Performance Analysis of Color Images Using Thresholding Techniques for Image Segmentation in Ziehl- Neelsen Sputum Slide Images

D. Nithiyapriya
Department of Computer Science
Periyar University
Salem
nithiyamphil9@gmail.com

I. Laurence Aroquiaraj
Department of Computer Science
Periyar University
Salem
laurence.raj@gmail.com

***Abstract* - Image segmentation is significant in investigation of clinical images. In this way, viable and accurate techniques are expected to get right diagnosis of quantitative clinical samples. This paper examines the comparison among color thresholding and extensive thresholding methods for Ziehl-Neelsen TB bacilli slide images in sputum samples. Color Thresholding utilizes RGB as input image while Global Thresholding used YCBr as input image. Also we applied Bayesian segmentation to predict the probability of a pixel depicting a 'TB entity' by make use of knowledge of Ziehl-Neelsen stain colours and shape size analysis. This technique is testing on N number of different images. The sensitivity and specificity of all tested classifiers is above 99% for the identification of bacillus objects represented by features. Our results may be used to reduce technician involvement in screening for TB, and would be particularly useful in laboratories in countries with a high burden of TB, where, typically, ZN rather than Auramine staining of sputum smears is the method of choice. The experiment results demonstrated the efficiency of the proposed system.**

1. INTRODUCTION

Biomedical image processing has encountered dramatic extension, and has been an interdisciplinary research field pulling in ability from connected arithmetic, computer sciences, designing, statistics, material science, life sciences and medicine. Sputum slide samples that are smeared with Ziehl-Neelsen method make the acid-fast TB bacilli appears to be red in color while the other cells and organisms retained in blue color as background (Rafikha Aliana A. Raof, Mohd Yusoff Mashor, R.Badlishah Ahmad,). These are important features that could be used to differentiate between TB bacilli and background on Ziehl Neelsen TB slide images.

Many studies have been conducted to detect TB bacilli. Veropoulos et al. (1998) and Forero et al. (2003) proposed pattern recognition technique for TB bacilli identification in images of auramine-stained sputum smears using fluorescence microscopes. Yongping et al. (2010) proposed an autofocus algorithm to automate image capturing for TB identification. Recent work by Osman et al. (2010) used moving k-means clustering method to segment TB bacilli from tissue slide images.

Such as feature extraction and classification. The main objective of segmentation stage is to partition an image into regions or objects for further analysis Sezgin et al (2004) categorized thresholding techniques into six which are histogram-based, clustering based, entropy-based, attribute similarity, spatial and locally adaptive (Rafikha Aliana A. Raof, Mohd Yusoff Mashor, R.Badlishah Ahmad,).

Rafika Aliana raof et al proposed Image segmentation technique for Ziehl-Neelsen sputum slide image is an important factor in extracting tuberculosis (TB) bacilli from the background and sputum cells. Color threshold using Red, Green, Blue (RGB) information has been used widely in many researches. This paper provides new approach to perform thresholding using Hue, Saturation, and Intensity (HSI) information (Rafikha Aliana A. Raof, Mohd Yusoff Mashor, R.Badlishah Ahmad, mohammad, 2010).

2. METHODOLOGY

This session was proposed to compare segmentation between color thresholding technique and global Thresholding technique on Ziehl-Neelsen TB slide images. The image acquisition was done using a light microscope with a 40x digital magnification without oil immersion. Before implementation of main

ISSN: 2349-6363

segmentation process using global and color thresholding methods, the contrast of the images were enhanced using a linear contrast stretching technique and eventually Bayesian theorem has been applied to diagnose the TB. Thresholding method is used to enabling differentiation between TB bacilli and background so that the input images can be easily processed for the next stage which involve with feature extraction and classification. The overview of the proposed approach is shown in figure 1.

