■ 2048

Inflammatory Cell Extraction in Pap smear Images: A Combination of Distance Criterion and Image Transformation Approach

Rahadian Kurniawan*, Arrie Kurniawardhani, Izzati Muhimmah
Department of Informatics, Faculty of Industrial Engineering, Universitas Islam Indonesia,
Jl. Kaliurang Km. 14.4, Besi-Sleman,Yogyakarta 55584 DI Yogyakarta,Indonesia
*Corresponding author, e-mail: rahadiankurniawan@uii.ac.id[1], arrie.kurniawardhani@uii.ac.id[2],
izzati@uii.ac.id[3]

Abstract

In order to obtain a diagnosis of cervical cancer information, the characteristics of each cell nucleus must be identified and evaluated properly through a Pap smear test. The presence of inflammatory cells in Pap smear images can complicate the process of identification of cell nuclei in the early detection of cervical cancer. Inflammatory cells need to be eliminated to assist pathologists in reading Pap smear slides. In this work, we developed a novel method to extract the inflammatory cells that allow detection of cell nuclei more accuracy. The proposed algorithm consists of two stages: extraction of inflammatory cells using the distance criterion and image transformation. This experiment applied to the 1358 cells comprising 378 nuclei cells and 980 inflammatory cells from 25 Pap smear images. The results showed that our method can significantly reduce the amount of inflammation that can disrupt the cell nuclei in the detection process. The proposed method has promising results with a sensitivity level of 97% and a specificity of 84.38%.

Keywords: cervical cancer, inflammatory cells, distance criterion, image transform

1. Introduction

Study about medical image processing have been conducted widely [1,2]. It is expected can help health worker in diagnosing a disease in remote area, reducing misdiagnosis, and quickening disease identification process. One of medical issues that can utilize image processing techniques to help in diagnosing is interpreting of pap smear images. Visual interpretation of Pap smear images is still having a lot of problems because of its complexity, such as inconsistency of the coloring process, low contrast, overlapping seals, and the existence of inflammatory cells. Those affect the interpretation of cervical cells become a difficult task, even for a trained pathologist. Moreover, on automatic detection of Pap smear image, the existence of inflammatory cells can mislead identification of cell nuclei. That is can be happened because inflammatory cells have similar size, shape, color intensity and texture to cell nuclei. Figure 1 shows the similarity between inflammatory cells and cell nuclei, especially their intensity and shape.

Some research has been conducted to the interpretation of Pap smear images automatically. One of the focus on that research is cell nuclei detections which used various method, such as Morphological Reconstruction and Clustering [3], Fuzzy C-Means Clustering [4], edge detectors [5,6]. Furthermore, the cell nuclei were segmented using deformable templates [7], pixel classification schemes [8], morphological operation and watershed transformation [9] and [10]. Several research have been conducted on the classification of Pap smear images is based on the features of nuclei: [11,12]. This research was conducted to classify types of cervical images and differentiate between normal cells and abnormal cells based on features. Most of research, however, did not describe how to extract the inflammatory cells on Pap smear images. Some of the researches about inflammatory cell extraction on Pap smear images that have been conducted such as [13] using a feature extraction method gray-Level Run-Length Matrix (GLRLM), [14] using shape feature extraction, texture feature extraction, and intensity to differentiate inflammatory cell and cell nuclei.
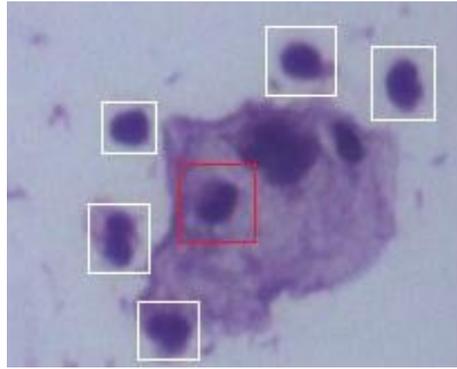
Figure. 1. A part of Pap smear image, nuclei (red box) and inflammatory (white box)

