Proceeding of International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2014), Yogyakarta, Indonesia, 20-21 August 2014

# *Detection Technique of Squamous Epithelial Cells in Sputum Slide Images using Image Processing Analysis*

Fatin Izzwani Azman
Universiti Malaysia Pahang
Faculty of Electrical and Electronics Engineering
Pahang Darul Makmur, Malaysia
fatinizzwani@yahoo.com

Rosyati Hamid
Universiti Malaysia Pahang
Faculty of Electrical and Electronics Engineering
Pahang Darul Makmur, Malaysia

Kamarul Hawari Ghazali
Universiti Malaysia Pahang
Faculty of Electrical and Electronics Engineering
Pahang Darul Makmur, Malaysia

Noor Amira Syuhada Mahamad Salleh[1]
Universiti Malaysia Pahang
Faculty of Electrical and Electronics Engineering
Pahang Darul Makmur, Malaysia

*Abstract:* **A good quality sputum is important to detect diseases. The presence of squamous epithelial cells (SEC) in sputum slide images is important to determine the quality of sputum. The presence of overlapping SEC in sputum slide images causes the process become complicated and tedious. Therefore this paper discusses on technique of detection and summation for Squamous Epithelial Cell (SEC) in sputum slide image. We addressed the detection problem by combining K-means and color thresholding algorithm. The design of aided system is evaluated using 200 images and the proposed technique is capable to detect and count each SEC from overlapping SEC image. Total of 200 images were clustered to 10 groups, labelled as Group Cell 1 to group Cell 10 that correspond to the number of cells in the image. Therefore, each group will contain 20 images. The accuracy of the algorithm to detect SEC was also measured, and results show that in 91% which provides a correct SEC detection and summation.**

***Keywords: Squamous Epithelial Cell (SEC); K-means algorithm; color thresholding.***

## I. INTRODUCTION

Sputum is a thick fluid produced in the lungs and in the airways leading to the lungs. Diseases relate to lung such as Moraxella catarrhalis, Mycobacterium tuberculosis and others can be determined from sputum. In sputum sample, it contains pus cells (neutrophils), squamous epithelial cells (SEC), gram-positive and gram-negative organisms. Sputum with good quality is important to detect diseases and able to provide the information needed in a sputum sample. Good quality sputum is obtained when it is cough up from the lower respiratory tract (LRT) and non-quality sputum obtained from upper respiratory tract (URT) which consists of only saliva and normally consists of more epithelial cells. Sputum with good quality is important to detect diseases.

There are six different criteria for judging the acceptability of sputum specimens [1]. All of the criteria,

Modified Barlett's criterion is chosen due to its easiness in interpretation and lower rejection rate than other criteria

[2]. The quality of sputum is determined using Bartlett"s Criteria by considering the score of SEC, pus cell (neutrophils) and macroscopy. The score of each criterion is summarized in the Table 1 [3].

Based on Modified Barlett's criteria, the grading of the sputum can be determined using the expression in Eq. (1):

$$Total\ score = Score\ A + Score\ B + Score\ C \quad (1)$$

If the total score is 1 and above, the sputum will be cultured and the specimens will be proceed accordingly. Whereas if the total score is 0 and below, the process of sputum will stop.

TABLE I. MODIFIED BARLETT'S CRITERIA

| | Criteria | Score |
|---|---|---|
| **Neutrophils (pus cells) count (Score A)** | < 10 neutrophil/10x field | 0 |
| | 10-25 neutrophils/10x field | +1 |
| | >25 neutrophils/10x field | +2 |
| **Macroscopy (Score B)** | Mucoid, Mucopurelent, Purelent, or Blood stained | +1 |
| **Squamous epithelial cell count (Score C)** | < 10 Squamous epithelial cell/10x field | 0 |
| | 10-25 Squamous epithelial cell /10x field | -1 |
| | >25 Squamous epithelial cell /10x field | -2 |

Currently, the detection cells in sputum slide are manually done by microbiologists using biological microscope. It's purely using user's vision which can be daunting and the error probability can be high due to

limitation of human vision. Since human might have eye drowsiness while dealing with this task, thus automated detection and summation of SEC for sputum slide images need to be developed to curb this problem.

However in this study, the focus is only on the method to detect and count Squamous Epithelial cells (SEC). Squamous Epithelial Cells (SEC) is presence in sputum of people who have diseases like pneumonia[4], tuberculosis (TB) [5] or lung cancer[6]. In lung cancer for example, Qi Qiu et al. [6] propose to obtain concentrated and purified bronchial epithelial cells to improve early detection of lung cancer in sputum samples. The presence of SEC was the most universal criterion for judging specimen quality and acceptability for culture, used by 98% of participant laboratories [7]. In comparison, the image for SEC is easy to see because the sizes of SEC are larger than other elements as can be refer from Figure 1 as it shows an example of one field of sputum slide image using biological microscope.

