# Segmentation of cell clusters in Pap smear images using intensity variation between superpixels

Marina E. Plissiti[1], Michalis Vrigkas[1] and Christophoros Nikou[1]
[1]Department of Computer Science & Engineering, University of Ioannina, Ioannina 45110, Greece
*marina@cs.uoi.gr*

***Abstract*—The automated interpretation of Pap smear images is a challenging issue with several aspects. The accurate segmentation of the structuring elements of each cell is a crucial procedure which entails in the correct identification of pathological situations. However, the extended cell overlapping in Pap smear slides complicates the automated analysis of these cytological images. In this work, we propose an efficient algorithm for the separation of the cytoplasm area of overlapping cells. The proposed method is based on the fact that in isolated cells the pixels of the cytoplasm exhibit similar features and the cytoplasm area is homogeneous. Thus, the observation of intensity changes in extended subareas of the cytoplasm indicates the existence of overlapping cells. In the first step of the proposed method, the image is tesselated into perceptually meaningful individual regions using a superpixel algorithm. In a second step, these areas are merged into regions exhibiting the same characteristics, resulting in the identification of each cytoplasm area and the corresponding nuclei. The area of overlap is then detected using an algorithm that specifies faint changes in the intensity of the cytoplasm of each cell. The method has been evaluated on cytological images of conventional Pap smears, and the results are very promising.**

***Keywords*—*Pap smear images, superpixels, cell cluster separation, overlapped cell segmentation.***

## I. Introduction

Cervical cancer is considered to be the fourth most common cause of cancer death in women worldwide [1]. The impact of the well known Pap-test examination in the prevention of cervical cancer is undoubtedly significant. Through this examination precancerous conditions and abnormal changes that may develop into cancer are recognized and they are early treated preventing the development of cervical cancer.

The Pap smear slides are examined under a microscope to identify abnormalities in the structure and morphology of cells. In practice, the interpretation and the characterization of the findings in these microscopical images are usually obtained by expert cytologists. The correct classification of a slide and the accurate diagnostic conclusion are crucial for the effective treatment of each incident. However, the limitations exhibited by these images in combination with human error may lead to misclassifications. Thus, many efforts have been made by several researchers for the development of automated methods for the analysis of Pap smear images, in order to assist the diagnostic procedure.

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).

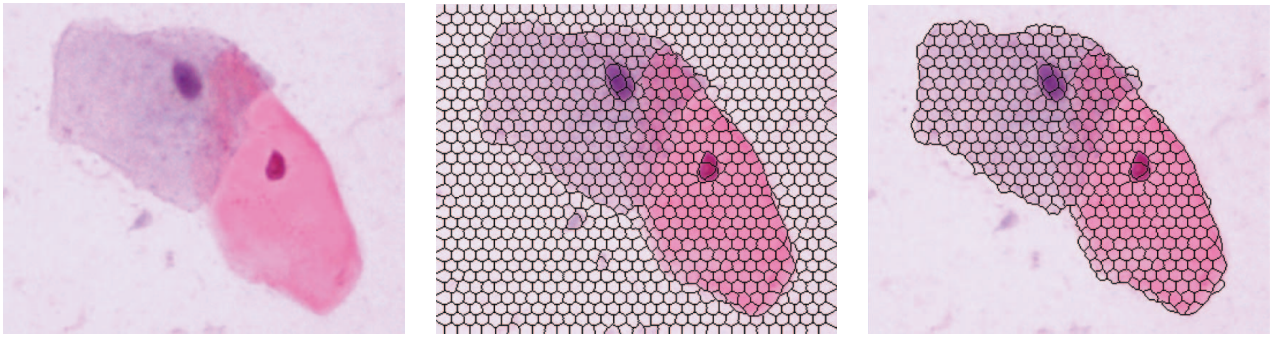
The structural elements of each cell provide significant information about the pathological condition of the smear. The nuclei features have shown high discriminative ability in recognizing pathological cases [2]. For this reason, many methods are focused on the accurate segmentation of the nuclei. Some of the general approaches that were used for this purpose are morphological analysis [3], watershed transform [4] and contour detectors [5].

