Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 6 (2012) 405 – 410

Procedia Technology

2nd International Conference on Communication, Computing & Security [ICCCS-2012]

Automatic System for Classification of Erythrocytes Infected with Malaria and Identification of Parasite's Life Stage

S.S.Savkare[a]*, S.P.Narote[b]

[a]*JSPM- NTC, RSSOER, University of Pune, Pune, India*
[b]*Sinhgad College of Engineering, University of Pune, Pune, India*

Abstract

Malaria is a serious disease caused by a blood parasite named Plasmodium spp. The World Health Organization estimates 300-500 million malaria cases and more than 1 million deaths per year (Tek F. B. et al., 2006). Manual counting and classifications of infected erythrocytes is a time-consuming and laborious process (Selena W.S. Sio et al., 2007).The aim of our study to develop a fully automatic system for counting and classification of Malaria parasite infected erythrocytes and detection of life stage of parasites. The system uses Giemsa stained thin blood images for processing; using Otsu's threshold erythrocytes are segmented form pre-processed images; watershed algorithm is used to separate overlapped cells. Statistical and colour features are extracted and given to the SVM binary classifier which classifies Malaria infected erythrocytes and SVM multi classifier is used for detection of parasite life stages.

Keyword: OTSU's threshold; Watershed Transform; Feature Extraction;SVM classifier.

1. Introduction

Malaria is a leading cause of morbidity and mortality in tropical and sub-tropical countries, with an estimated of 1–2 million deaths per year (Gloria Diaz et al., 2009). In blood sample visual detection and recognition of Plasmodium spp is possible and efficient via a chemical process called Giemsa staining (Silvia Halim et al., 2006). The staining process slightly colorizes the erythrocytes but highlights Plasmodium spp

* Corresponding author. Tel.: +919822237125; +919422085722.
E-mail address: swati_savkare@yahoo.com

doi:10.1016/j.protcy.2012.10.048

parasites, white blood cells (WBC), and artefacts. It has been shown in several field studies that manual microscopy is not a reliable screening method when performed by non-experts and also a time consuming. In terms of the evaluation of antimalarial therapies it is also important an estimation of the life parasite stage per infected erythrocyte. This paper discusses a fully automatic system for detection of infected erythrocytes from blood images and parasite life stage identification. Statistical, Textural and colour features are extracted for training of SVM classifier. System is supervised. Total 71 images are processed and result is analyzed for Linear kernel, Polynomial kernel and RBF kernel.

The paper is organized as follows section 2 summarizes literature related to detection of Malaria parasite infected cells, classification of blood cells. Section 3 describes system architecture which includes pre-processing, segmentation, feature extraction and SVM classifier Section 4 and 5 includes results and conclusions of this paper.

2. Literature Survey

A technique was proposed (Silvia Halim et al., 2006) for estimating parasitemia from blood smear images by extracting healthy and parasite infected red blood cells. Illumination correction is performed by taking a background image using a blank slide or portion of a slide. Based on pattern matching with parameter optimization and cross-validation against the expected biological characteristics, red blood cells are determined. Using unsupervised and color co-occurrence based technique red blood cells are classified. Red blood cells detection resulted in precision and recall rates of 80–88% and 92–98%, respectively.

(Gloria Diaz et al., 2009) proposed visual quantification of parasitemia in stained thin blood films infected with Plasmodium Falciparum. Images captured with proper adjustment of microscope with 1000× magnification. Luminance correction process applied to RGB image using YCbCr transform Low Pass Filter and then smoothening of matrix done. Using color space, K-NN, and normalized RGB pixels are classified as erythrocyte or background. Using color histogram, Saturation Histogram, Grayscale, Tamura Texture, and Sobel Histogram used for feature extraction. Only infected cells are forwarded for infection stage classification using trained SVM classifier. Automatic identification of infected erythrocytes showed a specificity of 99.7% and a sensitivity of 94%. The infection stage was determined with an average sensitivity of 78.8% and average specificity of 91.2%.

A method is proposed (Minh-Tam Le et al., 2008) for estimation of malaria parasitemia in thin blood image using comparison based analysis. Acquired image smoothed by energy normalized averaging filter and then analyzed for nucleated component using properties and location. Zack's thresholding algorithm, Skewness and moment used for nuclei segmentation. Image is decomposed by segmentation using bimodal histogram and Otsu algorithm. Erythrocytes are detected using granulometry and prior knowledge of size. Locating local maxima of Euclidean transform of individual erythrocyte in cluster its potential position is determined. Overlapping binary mask of cells and parasites infected cells are identified for parasitemia.

