

A fast and reliable approach to cell nuclei segmentation in PAP stained cervical smears

N. B. Byju · Vilayil K. Sujathan · Patrik Malm ·
R. Rajesh Kumar

Received: 10 June 2013 / Accepted: 23 September 2013 / Published online: 8 October 2013
© CSI Publications 2013

Abstract Fast and reliable segmentation of cervical cell nuclei is one of the crucial steps of an automated screening system that aims early detection of cervical cancer. In this paper, we propose an edge based approach using customized Laplacian of Gaussian (LoG) filter to segment free lying cell nuclei in bright-field microscope images of Pap smear. The LoG is generally employed as a second order edge detector in image processing. The images may have the challenges of inconsistent staining, overlapping and folded cells. Experimenting proposed method over all types of cervical images including sufficient number of high grade lesions of cervical cancer shows that our method performs well for stain varied images containing focused nuclei.

Keywords Segmentation · Edge detection ·
Pap smear image · Cervical cancer

1 Introduction

Cancer of the cervix is the second most common cancer in women worldwide, with about 500,000 new cases and 250,000 deaths each year. It has been the most prevalent cancer in women in India over the past two decades. As given by Globocan 2008, one woman dies every 7 min of cervical cancer in India. Cervical cancer causes early death as well as prolonged disability and impoverishes individuals and communities due to the high medical costs. It has the highest total cost of secondary care relative to all other cancers.

Cervical cancer is preventable and curable if detected at an early stage. Since early detection leads to better prognosis, one of the most effective ways of preventing and controlling cervical cancer is regular screening and early diagnosis. The most effective method of screening employed in the developed world has been cytology based using Pap smears. Pap smear test [1] is a manual screening procedure used to detect cervical cancer or precancerous changes in uterine cervix by microscopic analysis of cell sample (Pap smear) collected from the cervix.

In India there is no national screening programme in place for cervical cancer prevention. According to the 2003 WHO World Health Survey, only 2.6 % of the Indian female population had ever been screened for cervical cancer in the past 3 years. This is mainly because of lack of systematic primary screening programmes coupled with shortage of qualified cyto-technologists in relation to the large population in India. A computer assisted screening system for Pap smear tests will be quite useful in this scenario. A high throughput low cost automated screening system can improve the efficiency of the screening process. The system should be fast, reliable and deployable to extend the coverage of the screening process. Liquid based

N. B. Byju (✉) · R. R. Kumar
Health Informatics Section, C-DAC, Vellayambalam,
Trivandrum 695033, Kerala, India
e-mail: byju.nb@gmail.com

R. R. Kumar
e-mail: rajeshkumar@cdac.in

V. K. Sujathan
RCC, Trivandrum, India
e-mail: sujathan@rcctvm.org

P. Malm
Centre for Image Analysis, Uppsala University,
Uppsala, Sweden
e-mail: patrik.malm@cb.uu.se

cytology (LBC) slides are used in our work since conventional smears produce uneven layering, heavy overlapping of cells and may contain red blood cells, mucus and bacteria. In LBC slides, red blood cells and cell overlapping will be comparatively less than that of conventional smears, suitable for automated screening process.

In manual screening method, a cytologist may take an average of 10 min for analyzing a smear. The cytologists may screen an average of 40–50 PAP smears a day. The annual report of RCC Thiruvananthapuram [2] shows between 5,000 and 10,000 women are only screened by well established screening labs in a year. Thus to cover 400 million women in India we would need between 30,000 and 40,000 labs employing around one lakh cytologists. Statistically 90 % of all screened slides [2] are actually negative. Therefore, if an automated system is able to screen out maximum number of negative cases, the cytologist can focus their time and expertise on the remaining critical positive cases.

For an automated system the digitization of the Pap smear may generate on an average of 1,850 images at 40 \times magnification for a typical 2/3 CCD sensor and a video adapter of 0.63 \times . With the processing time of 1 s/image in a sequential way on an average configured PC, analysis of one smear may take 30 min excluding the image acquisition time. However this can be compensated by batch processing, scheduling during night time and using parallel processes or Graphics Processing Unit (GPU) computing. The system is not meant to replace the cytologists, rather to be used as a prescreening system, in effect boosting up the screening workflow efficiency by over five times. The available skilled cytologists together with the prescreening system can make the population screening programs for cervical cancer in India viable. Thus the system can overcome the major constraint of lack of trained personnel for the introduction of population screening programs in India.