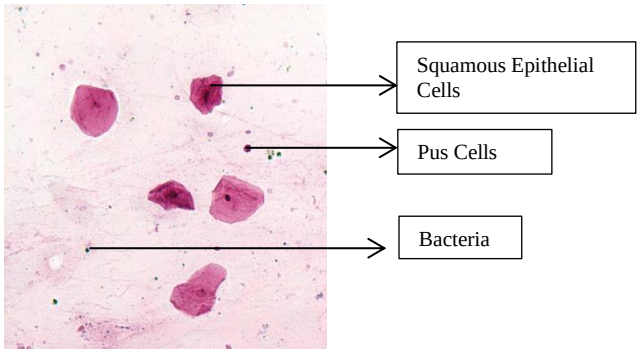



Figure 1 Sputum image under x10 magnification

The sputum cells are characterized by uncertainty cells pattern that make the segmentation and detection of the cells very problematic, so it is difficult to segment the foregrounds from the image automatically and perfectly. In few studies, color image processing has been used to detect tuberculosis in sputum samples by a few researchers namely Sotaquira *et al* [7], R.A.A. Raof *et al* [8], Rohit Nayak *et al* [9] and Vishnu Makkapati *et al* [10]. FatmaTaher *et al* [11] has proposed a method to segment the sputum color images by using Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm.

## II. METHOD AND ANALYSIS

Basically methods to detect and count SEC within a sample, comprising the step of: data acquisition, image conversion, enhancement process, noise elimination, segmentation and detection process.

In this paper, sputum images were collected as an input data from Hospital Universiti Sains Malaysia (HUSM), Kelantan, Malaysia. Images were collected using a microscope which is attached to digital camera and connected to PC with special software (Camera: Olympus XC50, U-CMAD3, Japan. Microscope: Model-BX41TF-FL_CCD, Olympus, serial number: BG22578. Software: AnalySIS docu, copyright 1986-2007, Olympus Soft Imaging Solutions GmbH). Then we completed the entire algorithm using computing software called MATLAB. For this paper, input data of 200 images were clustered to 10 groups which are group Cell 1 to group Cell 10 respects to their number of cells in