Nevertheless, the cytoplasm area may adduce informative facts about the cell, such as the cytoplasmic transparency and the thickness of the cytoplasmic membrane. Recent works are focussed on the segmentation of both the nuclei and the cytoplasm. The first attempts were performed on free-lying cells, where no cell overlapping is occurred [5], [6], [7]. However, as the cell overlapping is a widespread phenomenon in Pap smear images, more sophisticated methods that are able to achieve reliable segmentation of the cytoplasm of each cell in cell clusters were proposed. More specifically, in [8] a geometric active contour based method for the localization of cells and the detection of the nucleus and cytoplasm boundaries is developed. In the method proposed in [9], the cell segmentation is performed using a joint level set optimization on all detected nuclei and cytoplasm pairs. In addition, the locally constrained watershed transform is performed for the separation of overlapped cells in [10].

In this work, we present a novel method for the definition of the area of each individual cell in clusters containing two overlapping cells. The method exploits the similarity in intensity of small area fractions belonging to the same structuring element of the cell, in order to concatenate them in a single aggregated item. The area fractions are obtained using a superpixel algorithm [11]. The final boundaries of each cytoplasm are obtained after the estimation of the boundaries segments in overlapping areas. This is achieved with the detection of the cytoplasm subareas presenting intensity disparity compared to the cytoplasm area without overlap. The experimental results indicate that the proposed method is robust and it provides accurate cell delineation.

## II. Method

In order to take advantage of the intensity characteristics of each region of interest, it is convenient to separate the image into perceptually meaningful individual regions. This is achieved with the application of SLIC segmentation algorithm [11]. Thus, each image is tesselated into approximately equally sized subregions, presenting homogeneous intensity characteristics (Fig. 1). Next, we classify the superpixels of the cell



(a) (b) (c)

Fig. 1. (a) Initial image. (b) The tessellation of the image into superpixel using the SLIC algorithm [11]. (c) Selection of the superpixels of the cell cluster after the extraction of the image background.

cluster in three discrete categories: the nuclei superpixels, the cytoplasm superpixels and the overlapping cytoplasm superpixels. This is described in the next paragraphs.

### A. Automated determination of the nuclei boundaries

The detection of the nuclei is a relatively easy procedure, because the nuclei are usually ellipse-like areas presenting lower intensity than the surrounding cytoplasm. In this work, we use the algorithm presented in [3]. Firstly, the background of the image is extracted and the area of the cell cluster is detected (Fig. 1(c)). Next, the centroids of the candidate nuclei are calculated and then, in a second step, the k-means algorithm is used for the separation of the detected points in two classes. The class of the nuclei is determined to be the one with the lower mean intensity. In this way, the number and the position of the nuclei are calculated.

The final boundary of each nucleus is obtained with the determination of the superpixels lying in the area of the nucleus. The superpixels corresponding to the nuclei areas are defined in a circumference of each nuclei centroid with a simple thresholding procedure, comparing the mean intensity value of the superpixel and the intensity of the nuclei centroid. If the difference in the intensity is lower than a threshold $h$ then the superpixel is incorporated in the nucleus superpixels, otherwise it is considered to belong to the cytoplasm. This procedure is repeated radially until all the nuclei superpixels are defined. The boundary of the area after the merge of the superpixels is a reliable boundary of the nuclei area, as it is depicted in Fig. 2(a).

It must be noted that the detection of the position of each nuclei ($N_1$ and $N_2$) is very important, as the surrounding area of each nuclei will indicate the intensity level of the part of the cytoplasm that is not overlapped. This part is used as a reference, in order to detect smooth transitions in the intensity that are presented due to cell overlapping.

### B. Extraction of cytoplasm area of each cell

In this step we search for the non-overlapping cytoplasm area of each cell. The overlapping area will be addressed in the next step. Thus, the estimation of the boundary of each cell is roughly calculated by classifying the superpixels into categories, which equals to the number of cells in the image. In this stage, we do not focus on the detailed definition of the boundary, but we separate the region of interest into individual parts, each one belonging to a single cell.