The first and the most crucial step of such a system is the segmentation of cervical cell nuclei. Here we propose a fast and reliable method of segmenting cell nuclei suitable for implementing a deployable automated cervical cancer screening system.

2 Problem definition and related works

Segmentation of cervical cell nucleus is an area of research since many years. Still this is a research problem due to the complexities arising from stain variation, overlapping of cells [3, 4] and the standardization of imaging aspects. There are several methods which segments cervical cells. Earlier attempts to detect and segment cells in cervical microscopic images were based on image-thresholding techniques [5]. Pixel classification was also tried for the segmentation of cervical images [6]. Morphological watershed algorithms

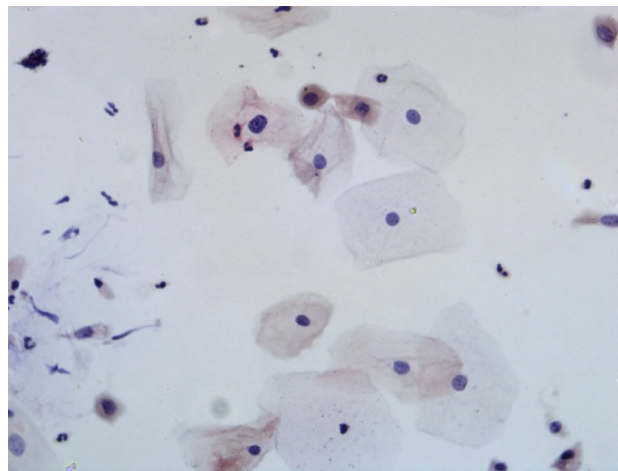


Fig. 1 Digitized cervical image

were used for delineating boundary of cell nuclei for small images [7, 8]. Active contours were used to trace the boundaries of cell nuclei [9, 10]. Region growing methods were used of getting the cell nucleus and cytoplasm [11, 12]. Marker controlled water shed algorithms are quite good for accurately delineating the boundary of cell nucleus but suffers from performance problems [8].

For a deployable solution, the segmentation method should have high sensitivity and high performance. Since the focus is on getting higher sensitivity in segmentation, artifacts like red blood cells and other biological debris will also be segmented. The artifacts present in massive orders which are similar in size and shape to abnormal cells would cause the misclassification of cytology images in the screening process. This increases the false positive rate which can hinder the mass screening. Therefore the elimination of these artifacts plays a key role in designing a proper classification strategy for the malignancy detection from cytology images. These artifacts can be removed later using our own original work based on pattern recognition strategies [13].

In this paper we focus on the problem of detecting and segmenting cervical cell nucleus attained from the digitized LBC slide images of Pap smears using fast and reliable edge based approach. We also focus on filtering red blood cells from the segmented objects. The evaluation for the proposed segmentation method has covered large set of cells including different cancerous and precancerous grades of cervical cells. A sample digitized LBC Pap smear image is given in Fig. 1.

3 Methodology

The flowchart of the proposed method for detection of cervical cell nucleus from Pap smear images is given in Fig. 2.

