Automatic Classification of Normal and Infected Blood Cells for Parasitemia Detection

S. S. Savkare[†], *S. P. Narote*[††]

[†]*Moze College of Engineering, University of Pune, Pune, India*
[††]*Sinhgad College of Engineering, University of Pune, Pune, India*

Abstract: Malaria, Thalassaema and Babesia are serious global health problem and rapid, precise diagnosis and determination of parasitemia is necessary for accurate medication. Visual quantification of parasitemia in thin blood films is a very tedious, subjective and time-consuming task. Manual counting by light microscopy is the most widely used technique for parasitemia determination but it is a time-consuming and laborious process and requires expertise. This work resents an automatic method for quantification and classification of erythrocytes in Giemsa stained thin blood films infected with Plasmodium Falciparum or Protozoan Parasite. The features are extracted using statistical parameters and SVM classifier used for classification of normal or infected blood cells.

Key words:
Erythrocytes, thrsholding, watershed transform, segmentation, feature extraction, SVM classifier.

1. Introduction

For both animals and humans, slides of stained peripheral blood smears are examined to aid diagnosis. Diagnosis using a microscope requires special training and considerable expertise [1], [2]. It has been shown in several field studies that manual microscopy is not a reliable screening method when performed by non-experts.

The very important step is automatically counting and segmenting cytological image. In parasitemia it is necessary to segment blood cells and classify them as infected or normal. For this purpose Giemsa stained thin blood images used. In thin blood images morphology of cells can be observed clearly. The staining process slightly colorizes the red blood cells (RBCs) but highlights spp parasites, white blood cells (WBC), and platelets or artifacts. Giemsa stains nuclei, chromatin in blue tone and erythrocytes in pink color. The present paper describes the techniques used in counting of infected and normal erythrocytes for purpose of estimating parasitemia (number of infected blood cells over total red blood cell count by correctly segmenting and separating erythrocytes.

The rest of the paper is organized as follows: Section 2. summarizes literature related to the problem of automatic blood analysis and different methods to segment RBCs. Section 3. illustrates the methodology used for pre-processing, feature extraction and classification. The results and conclusions are presented in Section 4 and Section 5 respectively.

2. Literature Review

D. Ruberto et. al. [3] 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. [4], proposed a technique for estimating parasitemia. An approach of 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 which is based on the individual color index of pixel and color indices of its eight neighboring pixel.

Tomasz Markiewicz et. al. [5], presents the feature characterization and assessment of the blasts that leads to the best performance of the recognizing and classifying system. The cells are classified by using the geometrical, textural and statistical features. The features categorized by using linear SVM network using SVM classifier.

Nicola Ritter et. al. [6] used stained blood images to present unsupervised blood cell segmentation. Algorithm finds all objects cells, cell groups and cell fragments that do not intersect the image border, and identifies the points interior to each object, finds an accurate one pixel wide border for each object, separates objects that just touch. Statistical analysis based by borders that have clusters of pixels is used to refine the borders by pruning stubs and thinning the border to one pixel width.

Gloria Diaz et. al. [7], presents a method for quantification and classification of erythrocytes in stained thin blood films infected with Plasmodium Falciparum. It uses three main phases a preprocessing step, which corrects luminance differences. A segmentation step that uses the normalized RGB color space for classifying pixels

Manuscript received February 5, 2011
Manuscript revised February 20, 2011

either as erythrocyte or background followed by an Inclusion-Tree representation that structures the pixel information into objects, from which erythrocytes are found classified as infected or normal erythrocytes and differentiates the infection stage, using a trained bank of classifiers.

3. Methodology

Algorithm used for cell identification includes following steps: Image Acquisition, Pre-processing, Cell segmentation, Feature Extraction, and Classification. Its block diagram and algorithm is explained as follows:

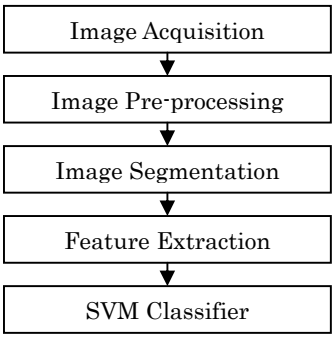

Fig. 1: Block Diagram

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. 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 to get good resolution. A popular stain, Giemsa, was used to prepare thin blood slides which slightly colors red blood cells (RBCs) but highlights the parasites, white blood cells (WBC), platelets, and various artifacts [8].