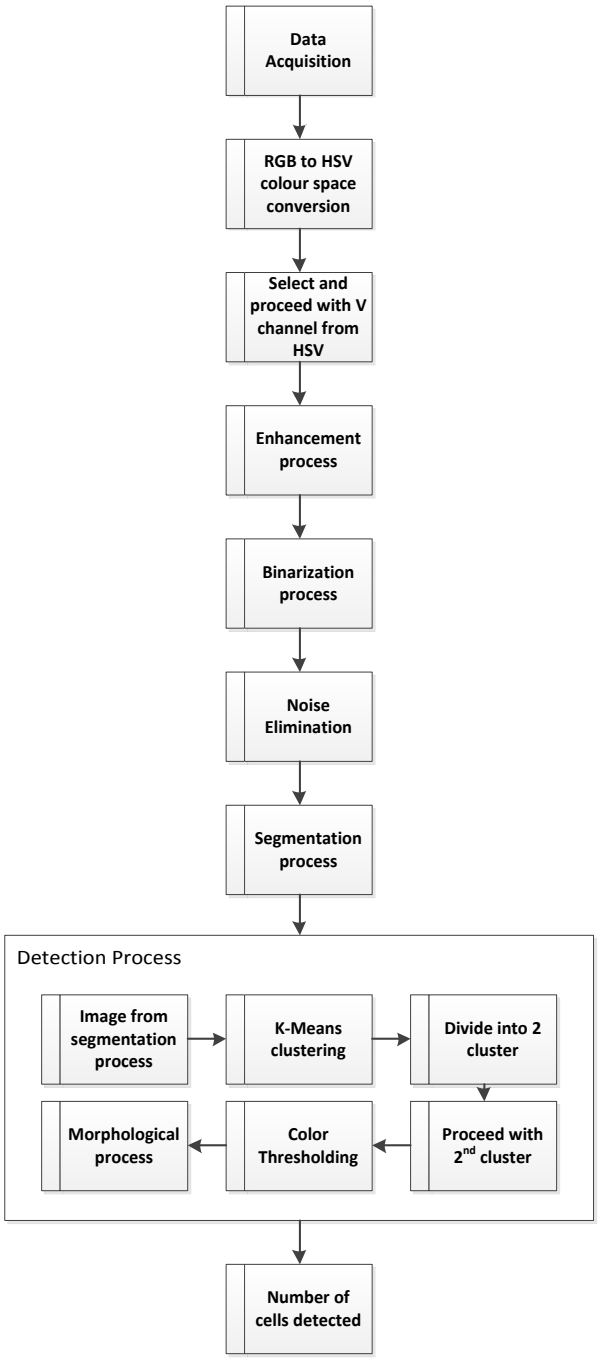
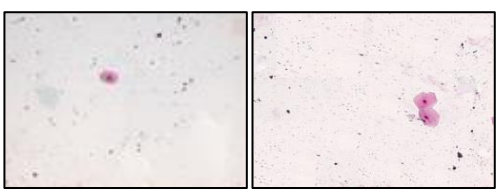
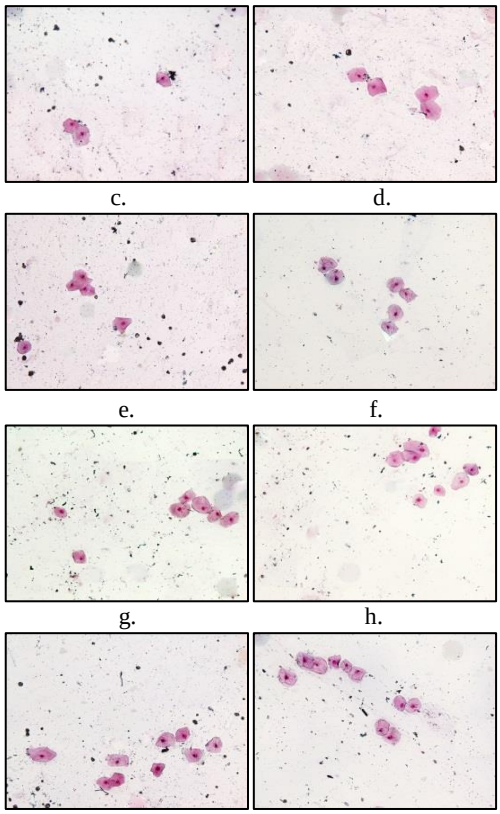



Figure 2. SEC detection algorithm

each image. As a result each group contains 20 images. This clustered step is important to analyse the detection and counting accuracy. Figure 3 shows samples of data in each group.



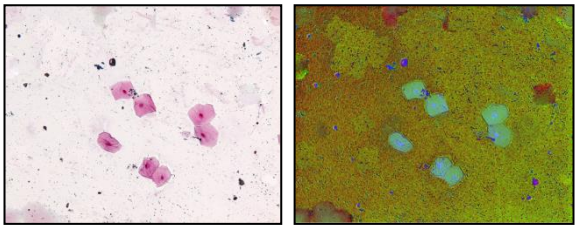
a. b.



c. d. e. f. g. h. i. j.

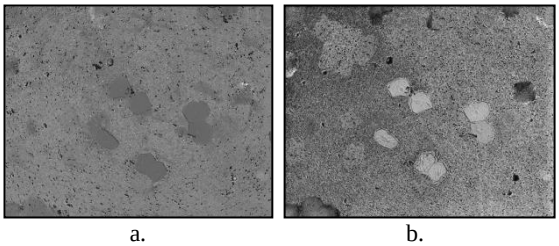
Figure 3. Sample of data for each clustered group: a. Group Cell 1, b. Group Cell 2, c. Group Cell 3, d. Group Cell 4, e. Group Cell 5, f. Group Cell 6, g. Group Cell 7, h. Group Cell 8, i. Group Cell 9, and j. Group Cell 10

The RGB sputum images will be converted to HSV color space. HSV color model is a method which defines color according to the three feature of the color hue, saturation, and intensity or value [12]. After conversion, it represents hue, saturation, and value respectively. In our system, the reason we convert RGB sputum image to HSV color space is because components of value are so closely link with the pattern of human visual perception.

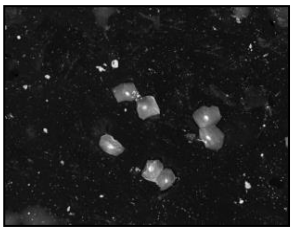


a. b.

Figure 4. HSV conversion: a. Original RGB image b. HSV image



a. b.



c.

