

Color segmentation of multi variants tuberculosis sputum images using self organizing map

Riries Rulaningtyas^{1,*}, Andriyan B Suksmono², Tati L R Mengko and Putri Saptawati

¹ Biomedical Engineering, Department of Physics, Faculty of Science and Technology, University of Airlangga, Jl. Mulyorejo, Kampus C, Surabaya 60115, Indonesia

² School of Electrical Engineering and Informatics, Bandung Institute of Technology Jl. Ganesha, Bandung 40132, Indonesia

*Corresponding email address: riries-r@fst.unair.ac.id

Abstract. Lung tuberculosis detection is still identified from Ziehl-Neelsen sputum smear images in low and middle countries. The clinicians decide the grade of this disease by counting manually the amount of tuberculosis bacilli. It is very tedious for clinicians with a lot number of patient and without standardization for sputum staining. The tuberculosis sputum images have multi variant characterizations in color, because of no standardization in staining. The sputum has more variants color and they are difficult to be identified. For helping the clinicians, this research examined the Self Organizing Map method for coloring image segmentation in sputum images based on color clustering. This method has better performance than k-means clustering which also tried in this research. The Self Organizing Map could segment the sputum images with y good result and cluster the colors adaptively.

1. Introduction

Pulmonary tuberculosis is a deadly infectious disease occurred in many countries in Asia and Africa. In Indonesia, many people with tuberculosis disease are examined in the community health centre. Examination of tuberculosis through sputum smear is a result from Ziehl – Neelsen staining in microscopy. From the results, Ziehl – Neelsen staining will give red color effect to the tuberculosis (TB) bacteria and blue color for background images. The first step of examination was to detect the presence of TB bacteria from the red side, then from the shape of the TB bacteria itself.

The Ziehl – Neelsen staining in sputum smear gives the complex color images, so that the clinicians have difficulty in doing slide examination manually. To assist the clinicians when reading the sputum smear slide, this research did color segmentation method for the sputum images which had been digitized. The purpose was to extract the TB bacteria images and eliminate the background images. The red TB bacteria after Ziehl – Neelsen staining provided the different brightness red levels, from brown, dark red, until to pink. The right color segmentation is necessary for the accurate classification techniques in order to distinguish the image of TB bacteria, the image of non TB bacteria, and the image of background.

Examination of patients with pulmonary tuberculosis began with an examination of sputum. The examination of sputum was the main step, because it could identify the presence of TB bacteria so that the disease can be diagnosed as pulmonary tuberculosis. The Indonesian community health centers commonly use a light microscopic examination techniques using Ziehl - Neelsen staining for sputum



observation. This staining gives a red effect for TB bacteria and blue color for the background of the sputum sample. The examination of the TB bacteria is still done manually by clinicians to perform manually counting the number of bacteria in each field of view of the microscope. For pulmonary TB disease diagnose becoming easy and fast, this research made a diagnosis of pulmonary tuberculosis automatically. The first step was the enhancement process for digital sputum images. After that, the color segmentation was done to sputum images for clearing up the background images and leaved the TB bacteria images. Some researchs which had been done correlated with the segmentation of TB bacteria images were [1],[2],[3]. Research [1] did the color segmentation using adaptive color thresholding in RGB color space while research [2] did color segmentation of the TB bacteria images using pixel classifier method. In addition, research [3] did color segmentation for TB bacteria using adaptive color thresholding in HSV color space. The researcher has opinion that the color segmentation has important rule to support the success of TB sputum diagnose. To improve the accuracy of TB identification in [1][2][3], this research already did color segmentation using self-organizing map (SOM). Previously, this research examined k-means cluster for color image segmentation, but there were some shortcomings to choose the best centroid values. It needs long computational time. With SOM method, the color could be segmented more fast and accurately. After segmenting background, identification is continued with the shape identification. Research [4] did TB bacteria identification in binary images using neural network based on geometry feature only without using color feature as information for classification process. SOM with clustering method to segment the color of TB images from background could support the identification TB bacteria images from sputum more accurately.

2. Material and Method

2.1. Image acquisition

The number of sputum smear slides used was about 1000 image of Ziehl-Neelsen sputum smear provided by Bandung Health Department in West Java Province, Indonesia. Images of sputum smear slide were captured using web camera integrated with ocular lens and mounted onto ocular field and saved in JPEG format in RGB 24 bits. The total of microscope magnification was 1000 times. The image size was 640x360 pixels.