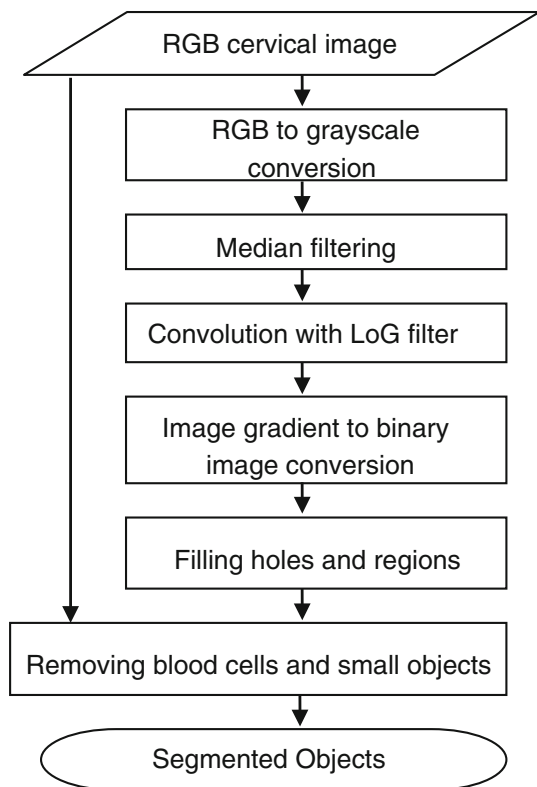


Fig. 2 Flowchart of proposed method

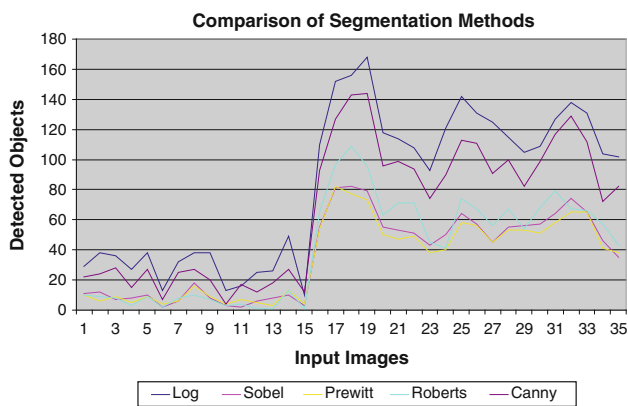


Fig. 3 Comparison of LoG with other methods

The true color RGB image is converted to gray scale image by eliminating the hue and saturation information while retaining the luminance. The grayscale image is used for segmenting the cervical cell nuclei.

Laplacian of Gaussian (LoG) filter is used for the detection of cell nuclei. The LoG [14] is widely used as a second order edge detector in image processing. The performance of LoG filter is better compared to other edge detection algorithms [15], which is crucial for practical deployable solution. Compared to other edge detection algorithms, LoG was giving good results for cell nuclei detection as shown in Fig. 3.

The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection. It is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce its sensitivity to noise. The operator normally takes a single gray level image as input and produces another gray level image as output.

The Laplacian $L(x,y)$ of an image with pixel intensity values $I(x,y)$ is given by:

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

In this approach the LoG operator is convoluted with an image and the resulting zero crossings are taken as edges in normal scenario. In fact, since the convolution operation is associative, we can convolve the Gaussian smoothing filter with the Laplacian filter first of all, and then convolve this hybrid filter with the image to achieve the required result. Doing things this way has two advantages:

- Since both the Gaussian and the Laplacian kernels are usually much smaller than the image, this method usually requires far fewer arithmetic operations, thus improves the performance.
- The LoG ('Laplacian of Gaussian') kernel can be pre-calculated in advance, so only one convolution needs to be performed at run-time on the image.

The 2-D LoG function centered on zero and with Gaussian standard deviation σ has the form:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

The cell nuclei are darker compared to the cytoplasm and background regions. Application of the LoG filter generates +ve gradient values inside the nucleus region and -ve values outside the boundary of cell nucleus. For the selection of prominent nucleus edges, the threshold used is 0.11 as the gradient value. This gradient threshold generates an average reduction of 1.5 % in area of segmented objects when experimented with 225 nuclei of various types. This method shrinks the boundary of segmented nucleus, providing better accuracy in classification of cells with more sensitivity in segmentation as compared with that of segmented nucleus where zero crossings are taken as the true boundaries in the case of LoG filter.

Heuristics are often employed to determine the required mask size. Larger sigma requires larger masks, and the performance of the edge detector increases with increasing mask size [16]. Ultimately the choice of mask size is a trade-off between computational considerations and precision. As the value of sigma decreases the detection focuses

on sharp edges. Sigma basically controls how “fat” the kernel function is going to be; higher sigma values blur over a wider radius. Bigger sigma also forces to use a larger kernel matrix to capture enough of the function’s energy. Large σ detects large scale objects and small σ detects fine features. For the specific case, we made the kernel to be big enough to cover most of the objects, but not so large such that it starts overlapping neighboring objects at a time. Based on the above factors we selected the value of sigma as 2 and the size of the kernel as 15×15 .

