cluster boundary, (c) the concavity points, (d)-(e) intermediate results with the selection of two concavity points, (f) the optimal separation achieved with our method. Notice that even in this case of many detected concavity points, our method results in the identification of reliable results for the boundary of each nucleus as it is depicted in (f). The figure is better seen in color.

Table 1. Comparative results in terms of accuracy and processing time (mean \pm std).

Method	Error (pixels)	Processing Time (sec)
Proposed Method	6.34 ± 2.51	3.15 ± 2.69
Concave Points ⁶	7.77 ± 4.01	1.49 ± 0.14
Watersheds ⁵	9.16 ± 8.09	11.44 ± 6.24
Gaussian Mixture ⁴	6.42 ± 3.00	827.20 ± 565.3

We have also compared our method with the methods proposed by Jung et al.^{4,5} and the method of Bai et al.⁶ in terms of the accuracy and the execution time required for the extraction of the results. The method of Bai et al.⁶ also exploits the concavity points of the boundary of the nuclei cluster and then the best fitting of ellipses with regard to the boundary segments is estimated, based on a number of rules that must be satisfied and some predefined thresholds. The method of Jung et al.⁵ uses the distortion evaluation function but the candidate ellipses are obtained by the watershed transform. In addition, the method proposed of Jung et al.⁴ separates the pixels in the nuclei cluster using Gaussian Mixture Models. The accuracy of the method was evaluated with the Euclidean distance from the ground truth, which was manually obtained by an expert cytopathologist. The comparative results are included in Table 1.

As it was verified by the results, our method outperforms the aforementioned methods in terms of the accuracy. The nuclei boundaries that are derived are closer to the ground truth in all cases. Thus, a more reliable approximation of each nucleus boundary is feasible by our method. This is also achieved in the case of nuclei clusters that contain more than two individual nuclei, as it is depicted in Fig. 4. It must be noted that our method produces the best results when the approximated nucleus boundary follow an elliptical shape (Fig. 4(a)). However, in some degenerate cases such as in Fig. 4(b) (middle nucleus) the method is limited in producing the best ellipse fitting in the part of the boundary that is identified as a single nucleus, since the rest of the nucleus area does not follow elliptical shape. This is the main common limitation for all the methods that perform ellipse fitting for the determination of the nuclei boundaries.

From Table 1 we can observe that the method proposed by Bai et al.⁶ requires less processing time than our method. However, since this method requires the predefinition of several thresholds which are obtained empirically and they depend on the information included in each image, in combination with the fact that our

method is more accurate, we can conclude that our method is more efficient. The results also indicate that our method is faster compared to both methods of Jung et al.,^{4,5} since it produces the optimal results in 3.15 ± 2.69 seconds (mean \pm standard deviation) in contrast to these methods which requires 11.44 ± 6.24 and 827.2 ± 565.3 seconds respectively. This can be explained as the iterative application of the watershed transform in Jung et al.⁵ for the determination of the best segmentation of each nuclei introduces a computational overhead. Furthermore, the estimation of each distribution parameters through the EM algorithm in the method of Jung et al.⁴ results in a remarkable increase in computational time. This is avoided by our method by exploiting the concavity points, which are computed once and there are no additional parameters to be calculated. Note that the relatively high standard deviation with respect to the mean value for our method is due to the skewed histogram of the execution times (some images present many candidate concavity points which should be examined).