2.1. Binarization and Thresholding

A binary image is a digital image that has only two possible values for each pixel. A binary image can be processed well than a gray-scale image. The basic idea in Thresholding is to select a threshold (T) to extract an object or several objects with the same value from background. Color Thresholding method selects the threshold by minimizing the within-class variance of the two groups of pixels separated by the Thresholding operator (Rafikha Aliana A. Raof, Zaleha Salleh, Shahiana Izma Sahidan, H, 2008).
Thresholding is used for separation of light and dark regions in a given image Figure 6. The basic global threshold, T is calculated as in figure 2

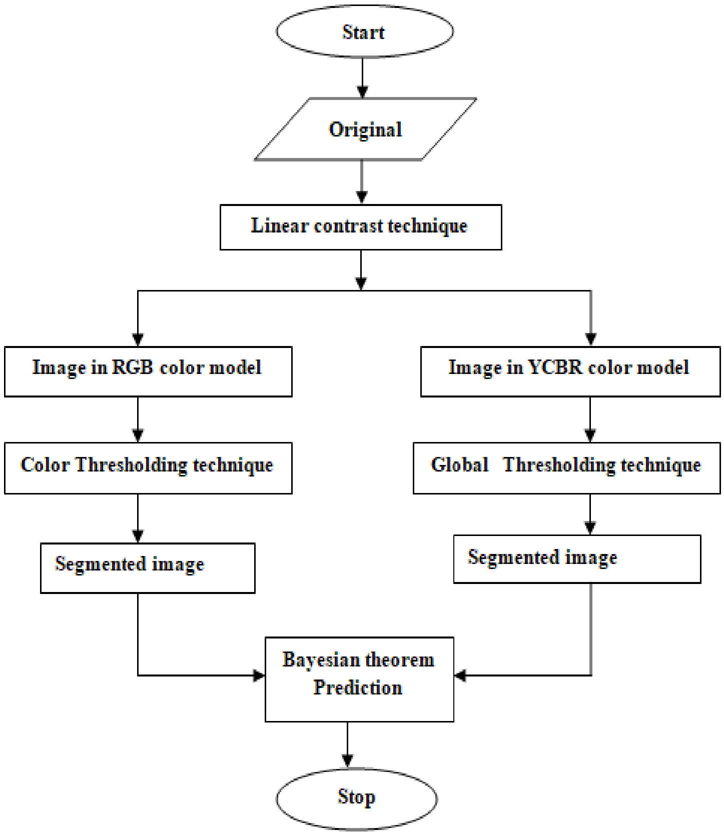

Figure 1. Segmentation process

Step 1. Calculate the average gray level in the image and assign it as T.
Step 2. Separate the image with respect to T to obtain two groups of pixels: G_1 consisting of pixels with grey levels >T and G_2 consisting pixels with grey levels $\leq$ T. 3. Compute the average grey levels of pixels in G_1 to give 1 and G_1 to give 2. 4. Calculate a net threshold value as follows,
Step 3. Compute the average grey levels of pixels in G1 to give μ_1 and G_2 to give μ_2.
Step 4. Calculate a net threshold value as follows,k

$$T = \frac{\mu_1 + \mu_2}{2} \qquad (1)$$

Repeat steps 2-4 until the difference in T in successive iterations is less than a predefined limit.

Figure 2. Algorithm for Global threshold T calculated

This color thresholding method was performed depend on the color information of the image pixels which consists TB, sputum and other objects. RGB color space was detected to be the most opt color paradigm to be implemented with color thresholding technique compared to the other color spaces such as HSI and HSV. The threshold value that was gained from the range of intensities of RGB component was made use of thresholding purpose. If the pixel value where detected not in the intensity range then it will be excluded from the next step process. In equation 1, common algorithm for gray level thresholding was modified as in equation 2. Therefore it can also be applied for color thresholding which have more than one intensity value as in figure 3.

$$g(x,y) = \begin{cases} 0, f(x,y)x < T, \\ 1, f(x,\ \&y) \geq T, \end{cases} \tag{2}$$

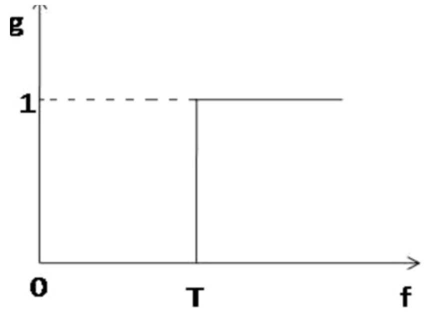

Figure 3. Thresholding mapping of input grey level to output grey level

2.2. Global Thresholding

Global Thresholding is another thresholding technique used for segmentation purpose. In this method YCBr color space which consists three parts was used as input. The three parts in YCBR color space are luminance. Y, R_y, and B_y. These parts can be calculated from RGB as in equation 3.

