Various Techniques for Classification and Segmentation of Cervical Cell Images - A Review

Bharti Sharma
Punjabi University Regional Center for
Information Technology & Mgmt.
Mohali, India

Kamaljeet Kaur Mangat
Punjabi University Regional Center for
Information Technology & Mgmt.
Mohali, India

ABSTRACT

Pap smear test plays an important role for the early diagnosis of cervical cancer in which human cells taken from the cervix of patient are analysed for pre-cancerous changes. The manual analysis of these cells by expert cytologist is labor intensive and time consuming job. The automatic and accurate detection of cervical cells are two critical preprocessing steps for automatic Pap smear image analysis and also for diagnosis of pre-cancerous changes in the uterine cervix. Similarly, the reliable segmentation of abnormal nuclei in cervical cytology is of utmost importance in automation-assisted screening techniques. This paper presents, the existing automated methods for the detection, segmentation and boundary determination of cells nuclei in conventional Pap stained cervical smear images. The majority of cytoplasm segmentation uses K-means algorithm, edge detection method, thresholding approach, graph cut and active contours technique. Most of existing work is done on images of isolated cells, especially for those which are in the Herlev data set. For segmentation of images which contains multiple cells, level set and thresholding techniques have been used. The nucleus segmentation varies as: single-nucleus segmentation, touching-nuclei splitting and multiple-nuclei segmentation. However, many segmentation methods incorporates shape priors, usually enforcing elliptical shapes in order to overcome cell occlusion and noise. The main focus of this paper is comprehensive literature survey of various existing classification and segmentation techniques. The shortcomings and failures of the existing work are also provided for further enhancement and improvement of overall performance and accuracy.

Keywords
Cervical cell classification, Cervical cancer, Unsupervised segmentation, Radiating GVF Snake, Global and local scheme, Multiscale convolutional network

1. INTRODUCTION
Cancer of the uterine cervix is the second most common cancer found among women worldwide and is most frequent in developing countries. Persistent infection of cervical epithelial cells along with certain high-risk types of human papillomavirus (HPV) is considerd as most important cause of cervical cancer. Recent studies have suggested that HPV-DNA positive women has great risk of cervical cancer in case of using oral contraceptives for period longer than 5 years. While in other types of cancer diseases, many distinguishable symptoms starts to appear at initial stage but in case of cervical cancer there are no specific symptoms which indicate its existence. However, early detection and treatment of precancerous lesions symptoms in patient can prevent its progression to cervical cancer.

The most effective way of detecting the cervical cancer is Pap smear screening test which was introduced by Dr. Georges Papanicolaou in 1940s. This Pap test is a manual screening procedure which is used to detect pre-cancerous changes in cervical cells on the basis of color and shape properties of cell nuclei and cytoplasm regions. Pap test has been used to screen out abnormal cervical cells. Abnormal cervical cells that have undergone the pre-cancerous changes are called as dysplastic cells, which have three main phases. In the first phase known as mildly dysplastic, the nucleus becomes larger and brighter than normal one. In the second phase known as moderate dysplastic, the nucleus is much larger and darker. In the final phase known as the severe dysplastic, the nucleus and cytoplasm both change their size and texture. The nucleus becomes larger and darker with grosteque shape but the cytoplasm is usually darker and smaller[9]. The last type of abnormal cells is Carcinoma-in-situ which posses more serious problem than severe dysplastic.

The main factors which affect the sensitivity of the Pap smear test are number of cells and type of cells sampled on the glass slide and the presence of inflammatory cells, mucus and blood in the image. Additional factors which complex the problems are nature of the tissue, overlapping of cells with variability, poor contrast and inconsistent staining of cell images. These challenges in the field of cervical cancer has lead to automation and computerization of the diagnosis process. Therefore image processing and automated assisted reading techniques[10] can be used for automatic diagnosis of cervical cells. Many such techniques had been proposed in existing work for the automated segmentation of nucleus in Pap smear images. The completely Automated-assisted techniques for detection and diagnosis of the cervical cancer is still a challenging problem due to the complexities of cell structures.

This paper is divided into four sections. Section 2 is concerned with study of the various existing algorithms for segmentation and classification of cervical smear images. This is presented in the form of literature survey. The research gaps and flaws in existing techniques of classification and segmentation are provided in Section 3 . The conclusion of the paper is provided in section 4. This section also highlights the important aspects from the existing study and future projections.