Even though those researchers could extract cell nuclei and inflammatory cells pretty well using feature extraction method, there are some inappropriate features involved to identify cell nuclei and inflammatory cells. Another challenge to identify cell nuclei based on feature extraction is Pap smear images taken by the low resolution camera. Furthermore, a feature extraction method takes longer execution time. Therefore, this paper proposes a novel method to extract and eliminate inflammatory cells using distance criteria and image transformation. In this paper, there are three main phases to extract the inflammatory cell, among others: 1) Segmentation of cytoplasm; 2) Segmentation of nuclei candidate, and 3) inflammatory cell extraction. Further explanation will be described in subsequent subsections.

2. Method
2.1. Segmentation of Cytoplasm

The cytoplasm segmentation phase is needed to reduce the search area in the image. In the first step, the h-minima transform [15] is applied individually for each color component of initial image I (called Hm). The h value of image Hm at each layer of color (r, g, b) is given by,

$$h_r = \frac{\mu(I_r)}{2}; \qquad h_g = \frac{\mu(I_g)}{2}; \qquad h_b = \frac{\mu(I_b)}{2} \qquad (1)$$

where μ is value of average intensities and I is intensity value in the original image. Next, from each filtered image, a grayscale image is produced through grayscale transformation as follows:

$$G = (\,0.2989 \times h_r) + (0.5870 \times h_g) + (0.1140 \times h_b) \qquad (2)$$

A morphological open is then performed in order to flatten the intensity of the region of interest using flat disk-shaped structuring element with radius of 5. Next, we perform sum of the image of the morphological open processes (called M) and the image of the grayscale transformation process (G) define as:

$$J(x, y) = G(x, y) + M(x, y) \qquad (3)$$

After this operation, contrast enhancement filter is applied to complement image of J. Furthermore, in order to expand the region of interest, we perform a morphological dilation using flat disk-shaped structuring element with radius of the derived image (called D). In order to obtain the boundaries of the cytoplasm, we perform the subtraction of two images. The first image A is constructed through grayscale transformation of original images, image B is the outcome of the application subtraction between image D and G. The result of the subtraction of these two images is an image with all cytoplasm region sharp. Finally, the binary mask BW, with the regions of cytoplasm included, is obtained by finding the pixel intensity that less than t value. The value of t is defined as follows:

$$t = \frac{\mu(G)}{2} \tag{4}$$

where μ is value of average intensities. Furthermore, we do some noise reduction in *BS* image by performing the morphological open using flat disk-shaped structuring element with radius of 15. The resulted binary image is used as a mask to indicate the cytoplasm regions. The final result of this part is shown in Figure 2.
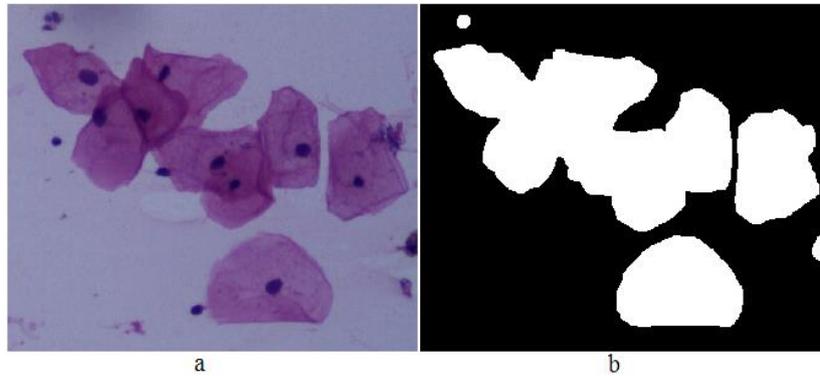
Figure 2. (a) Original Pap smear image, and (b) binary mask, which is obtained after the segmentation of cytoplasm process