The application of a genetic algorithm (GA) and a support vector machine (SVM) is presented (Stanislaw Osowski et al., 2009) 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. SVM is used for final recognition and classification of cells.

3. System Architecture

Block diagram of an automatic system for classification of malaria infected blood cells and detection of life stage of parasite is shown in Fig.. 1; which involves following steps: image acquisition, pre-processing, erythrocyte segmentation, feature extraction, and classification.

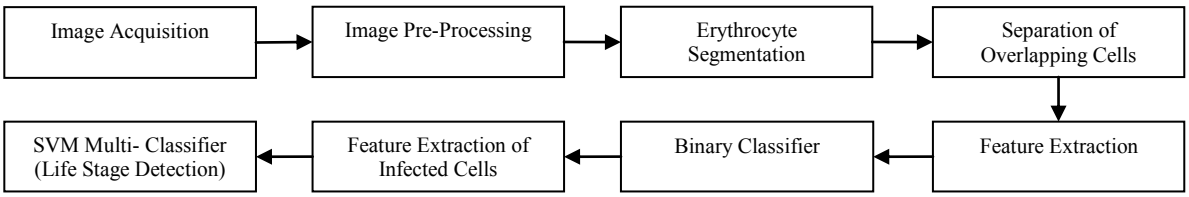


Fig.1. Block diagram of an automatic system for classification of malaria infected cells and life stage detection of parasites.

3.1. Image Acquisition

The slides are prepared using working solutions of Giemsa (Gloria Diaz et al., 2009). Image was captured by connecting high resolution digital camera to microscope. Images used for processing is having different magnification of microscope, light source, and staining effects.

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 (Minh-Tam Le et al, 2008) 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 erythrocytes segmentation block (S.S. Savkare and S.P. Narote, 2011).

3.3. Erythrocyte Segmentation

As Malaria parasite lies in erythrocytes so first step is to segment erythrocytes from blood image. To segment foreground from background Global threshold and Otsu threshold is used on grayscale enhanced image (N. Otsu, 1979). 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. Average area of cells is calculated which helps to remove small artifact from image. Unwanted pixels are removed by using Median Filter. For separation of overlapping cells distance transform is applied separately on each cluster of cells. After that Watershed transform is applied on to get separation of overlapping cells (S.S. Savkare and S.P. Narote, 2011). This final binary image of cells is given to next block.

3.4. 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 (T. Markiewicz and S. Osowski, 2006). 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: *Area* - the number of pixels on the interior of the cell. *Compactness* - is the ratio of (perimeter)^2 by area, *Metric* – is (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 (S.S. Savkare and S.P. Narote, 2011).

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

$$\text{Standard Deviation} = \left[\sum_{b=0}^{l-1}\left(b-\bar{b}\right)\right]^{1/2} \tag{2}$$

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 third momentum, skewness, standard deviation are treated as the features and calculated using above equations.

3.5. SVM Classifier

The SVM is a powerful solution used for classification problems. 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 discriminate 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 (Marti A. Hearst, 1998). 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 (S.S. Savkare and S.P. Narote, 2011).

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 binary classifier is used for classification of normal and Malaria infected erythrocytes.

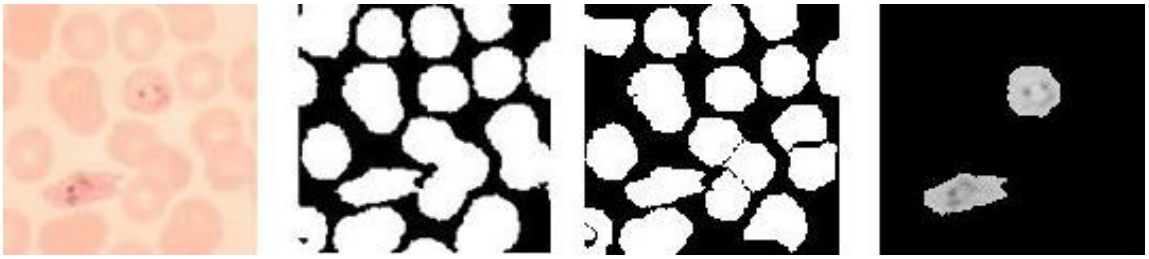
3.6. Feature Extraction of Infected Cells

Malaria parasites show three life stages in bloodstream: Ring stage, Schizont Stage and Gametocyte Stage; Gametocyte is sexual stage. For detection of life stage of parasites color, geometrical and statistical features are extracted from infected erythrocytes. Using saturation plane parasites are segmented from image and by overlapping marker image of parasite with infected erythrocyte total area occupied by parasite is calculated. Parasites stains in dark blue or red color, number of stained pixels are counted. In schizont stage more number

of merozoits is present which are segmented from saturation plane and counted. By using morphology in different life stages rich group of feature set is extracted and SVM three- class classifier is trained.