2. LITERATURE SURVEY
Genctav *et al.*[1] proposed an unsupervised approach for the segmentation of pap smear cells from Herlev and Hacettepe Pap smear dataset. This approach can also deal with images comprises a single cell as well as overlapping cells. The two stage segmentation is proposed where first stage undertakes multi-scale hierarchical egmentation algorithm to partition the cell image into regions depending on homogeneity and circularity.The morphological operations along with automatic

thresholding algorithm separates the cell regions from the background. Second phase uses a binary classifier to differentiate these regions into nucleus and cytoplasm. They used five various types of classifiers to analyse the performance. These classifiers are: Decision tree classifier, Bayesian classifier, Support Vector Machine and combination of above three classifiers using product and sum of individual posterior probabilities.

Li *et al.*[2] proposed both nucleus and Cytoplasm segmentation in cervical smear images using Radiating GVF Snake. After preprocessing, the regions in the image are roughly clustered into nucleus, background and the cytoplasm by a spatial K-means clustering algorithm. The initial contours are extracted, then, the cell image is segmented out using RGVF. RGVF involves a stack-based refinement and a new edge map computation method, which is robust to contaminations and is able to locate the obscure boundaries. Also,the boundaries can be correctly traced even if there exists interferences near the cytoplasm and nucleus regions. Experiments are performed on the public available Herlev dataset.

Cengizler *et al.*[3] proposed fluid dynamics based deformable model for the segmentation of cervical cell image.In this approach, simulation starts inside the cytoplasmic region and the simulate fluid is attracted toward the cell contours. The gradient magnitude of the image is considered as the source for extracting particles. Direction of propagation of the flow is determined by an interaction mechanism which is based on the permeability rate of these particles. Interaction between fluid and particles guides the advancing fronts of the fluid towards the object boundaries. Redefinition of complex topologies with particle groups provides potential of improved segmentation capability and flexibility to the model. Results showed that proposed method may be more adaptive as compared to watershed algorithm and have an improved performance on regaining shape and boundary data of cervical cells.

Zhang *et al.*[4] proposed global and local scheme which is based on graph cut approach for the segmentation of cervical cell images. The global scheme is implemented for the separation of cytoplasm regions from the background by using multi-way Graph Cut approach on a* channel enhanced image. The local schema is enforced for the segmentation of nucleus areas by applying Local Adaptive Graph Cut (LAGC) approach which results in robust binarisation of nuclei with a variety of chromatin distributions, morphologies and low contrasts. Then, splitting method is proposed to break down touching nuclei with combination of two concave points-based approaches. The 21 Pap smear image dataset of cervical cells are acquired for experiments which is not available publically.

Chankong *et al.*[5] used patch-based fuzzy C-means clustering method for the segmentation of cytoplasm and nuclei regions from Herlev dataset only. The four step segmentation approach is proposed in which first input single cell image is changed into grayscale and then median filter of size 7×7 is used to eliminate the noise and smoothen the image. The third step involves the segmentation of processed image in 3 regions, i.e. cytoplasm, nucleus and background using the FCM clustering method. Step four identifies the cytoplasm and nucleus regions by sorting the patches by determining its center value. The patch with the value of centers less i.e. darker than the nucleus threshold is labeled as nucleus. In the same way, other patches are classified as cytoplasm, if the value of any of the patches center lies between the nucleus threshold and cytoplasm threshold.

Oscanoa *et al.*[6] The proposed work develops an algorithm based on adaptive thresholds, which can be used in Pap smear assisted quality control software. The first phase of the method is a pre-processing step, which includes noise and background removal .Then a block is segmented for each one of the points selected as non background. Local threshold per block is evaluated to search for cell nuclei. If a nucleus is detected, an artifact rejection is followed, where only inflammatory cells and cell nuclei are left for the doctors to interpret. The proposed method was validated with a set of 55 images containing 2317 cells.

Guan *et al.*[7] proposed dynamic sparse contour searching and GVF snake model for the accurate segmentation of partially overlapping cervical cells. First, background and the cell nucleus are extracted by using morphological filtering-based K-means clustering algorithm. Then, a gradient decomposition-based edge enhancement method is applied for enhancing the true edges belonging to the center cell. Dynamic sparse contour searching algorithm is used for least squares estimation and a dynamic searching principle is proposed to gradually locate the weak contour points in cell overlapping regions on the basis of strong contour points. Using the located contour points, the Gradient Vector Flow Snake model is finally employed to extract the accurate cell contour. Experiments have been done on two cervical smear image datasets containing both partially overlapping cells and single cells.