2.2. Segmentation of Nuclei Candidate

This stage is done to find the area of candidate nuclei that is region consisting of nuclei cell and inflammatory cell. First, Binary mask as a result from the previous phase uses to remove the background image. This step is performed to remove noise occurred in the background, while narrowing the search space nuclei candidates. This phase starts with each color layer of original image *I* is transformed to RGB image using h-minima transformation with a value of *h*=115, called *MT* image. *MT* image is added to *I* image to generate image *A*, as defined in equation (5). Figure 3(b) shows the result of the adding process, A image.

$$A(x,y) = MT(x,y) + I(x,y) \tag{5}$$

Furthermore, image *A* is subtracted with *Hm* image obtained from the cytoplasm segmentation phase. Image resulting from the subtraction process is called image *S*. Erosion, one of morphology process, is applied to image S using flat disk-shaped structuring element with radius of 5, in order to enlarge the regions of candidate nuclei. Image resulting from the erosion process called image *C* and shown in Figure 3(c), is enhanced using contrast enhancement filter.

The average value of the grayscale intensity of each color layer in image *C* is calculated. The color layer with the lowest average value, called image *K*, is chosen to be subjected of global thresholding processes using the Otsu Thresholding method [16]. Figure 3(e, f, g) shows the grayscale image of each color layer in image *C*. As shown in those Figure 3(e), red layer of image *C*, has the lowest average value of grayscale intensity among another layer. Finally, the binary mask *BM*, with the regions of interest of the image included given by,

$$BM = BS \cap TH \tag{6}$$

where the image *BS* is an image obtained from the final result of cytoplasm segmentation process and image *TH* is an image obtained from a global thresholding process using the Otsu Thresholding method, in the previous step. Figure 3 (h) shows image *BM*. The last step of the nuclei candidate segmentation process is cell separation process using the modified watershed method [17] to handle the overlapping cell nuclei.
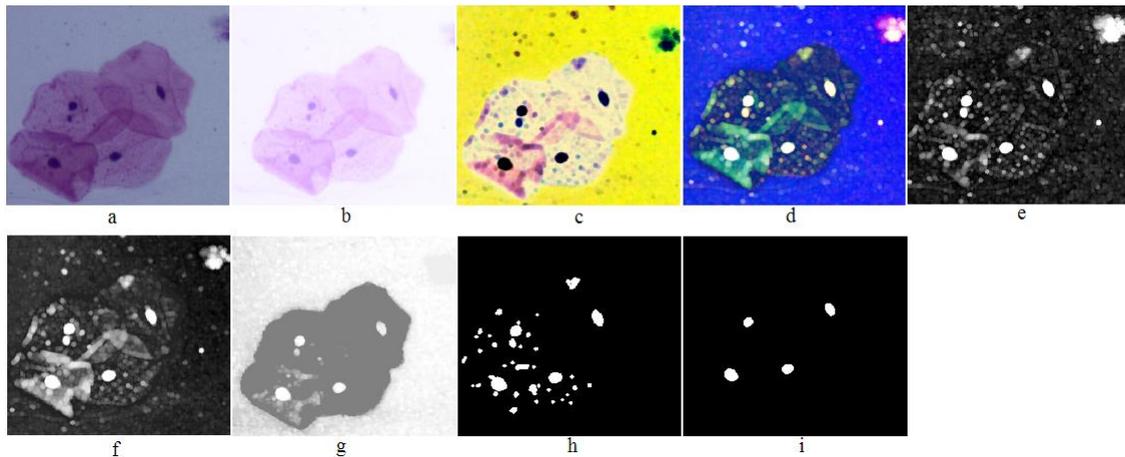
Figure 3. (a) Original Pap smear image, (b) Image A, the output image of addition process,
(c) Image C, (d) Complement image of image C, (e) Red color layer image in Figure 3(d),
(f) Green color layer image in Figure 3(d), (g) Blue color layer image in Figure 3(d),
(h) Image BM, (i) Output image of Nuclei Candidate Segmentation

2.3. Inflammatory Cell Extraction

After successfully obtaining the region of nuclei candidates, the next stage is to eliminate the inflammatory cell area. This stage consists of 2 processes, namely: Distance Criterion, and Image Transform.