$$\begin{bmatrix} Y \\ R_y \\ B_y \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.701 & -0.587 & -0.114 \\ -0.299 & -0.587 & 0.886 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3}$$

In Equation 3, R, G, and B refer to respective red, green and blue components of RGB color model. As for hue, θ and saturation S, they can be derived as in equation 4 and 5.

$$\theta = \begin{cases} tan^{-1}\frac{R_y}{B_y}\ for\ S\ \neq 0 \\ underdefined, for\ S = 0 \end{cases}, \tag{4}$$

$$S=\sqrt{{R_y}^2 + {B_y}^2} \tag{5}$$

$$\begin{bmatrix} Y \\ C_y \\ B_y \end{bmatrix} = \begin{bmatrix} 0.319 & 0.677 & 0.124 \\ 0.708 & -0.677 & -0.184 \\ -0.399 & -0.597 & 0.896 \end{bmatrix} \begin{bmatrix} Y \\ C \\ B_r \end{bmatrix} \tag{6}$$

From Equation 6 refers to a pixel of an input image. The pixels which located in the range will be retained as original. Color and the other background pixels will be assigned as white. This resultant image then will be input image for the next step of image processing.

3. THE COLOR AND GLOBAL THRESHOLDING WITH BAYESIAN APPROACH

3.1. Select the Threshold

A brief study on the color information of the mycobacterium and sputum was carried out in order to get the most suitable values to be the selection range of the threshold. The study was carried out on five different types of images, which comprise of normal images, low quality images with low light (under exposed), images which are too bright (over exposed) and blur images.

3.2. Color Thresholding Approach

The color thresholding technique was carried out based on the color information of the bacterium to extract it images from the sputum and other objects. This technique specifies the range of RGB intensities for thresholding. The properties of the RGB pixels are being studied to extract the important features from the image. Based on the color information, the color thresholding algorithm should be able to extract the pixels of mycobacterium and reject pixels of other objects. In order to view the important properties of each segment so that necessary features and accurate value of threshold can be obtained from the result, the information is being gathered in a table.

In this table, among the features that are noted are the maximum, minimum and average values for each of the RGB components in mycobacterium and sputum respectively. The maximum, minimum and average values for each of the pixels are also noted to extract important characteristic of the RGB values that may be converted into threshold values. Summary of the findings from the study can be visualized in Table 1.

Table 1. RGB information in Ziehl-Neelsen TB Bacilli Slide

	Min	**Max**	**Average**
Blue	29	218	133
Red	32	255	158
Green	6	224	141

3.3. RGB Components using thresholding

Considering the minimum and maximum values of RGB components, Equations (7), (8), and (9) have been formulated to get the thresholding algorithm. However, the original values have been modified to cater up to 10% of difference.

$$g(x,y) = \begin{cases} 255, & Red\ (x,y) \leq 26, \\ g(x,y), & 26 < red(x,y) < 239, \\ 255, & red(x,y) \geq 239 \end{cases} \tag{7}$$

$$g(x,y) = \begin{cases} g(x,y), & green\ (x,y) \leq 246, \\ 255, & green(x,y) \geq 246, \end{cases} \tag{8}$$

$$g(x,y) = \begin{cases} 255, & blue\ (x,y) < 121, \\ g(x,y), & blue(x,y) \geq 121, \end{cases} \tag{9}$$

$$g(x,y) = \begin{cases} 0, & green(x,y) < red(x,y), \\ 255, & green(x,y) \geq red(x,). \end{cases} \tag{10}$$

Note that since Equation (10) is the last rule for the whole algorithm, then the final value is either 0 (black pixel) or 255 (white pixel).

Where red (x,y) , green (x,y) and blue (x,y) are the pixel value for each of the red and green components respectively. From Equations (7), (8), and (9), it can be seen that the original equation that has been mentioned in Equations (7) and (8) have been slightly modified to adopt the method of grey level thresholding to color thresholding.