3.2 Image Pre-processing

Pre-processing step includes noise reduction, smoothening of image. Here the Laplacian filter is used for smoothening the color image. The median filter [9] is a non-linear digital filtering technique, often used to remove noise from images or other signals. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order. For edge detection the significant operator here is Prewitt that convolves the image. After pre-processing image is send to segmentation block to segment erythrocytes.

3.3 Cell Segmentation

The goal of the segmentation process is to define areas within the image that have some properties that make them homogeneous. To segment foreground from background thresholding is used [9] and also Otsu [10] thresholding is used on grayscale enhanced image. Green channel is used for low contrast image. The edge enhancement is done using Laplacian filter. Result of thresholding is added to get binary image of erythrocytes. White blood cell has the biggest size and has nucleus in it; the plasma and the platelet are considerably smaller compared to red blood cells are removed from binary image of cells.

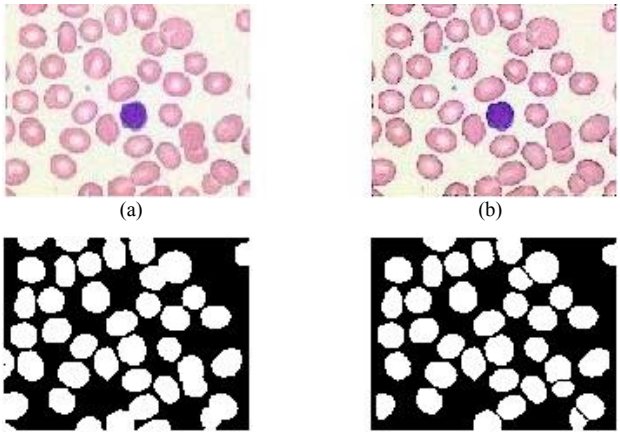

Fig. 2: a) Original Image, b) Enhanced Image, c) Binary Image of erythrocytes, d) Separation of overlapping cell.

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 in the red blood cells and background. In order to fill all holes in the cells, close holes operator is applied. For separation of overlapping cells watershed transform is applied on distance transform of binary mask of cells having larger area. By labeling this binary image total number of cells is calculated. Fig. 2 shows original image, enhanced image, and binary mask of erythrocytes and separation of touching cells after applying watershed transform.

3.4 Feature Extraction

The features mainly fall into two categories: shape based feature and statistical based feature. The mathematical morphology provides an approach to the processing of image based on shape. Mathematical morphological operations tend to extract their essential shape characteristics and to eliminate irrelevancies. Mathematical morphology represents the shapes which are manifested on binary or grey tone images. The set of

parameters corresponds to the geometrical shapes of the whole cell [11]. For the characterization of the blood cells the following geometrical parameters are used:
Radius -measured by averaging the length of the radial line. Segments defined by the centroid and border points, **Perimeter** - the total distance between consecutive points of the border, the ratio of the perimeter and radius, Area - the number of pixels on the interior of the cell. **Compactness** –is the ratio of perimeter^2 by area.

The statistical parameters refer to the color distribution contained in the cell image. Features are created on the histograms of the image matrix and the gradient matrix of the image for three color components: red, green, and blue. 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.

The mean value, variance, Skewness, standard deviation and kurtosis of both histograms are used as the features which are determined as given below.

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

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

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

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.

3.5 SVM Classifier

Since the chosen features affect the classifier performance much, deciding on which features to be used in a specific data classification problem is as important as the classifier itself.

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.

The SVM is a linear machine works in the high-dimensional feature space formed by the nonlinear mapping of N-dimensional input vector x into a K-dimensional feature space (K>N) through the use of function $\phi(x)$. The separation of two classes is performed by the hyper plane defined in the form $g(x)=w^T \phi(x)+b=0$, with $\phi(x) =[\phi1(x),\phi2(x), \ldots, \phi K(x)]^T$, w as the weight vector of network $w=[w1, w2, \ldots, wK]^T$, and b as the bias [11], [12].

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.

4. Results

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. Thus all extracted feature of each cell is sends for classification.

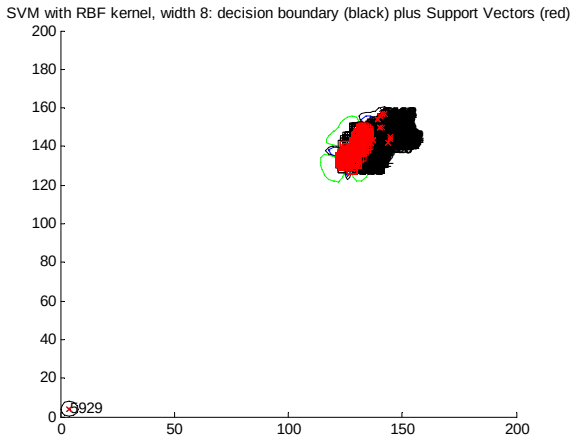

Fig. 3: Training of data using RBF kernel.