The hole filling of the binary image yield segmented objects. Very small objects can be removed based on size measurements. The input RGB image is shown in Fig. 4. The preprocessed gray image is shown in Fig. 5. The gradient image using LoG filter is shown in Fig. 6. The binary image created from the gradient images is shown in Fig. 7. The holes filled image is shown in Fig. 8. The binary image containing segmented objects after filtering

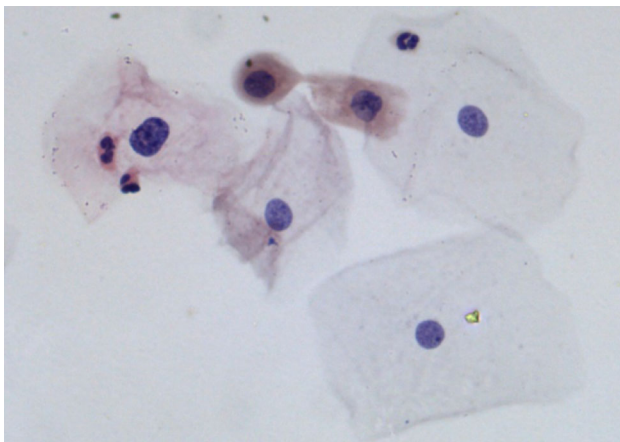


Fig. 4 Input RGB image

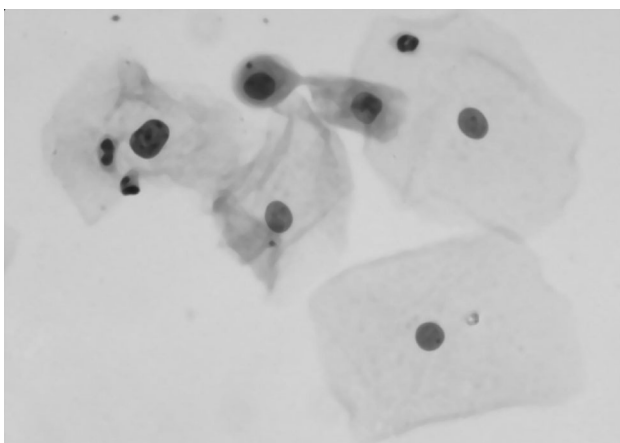


Fig. 5 Median filtered gray image

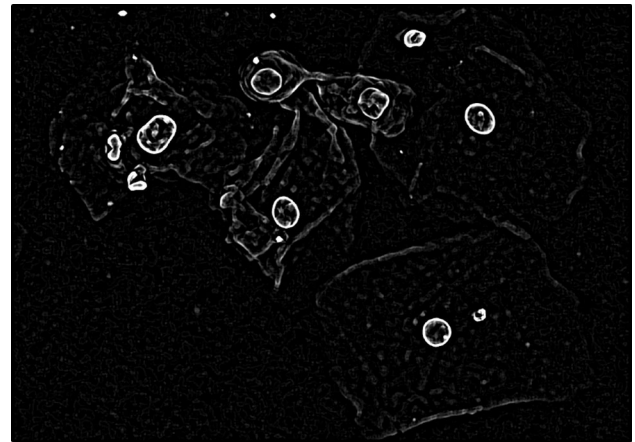