2.3.1. Distance Criterion

The general characteristics of inflammatory cell are crowded and its cytoplasm are not as large as cell nuclei. Therefore, in this step, the distance of inter-nuclei candidates are calculated. We developed an algorithm to identify the clustered nuclei cell candidates described in Table 1.

Table 1. Distance Criterion Algorithm

Pseudo Code
Input: Pap smear images contained cell nuclei and inflammatory cells (nuclei candidate).
Output: Pap smear images contained cell nuclei and inflammatory cells that have been reduced.

1. Label connected components of Image *BW* (Segmentation of Cytoplasm step)
2. do loop as many as the number of labels
    2.1 Count the number of nuclei candidates on each label
    2.2 If there is 1 nuclei candidate at a single label
    Then check whether the ratio of the segmentation area of nuclei candidate to segmentation area of cytoplasm smaller than the value *nc*
        2.2.1 If ratio of the segmentation area of nuclei candidate to segmentation area of cytoplasm less than *nc*
        Then marked as cell nuclei
    2.3 If there are 2 to *n* nuclei candidate
    Then marked as cell nuclei
    2.4. If there are more than *n* nuclei candidate
    Then
        2.4.1 Measure the centroid distance of each nuclei candidate using euclidean distance.
        2.4.2 Determine the value of each nuclei candidate according to its shortest distance.
        2.4.3 Normalize the distance using min-max normalization
        2.4.4 Mark as cell nucleus at nucleus candidate that have normalize distance more than *ndist*
  3. end

If it is known that there are cell nuclei candidates crowded in a particular area, then the relative distance of each cell nuclei candidate is sought to the nearest neighbors. If the value of its relative distance is smaller than a predetermined threshold value, it will be considered as inflammatory cells. Algorithm of inflammatory cell extraction using distance criterion is described

in Table 1. In Table 1, there are three parameters, namely *nc*, *n* and *ndist*. The value of each parameter can be seen in Table 3. The value of nc used in this study is determined based on average of nc value from a series of studies conducted previously to 521 normal nuclei different types (superficial, intermediate, parabasal, and endocervix cells). Value of nc is obtained as follows:

$$nc = \frac{Nuclei\ Area}{Cytoplasm\ Area} \tag{7}$$

The value of n is the relative value of the number of nuclei permitted either adjacent or overlap in a Pap smear image. Value of n used in this study is determined based on observation of 421 Pap smear image which there are not overlapping nuclei more than four cells. It means the value of n used for our dataset is 4. Next, the Normalization process is done using min-max normalization method defined as,

$$ndist = \frac{MDn - MDa}{MXa - MDa} \tag{8}$$

where *MDn* is the shortest distance of each nuclei candidate, *MDa* is the shortest distance of nuclei candidates in a cytoplasm, and *MXa* is the longest distance of nuclei candidates in a cytoplasm. Figure 4 illustrates the proposed distance criterion method.

Box A in Figure 4(a) signifies the existence of crowded nuclei candidate while Box B signifies nuclei candidate which do not have cytoplasm. It is indicated that in Box A, inflammatory cell successfully identified and reduced based on a concept of the adjacent nuclei candidate. In Box B, inflammatory cell successfully identified and reduced because the ratio of nuclei candidate area to cytoplasm area is less than value of *nc*.
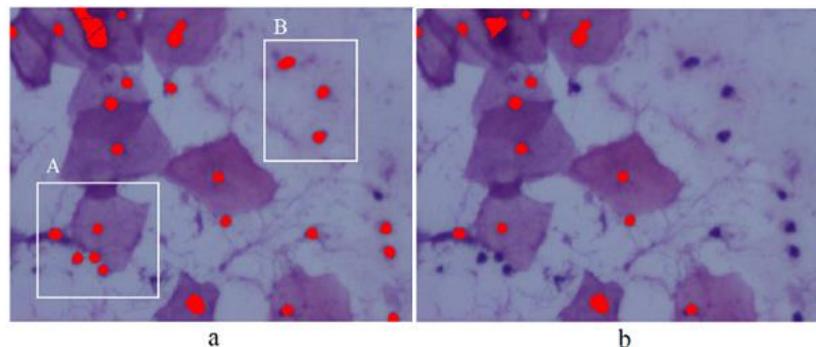
Figure 4. (a) The output of nuclei candidate segmentation (red), (b) The output of inflammatory extraction using distance criterion