3.4. YCBr Color Conversion

YCBr is an encoded non-linear RGB signal, commonly used by European television studios and for image compression work. YCBr values can be obtained from RGB color space according to equations uses YCBr space for skin detection. YCBr is a color space used as part of the color image pipeline in video and digital photography systems. YCBr is not an absolute color space. It is a way of encoding RGB information. The actual color displayed depends on the actual RGB primaries used to display the signal. It represents color in terms of one luminance component (Y) and two chrominance components (Cb and Cr), where Cb is the chrominance-blue component and Cr is the chrominance –red component. The YCBr image can be converted to/from RGB image. To convert from RGB to YCBr, one variant of this color space (According to ITU-R BT.709): Y = 0.2126 *red+0.7152*green+0.0722*blue. The color model is shown in figure 4

Cb = 0.5389*(blue-Y)
Cr = 0.6350*(red-Y)

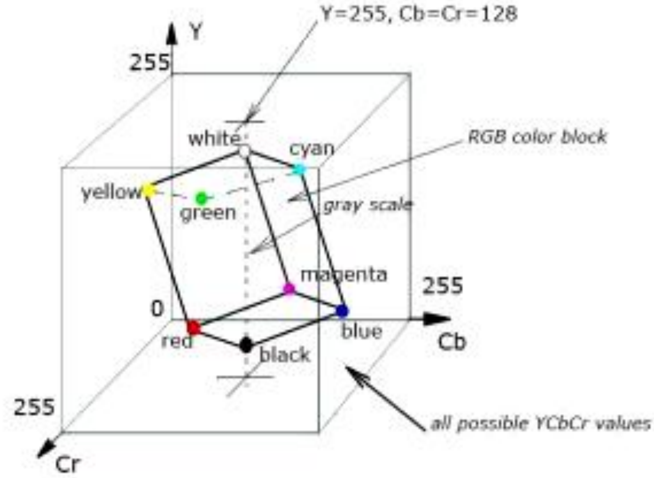

Figure 4. Color Models

3.5. YCBR Color Model

In this format, luminance information is represented by a single component, Y, and color information is stored as two color-difference components, Cb and Cr. Component Cb is the difference between the blue component and a reference value, and component Cr is the difference between the red component and a reference value. The YCBr color model was developed as part of ITU-R BT.601 during the development of a worldwide digital component video standard. YCBr is a scaled and offset version of the YUV color model. Y is the luma component defined to have a nominal 8-bit range of 16 – 235; Cb and Cr are the blue-difference and red-difference chroma components respectively, which are defined to have a nominal range of 16 – 240. The transformation used to convert from RGB to YCBr color space is shown in the equation (11).

$$\begin{bmatrix} Y \\ C \\ B_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 128 & -93.768 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{11}$$

In contrast to RGB, the YCbCr color space is luma independent, resulting in a better performance. The corresponding skin cluster is given as :

Y > 80
85 < Cb < 135
135<Cr < 180, Where Y, Cb, Cr = [0, 255].
MATLAB is one of the greatest image processing tools in the world nowadays. For that reason, we use this tool as our solving equipment. We also use some of the noteworthy methods of MATLAB.

3.6. Bayesian Approach

Combination of a Bayesian approach and a generalized linear mixed model (GLMM) was used to assess spatial heterogeneity in the TB standardized incidence ratio (SIR) and to investigate associations between the three- year average TB incidence rates and the following five variables: number of patients undergoing retreatment (X_{1i}), number of patients with treatment failure and those suffering relapse after the completion of treatment (X_{2i}), number of patients stopping treatment within two months (lost to follow-up) (X_{3i}), number of households with more than one case (X_{4i}), and distance from the patient's residence to the DTC (X_{5i}). All these X_{ki} (with $k = 1, 2, ..., 5$) , calculated for each neighbourhood "i" (i = 1, 2, ..., 192), were obtained from TB registries and incorporated into the national TB control program.

3.6.1. Applying Bayesian Networks Algorithms

In this section, Bayesian Networks algorithms were applied to the Ziehl-Neelsen Bacilli Slide TB dataset. The results using both the Bayesian Networks were presented respectively. There are five steps in this work has illustrated them.

Step 1: Gathering data Step 2: The data were processed before the analysis process; the operations include attribute's selections, data type transformation from one type to another type such as a number to be nominal, process the missed data and so on. Step 3& 4: The third and fourth steps include the actual data mining tasks. Step 5: Learning of the relationship between overall attributes was done Step 6: It is targeted algorithms were implemented in order to construct the prediction model. Step 7: Evaluate and comparing the results based on sensitivity analysis. If the model achieves the acceptable accuracy then the process will terminate otherwise the third step will be repeated.