Figure 5. HSV channels: a. Hue component, b. Saturation component, c. Value component

From the HSV image, only value component will be proceed for further analysis. A satisfying segmentation can be obtained from value component rather than hue and saturation component. An enhancement process is applied to enhance the image by using 2-D adaptive noise removal filtering. It lowpass-filters a grayscale image that has been degraded by constant power additive noise.

$$\mu = \frac{1}{NM}\sum_{n1,n2\in\eta} A(n_1, n_2) \quad (2)$$

and

$$\sigma^2 = \frac{1}{NM}\sum_{n1,n2\in\eta} a^2(n_1, n_2) - \mu^2 \quad (3)$$

where $\mu$ is the *N*-by-*M* local neighborhood of each pixel in the image A.

$$b(n_1, n_2) = \mu + \frac{\sigma^2 - v^2}{\sigma^2}(A(n_1, n_2) - \mu) \quad (4)$$

where $v^2$ is the noise variance. If the noise variance is not given, it uses the average of all the local estimated variances. This enhancement process results an image with a smooth and blur surface as it reduces a small noises.

Then, the enhance image is changed into black and white image. The noises which are smaller objects compared to SEC are eliminated by removing objects which are less than 300 pixels. Image without noise then will be multiply with original image so the SEC will be extracted from the background.

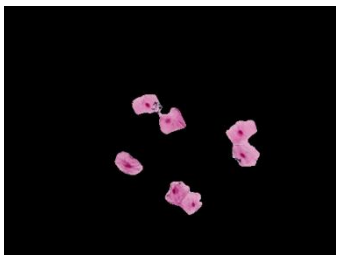


Figure 6. Segmented SEC

After we segment the SEC, we proceed to the detection stage. K-Means clustering was performed, so the brightest pixel values of nucleus are clearly visible.

K-Means algorithm classifies the input data points into multiple classes based on their inherent distance from each other. K-means clustering is used because it is simple and has relatively low computational complexity [13]. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them [14].

$$V = \sum_{i=1}^{k} \sum_{x_j \varepsilon s_i} \left(x_j - \mu_i\right)^2 \quad (5)$$

For smaller values of k the k-Means algorithms give good results [14]. For larger values of k, the segmentation is very coarse, many clusters appear in the images at discrete places [14]. This is because Euclidean distance is not a very good metric for segmentation processes. The result after undergo k-Means clustering is shown in Figure 7.

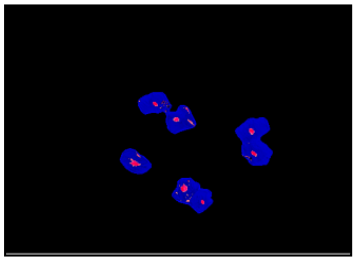
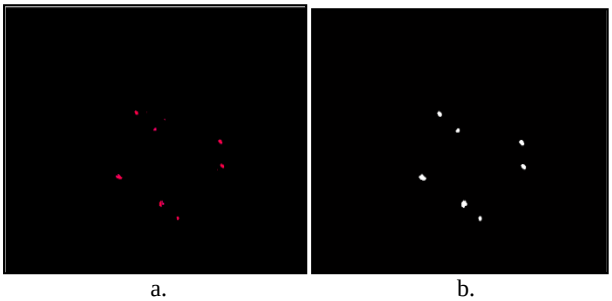


Figure 7. Image after k-Means clustering (two clusters, k=2)

As can be seen in Figure 7, the pink color represents the nucleus of each cell is clearly visible. To extract the nucleus, this system undergoes color thresholding to get the pink color. Then, the image will go through some morphological process to remove small noises.



a. b.

Figure 8. a. Results after color thresholding, b. Results after some morphological process

## III. RESULT AND DISCUSSIONS

Then, the system will count the connected object in Figure 8b and stored the results. The example of result is recorded in Table 2.

TABLE II. EXAMPLE OF RESULTS

| | Image | Number of SEC in the Image | Number of SEC detected |
|---|---|---|---|
| Group Cell 1 | 1 | 1 | 1 |
| | 2 | 1 | 1 |
| | . | 1 | 1 |
| | . | 1 | 1 |
| | 20 | 1 | 1 |
| Group Cell 2 | 1 | 2 | 2 |
| | 2 | 2 | 2 |
| | . | 2 | 2 |
| | . | 2 | 2 |
| | . | 2 | 2 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| Group Cell 20 | 1 | 20 | 19 |
| | 2 | 20 | 20 |
| | . | 20 | 19 |
| | . | 20 | 20 |
| | 20 | 20 | 20 |

We conducted a set of experiments to analyse and discuss the results of our work in detection process. The performance of detection method is evaluated on detection and counting accuracy. Input data of 200 images were clustered to 10 groups which are group Cell 1 to group Cell 10 respects to their number of cells in each image. As a result each group contains 20 images. For example, in group Cell 3, each image contains three SEC. This clustered step is important to test the accuracy of the system. The result is shown in Figure 9. The number of image correctly detected represent images that detect the number of SEC correctly in each image.