Song *et al.*[8]proposed a multiscale convolutional network (MSCN) and graph-partitioning-based method for accurate segmentation of cervical nuclei and cytoplasm. Firstly, the initial segmentation is done by extracting the scale invariant features and then, segment out regions centered at each pixel. Then, the coarse segmentation is refined by an automated graph partitioning method which is based on the pretrained feature. For further refinement, a coarse-to-fine nucleus segmentation framework is developed. Computational complexity of segmentation is decreased by using superpixel rather than raw pixels.

This section has described the literature survey over the various segmentation and classification methods of cervical smear images .The next section describes the research gaps of the existing methods .

3. RESEARCH GAPS
The existing approach works either on single cervical cell images or multiple cervical smear images. The algorithm can be proposed to cope with both single cell images and multiple cell images at the same time . FCM clustering method has been used in existing work with fixed number of clusters .The varying number of clusters may lead to further improvement in overall accuracy and performance of the method. Integrating the existing segmentation method with the cell region of interest locating method, can help to extract the cell nuclei and the whole cell regions in low-resolution cervical smear images also. This will help in more efficient and accurate diagnosis of precancerous changes in the patient. The existing model offers the accuracy of the nearly 93.78% in Herlev dataset images which can be improved up to the higher level. The accuracy of the system can be improved by various parameters like variation in features to be extracted ,improvement in noise removal methods and segmentation techniques etc.

This section has described the gaps and flaws in various existing methods. The next section describes the conclusion.

Table 1. Literature Review

AUTHORS	PAPER TITLE	PROPOSED WORK	MERITS	DEMERITS
A. Gençtav, S. Aksoy, and S. Önder [1]	Unsupervised segmentation and classification of cervical cell images	Multi-scale hierarchical segmentation algorithm is proposed to partition the image into regions. Binary classifier is used to discriminate regions into the nucleus and cytoplasm.	Both the segmentation and the classification algorithms are parameter-free and generic. Accurate segmentation and classification of cervical cell images having inconsistent staining, poor contrast, and overlapping cells.	Further improvements in overall accuracy can be achieved by using accurate cytoplasm features. Columnar cells are not included in input data.
K. Li, Z. Lu, W. Liu, and J. Yin [2]	Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake	Firstly ,preprocessing is done . Spatial K -means clustering algorithm is used to roughly cluster image into nucleus, cytoplasm and the background areas. After initial contours are extracted, image is segmented using RGVF which involves a new edge map computation method and a stack-based refinement.	Have potential ability to locate the obscure boundaries. Can diminish the contaminations caused by inflammatory cells and blood stains.	Can segment single cell images only. Not able to extract boundaries of cytoplasts from overlapping cells images.
C. Cengizler, M. Guven, and M. Avci[3]	A fluid dynamics-based deformable model for segmentation of cervical cell images	Simulation starts inside the cytoplasm region, simulated fluid is attracted toward the cell contours. Direction of propagation of the flow is determined by an interaction mechanism based on the permeability rate of particles.	Performs well on most of the overlapped regions. Can cope up with complicated topologies as an effect of particle interactions. More suitable method for automated segmentation tasks.	Fluid leakage through cell borders. Optimization of initial parameters is not done . Uncontrolled expansion rate of fluid.
L.Zhang, H.Kong,C.T. Chin, S. Liu, T. Wang, and S. Chen *et al.* [4]	Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts	Global and local scheme based on graph cut approach for the segmentation of cervical cells. The splitting method is used to separate touching nuclei.	Achieved cytoplasm segmentation accuracy of DSC = 93% and an abnormal nuclei binarization accuracy of F-measure = 88.4%.	Does not attempt to delineate the cytoplasm boundary for each cell. Fail to delineate the cytoplasm boundary caused by intensively bright illumination.
T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul[5]	Automatic cervical cell segmentation and classification in Pap smears	Patch –based fuzzy C-means clustering is used for segmentation of nuclie and cytoplasm . FCM clustering is used for segmentation of processed images into nucleus, cytoplasm, and background.	For the Herlev dataset, yielded the accuracies of 93.78% and 99.27% for the 7-class and 2-class problems. Provide better sensitivity than hard C-means and watershed segmentation methods. Provides a better set	Not applicable for cervical cell image with multiple cells. Number of clusters are fixed in clustering process.