Figure 5. Bayesian Algorithm

4. EXPERIMENTAL ANALYSIS AND RESULTS

The experiment analysis to perform the comparison between color thresholding and global thresholding, 99 slide images were used. The slide images were captured from a set of positive control TB slide sample for sputum which stained with Ziehl-Neelsen method. The slides were prepared by experience technologists from Department of Microbiology and Parasitology of HUSM, Kelantan. These original images were captured using a Infinity-2 digital camera that was attached to a LEICA DM-LA microscope. It can be seen that TB bacilli appear as red in color while the sputum background as blue color.

The rules that have been formulated are applied to the original raw image of Ziehl-Neelsen sputum slide. Five images (of different types) have been considered for initial studies and other fifty images for testing. For the testing images, the images that are not involved in the previous study have been chosen to test the formulated rules. The original images that have been used for testing, in which one is taken from each category that have been mentioned in table 2.

Accuracy = (TP+TN))/ ((TP+TN+FP+FN)) × 100
TB segmentation rate = TP/ ((TP+FN)) × 100
Background segmentation = TN/((TN+FP)) × 100

Table 2. Confusion Matrix 2*2

	Negative (Predicted)		***Positive (Predicted)***	
Negative (Actual)	True negative	**0**	False positive	**39**
Positive (Actual)	False negative	**0**	True positive	**101**

4.1. Sensitivity

Sensitivity is the measure that is commonly used in two class problems, where one is more interested in a particular class. Sensitivity corresponds to the true positive rate is using the Equation 12.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{12}$$

4.2. Specificity

Specificity is the measure that is commonly used in two class problems, where one is more interested in a particular class. Specificity corresponds to the true negative rate is using the Equation 13.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{13}$$

The accuracy measures the number of correctly segmented TB and background pixels. The TB segmentation rate refers to the sensitivity of segmenting of TB pixels, while the background segmentation rate is the ability of the segmentation in rejecting the background pixels. For these measurements, their values range between zero and one, with higher measurement number indicates better segmentation performance. Table 3 shows segmentation performance of clustering and thresholding algorithms.

Table 3. Segmentation performance of Color thresholding and global thresholding algorithms

Segmentation Method	**Accuracy**	**Sensitivity**	**Specificity**
Color Thresholding	88.9023	0	99.1955
Global Thresholding	99.0438	100	98.933

Color thresholding algorithms are able to segment the TB bacilli well, with an accuracy of up to 88.90%. The performance of the global thresholding is acquiring 99.04% accuracy. So it is found to be better than color thresholding for identification of bacilli in ZN stain sputum images because it gives clear outcome. Global thresholding method is very sensitive to image contrast and output of TB detecting using Bayesian, hence global thresholding is better one. The input image, segmentation results and comparison analysis are shown in the figures 6 to 14.

