# Segmentation of overlapping cells in cervical smears based on spatial relationship and Overlapping Translucency Light Transmission Model

Jianwei Zhang *, Zhenpeng Hu, Guoqiang Han, Xiaozhen He

*School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China*

## ARTICLE INFO

## ABSTRACT

Overlapping cell segmentation in cervical smear images is a difficult task due to the shape multiformity and color proximity of the cells. In this paper, we propose a segmentation approach by using the spatial relationship of the non-overlapping and overlapping areas as well as Overlapping Translucency Light Transmission Model (OTLTM) to segment the overlapping cells in these images. The spatial relationship, which denotes the overlapping area locates in the middle ground of the non-overlapping areas, reflects the overlapping area can be accurately gained by the precise non-overlapping areas. After removing the background by threshold technique, a fragmentation method by using mean shift and watershed is adopted to divide the overlapping cells into fragments according to the similarity of their colors. The fragments belong to a single-tier individual cell, or to the overlapping area between two cells. We firstly construct the initial fragment collections of non-overlapping areas based on the Voronoi diagram, then the initial collections are optimized by using the initial cell overlapping matrix based on the spatial relationship, and OTLTM based on Beer–Lambert law, which states the relationship between the transmittance, attenuation coefficient of a kind of material and the distance the light travels through it. The cell overlapping matrix is accurately reconstructed by the optimized set of the non-overlapping areas. We obtained the segmentation result by combining the cell overlapping matrix and the optimized set of the non-overlapping areas. The experimental results show that the proposed method can give an impressive performance. Besides cervical smear images, these proposed techniques can be utilized in segmenting translucent objects from other kinds of images.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

According to the statistics of the World Health Organization, cervical cancer is the second most common of malignancies among women, more than 530 thousands new cases of cervical cancers appear a year, and the incidence rate is on the rising [1]. Fortunately, cervical cancer can be detected and preventable during its early stage, so screening and early treatment is important to lessen morbidity and mortality [1,2]. Currently, cervical smear screening is the most popular method to detect the presence of abnormal cervical cells. At first, Pap smear was frequently used in cervical cytology test, then with the development of smear preparation techniques, thin-prep liquid-based cytology test (TCT) and liquid-based cytology test (LCT) have been widely used.

The clinical cytopathologists observe the smear through microscopes, try to find diseased cells according to the difference of characteristics between normal and abnormal cells, and finally issue the test report. In general, morphological and architectural properties, such as the color, size, shape and texture of the nuclear and cytoplasm, the nuclear-cytoplasmic ratio, and the arrangement pattern of the cell clustering are used to make the diagnosis [3].

However, due to the large number of smear samples in the tedious smear screening, cytotechnologists are easily prone to fatigue, and it might lead to wrong test reports. Also, the accuracy of the diagnosis is highly depended on the experience of the cytotechnologists. Therefore, the development of computer assisted screening and diagnosing system has application prospects [4].

Computer assisted screening and diagnosing system involves a critical procedure: image segmentation, which is mainly designed to segment the nuclei and cytoplasm from the smear images. Numerous segmentation algorithms for cervical smear images have been proposed. Earlier methods usually detect and segment cells by using relatively simple methods, such as thresholding techniques [4]. However, these earlier methods cannot deal with the complex situations in cervical smear images. Recent years, extracting contour of nucleus or cytoplasm of each cell from

cervical smear images has attracted more attention. Harandi et al. [5] used thresholding to identify the nucleus within each cell, and applied the active contour algorithm to obtain the cervical cell boundaries. Specially, a separate active contour for each nucleus is used to delimit the corresponding cytoplasm within connected cell groups. Li et al. [6] carried out a K-means clustering to roughly segment an image into background areas, cytoplasm, nucleus, and then used a snake contour to refine the nucleus and cytoplasm boundaries. Pai et al. [7] utilized a NCC detector to automatically detect the nucleus and cytoplasm boundaries, and applied adaptive thresholding to the segmentation of the cervical cell, then used the maximum gradient difference for the segmentation of the nucleus from the cell areas. Chen et al. [8] implemented a semi-auto cervical cell classification system, by using a NCC detector to automatically detect the nucleus and cytoplasm boundaries, and then selecting the single cell by manual ways. Yang-Mao et al. [2] removed impulse and Gaussian noises by using a trim-meaning filter, made a clear-cut separation of the pixels lying in-between two objects with a bi-group enhancer, and used a contour detector to identify the edge pixels corresponding to nucleus and cytoplasm boundaries. Xu et al. [9] transformed an image into the polar coordinate system, applied the Sobel operator to calculate the level contour, then used a sand inhibition method to resist the interference factors of the contour position, finally applied the GVF method for the segmentation of the cervical cells. Fan et al. [10] used an automatic double threshold approach to segment the single cell from the image, and then regarded the cervical image as a vector-value image, applied a modified CV model for the segmentation of the nucleus and cytoplasm. Tsai et al. [11] used the K-means algorithm to cluster pixels into two partitions and applied the maximal color difference method to extract the nucleus contour. Plissiti et al. [12,13] used the local minima of image gradient to detected the locations of nuclei centroids in Pap smear images, eliminated the candidate centroids that were too close to each other, and then applied a support vector machine (SVM) classifier for the final selection of points by using color values in square neighborhoods. Finally, they regarded the detected centroids as markers in marker-based watershed segmentation to find the nuclei boundaries and used a binary SVM classifier with texture, shape, and intensity features to eliminate the false-positive areas.