Fig. 6 Gradient image using LoG filter

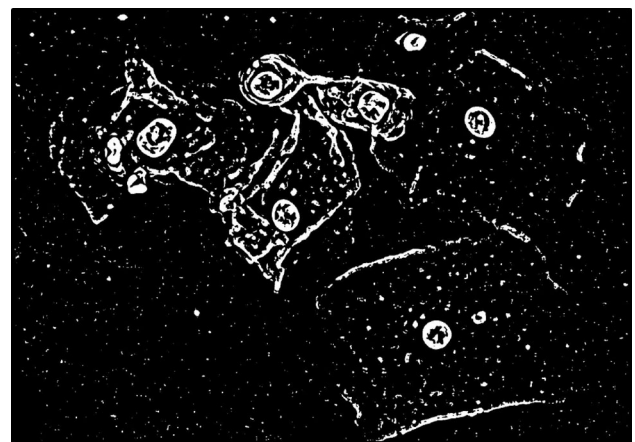


Fig. 7 Binary image from gradient image

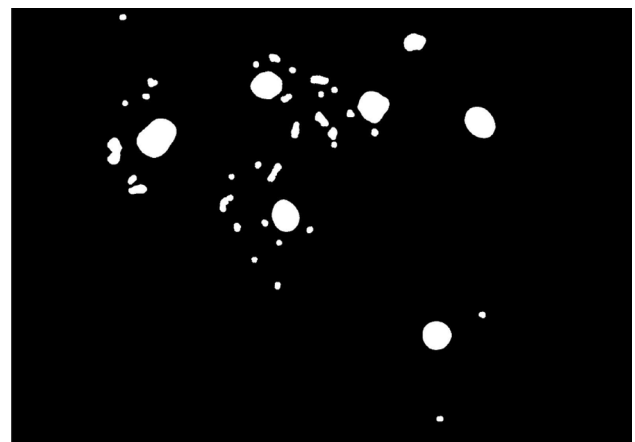


Fig. 8 Holes filled image

out small objects is shown in Fig. 9. The true color image consisting of segmented objects outlined in pink color is shown in Fig. 10.

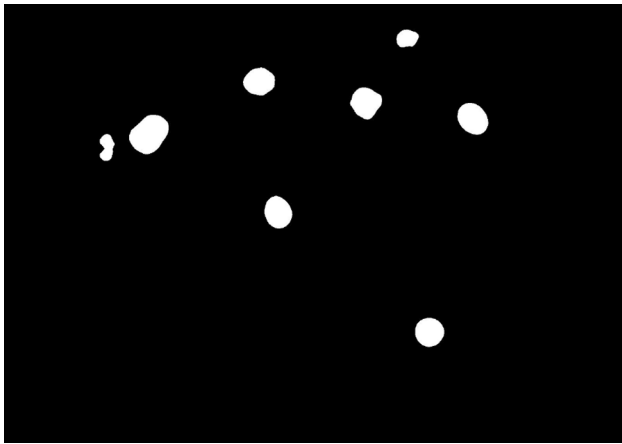


Fig. 9 Segmented objects

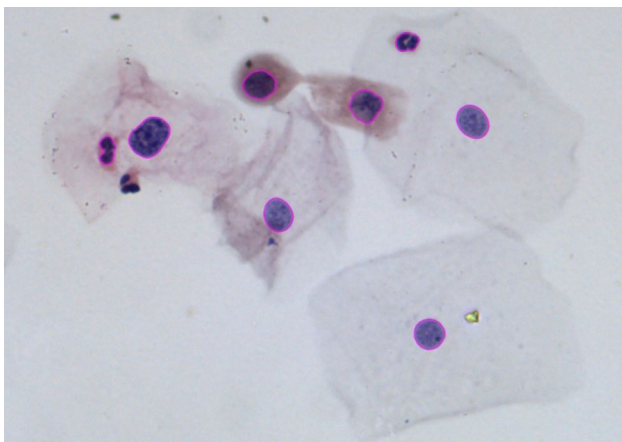


Fig. 10 Segmented nuclei outlined with *pink color*. (Color figure online)

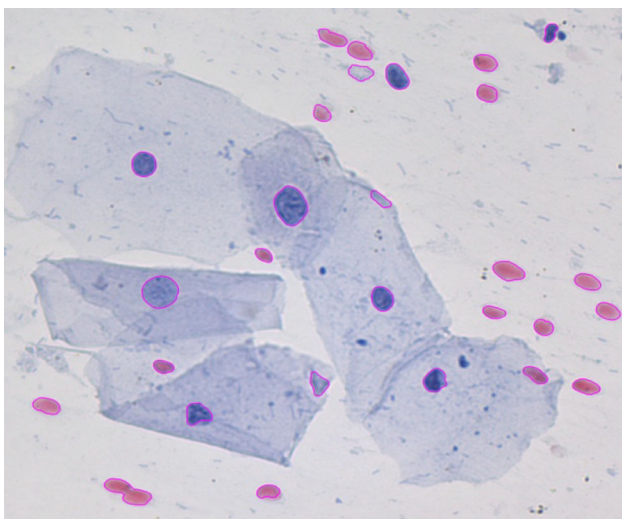


Fig. 11 Segmented objects outlined before red blood cells removal. (Color figure online)