2.3.2. Image Transform

In this step, the remaining inflammatory cells from the previous step are reduced utilizing image transformation. We propose a process for reducing the number of inflammatory cells in the image based on image transformation. Figure 5 shows the sequence of the proposed image transformation process.

Firstly, Grayscale transformation is performed in each of nuclei candidates, then median filter is performed to attenuate noise in that area. Gaussian filter is performed in an image obtained from median filter process to take the background area in the bounding box of nuclei candidate. Subtraction process is performed between images from median filter and images from Gaussian filter multiplied by a scaling factor for sharpening the contrast in nuclei area in, this study the scaling factor is 0.9. Later, the thresholding process using the method proposed by Ridler [18] is performed to segment image based in nuclei candidate texture. Synthesis process for the texture pattern (geometrical structure) of nuclei candidate in the output image from segmentation process is performed using Fast Fourier transformation [19]. Log transformation is performed for compressing bright intensity of the pixel and enhancing the dark

intensity of pixels in an image. The last step, min-max normalization process is performed to Log transformation output using equation defined as,

$$mx = \frac{mean_{log} - min_{log}}{max_{log} - min_{log}}$$ (9)

Each Nuclei Candidate (in Grayscale) → Median Filter → Subtract Mask

normalization ← Log Transform ← Fourier Transform
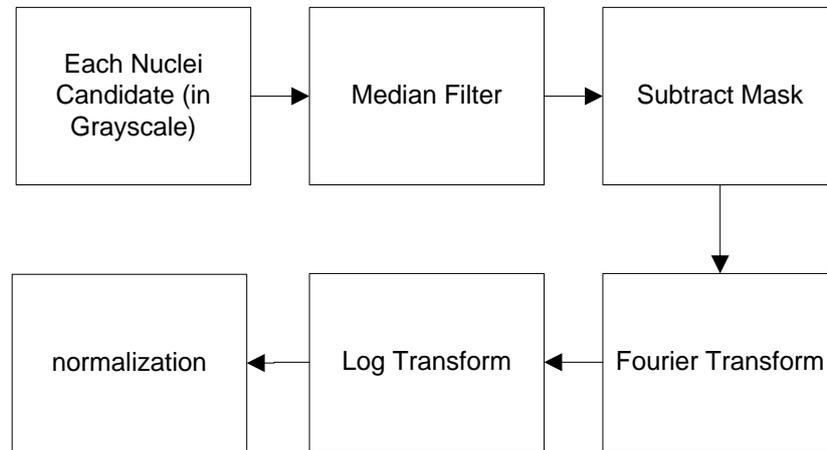
Figure 5. Flowchart for image transformation algorithm