In these above-mentioned methods, some focus on the accurate segmentation of cervical nuclei, and the others are almost interested in segmenting the individual cervical cells from the background. However, in practice the overlapping cells often appear in cervical smears, and the absence of the effective segmentation methods for them will limit the subsequent quantitative analysis. If these overlapping cells are accurately segmented, there will be a significant improvement for the computer assisted screening and diagnosing system. To overcome this challenge problem, a few methods and algorithms have been presented. Sulaiman et al. [14] combined the edge information with the pseudo-color feature to split the overlapping cells, but this method can only split the overlapping cervical cells that have different colors. Guan et al. [15] used a morphological filtering-based K-means clustering algorithm to extract the cell nucleus and the background, applied a gradient decomposition-based edge enhancement method for enhancing the true edges belonging to the center cell, and used a dynamic sparse contour searching algorithm to locate the weak contour points. Finally, the GVF snake model was employed to extract the accurate cell contour. In this proposed framework, it is not easy to ensure the stable performance of two procedures: edge enhancing and weak contour locating. Lu et al. [16] presented an approach to segment the cell clumps and their nuclei, by using a joint level set optimization. In this method, some features were used, such as, the length and area

of each cell, a prior on cell shape, the amount of cell overlap and the expected gray values within the overlapping regions. This algorithm can segment all cells in the meantime, but it sometimes obtained the inaccuracy contours, specially the overlapping contours.

Some methods have been presented to address the segmentation of the overlapping cells in other type images, but it is difficult to use these algorithms to segment the overlapping cells in cervical smear images because of the significant diversity between different kinds of images.

In this paper, we proposed a novel method to address the segmentation problem of partially overlapping cell mass. We found the vast majority of overlapping cervical cells are two-layer overlapping after observing a large of images. That is to say the most of overlapping areas are formed by two overlapping cells. And we find it seems to be directly observed that an overlapping area is located in the middle of the two non-overlapping areas of the two cells. Therefore, we can obtain the full zone of each cell after obtaining its accurate overlapping area and non-overlapping area. First of all, a mixture Gaussians model is adopted to remove the background. Then, we use a mean shift algorithm to complete the color clustering based on the adjacent color similarity, and the watershed method is applied to split the foreground into fragments. Theoretically, the overlapping area can be gained by utilizing the spatial relationship, which shows that the overlapping area is placed in the middle of the two non-overlapping areas. We construct the initial fragment set of the non-overlapping area for each cell based on the Voronoi diagram, then, construct the initial cell overlapping matrix and proposed a light transmission model, which can be used to estimate the color of the overlapping area in order to evaluate the possibility that fragments belong to the overlapping area, to optimize the initial set of the non-overlapping areas. For acquiring the final segmentation result, we reconstruct the cell overlapping matrix based on the optimized set of non-overlapping areas. The experimental results can prove the effectiveness of the proposed method.

The rest of the paper is organized as follows: the overview of our method is presented in Section 2. Section 3 discusses the preprocessing, which includes foreground extraction and fragmentation of overlapping cells by color clustering and watershed. Segmentation of the overlapping cells is proposed in Section 4. The results of the experiment are discussed in Section 5. Finally, conclusions are presented in Section 6.

## 2. Materials and framework

### 2.1. Data set

The datasets used in our experiments were from The First Affiliated Hospital of Guangzhou University of Chinese Medicine. The images were obtained by an automatic microscopic image acquisition system, and the cervical smears were made by Automatic Liquid-based Cytology Test System, manufactured by LBP Medicine Science & Technology Co., Ltd., China. We randomly chose 112 overlapping cell mass, which are clipped them down from more than 100 cervical smear images. The overlapping cell mass are all composed by two or more cells.

### 2.2. Framework

Our approach can be summarized by a processing framework that includes six parts, showed by Fig. 1. The first step is removing the background from the cervical image to obtain the cell mass. We fit the luminance histogram using the Gaussian mixture model, and set the trough which has the highest brightness