Automatic Detection of Malaria Parasites for Estimating Parasitemia

S. S. Savkare *swati_savkare@yahoo.com*
Moze College of Engineering,
University of Pune,
Pune, India

S. P. Narote *snarote@rediffmail.com*
Sinhgad College of Engineering,
University of Pune,
Pune, India

Abstract

Malaria parasitemia is a measurement of the amount of Malaria parasites in the patient's blood and an indicator for the degree of infection. In this paper an automatic technique is proposed for Malaria parasites detection from blood images by extracting red blood cells (RBCs) from blood image and classifying as normal or parasite infected. Manual counting of parasitemia is tedious and time consuming and need experts. Proposed automatic approach is used Otsu thresholding on gray image and green channel of the blood image for cell segmentation, watershed transform is used for separation of touching cells, color and statistical features are extracted from segmented cells and SVM binary classifier is used for classification of normal and parasite infected cells.

Keywords: OTSU Thrsholding, Watershed Transform, Feature Extraction, SVM Classifier.

1. INTRODUCTION

Malaria is a serious disease caused by a blood parasite named Plasmodium spp. It affects at least 200 to 300 million people every year and causes an estimated 3 million deaths per annum. Diagnosis and medication of it is necessary [1], [2]. In blood sample visual detection and recognition of Plasmodium spp is possible and efficient via a chemical process called (Giemsa) staining [4]. The staining process slightly colorizes the RBCs but highlights Plasmodium spp parasites, white blood cells (WBC), and artifacts. Giemsa stains nuclei, chromatin in blue tone and RBCs in pink color. It has been shown in several field studies that manual microscopy is not a reliable screening method when performed by non-experts. Malaria parasites host in RBCs when it enter in blood stream. In Malaria parasitemia count it is important step to segment RBCs from blood image and classify it as parasite infected or normal. In thin blood images morphology of cells can be observed clearly. The present paper describes the techniques used in segmenting normal and infected RBCs for purpose of Malaria parasitemia (number of infected blood cells over total red blood cell) count.

This paper is organized as follows: Section 2 summarizes literature related to segmentation of cells and count Malaria parasitemia. Section 3 illustrates the system architecture which includes pre-processing, cell segmentation, RBCs segmentation, feature extraction and classification. Section 4 and 5 include results and conclusion of this paper.

2. RELATED WORK

Minh-Tam Le et. al. [3], proposed a comparison-based analysis, which differentiates solid components in blood smears. The semiautomatic method uses statistical measures and cross-referencing validations yields a reliable detection scheme. The nucleated components are identified using adaptable spectral information. Cells and parasites are isolated from the background, by comparing the input image with an image of an empty field of view. The range of erythrocyte sizes is determined by input of isolated RBC.

Jesus Angulo et. al. [4], presents a technique to automatically detect the working area of peripheral blood smears stained with Giemsa. The approach consists of two stages. First, an image analysis procedure using mathematical morphology is applied for extracting the erythrocytes, the centers of erythrocytes and the erythrocytes with center. Second, the number of connected components from the three kinds of particles is counted.

D. Ruberto et. al. [5] follow morphological method for detection of parasites in Giemsa stained blood slides. Different objects in blood are identified using their dimensions and color. The parasites are detected by means of an automatic thresholding based on morphological approach, using Granulometrices to evaluate size of RBCs and nuclei of parasite. A segmentation method using morphological operators combined with the watershed algorithm.

Silvia et. al. [6], proposed a technique for estimating parasitemia. Template matching is used for detection of RBCs. Parasites are detected using variance-based technique from grayscale images and second approach is based on color co-occurrence matrix. Support Vector Machine (SVM) as the classifier which exploits the texture, geometry and statistical features of the image.

Stanislaw Osowski et. al. [7], presents the application of a genetic algorithm (GA) and a support vector machine (SVM) to the recognition of blood cells on the image of the bone marrow aspirate. GA is used for the selection of the features for the recognition of the neighboring blood cells belonging to the same development line. The SVM is used for final recognition and classification of cells.

3. SYSTEM ARCHITECTUR

System architecture used for Malaria parasite detection involves following steps: Image Acquisition, Pre-processing, cell segmentation, Feature Extraction, and Classification. Block diagram of system architecture is shown in Figure1.