From a series of experiments, mx value for nuclei area is around 0.13-0.29. Therefore, nuclei candidate areas that have mx value around 0.13-0.29 is considered as nuclei area, except that is considered as inflammatory cell. Figure 6 shows mx value of each nuclei candidate.
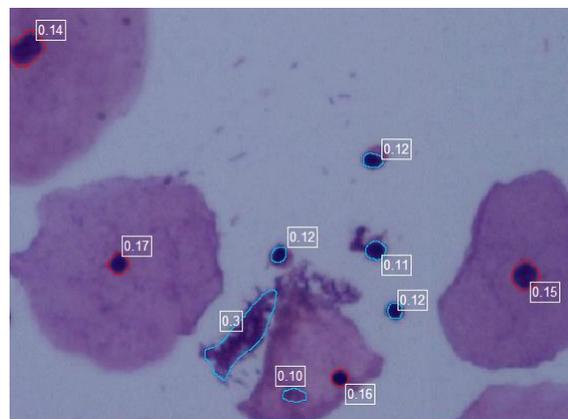
Figure 6. Nuclei boundary (red), inflammatory boundary (blue)

3. Experiment
3.1. Data

Pap smear images used in this study are captured from the Laboratory CITO collection of 25 Pap smear slides using NIKON D100 Microscopy. In those images, there are 1358 cell nuclei candidate composed of 378 cell nuclei and 980 inflammatory cells. Those numbers have been confirmed by our experts. Our experts interpreted those images twice at eight-day interval randomly. That protocol needs to be done to prevent bias which is causing the experts still remember the previous interpretation when the data are neither randomized nor given interval

time to another interpretation. The images are saved in JPEG format, 1280x960 pixel dimensions.

3.2. Evaluation
3.2.1. Execution Time Testing

Table 2 shows time required to execute each step in order to test the efficiency of proposed method using MATLAB software, 2.66 GHz Intel Pentium, and 4 GB RAM. Execution time of distance criterion and the image transformation stage is diverse depending on the number of nuclei candidates in a single image. Beside that the proportion of noise in the background and the bad staining result are other factors that can affect the execution time. At distance criterion stage, the number of crowded cells in a single image gives greatly effect at execution time. The more cells crowded, the longer execution time needed.

Table 2. Execution Time

Step	Time (sec)(mean $\pm$ std)
A. Distance Criterion	4.46$\pm$2.10
B. Image Transformation	2.75$\pm$1.12

3.2.2. Numerical Evaluation

This step compares the results between the proposed method of nuclei detection and manual interpretation of the nuclei area by the experts (expert truth). System validity testing used to calculate sensitivity and specificity level using Single Decision Threshold composed of:

a. TP (True Positive) when the system and the experts state "Cell Nuclei".
b. TN (True Negative) when the system and the experts state "Inflammatory cell".
c. FP (False Positive) when the experts state "Inflammatory cell" whereas system state "cell nuclei".
d. FN (False Negative) when the experts state "cell nuclei" whereas system state "Inflammatory cell".
e. Sensitivity (Se)
   Formula for sensitivity is defined as,

$$\text{Sensitivity (\%)} = \frac{TP}{TP+FN} \times 100\% \tag{10}$$

f. Specificity (Sp)
   Formula for specificity is defined as,

$$\text{Specificity (\%)} = \frac{TN}{TN+FP} \times 100\% \tag{11}$$

The average sensitivity rate of our proposed method was 97% and specificity rate was 84.38% which are quite promising.

4. Discussion

Some parameter values used in proposed method can be shown in Table 3. In our experiment, each of those parameters are acquired empirically by trial and error to 17 Pap Smear images that are different images used for our dataset and 25 Pap Smear images used as our dataset.

Table 3. Values of The Parameters

Step of Method	Parameter	Value
Distance Criterion	the number of nuclei which may overlap (*n*)	4
	a minimum distance between the candidates nuclei (*dist*)	0.1
	the ratio between the nuclei and cytoplasm (*nc*)	0.37
Image Transformation	value of Log Transform(*mx*)	between 0.13 and 0.29

Table 4 shows the comparison between proposed method and the other proposed method previously. It must be noted that it is difficult to compare the methods directly due to differences of datasets and the unavailability of several parameters of data in studies, such as in [20]. Moreover, the image size used in each study is different. This can affect the execution time required for each method. Therefore, execution time testing cannot be done.