4. Results

SVM binary classifier is used to classify erythrocytes as infected or non-infected. Total 71 images are processed through an automatic system. Area of infected erythrocyte is larger than normal erythrocyte; standard deviation is very high as compare to normal erythrocyte, skewness of healthy cells is up to 2 and for infected cell it is above 2. In Table specificity, sensitivity for Linear, Polynomial and RBF kernel is given. For SVM multi classifier the described methods of feature extraction produce a very rich group of parameters. Area occupied by ring stage parasite in erythrocyte is 20-50%, by schizont stage is 40-80% and in gametocyte stage whole erythrocyte can be occupied by parasite. Fig. 2 shows output of segmentation of erythrocytes, separation of overlapping cells, and detection of infected cells. In command window total number of erythrocytes, total number of infected cells and life stage of parasite will get displayed. For given image system shows ring stage of parasite.



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

Fig.. 2. (a) Original Image, (b) Segmentation of Erythrocytes From Image, (c) Separation of Overlapping Cells, (d) Detection of Malaria Parasite Infected Cells.

Table 1 shows sensitivity and specificity of SVM binary classifier for detection of infected erythrocytes; and also summarizes correct identification of life stage of parasite for Linear, Polynomial and RBF kernel. Results are summarized in Table 1.

Table 1. Summary of Results of SVM Binary Classifier and Multi-Classifier for 71 Images.

Kernel	SVM Binary Classifier		SVM Multi-Classifier Correct Detection Rate (%)
	Sensitivity (%)	Specificity (%)	
Linear	94.85	89.96	93.87
Polynomial	96.26	99.09	90
RBF	96.26	99.09	96.42

5. Conclusions

The proposed automatic system for classification of infected erythrocytes and life stage identification of malaria parasite uses Otsu's threshold for segmentation of cells from Giemsa stained blood Images. After segmentation, watershed transform is used for separation of overlapping cells. Geometrical and statistical features are extracted from each cell and given to SVM binary classifier to classify erythrocytes as infected or normal. SVM binary classifier gives 96.26% sensitivity and 99.09% specificity. Color and geometrical features are extracted from infected erythrocytes and given to SVM multiclassifier which identifies life stage of Malaria parasite. SVM kernels such as linear, polynomial and RBF are used on 71 images. SVM multiclass classifier (i.e. RBF kernel) gives 96.42% accuracy for correct identification rate of life stage of parasite. Above algorithms are implemented using MATLAB.

References

Tek FB, Dempster AG, Kale I., 2006, Malaria Parasite Detection in Peripheral Blood Images, in Proceedings of British Machine Vision Conference, Edinburgh.

Selena W.S. Sio, Weiling Sun, Saravana Kumar, Wong Zeng Bin, 2007, Malaria Count: An image analysis-based program for the accurate determination of parasitemia, Journal of Microbiological Methods 68, Science Direct, pp 11-18.

Gloria Diaz, Fabio A. Gonzalenz, Edurdo Romero, 2009, 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.

Silvia Halim, Timo R. B., Yikun Li, 2006, Estimating Malaria Parasitaemia from Blood Smear Images, in Proceedings of IEEE international conference on control, automation, robotics and vision, pp 1-6.

Minh-Tam Le, Timo R Bretscheneider, 2008, A Novel Semi-Automatic ImageProcessing Plasmodium Falciparum Parasitemia in Approach to Determine Giemsa-stained Thin Blood Smears, Journal BMC cell biology, 9:15.

Stanislaw Osowski, Robert Siroic, Tomasz Markiwicz, Krzysztof Siwek, July 2009, 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.

World Health Organization What is malaria? Facts sheet no 94. http://www.who.int/mediacentre/factsheets/fs094/en/

DPDx Laboratory Identification of Parasite of Public Health Concern, Image Library and Diagnostic Procedure, http://www.dpd.cdc.gov/dpdx/

S. S. Savkare, S. P. Narote, July 2011, Automatic Detection of Malaria Parasites for Estimating Parasitemia, International Journal of Computer Science and Security (IJCSS), Vol. 5, No. 3,pp 409-414.

S. S. Savkare, S. P. Narote, Feb 2011, Automatic Classification of Normal and Infected Blood Cells for Parasitemia Detection, International Journal of Computer Science and Network Security (IJCSNS), vol. 11, No. 2, pp 94-97.

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

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

Marti A. Hearst, July-Aug1998, Trends & Controversies Support vector machines, IEEE Intelligent System, pp 18-28.