The binary classifier using RBF kernel is used for classification. The sample result of training data is shown in Fig. 3. Image processed through automatic classifier counts total number of erythrocytes, total number of normal cells and infected cell in command window. 20 images processed through automatic system. Table 1 summarizes result of manual and automatic classification of normal and infected cells for 20 images.

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 further estimation of parasitemia. The extraction of red blood cells achieves a reliable performance and the actual classification of

infected cells. 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.

Table 1. Result of manual and automatic classification of cells

Image no	Total no. of cells		Infected Cells	
	Manual Classification	Automatic Classification	Manual Classification	Automatic Classification
1	8	8	2	2
2	15	15	2	1
3	9	9	1	1
4	8	8	1	1
5	15	14	1	1
6	12	12	2	3
7	33	33	1	1
8	21	21	1	1
9	11	11	2	2
10	36	36	1	1
11	18	18	0	0
12	25	25	1	1
13	10	10	2	2
14	10	11	1	2
15	34	34	1	1
16	19	20	2	2
17	18	18	1	1
18	18	18	1	1
19	18	18	2	1
20	14	16	1	1

References

[1] World Health Organization What is malaria? Factssheetno94.http://www.who.int/mediacentrefactsheetsfs 094/en/./factshee.*J.* Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Makler MT, Palmer CJ, Alger AL,"A review of practical techniques for the diagnosis of malaria", Ann Trop Med Parasitol 92(4):419–433, 1998.

[3] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Automatic thresholding of infected blood images using granulometry and regional extrema", 441-444, 2000 IEEE.

[4] Silvia Halim, Timo R. Bretschneider, YikunLi, "Estimating Malaria parasitaemia from Blood Smear Images", 1-4244-0342-1/06, IEEE, ICARCV 2006.

[5] Tomasz Markiewicz, Stanislaw Osowski, "Automatic Recognition of the Blood Cells of Myelogenous Leukemia Using SVM", Proceedings of International Joint Conference on Neural Networks, Canada,2496-2501, Aug. 2005.

[6] Nicola Ritter, James Cooper, "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", The Thirtieth Australasian Computer Science Conference (ACSC2007), Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 62, 2007.

[7] Gloria Diaz, Fabio A. Gonzalez, Eduardo Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images", Journal of Biomedical Informatics 42 (2009) 296–307.

[8] Selena W.S. Sio, Weiling Sun, "MalariaCount: An image analysis-based program for the accurate determination of parasitemia", Journal of Microbiological Methods 68 (2007) 11 – 18, 2007.

[9] Dwi Anoraganingrum, "Cell Segmentation With Median Filter and Mathematical Morphology Operation", 21-23, D-28359 Bremen, Germany.

[10] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man, and Cybernetics, VOL. SMC-9, No. 1, Jan 1979.

[11] Jesus Angulo, rue Saint-Honore, " Automated detection of working area of peripheral blood smears using mathematical morphology", Centre de Morphologie Mathematique, Paris, Fontainebleau, France, IOS Press, 17 Oct 2002.

[12] Stanislaw Osowski, Robert Siroic, "Application of Support Vector Machine and Genetic Algorithm for Improved Blood Cell Recognition", IEEE transactions on instrumentation and measurement, VOL. 58, NO. 7, 2159-2168, July 2009.

Swati S Savkare was born in India in 1982. She received the B.E. (E & TC) degree from North Maharashtra University, Jalgaon in 2004. She is currently working as a Lecturer in Electronics and Telecomm. Engineering, with the Moze College of Engg., Pune. Her research and teaching interests include image processing.

Sandipan P Narote was born in India in 1975. He received the B.E. (E & TC), M.E.(Electronics) from Dr. Babasaheb Ambedkar University Aurangabad and Ph.D. from Ph.D. (E & TC) from Swami Ramanand Tirth Marathwada University Nanded in 2000, 2002 and 2010, respectively. He is currently a Professor of Electronics and Telecommunication Engineering with the Sinhgad College of Engineering, Pune, India. His research and teaching interests include signal processing, image processing and biometrics.