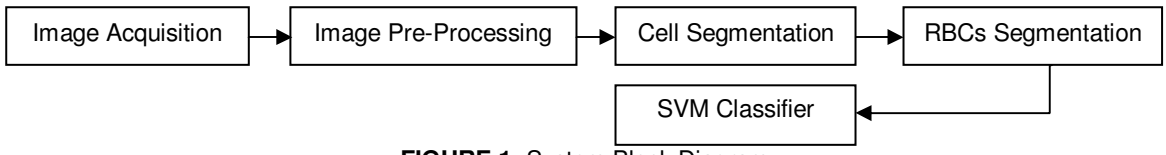


FIGURE 1: System Block Diagram.

3.1 Image Acquisition

For slide preparation working solutions of Giemsa were made by adding 100 µl stock solution to each milliliter of distilled water. Dried thin blood films were fixed with methanol for 30 s, poured off and stained with Giemsa for 20 min [4], [8]. The stain was rinsed off with tap water for 10 s. Upon drying, slides were used immediately or stored for future use. Image was captured by connecting high resolution Digital camera to microscope. By adjusting microscope magnification image is captured.

3.2 Image Pre-Processing

Pre-processing step includes noise reduction, smoothening of image. In this paper we used median filter for smoothening of color image and Lapalcian filter is used for edge sharpening. This result is subtracted from original to enhance the image. The median filter [8] is a non-linear digital filtering technique, used to remove noise from images. In median filtering pixel replaces with the median of its neighboring pixel values. Lapalcian filter takes second order derivative of pixel. After pre-processing image is send to cell segmentation block to segment cells.

3.3 Cell Segmentation

To segment foreground from background Global threshold and Otsu threshold [10] is used on grayscale enhanced image. For low contrast image segmentation applied on enhanced green channel of the image. Result of thresholding on both images is added to get binary image of cells. A 3 x 3 median filter was applied on this binary cell mask to fill the holes in blood cells and to remove the unwanted points from binary image of cells and background [11]. Using morphological operation cells having larger area is identified which is overlapping of the cells.

Distance transform is applied on it followed by watershed transform [5]. This gives separation of overlapping cells. This final binary image of cells is given to next block.

3.4 RBCs Segmentation

First rule check for White blood cells which are bigger than the RBCs, and second check for platelets which are smaller than RBCs. Using morphological operation platelets are removed from binary image. By labeling this binary image total number of cells is calculated.

3.5 Feature Extraction

Since the chosen features affect the classifier performance, selection of feature which is to be used in a specific data classification problem is as important as the classifier itself [12]. The features which give predominant difference between normal and infected cells are identified and used for training purpose. The selected features are geometrical, color and statistical based. The mathematical morphology provides an approach to the processing of image based on shape. The set of parameters corresponds to the geometrical features are as follows:

Radius -measured by averaging the length of the radial line. *Perimeter* - the total distance between consecutive points of the border, *Area* - the number of pixels on the interior of the cell. *Compactness* - is the ratio of perimeter^2 by area, *Metric* – (Perimeter)^2/4π·Area which is 1 for circle.

The values of saturation histogram is used for classification it is spread for infected cell and lye towards left if normal cell. Histogram of green plane of normal cell is spread and for infected cell it lies towards right [7].

$$Skewness = \frac{1}{\sigma^2}\sum_{b=0}^{L-1}(b-\bar{b})^3$$

$$Kurtosis = \frac{1}{\sigma\ b^4}\sum_{b=0}^{L-1}(b-\bar{b})^4 P(b) - 3$$

$$Energy = \sum_{b=0}^{L-1}[p(b)]^2$$

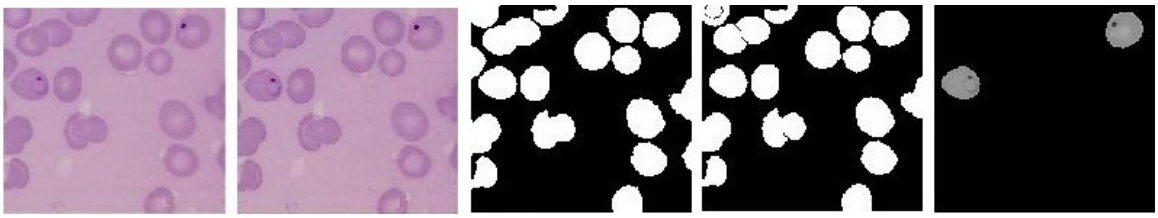
$$StndardDev\quad iation = \left[\sum_{b=0}^{l-1}\left(b-\bar{b}\right)\right]^{1/2}$$