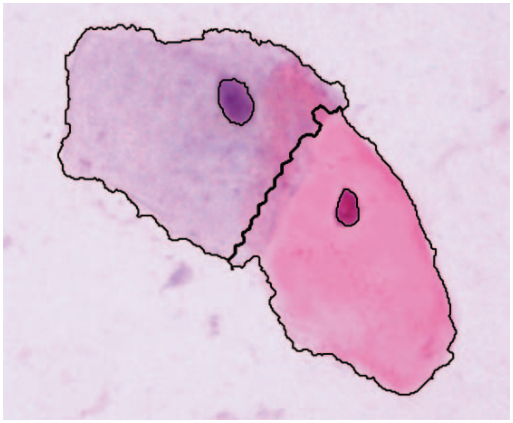
For this reason, from each superpixel we calculate the centroid and a set of features concerning the intensity in all of the color channels. More specifically, we construct a feature vector containing the coordinates of the centroid of the superpixel, the mean intensity in the red, green and blue channel, and finally the mean intensity in all channels. These vectors are used as input in k-means clustering algorithm and the superpixels are classified into two classes $C_i, i = 1, 2$, each one corresponding to a cell.

As it is depicted in Fig. 2(a), the result of this procedure is to define the cytoplasm area and to obtain a borderline between the two cells. However, this line is not the actual boundary, as both of the cells expand beyond this border line. For a more detailed definition of the borders of the cells we must search in the overlapping area, in order to detect differences in the intensity that indicate the existence of a cytoplasm border. This is feasible in the next step of the method.

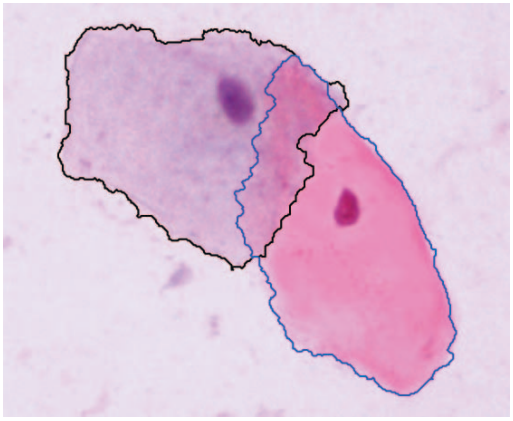
### C. Detection of the overlapping area and final cell boundaries

The region of interest in the image is composed by two cells that are partially overlapped. The cytoplasm area of each cell exhibits specific characteristics, while the area of overlap differs from the cytoplasm areas of each single cell. This is a consequence of the differentiation in the background. In the case of a free cell in the image, the background is homogeneous and transparent. However, when two or more cells overlap, the cytoplasm lying in different layers of the microscopic slide produce an area in the image where the intensity level present lower values. This is the result of the partial obstruction of the cell of the upper layer to the cell lying underneath.

For the detection of intensity differences that occur due to overlapping, we start searching from the centroid of each nucleus towards the direction of the other nucleus. The line connecting the centroids of the two nuclei (Fig. 3(a)) is expected to present a given intensity profile, as it is depicted in Fig. 3(b). As we observe, the intensity level follows a bimodal profile. More specifically, the centroid of the first nucleus ($N_1$) presents a minimum and as the distance from the nucleus increases, the intensity reaches higher values. This



(a)



(b)

Fig. 2. (a) The boundaries of the cell cluster and the nuclei after the nuclei detection step. The borderline between the cells is obtained after the extraction of cytoplasm area of each cell. (b) The cytoplasm boundaries after the detection of the area of overlap.

is typical, since the cytoplasm is brighter than the nucleus. However, at the area of overlap (between the points A and B), we observe a valley in the intensity of the image. Finally, as we move towards the cytoplasm of the second cell, the intensity is maintained in high levels, and it rapidly decreases in the area of the nucleus ($N_2$).