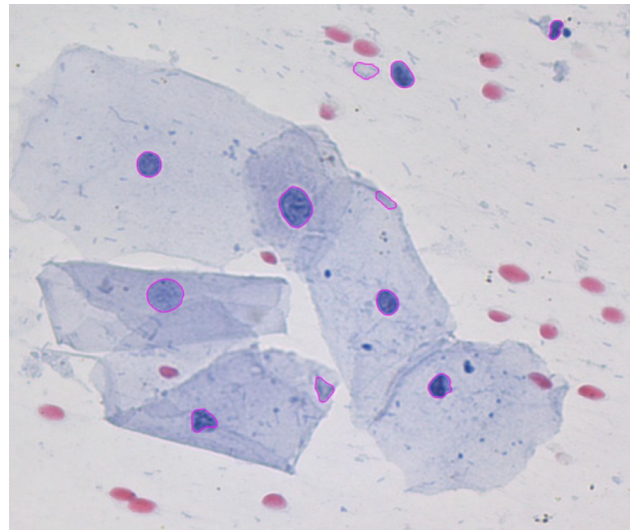


Fig. 12 Segmented objects outlined after red blood cells removal. (Color figure online)

Red blood cells can be removed from the segmented objects set by using true color RGB input image. The method envisaged is to consider each segmented object and find the total red component value and total blue component value of the segmented object. If the total red component value is greater than the total blue component value, it assumes to be a red blood cell and removes that object from the segmented object set.

Figure 11 given below shows the output of segmentation before red blood cell removal. Figure 12 given below shows the output of segmentation after red blood cell removal. The segmented objects are outlined in pink color.

Nucleus overlapping is another issue in segmentation. The overlapped cells can be detected as similar to non overlapped cell detection. The overlapped cells can be extracted out separately based on analysis of perpendicular distance curve [17].

4 Experiments and results

The dataset consists of digitized images of pap smears at 40 \times magnification. The images are taken at 8 MP resolution in three color channel with eight bits/channel using LEICA DFC495 camera with pixel size 2.7 μ m attached to LEICA DM2500 microscope.

A marking tool (GUI) was developed to mark the well focused cell nuclei as ground truths for segmentation. The nuclei marked for both typical and atypical smears. The marking was done for all types of cells, like superficial, parabasal, intermediate etc. A total of 14,704 cells were manually marked as ground truths from a set of normal and

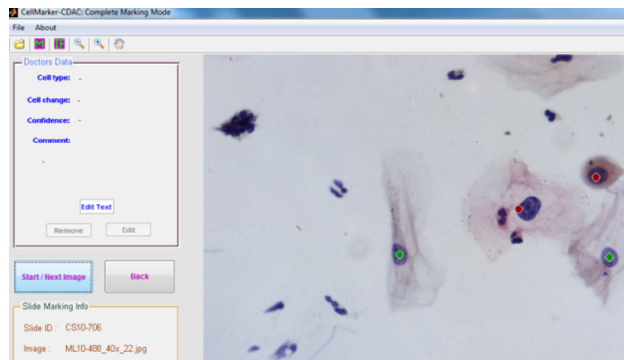


Fig. 13 Ground truth marking for cell nuclei



Fig. 14 Segmented nucleus selection

Table 1 Segmentation sensitivity (S_n)

Smear category	Smear grade	S_n	S_n	S_n
Abnormal	ASC-H	86.22	79.24	80.94
	ASC-US	88.91		
	HSIL	81.17		
	LSIL	85.28		
	SCC	72.28		
Normal	NILM	94.01	94.01	

abnormal smears using the tool. After marking, segmentation was done for the PAP smear images using the proposed algorithm. The average total time taken for segmenting the cervical image is 3 s.

The output of the segmentation was outlined with red color for abnormal cells and green color for normal cells. This tool is used to analyze the sensitivity of segmentation algorithm by selecting the properly segmented objects.