P(b) is the first-order histogram estimate, Parameter *b* is the pixel amplitude value. *L* is the upper limit of the quantized amplitude level. The above parameters are used for feature extraction. The statistical features use gray level histogram and saturation histogram of the pixels in the image and based on such analysis, the mean value; angular second momentum, Skewness, Standard deviation, Kurtosis are treated as the features [14] and calculated using above equations.

3.6 SVM Classifier

The SVM is a powerful solution to the classification problems. In this paper, it has been used for the recognition and classification of cells. The main advantage of the SVM network used as a classifier is its very good generalization ability and extremely powerful learning procedure, leading to the global minimum of the defined error function. Linear SVM is a linear discriminant classifier working on the principle of maximum margin between two classes. The decision function of the N-dimensional input vector x for K-dimensional feature space (K>N) is defined as $D(x) = w^T \varphi(x) + b$ through the use of function $\varphi(x)$. Where $\varphi(x) = [\varphi 1(x), \varphi 2(x), \ldots, \varphi K(x)]$, w as the weight vector of network $w=[w_1, w_2, \ldots, w_k]^T$, and b as the bias weight [12]. All values of weights have been arranged in decreasing order and only the most important have been selected for each pair of classes and then used in the final classification system.

The learning of the SVM network working in the classification mode is aimed at the maximization of the separation margin between two classes. Simple classification algorithm is proposed that classifies points by assigning them to the closer of two parallel planes (in input or feature space). Standard support vector machines (SVMs), which assign points to one of two half spaces. SVM classifier is used for classification of normal and infected cells. Results pre-processing, Otsu's threshold to get binary image of cells, separation of overlapping cells and finally detection of infected cells is shown in Figure 2.



(a) (b) (c) (d) (e)

FIGURE 2: a) Original Image, b) Pre-processed Image, c) Binary Image of Cells, d) Separation of Overlapping cells, e) Detected parasite infected cells

4. RESULT

The described methods of feature extraction produce a very rich group of parameters. Skewness of healthy cells is up to 2 and for infected cell it is above 2. Kurtosis of normal cell is below 3 and for infected it is up to 9. Standard deviation of infected cell is very high as compare to normal cell. Thus all extracted features are sends to next block for classification. The binary classifier using RBF kernel is used for classification.

Image	Manual Parasitemia	Automatic Parasitemia
1	25.00	25.00
2	13.33	6.67
3	11.11	11.11
4	12.50	12.50
5	6.67	7.14
6	16.67	25.00
7	3.03	3.03
8	4.76	4.76
9	18.18	18.18
10	2.78	2.78
11	0.00	0.00
12	4.00	4.00
13	20.00	20.00
14	10.00	18.18
15	2.94	2.94

TABLE 1: Summary of Manual and Automatic Parasitemia.

The cost parameter C and Lagrange multiplier λ are taken 1000 and 10^{-7} respectively. Image processed through automatic system segments RBCs from input image, separate overlapping cells, counts total number of erythrocytes and SVM binary classifier detect infected cells. Finally system gives number of normal cells and infected cell, and percentage parasitemia in command window. 15 images processed through automatic system. Table 1 summarizes result of manual and automatic parasitemia for 15 images. Figure 3 shows graphical comparison of manual and automatic parasitemia count.