**Algorithm 1** Overlapping Area Detection Algorithm

**input**

$N_1, N_2 \rightarrow$ The centroids of the detected nuclei
$C_1, C_2 \rightarrow$ The set of the superpixels of each class

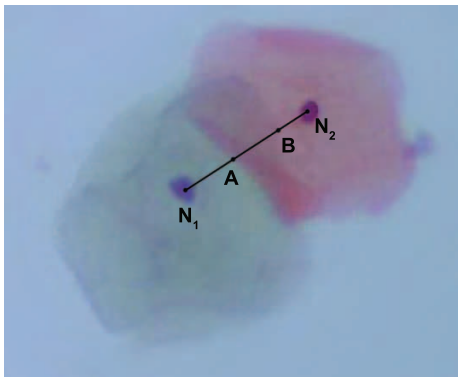
**output**

$S \rightarrow$ The set of superpixels at the area of overlap

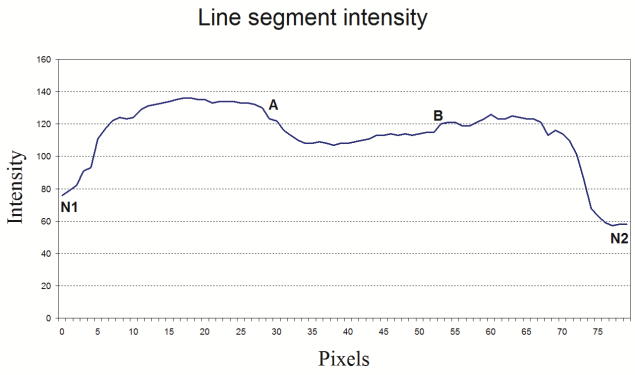
**for** $n = 1 : 2$ **do**
  $S_n \leftarrow \emptyset$
  **for** $i = 1 : \text{size}(C_n)$ **do**
    Define the line segment $\epsilon_i$ connecting $N_n$ and $C_n^i$
    Find the maximum intensity $I_{\max} \in \epsilon_i$
    **if** $|I_{max} - I(C_n^i)| > h_c$ **then**
      $S_n \leftarrow S_n \cup C_n^i$
$S \leftarrow \bigcup_n S_n$

The procedure we follow for the detection of the overlapping area is summarized in Algorithm 1. Firstly, the sets of the superpixels between each nucleus and the cytoplasm borderline obtained in the previous step is defined. This is necessary to define the region of interest and to reduce the searching space, as the overlapping area occurs between the two nuclei. Then, we construct the line segments $\epsilon_i$ that connect the nucleus centroid and the centroid of every other superpixel of the same class. If a line segment is lying in the area of the cytoplasm, the intensity of the pixels gradually increases. However, if the other end of the line segment is a pixel in the overlapping area, the intensity reaches a maximum and then decreases at the end point. If the difference of the maximum intensity and the intensity of the extreme superpixel centroid outreaches a limit, then the extreme superpixel is considered to belong to the area of overlap. The superpixels that satisfy this rule constitute the final set of the superpixels in the overlapping area.

This procedure is repeated for the second cell and the overlapping area is defined as the union of the final sets of the superpixels. The boundaries of each cell are then refined as the union of the initial boundaries and the boundaries of the overlapping areas. This is clearly a more reliable approximation of the true cell boundaries, as it is depicted in Fig. 2(b). However, since the boundaries are based on the borders of the the superpixels, a post processing is required to obtain more smooth boundaries (Fig. 4). This is feasible after an opening procedure.




(a)




(b)

Fig. 3. (a) The line segment connecting the two nuclei centroids ($N_1$, $N_2$). The pixels of the area of overlap are delimited by the points A and B. (b) The intensity profile of the line segment.

## III. Results and Discussion

Our method was evaluated on a data set of conventional Pap smear images containing two overlapped cells each. The images were acquired through a CCD camera adapted to an optical microscope (Olympus BX51). We obtained 25 images of $40\times$ magnification, which were stored in JPEG format. As it was verified by the results, the method is able to achieve a reliable estimation of the boundaries of each cell, overcoming the limitations introduced by the cell overlapping.

The parameters of the steps of the method were selected after several tests. The number of the superpixels in each image is proportional to its size and each superpixel equals to a hundredth of the total area of the image. The thresholds were set to $h = 30$ and $h_c = 15$ respectively.

To evaluate the performance of our method, we have compared the extracted boundaries of the cells to the ground truth, which was given by an expert cytologist. Thus, we have calculated the Euclidean and the Hausdorff distances between the estimated cytoplasm boundaries and the ground truth, which were correspondingly $2.29 \pm 0.71$ and $7.69 \pm 1.89$ pixels (mean $\pm$ standard deviation). Both of these measures verify the observations that the results of our method are very close to the real cells boundaries.