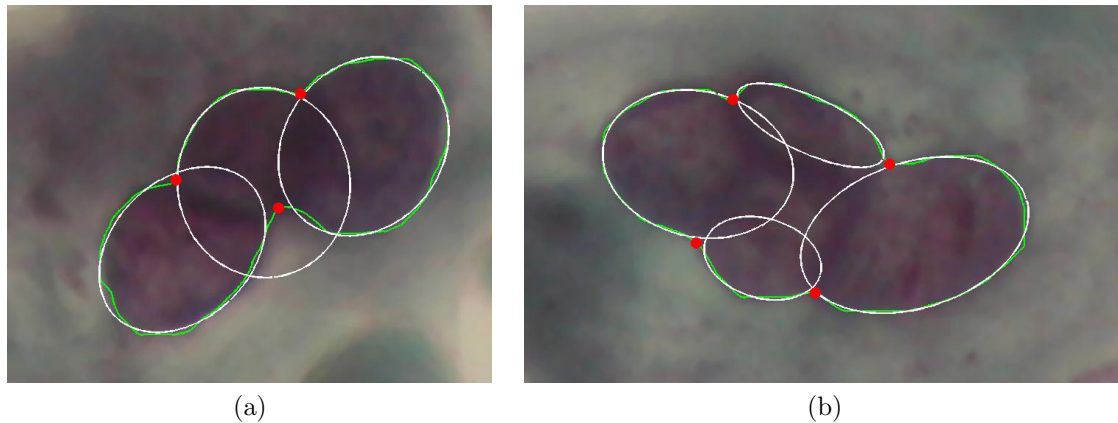


Figure 4. Results obtained in cell clusters containing (a) three and (b) four individual nuclei. The figure is better seen in color.

4. CONCLUSION

The correct identification of the nucleus boundary is a crucial task for the reliable interpretation of Pap smear images, as the nuclei provides significant information about the cell pathology. The existence of overlapping nuclei in these images is a challenging issue for every automated method aiming at the segmentation of the individual nuclei, as these images present several limitations. The method presented herein exploits the concavity points of the global boundary of the cluster and identifies the ellipses that best fit to the boundary of each nucleus. As it was verified by the results, the method is accurate, fast and robust, even in the case where the actual global boundary of the cluster has not been detected precisely, and many concavity points are produced. A perspective of this work is the extension of the method to the extraction of the nuclei boundaries using a more flexible model than the ellipse fitting, in order to improve the approximation of the nuclei boundaries that do not follow an elliptical shape.

ACKNOWLEDGMENTS

This work is co-financed by the European Union (European Regional Development Fund- ERDF) and Greek national funds through the Operational Program “THESSALY-MAINLAND GREECE AND EPIRUS-2007-2013” of the National Strategic Reference Framework (NSRF 2007-2013).

REFERENCES

1. M. E. Plissiti and C. Nikou, “Cervical cell classification based exclusively on nucleus features,” in *Proceedings of the 9th International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science*, **7325**, pp. 483–490, 2012.
2. E. Bengtsson, O. Eriksson, J. Holmquist, T. Jarkrans, B. Nordin, and B. Stenkvist, “Segmentation of cervical cells: Detection of overlapping cell nuclei,” *Computer Graphics and Image Processing* **16**, pp. 382–394, 1981.

3. M. E. Plissiti and C. Nikou, "Overlapping cell nuclei segmentation using a spatially adaptive active physical model," *IEEE Transactions on Image Processing* **21**(11), pp. 4568–4580, 2012.
4. C. Jung, C. Kim, S. W. Chae, and S. Oh, "Unsupervised segmentation of overlapped nuclei using Bayesian classification," *IEEE Transactions on Biomedical Engineering* **55**(12), pp. 2825–2832, 2010.
5. C. Jung and C. Kim, "Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization," *IEEE Transactions on Biomedical Engineering* **57**(10), pp. 2600–2604, 2010.
6. X. Bai, C. Sun, and F. Zhou, "Splitting touching cells based on concave points and ellipse fitting," *Pattern Recognition* **42**, pp. 2434–2446, 2009.
7. A. Garrido and N. P. de la Blanca, "Applying deformable templates for cell image segmentation," *Pattern Recognition* **33**, pp. 821–832, 2000.
8. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics* **9**(1), pp. 62–66, 1979.
9. A. W. Fitzgibbon, M. Pilu, and R. B. Fischer, "Direct least squares fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(5), pp. 476–480, 1999.