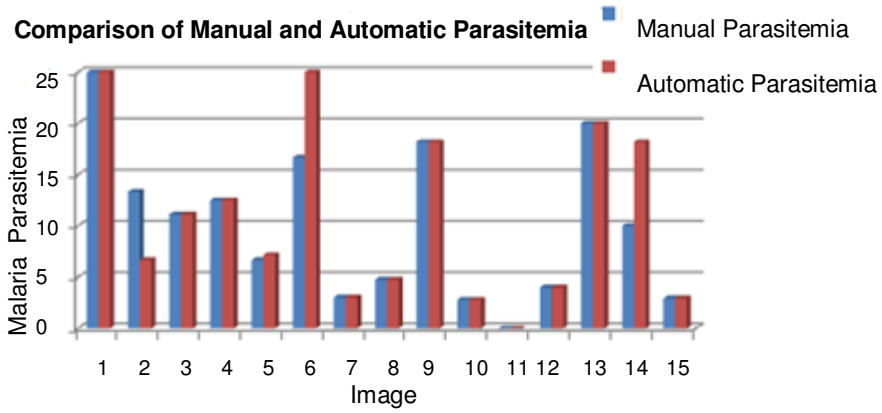


FIGURE 3: Graphical comparision of Manual and Automatic Parasitemia.

5. CONCLUSION

The proposed automated method of segmentation and classification of cell is simple. An approach is proposed to detect red blood cells with consecutive classification into parasite infected and normal cells for estimation of parasitemia. The extraction of red blood cells achieves a reliable performance and the actual classification of infected cells. Sensitivity of system is 93.12%, and Specificity is 93.17%.

Shape based and statistical features are generated for classification. The features are selected for recognition of two classes only. This approach leads to the high specialization of each classifier and results in an overall increase in accuracy. The above algorithms are implemented using MATLAB.

6. REFERENCES

[1] Shiff, C., 2002. Integrated approach for malaria control. Clin. Microbiol. Rev.15, 278–293.

[2] World Health Organization What is malaria? Factssheetno94. http://www.who.int/mediacentrefactsheetsfs094/en/./factshee. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp 68–73.

[3] M. Tam Le, T. Bretschneider,"*A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears*", Research article, BMC Cell Biology, 28 March 2008.

[4] J. Angulo, G. Flandrin, "*Automated detection of working area of peripheral blood smears using mathematical morphology*", U. S. National Library of Medicine, Analytical Cellular Pathology 25(1), pp 39-47, 2003.

[5] C.D. Ruberto, A.G. Dempster, S. Khan, and B. Jarra, "*Automatic Thresholding of Infected Blood Images Using Granulometry and Regional Extrema*", in Proceedings of International Conference on Pattern Recognition, pp 3445-3448, 2000.

[6] S. Halim et al., "*Estimating Malaria Parasitaemia from Blood Smear Images*", in Proceedings of IEEE international conference on control, automation, robotics and vision, pp 1-6, 2006.

[7] S. Osowski et al., "*Application of Support Vector Machine and Genetic Algorithm for Improved Blood Cell Recognition*", in proceedings of IEEE transaction on Instrumentation and Measurement, Vol. 58, No. 7, pp 2159-2168, July 2009.

[8] S.W.S. Sio et al., "*Malaria Count: An image analysis-based program for the accurate determination of parasitemia*", Journal of Microbiological Methods 68, Science Direct, pp 11-18, 2007.

[9] D. Anoraganingrum et al.' "*Cell Segmentation with Median Filter and Mathematical Morphology Operation*", in proceedings of on Image Analysis and Processing, Italy, pp 1043-1046, 1999.

[10] N. Otsu, "A threshold selection method from gray-level histograms", in proceedings of IEEE Transactions on Systems, Man and Cybernetics, 9(1), pp 62-66, 1979.

[11] K. Kim et al., "Automatic Cell Classification in Human's Peripheral Blood Images Based on Morphological Image Processing", Lecture Notes in Computer Science, vol. 2256. pp 225-236, 2001.

[12] T. Markiewicz, S. Osowski, "*Data mining techniques for feature selection in blood cell recognition*", European Symposium on Artificial Neural Networks, Bruges (Belgium), 26-28 April, pp 407-412, 2006

[13] N. Ritter, J. Cooper, "*Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides*", in Proceedings of Thirtieth Australasian Computer Science Conference (ACSC2007), CRPIT, 62, 161-169, 2007.

[14] G. Diaz et al., "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images", Journal of Biomedical Informatics 42, Science Direct, pp 296–307, 2009.