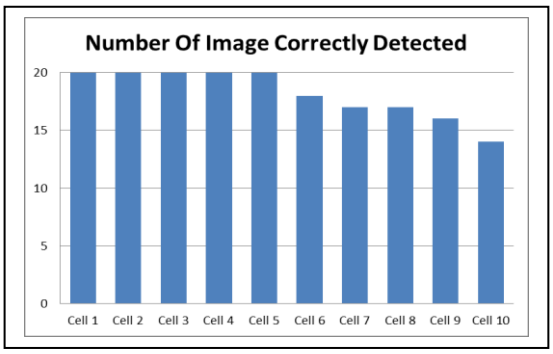



Figure 9. Detection and summation results

Figure 9 summarizes the number of image successfully detected. As can be seen from Figure 9 above, all images in group Cell 1 to Cell 5 correctly detect total number of cells in each image. However with the increasing number of cell in images, the detection and counting result become less accurate. The detection performance is measured by percentage of number of images successfully detected. From the results we can see that the system reliability and accuracy is above 90%.

TABLE III. RESULT OF ACCURACY

| Algorithm | Accuracy |
|---|---|
| Detection Process | 91% |

## IV. CONCLUSIONS

In this paper, a detection technique for SEC detection in sputum slide images was presented. The technique is a combination of K-Means algorithm and color thresholding method. The detection technique to detect SEC has proven successfully able to detect SEC in sputum slide images. It has ability to detect and count overlapping SEC in sputum images. Even though there some error occurred regarding to unsuccessfully cell detected, but this system reliability is above 90% from the validation results.

## References


1. Lester K. Wong, A.L.B., Susan M. Horgan, *Comparison of Six Different Criteria for Judging the Acceptability of Sputum Specimens.* Journal of clinical microbiology, October 1982. 16(4): p. 627-631.
2. Zakuan Z. Deris , S.A.H., Zeehaida Mohamed, Habsah Hassan, *Implementation of routine sputum rejection criteria to improve the outcome of sputum culture result.* International medical journal, September 2008. 15: p. 287-290.
3. *Medical Microbiology and Parasitology Laboratory Hospital HUSM* in *Sputum Culture*.
4. *TML\MSH Microbiology Department Policy & Procedure Manual*
5. Konstantinos Veropoulos, C.C., and Genevieve Learnmonth,, *Image Processing and Neural Computing Used in the Diagnosis of Tuberculosis*.
6. Qi Qiu, N.W.T., Ruiyun Li, Hong Peng, Zhenqiu Liu, Harris G.Yfantis, Ruth L.Katz, Sanford A.Stass, and Feng Jiang, *Magnetic Enrichment of Bronchiol Epithelial Cells from Sputum for Lung Cancer Diagnosis*, 2008.
7. Ron B. Schifman, T.V.A., and Frederick Meier, *Sputum Specimen Adequacy Data Analysis and Critique*, 1991.
8. R.A.A Raof, Z.S., S.I. Sahidan, M.Y. Mashor, S.S. Md Noor, F. MohamadIdris and H.Hasan, *ColorTresholding Method for Image Segmentation Algorithm of Ziehl-Neelsen Sputum Slide Images*, in *5th International Conference on Electrical Engineering*2008:
9. Rohit Nayak, V.P.S., Ramesh R.G, *A New Algorithm of Automatic Assessment of the Degree of TB-infection Using Images of ZN-stained Sputum Smear,* in *Proceeding of 2010 International Conference on Systems in Medicine and Biology*2010.
10. Vishnu Makkapati, R.A., Raviraja Acharya,, *Segmentation and Classification of Tuberculosis Bacilli from ZN-stained Sputum Smear Images,* in *IEEE Conference on Automation Science and Engineering* 2009.
11. Sammouda, F.T.a.R., *Artificial Neural Network and Fuzzy Clustering Methods in Segmenting Sputum Color Images for Lung Cancer Diagnosis*, in *ICISP*2010. p. 513-520.
12. Woods, R.C.G.a.R.E., *Digital image processing*. 3rd ed. 2009: person prentice hall.
13. H.P. Ng, S.H.O., K.W.C. Foong, P.S. Goh and W.L. Nowinski, *Medical image segmentation using k-means clustering and improved watershed algorithm*, 2006.
14. Mehta, S.T.a.A., *Image segmentation using k-means clustering, EM and Normalized cuts.*