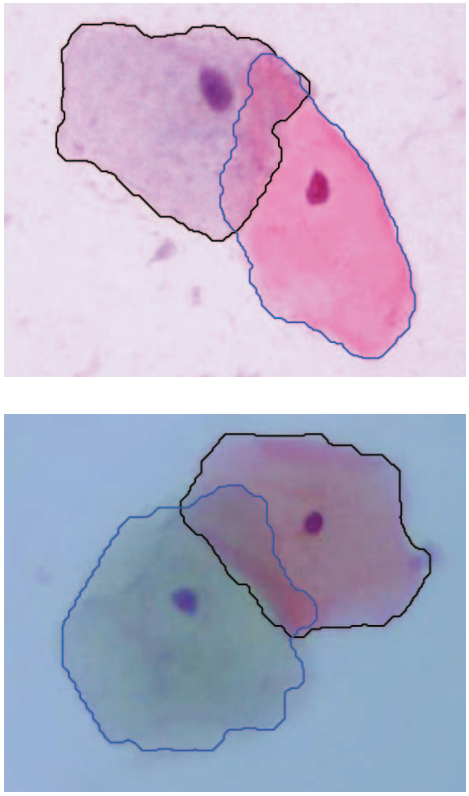


Fig. 4. Representative segmentation results of the method showing the accurate estimation of the area of overlap.

## IV. Conclusions

We have developed a fully automated segmentation method which combines the local intensity features of the cells obtained by the superpixels tessellation of the image and the a priori knowledge concerning the characteristics of the overlapping area of the cytoplasm. The method has been tested in terms of the accurate segmentation of the cytoplasm borders in images from Pap smear slides, and as it was verified by the results it presents a high performance. The main advantage of the proposed method is that it provides a flexible way for the simultaneous recognition of the nucleus and the cytoplasm, even in cases of cell overlapping. As a future work, we intend to examine the efficacy of the method in more complicated images containing a large number of overlapped cells, and to include a postprocessing scheme (e.g. a deformable model) for the refinement of the extracted boundaries.

## References


[1] B. W. Stewart and C. P. Wild, *World Cancer Report 2014*, World Health Organization, 2014.

[2] 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*, 2012, vol. 7325, pp. 483–490.

[3] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, 2011.

[4] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images," *Pattern Recognition Letters*, vol. 32, no. 6, pp. 838–853, 2011.

[5] M. H. Tsai, Y. K. Chan, Z. Z. Lin, S. F. Yang-Mao, and P. C. Huang, "Nucleus and cytoplasm contour detector of cervical smear image," *Pattern Recognition Letters*, vol. 29, pp. 1441–1453, 2008.

[6] N. Mustafa, N. A. Mat Isa, and M. Y. Mashor, "Automated multicells segmentation of ThinPrep image using modified seed based region growing algorithm," *Biomedical Soft Computing and Human Sciences*, vol. 14, no. 2, pp. 41–47, 2009.

[7] S. F. Yang-Mao, Y. K. Chan, and Y. P. Chu, "Edge enhancement nucleus and cytoplasm contour detector of cervical smear images," *IEEE Transactions on Systems, Man and Cybernetics -Part B: Cybernetics*, vol. 38, no. 2, pp. 353–366, 2008.

[8] N. M. Harandi, S. Sadri, N. A. Moghaddam, and R. Amirfattahi, "An automated method for segmentation of epithelial cervical cells in images of ThinPrep," *Journal of Medical Systems*, vol. 34, pp. 1043–1058, 2010.

[9] Z. Lu, G. Carneiro, and A. P. Bradley, "Automated nucleus and cytoplasm segmentation of overlapping cervical cells," in *Proceedings of Medical Image Computing and Computer-Assisted InterventionMICCAI 2013, Lecture Notes in Computer Science*, 2013, vol. 8149, pp. 452–460.

[10] N. Beliz-Osorio, J. Crespoand, M. Garcua-Rojo, A. Munoz, and J. Azpiazu, "Cytology imaging segmentation using the locally constrained watershed transform," *Mathematical Morphology and its Applications to Image and Signal Processing, Lecture Notes in Computer Science*, vol. 6671, pp. 429–438, 2011.

[11] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.