			of features for the classifiers.	
J. Oscanoa, M. Mena, and G. Kemper[6]	A Detection Method of Ectocervical Cell Nuclei for Pap test Images, Based on Adaptive Thresholds and Local Derivatives.	Algorithm based on adaptive thresholds is proposed for the nuclie detection . Local threshold per block is calculated to search for cell nuclei.	Successfully detected 92.3% of the total cell nuclei in all collected images.	This test shows some efficacy problems. Proper technological assistance is required for improvement.
T. Guan, D. Zhou, and Y. Liu [7]	Accurate Segmentation of Partially Overlapping Cervical Cells Based on Dynamic Sparse Contour Searching and GVF Snake Model	Morphological filtering-based K-means clustering algorithm is used to extract nucleus and the background. A dynamic sparse contour searching algorithm is proposed to locate the weak contour points in the cell . Using the located contour points, the Gradient Vector Flow Snake model is finally employed to extract the accurate cell contour.	Whole process is fully automatic. Can cope with the cell containing folds. Robust for segmentation of partially overlapping cells in cervical smear images.	Not applicable for two or more overlapping cells segmentation. Segmentation of low resolution cervical smear images cannot be done .
Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang [8]	Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multi-scale Convolutional Network and Graph Partitioning	Initial segmentation is performed by multiscale convolutional network (MSCN) .Then graph-partitioning-based method is used for further refinement in segmentation of cervical cytoplasm and nuclei.	Using MSCN , segmentation performance is improved by 27.42%, and the nucleus segmentation accuracy is improved by 35.09% compared with SSCN. This not only accurately segments the cytoplasm and nucleus, but also accurately separates the touching nuclei.	The ignorance of the overlapping cytoplasm splitting. Inaccurate segmentation of the atrophic cell. Overlapping nucleus with complicated shapes degrades the splitting performance.

4. CONCLUSION

Some of the existing work did not consider the presence of unwanted components like mucus, blood and inflammatory cells as well as variable overlapping of cells and inconsistent staining which may be present in the image. These components directly affects the overall performance and accuracy of the algorithm. Also,the selection of the number of clusters is the critical aspect for clustering algorithms which directly affects the segmentation results.

In the existing approach, the number of clusters are kept fixed which leads to degradation in the performance and accuracy. The more accurate features computed from the improved cytoplasm regions will also improve the performance. Therefore, above mentioned factors can be considered and selection of number of clusters is not kept fixed but it may vary from image to image for nucleus and entire cell segmentation. FCM clustering can be used for the segmentation of nucleus and cytoplasm from the Pap smear images because it can cluster data having uncertainty. By considering aforesaid factors in future work, the overall precision and accuracy of the segmentation and classification approach can be extended to higher level.

5. REFERENCES

[1] A. Gençtav, S. Aksoy, and S. Önder, "Unsupervised segmentation and classification of cervical cell images," Pattern Recognit., 2012.

[2] K. Li, Z. Lu, W. Liu, and J. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake," Pattern Recognit., 2012.

[3] C. Cengizler, M. Guven, and M. Avci, "A fluid dynamics-based deformable model for segmentation of cervical cell images," Signal, Image Video Process., 2014.

[4] L. Zhang, H. Kong, C. T. Chin, S. Liu, T. Wang, and S. Chen, "Automated segmentation of abnormal cervical cells using global and local graph cuts," in Biomedical Imaging (ISBI), IEEE 11th International Symposium on , 2014.

[5] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in Pap smears.," Computer methods and programs in biomedicine, vol.

113, no. 2, pp. 539–56, 2014.

[6] J. Oscanoa, M. Mena, and G. Kemper,"A detection method of ectocervical cell nuclei for pap test images based on adaptive thresholds and local derivatives," International Journal of Multimedia and Ubiquitous Engineering, vol. 10, no. 2, pp.37-50,2015.

[7] T. Guan, D. Zhou, and Y. Liu, "Accurate Segmentation of Partially Overlapping Cervical Cells based on Dynamic Sparse Contour Searching and GVF Snake Model,"IEEE J. Biomed. Heal. Informatics, vol. 19, no. 4, pp. 2168–2194, 2015.

[8] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multi-scale Convolutional Network and Graph Partitioning," IEEE Trans. Biomed. Eng., 2015.

[9] J. Jantzen and G. Dounias, "Analysis of Pap-smear image data," in Proceedings of 2nd Annual Symposium on Nature-Inspired Smart Information Systems , 2006.

[10] L. Zhang, H. Kong, C. Ting Chin, S. Liu, X. Fan, T. Wang, and S. Chen, "Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining," Cytom. Part A, 2014.