Table 4. Comparison of the Proposed Method and Other Methods

Method	Slides	Images	Image Size	Cells	Results	
					Sens	Spec
[3]	15	38	1536x2048	5617	90.57%	75.28%
[14]	11	11	2000x1600	179	95%	98%
[20]	*	50	*	122	49.82%	75.04%
This Work	25	25	1280x960	378	97%	84.38%

*) unknown

According to Table 4, in general, it can be concluded that the sensitivity and specificity of our proposed method are better than those in [20] and [3], but the specificity of our proposed method is worse than it in [14]. Even though having less specificity than it in [14], our proposed method has more numerous and various cells. This can affect the robustness of the model when is applied to a number of new data.

Figure 7 shows inflammatory cell extraction results contained FN and FP. According to Figure 7, FN occurs because the cell nuclei are considered to have a transformed image which tends to lower than the threshold value. On the contrary FP occurs because the inflammatory cells are considered to have a transformed image which tends to higher than the threshold value. A relative far distance between the centroids indicate the existence of true cell nuclei. The remaining nuclei candidates that have relatively shorter distance are eliminated and marked as inflammatory cells. Inflammatory cells are extracted and eliminated by changing their color become the same color as the cytoplasm. Cell nuclei detected are remained with their original color. Figure 8(b) shows the results of this procedure in the original image. According to Figure 8(b), the color of inflammatory cells is changed in accordance with the color of cytoplasm area.
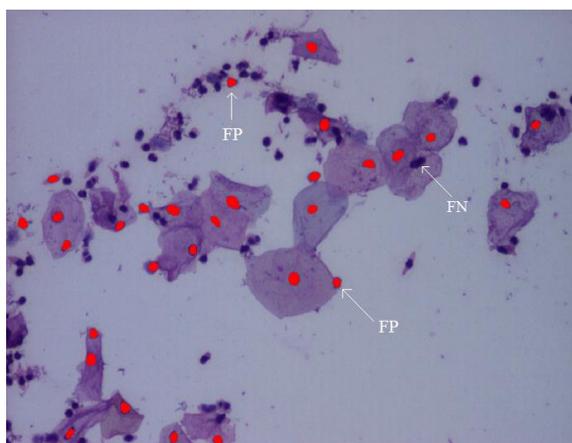
Figure 7. The result of detection of nuclei cells containing FP and FN
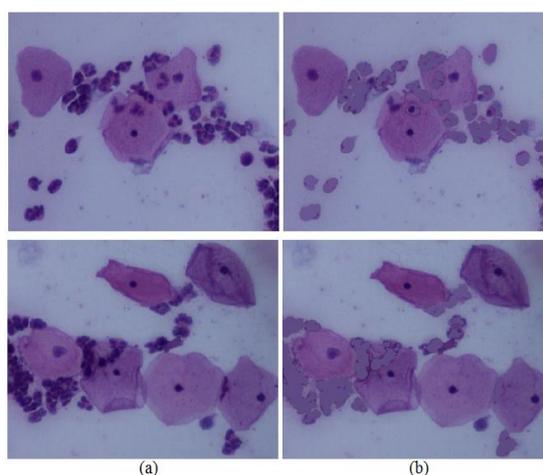
Figure 8. Application of the proposed method (a) Original images, (b) resulted images

5. Conclusion

In this study, we proposed a method to extract inflammatory cells in the Pap smear image. The main advantage of the proposed method is the extraction process is automated and accurate (Se. 97% and Sp. 84.38%). The high specificity percentage indicates that proposed

method is very suitable used for extracting inflammatory cell since the proposed method can distinguish cell nuclei and inflammatory correctly. The high sensitivity percentage indicates that proposed method can be used for further research to classify the type of epithelial cell since very few cell nuclei are not recognized. It means the possibility of loss of certain types of cell nuclei in a single image can be reduced.