Sensitivity, $S_n = T_s/T_g$

where T_s is total number of properly segmented cells, T_g is total number of ground truths

Figure 13 shows the marking of cervical cell nuclei using the tool and Fig. 14 shows the result of segmentation as outlined and then selecting the properly segmented objects. Table 1 gives the sensitivity of the proposed segmentation method.

5 Conclusion

In this paper we presented an automated method for detecting and segmenting cell nuclei in bright field microscope images of PAP smears having the challenges of stain variation. The result shows that the performance and sensitivity of the method for detecting cell nuclei from cervical images is suitable for a deployable solution. The evaluation of the method has covered a large data set consisting of 14,704 cells from various precursors of cervical cancer like ASC-H, ASC-US, HSIL, LSIL, SCC and NILM. Performance of the algorithm can still be improved using parallel processing.

Acknowledgments This work was carried out as part of the R&D project ‘Medical Image Analyser for Cervical Cancer’ funded by Department of Electronics and Information Technology, Ministry of Communications and Information Technology Govt. of India.

References

- Papanicolaou GN (1942) A new procedure for staining vaginal smears. *Science* 95(2469):438–439
- Regional Cancer Centre (2013) Thiruvananthapuram, http://www.rccvm.org/RCC_AR_2011-12.pdf. Accessed 10 July 2013
- Sulaiman SN, Isa NAM, Yusoff IA, Othman NH (2010) Overlapping cells separation method for cervical cell images. In: Proceedings of the 10th international conference on intelligent systems design and applications (ISDA), pp 1218–1222
- Shah S (2008) Automatic cell image segmentation using a shape classification model. *IEICE Trans Inf Syst* E91-D(7):1955–1962
- Wu HS, Gil J, Barba J (1998) Optimal segmentation of cell images. *IEE Proc Vis Image Signal Process* 145(1):50–56
- Bak E, Najarian K, Brockway JP (2004) Efficient segmentation framework of cell images in noise environments. In: Proceedings of the 26th international conference of the IEEE engineering in medicine and biology, vol 1, pp 1802–1805
- Lezoray O, Cardot H (2002) Cooperation of color pixel classification schemes and color watershed: a study for microscopic images. *IEEE Trans Image Process* 11(7):783–789
- Bamford P, Lovell B (1996) A water immersion algorithm for cytological image segmentation. In: Proceedings of the APRIS image segmentation workshop, Sydney, pp 75–79
- Krishnan N Dr, Naga Nandini Sujatha S (2010) Segmentation of cervical cancer images using active contour models. In: IEEE International conference on computational intelligence and computing research (ICCIC), pp 1–8
- Bamford P, Lovell B (1998) Unsupervised cell nucleus segmentation with active contours. *Signal Process* 71(2):203–213
- Sulaiman SN, Mat Isa NA, Yusoff IA, Othman NH (2010) Overlapping Cells Separation Method for Cervical Cell Images.

- In: 10th international conference on intelligent systems design and applications (ISDA), pp 1218–1222
12. Mat Isa NA (2005) Automated edge detection technique for Pap smear images using moving K-means clustering and modified seed based region growing algorithm. *Int J Comput Internet Manag* 13(3):45–59
 13. Rajesh Kumar R, Ajith Kumar V, Sharath Kumar PN, Sudhamony S, Ravindrakumar R (2011) Detection and removal of artifacts in cervical Cytology images using Support Vector Machine. In: *Proceedings of the IEEE symposium on IT in medicine and education (ITME)*, pp 717–721
 14. Marr D, Hildreth E (1980) Theory of edge detection. *Proc Royal Soc Lond B* 207:187–217
 15. Venkat Narayana Rao T, Govardhan A, Jahangir Badashah S (2011) Statistical analysis for performance evaluation of image segmentation quality using edge detection algorithms. *Int J Adv Netw Appl* 3(3):1184–1193
 16. Gunn SR (1998) Edge detection error in the discrete Laplacian of Gaussian. In: *Proceedings of the international conference on image processing (ICIP 98)*, vol 2, pp 515–519
 17. Mao KZ, Zhaol P, Koh TS, Tan PH (2003) Overlapping/touching cell nuclei segmentation based on analysis of perpendicular distance curve. In: *Proceedings of the IEEE EMBS Asian-Pacific conference on biomedical engineering*, pp 212–213