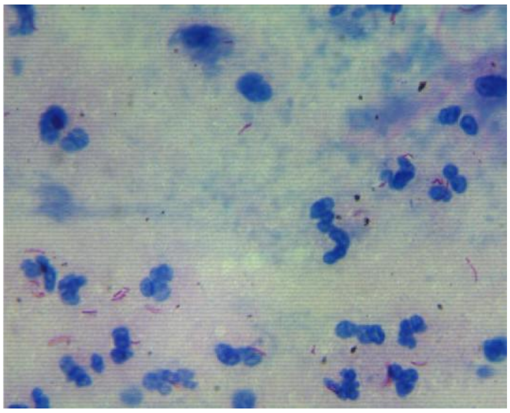

Figure 6. Original Image

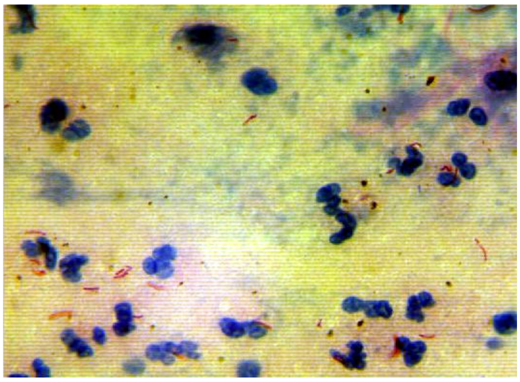

Figure 7. RGB to YCBr Conversion

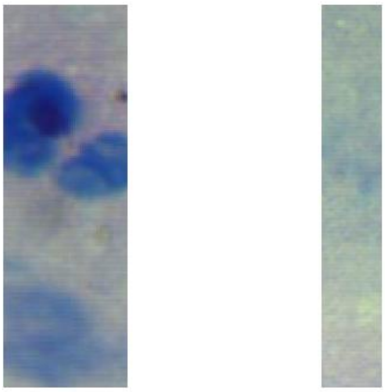

Figure 8. Selection of Positive and Negative Region

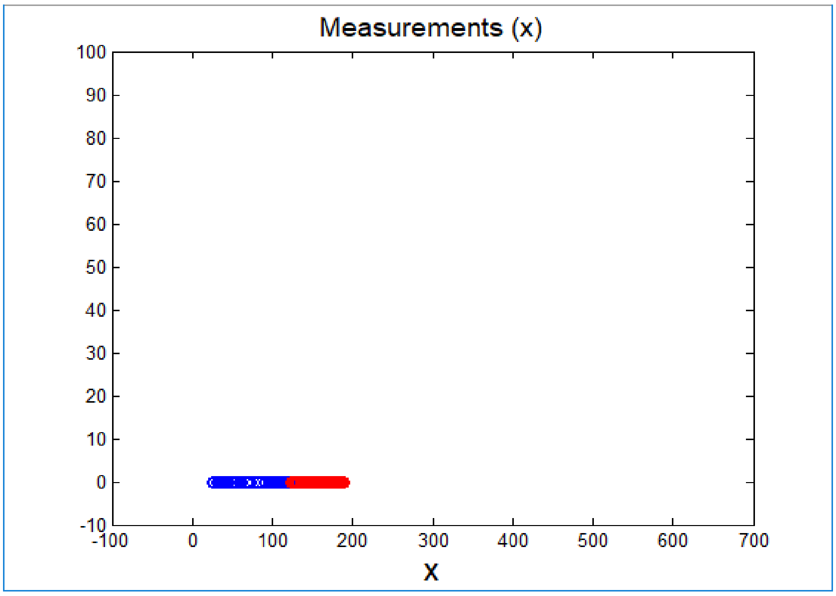

Figure 9. Measurement of Positive and negative region

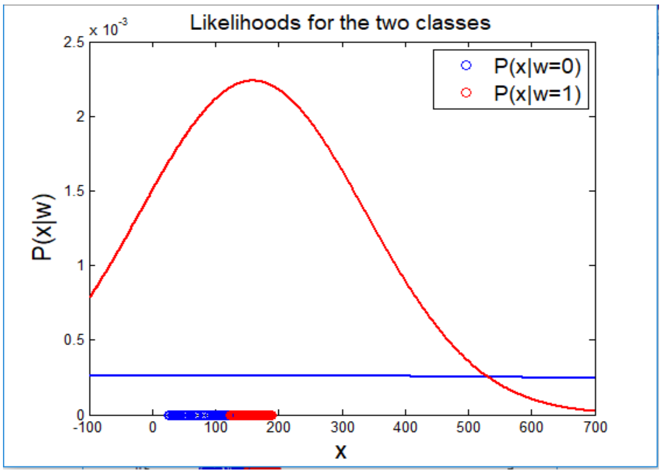

Figure 10. Likelihood for the two classes

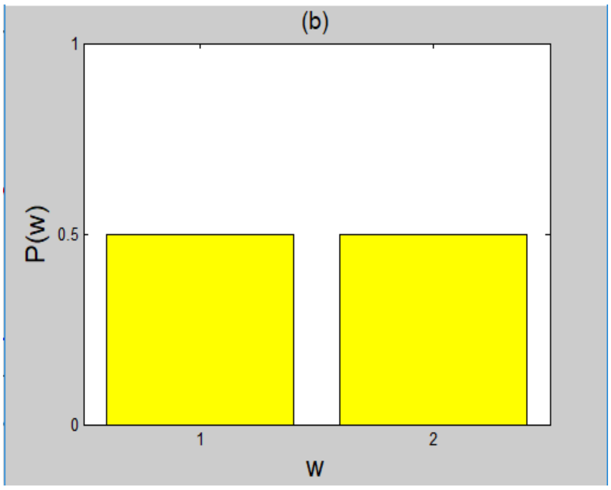

Figure 11. Vector Measurements

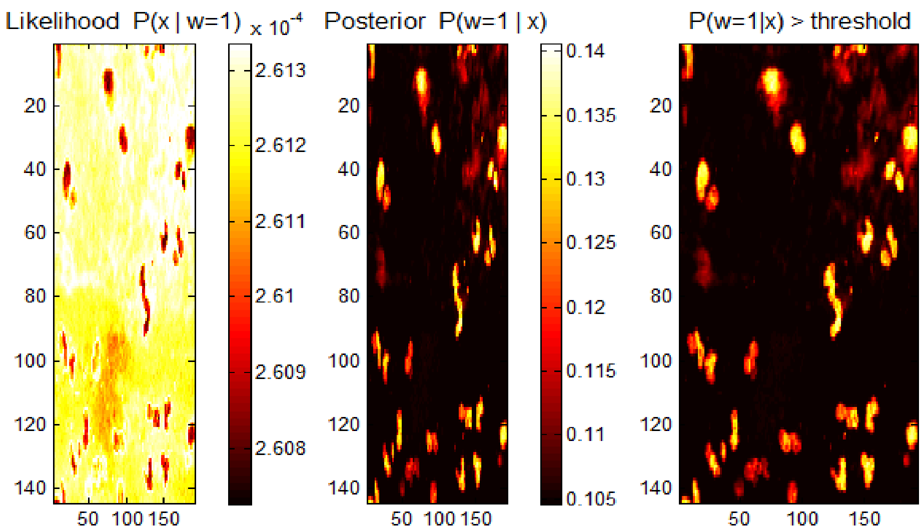

Figure 12. YCBr conversion of likelihood measurements

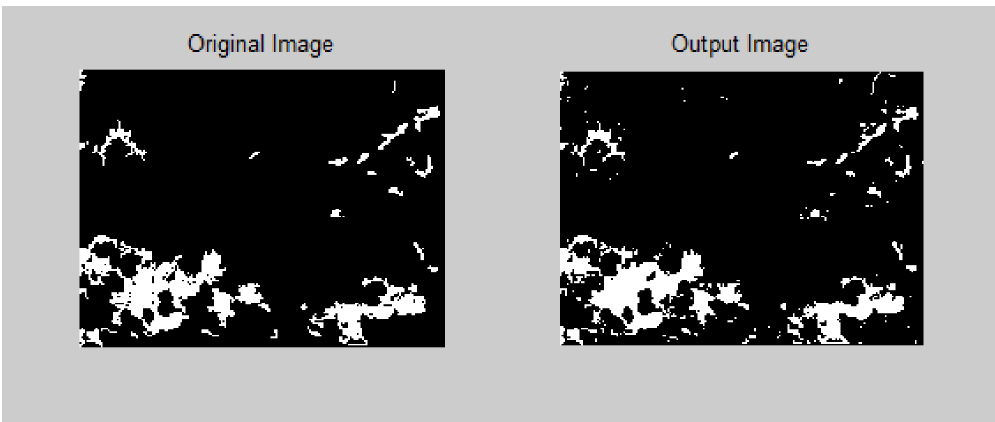

Figure 13. Result image

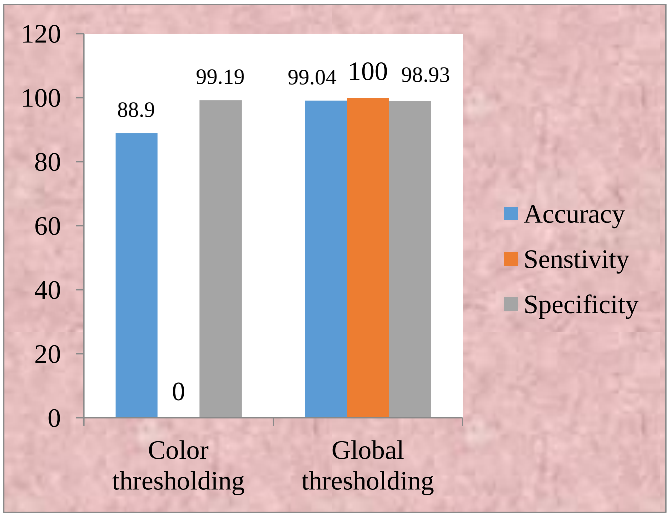

Figure 14. Comparison chart

5. CONCLUSION

In the conclusion, our research can detect Tuberculosis through thresholding methods and Bayesian algorithm and the system has a high degree of accuracy, specificity and better speed in detecting TB bacilli. The method is simple and inexpensive for use in rural/remote areas in the emerging economies. Segmentation algorithm is developed to automate the process of detection of TB using digital microscopic images of different subjects. A performance comparison of color thresholding and global thresholding along with Bayesian algorithms for segmenting TB bacilli in ZN-stained tissue slide images is carried out. These results also suggest that global thresholding is the best method for segmenting the bacilli, as it is highly sensitive to the TB pixels.