Acknowledgment

References

[1]   J Na'am, J Harlan, S Madenda, EP Wibowo. Image Processing of Panoramic Dental X-Ray for Identifying Proximal Caries. *TELKOMNIKA (Telecommunication Comput. Electron. Control).* 2017; 15(2): 702–708.
[2]   S Wardoyo, AS Pramudyo, ED Rizanti, I Muttakin. Exudate and Blood Vessel Feature Extraction in Diabetic Retinopathy Patients using Morphology Operation. *TELKOMNIKA (Telecommunication Comput. Electron. Control).* 2016; 14(4): 1493–1501.
[3]   ME Plissiti, C Nikou, A Charchanti. Automated Detection of Cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering. 2011; 15(2): 233–241.
[4]   I Muhimmah, R Kurniawan, Indrayanti. *Automatic Epithelial Cells Detection of Pap smears images using Fuzzy C-Means Clustering.* 4th International Conference on Bioinformatics and Biomedical Technology. 2012: 122–127.
[5]   CH Lin, YK Chan, CC Chen. Detection and segmentation of cervical cell cytoplast and nucleus. *Int. J. Imaging Syst. Technol.* 2009; 19(3): 260–270.
[6]   SF Yang-Mao, YK Chan, YP Chu. Edge enhancement nucleus and cytoplast contour detector of cervical smear images. *IEEE Trans. Syst. Man Cybern. B, Cybern.* 2008; 38(2): 353–366.
[7]   A Garrido, NP de la Blanca. Applying deformable templates for cell image segmentation. *Pattern Recognit.* 2000; 33: 821–832.
[8]   E Bak, K Najarian, JP Brockway, POB Davidson. *Efficient segmentation framework of cell images in noise environments.* Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE. 2004; 1: 1802–1805.
[9]   I Muhimmah, R Kurniawan. *Shape-based nuclei area of digitized pap smear images.* Fourth International Conference on Digital Image Processing (ICDIP 2012). 2012; 8334(Icdip): 83344J–83344J–5.
[10] I Muhimmah, R Kurniawan, Indrayanti. *Automated Cervical Cell Nuclei Segmentation using Morphological Operation and Watershed Transformation.* IEEE International Conference on Computational Intelligence and Cybernetics. 2012: 163–167.
[11] NM Harandi, S Sadri, NA Moghaddam, R Amirfattah. An automated method for segmentation of epithelial cervical cells in images of ThinPrep. *J. Med. Syst.* 2010; 34(6): 1043–1058.
[12] H Greenspan, et al. Automatic Detection of Anatomical Landmarks in Uterine Cervix Images. *IEEE Trans. Med. Imaging.* 2009; 28(3): 454–468.
[13] D Riana, ME Plissiti, C Nikou, DH Widyantoro, TLR Mengko. Inflammatory cell extraction and nuclei detection in Pap smear images. *Int. J. e-Health Med. Commun.* 2015; 6(2): 27–43.
[14] I Muhimmah, R Kurniawan, Indrayanti. *Analysis of features to distinguish epithelial cells and inflammatory cells in Pap smear images.* IEEE 2013 6th International Conference on BioMedical Engineering and Informatics (BMEI 2013). 2013: 519–523.
[15] P Soille. Morphological Image Analysis: Principles and Applications. New York. Springer-Verlag. 1999.
[16] N Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 1979; SMC-9(1): 62–66.
[17] R Kurniawan. Modified Watershed Algorithm Based on Distance- Metric Criterion for Nuclei Clustered Separation in Pap Smear Images. *Teknoin.* 2013; 19: 46–54.
[18] TW Ridler and S Calvard. Picture thresholding using an iterative selection method. *IEEE Trans. System, Man and Cybernetics.* 1978: 630–632.
[19] JW Cooley and JW Tukey. An algorithm for the machine computation of complex fourier series. *Math. Comput.* 1965; 19.
[20] D Riana, DH Widyantoro, TL Mengko. *Extraction and Classification Texture of Inflammatory Cells and Nuclei in Normal Pap smear Images.* 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME). 2015: 65–69.