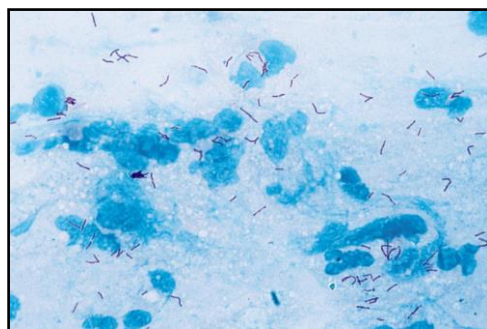


Figure 1. The Ziehl-Neelsen Sputum Image

2.2. K-Means clustering

K-means clustering is grouping data method which widely used. K-means clustering method has objective function based on the proximities of the data points to the centroid of cluster [8]. With $X = \{x_1, \dots, x_n\}$ be the data, $m_l = \sum_{x \in C_l} \frac{x}{n_l}$ be the centroid of cluster C_l , $1 \leq l \leq k$, where n_l is the number of the data objects in cluster C_l , and k is the number of clusters [3]. The objective function of K-means clustering is then formulated as the sum of squared errors [5] :

$$F_k = \sum_{l=1}^k \sum_{x \in C_l} \|x - m_l\|^2 \quad (1)$$

With $d(C_p, C_q) = \sum_{x_i \in C_p} \sum_{x_j \in C_q} \|x_i - x_j\|^2$, the sum of all pair wise distances of data objects within K clusters as follows [3]:

$$D_k = \sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|^2 = \sum_{l=1}^k d(C_l, C_l) + 2 \sum_{1 \leq i < j \leq k} d(C_i, C_j) \quad (2)$$

2.3. Self Organizing Map

Self-Organizing Maps (SOM) have unsupervised training to get the best architecture. With the concepts of mapping input to the output based on similarity, so that SOM suitable to the clustering process. This research tried to use SOM for color image segmentation based on clustering method.

Kohonen SOM is designed to group a set of Q continuous - valued vectors $s^{(q)} = [s_1^{(q)} s_2^{(q)} \dots s_N^{(q)}]$, $q = 1, \dots, Q$, into M ($< Q$) clusters. The SOM consists of an input layer having N neurons, and an output layer having M neurons (cluster units) arranged in some predetermined fashion. Each neuron in the input layer is connected to a neuron in the output layer. Thus, output neuron (cluster units) j is connected to each of the input neuron through weights $w_{ij}, i = 1, \dots, N$, which is referred to as the j-th weights vector. In vector or matrix notation, it is denoted the j-th weight vector by $W_j = [w_{1j} w_{2j} \dots w_{Nj}]$. It is given by the j-th column of the weight matrix. There is a total of M such weight vectors, one for each cluster units.

During the training process, each training vector is cyclically or randomly selected and presented to the network. Two vectors are considered the closest if the square of the Euclidean distance between them, $\|x - W_j\|^2$ was the smallest. The winning unit and its neighboring units (those located in its first and second neighborhoods) then update their weight vectors according to the Kohonen rule. This process is continued until the weight vectors changed by less than a preset amount. The square of the Euclidean distance between the input vector and each of the weight vectors is [6]:

$$d_j = \|x - W_j\|^2 = \|x\|^2 + \|W_j\|^2 - 2x \cdot W_j \quad (3)$$

The SOM algorithm:

1. Initialize M weight vectors. Set topological neighborhood parameters and learning rate, $\alpha (< 1)$
2. For step $k = 1, 2, \dots$, do steps a-d by cycling through training set until weight vectors converge
 - a. Set input vector $x = s^{(q)}$, one of the training vectors.
 - b. Compute for each cluster unit $j = 1, \dots, M$ the Euclidean distance

$$d_j = \sum_{i=1}^N (x_i - w_{ij}(k))^2 \quad (4)$$

c. Find the index j' such that $d_{j'}$ is a minimum

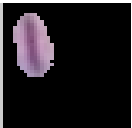
d. For all cluster units j within the specified neighborhoods of j' , update the weight vectors

$$w_{ij}(k+1) = w_{ij}(k) + \alpha [x_i - w_{ij}(k)], \quad i = 1, \dots, N \quad (5)$$

3. Result

Firstly, color segmentation was done to TB bacteria identification. Color segmentation had a purpose to identify the red object as TB bacteria candidate. This research examined some methods based on clustering concepts for color image segmentation. They were k-means clustering and SOM.

Table 1. The color segmentation results

Before Segmentation	K-Means Clustering	SOM
		
		
		
		