REFERENCES

G Tuberculosis Control 2010, World Health Organization, 2010.

M.K. Osman, M.Y. Mashor, and H. Jaafar (2011), Combining Thresholding and ClusteringTechniques for Mycobacterium tuberculosis Segmentation in Tissue Sections, Australian Journal of Basic and Applied Sciences, 5, (12), 1270-1279.

K. Veropoulos, C. Campbell, and G. Learmonth (1998), Image Processing and Neural Computing Used in the Diagnosis of Tuberculosis. Proc. IEE Colloquium on Intelligent Methods in Healthcare and Medical Applications (Digest No. 1998/514), York, UK, 20 8/1-8/4.

M.G. Forero, F. Sroubek, and G. Cristobal (2004), Identification of tuberculosis bacteria based on shape and color, Real-Time Imaging 10, (4), 251–262.

K. Veropoulos, C. Campbell, G. Learmonth, B. Knight, and J.Simpson (1998), The automated identification of tubercle bacilli using image processing and neural computing techniques, in Proc. 8th International Conference on Artificial Neural Network Skovde, Sweden.

M. K. Osman, M. Y. Mashor, Z. Saaad, and H. Jaafar (2010), Colour Image Segmentation of Tuberculosis Bacilli in Ziehl Neelsen Stained Tissue Images using Moving K-Mean Clustering Procedure, in Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation.

Kupper TH, Steffen K, Wekle G, Richartz G, Pfitzer P (1995), Morphologicla study of bacteria of the respiratpry system using fluorescence microscopy of Papanicolaou stained smears with special regard to identification of Mycobacteria in sputum. Cytopath; 6:388-402.

Sadaphal P, Rao J, Comstock GW, Beg M.F (2008). Image processing techniques for identifying Mycobacterium tuberculosis in Ziehl-Neelsen stains. The International Journal of Tuberculosis and Lung Disease: 12(5): 579-82. 42.

Rafikha Aliana A. Raof, Mohd Yusoff Mashor, R.Badlishah Ahmad, Siti Suraiya Md Noor, Image Segmentation of Ziehl-Neelsen Sputum Slide Images for Tubercle Bacilli Detection international conference on computer science engineering.

Rafikha Aliana A. Raof, Mohd Yusoff Mashor, R.Badlishah Ahmad, mohammad (2010), Comparison of colour thresholding method using rgb and Hsi information for ziehl-neelsen sputum slide images, 10th International Conference on Information Sciences, Signal Processing and their Applications, ISSPA 2010, Kuala Lumpur, Malaysia, 10-13.

Rafikha Aliana A. Raof, Zaleha Salleh, Shahiana Izma Sahidan, H (2008). Hasan Color Thresholding Method For Image Segmentation Algorithm Of Ziehl-Neelsen Sputum Slide Images, Conference: Electrical Engineering, Computing Science and Automatic Control, 2008. 5th International Conference on CCE 2008.

Forero-Vargas, M., Sroubek, F., Alvarez-Borrego, J., Malpica, N., Cristóbal, G., Alcalá,L., Alcala, L., Desco, M., and Cohen, L. (2002), Segmentation, autofocusing and signature extraction of tuberculosis sputum images. In Proc. of SPIE Photonic Devices and Algorithms for Computing IV. 4788:171–182.