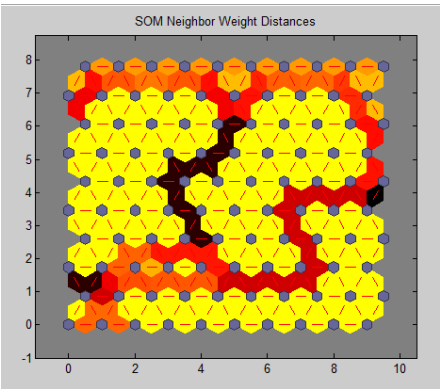


Figure 2. SOM neighbor weight distances

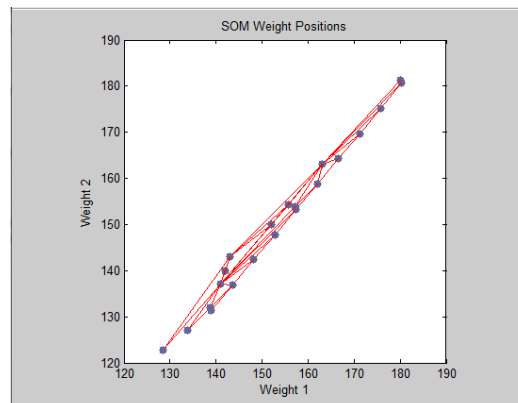


Figure 3. SOM weight positions

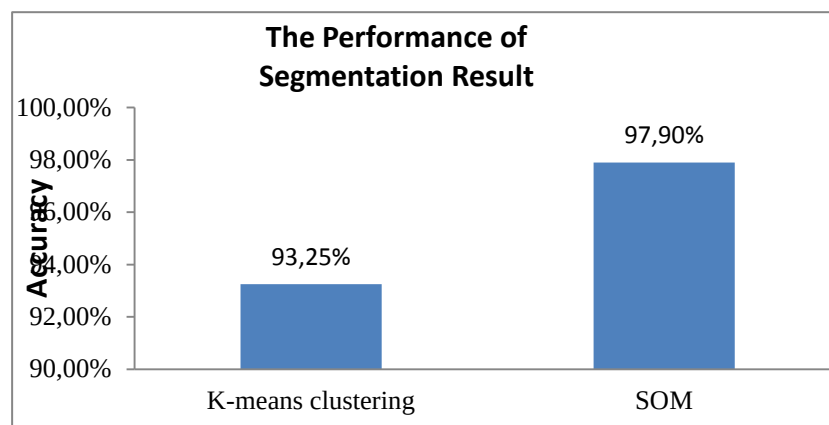


Figure 4. The Accuracy of Segmentation

4. Conclusion

SOM technique could do better segmentation in Ziehl-Neelsen sputum smear images which is very multi variants with the 97.90% of accuracy. The background images could be clear enough and remain TB bacilli candidate objects. Therefore, it could be easy of doing classification for the red objects to decide TB bacilli or not TB bacilli.

SOM technique could support the classification step to give the best accuracy in pulmonary disease diagnose. K-Means clustering still have some lacks in centroid finding. It remains use trial and error to find centroid and the data needs long computational time to achieve convergence. The SOM method could implement in real application for TB identification combining with the classification method.

Acknowledgment

This research was supported by the grant from Directorate General of higher Education Indonesia. We would like to thank to Indonesia Government who already supported and gave the fund for this research. Special thanks went to Bandung Health Department in West Java Province, Indonesia which already supported Ziehl-Neelsen sputum smear slide data.

References

- [1] M K Osman, M Y Mashor, H Jaafar 2011 Combining Thresholding and Clustering Techniques for Mycobacterium Tuberculosis Segmentation in Tissue Sections *Australian Journal of Basic and Applied sciences* 5(12): 1270-1279 **ISSN 1991-8178**
- [2] Khutlang, Rethabile, et al 2010 Classification of Mycobacterium tuberculosis in images of ZN-stained sputum smears *IEEE Transactions on Information Technology in Biomedicine* **14.4**

- 949-957.
- [3] V Makkapati, R Agrawal, R Acharya 2009 Segmentation and Classification of Tuberculosis Bacilli from ZN-stained sputum /smear Images *IEEE Conference on Automation Science and Engineering* Bangalore India.
 - [4] Rulaningtyas R 2011 Automatic Classification of Tuberculosis Bacteria Using Neural Network *IEEE Conference Publications* Pages: 1 - 4 DOI: 10.1109/ICEEI.2011.6021502
 - [5] J Wu 2012 Advances in K-Means Clustering-Data Mining Thinking *Springer Theses, Recognizing Outstanding Ph.D Research* London.
 - [6] Leung KM 2007 *Kohonen Self-Organized Maps* Polytechnic University Department of Computer and Information Science
 - [7] G Guo, C R Dyer 2007 Patch-based Image Correlation with Rapid Filtering Proceeding/CVPR *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* DOI: **10.1109/CVPR.2007.383373**.
 - [8] J Liu, YH Yang 1994 Multi resolution Color Image Segmentation *IEEE Transaction on Pattern Analysis and Machine Intelligence* Vol.**16**, No.7.
 - [9] K A Jain 1989 *Fundamental of Digital Image Processing* Prentice Hall
 - [10] MK Osman, MY Mashor, H Jaafar 2011 Tuberculosis Bacilli Detection in Ziel-Neelsen Stained Tissue Using Affine Moment Invariants and Extreme Learning Machine *IEEE 7th International Colloquium on Signal Processing and its Applications*
 - [11] MK Osman, MY Mashor, H Jaafar 2011 Improving Performance of Hybrid Multilayered Perceptron Network for Classification of Mycobacterium Tuberculosis in Tissue using Modified Recursive Prediction Error-Extreme Learning Machine *International Journal of Research and Reviews in Soft and Intelligent Computing (IJRRSIC)* vol.**1** No.3 ISSN : 2046-6412 Science Academy Publisher United Kingdom.
 - [12] P S Hiremath, P Bannigidad 2009 Automatic Classification of Bacterial Cells in Digital Microscopic Images *International Journal of Engineering and Technology* Vol.2 Num.4 Eashwar Publications.