Automated image processing method for the diagnosis of malaria on blood smears using SVM classifier

[1]Sayali N. Dharpal, [2]Dr.A.V. Malviya

[1]M.E. Student, [2]Assistant Professor
Department of Electronics and Telecommunication Engineering
Sipna College of Engineering and Technology, Amravati, Maharashtra, India.

***Abstract*: Malaria is a serious disease caused by a blood parasite named Plasmodium. The World Health Organization calculates 300-500 million intestinal sickness cases and in excess of 1 million passings for every year. Manual counting and classifications of contaminated erythrocytes is a time-consuming and relentless process. This proposed system help to develop a completely automatic system for classification of Malaria parasite infected erythrocytes are sectioned from the pre-processed images. Statistical and color features are extracted and given to the SVM binary classifier which characterizes Malaria infected erythrocytes on blood smears.**

***Index Terms*: Malaria, erythrocyte, Parasite, Feature Extraction, SVM classifier.**

I. INTRODUCTION

Intestinal sickness is a main source of morbidity and mortality in tropical and sub-tropical nations, with an expected of 1-2 million passings for each year. In blood test visual identification and recognition of Plasmodium is conceivable and productive by means of a chemical procedure called Giemsa staining. The staining procedure somewhat colorizes the erythrocytes however features Plasmodium parasites, white platelets (WBC), and relics. It has been appeared in a few field examines that manual microscopy is definitely not a dependable screening technique when performed by non-specialists and furthermore a tedious. Regarding the assessment of antimalarial treatments it is additionally critical an estimation of the existence parasite arrange per tainted erythrocyte.

This paper discusses a fully automatic system for location of contaminated erythrocytes from blood pictures and parasite stage life identification. Statistical, Textural and color features are removed for preparing of SVM classifier. System is supervised. 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 preprocessing, segmentation, feature extraction and SVM classifier Section 4 and 5 includes results and conclusions of this paper.

II. RELATED WORK

Minh-Tam Le et. al. [1], proposed a comparison-based analysis, which differentiates solid components in blood smears. The semiautomatic method uses statistical measures and crossreferencing 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. [2], 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. [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. 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. [5], 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.

Diaz et al., [6] developed a technique for detection, quantification of parasitemia and parasite life stages. Pixels' color features were extracted and used to train classifiers for detection and determination of parasite life stages. Clustered erythrocytes were resolved by use of template matching before parasitemia was estimated. The study reported a sensitivity of 94% for detection of infected erythrocytes and 79% for stages identification. The technique was not fully automatic as it called for human intervention during training of the classifier every time diagnosis had to be made.

Di Ruberto et al., [7] proposed a technique of automatically detecting and quantifying malaria parasites infection in blood images of patients. The method employed a modified watershed algorithm to segment erythrocytes. There were two alternatives proposed for classifying parasite stages. One was the use of morphological thinning, where skeletons of parasites images were used to

categorize parasites into their respective stages of infection. The second option was use of color histograms similarity. The efficiency of the segmentation algorithm proposed reduces with the degree of clustering of erythrocytes. Similarly the accuracy of colour histogram similarity for classification of parasites would depend on the imaging parameters and illumination conditions under which the image being probed is taken. The detection accuracy of parasitemia reported was relatively low, 50%.

III. METHODOLOGY

This method involves several processes namely; image acquisition, preprocessing, feature extraction, parasite classification as shown in Figure 1.

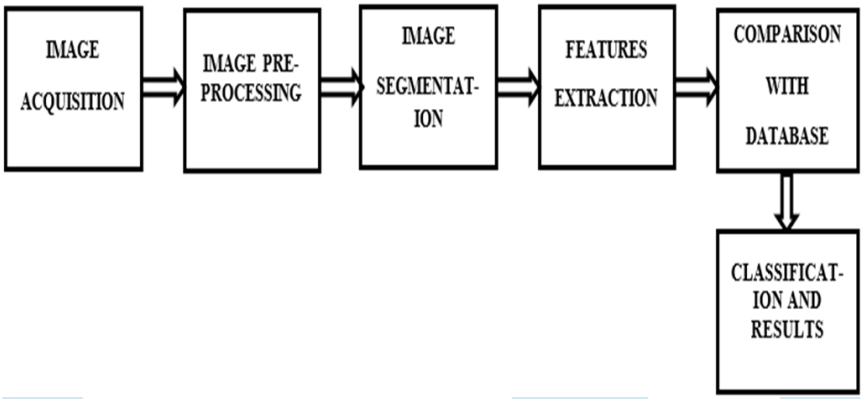

Fig -1: Block Diagram

3.1 Image Acquisition:

Thin blood smear pictures were procured from the Center for Disease Control (CDC) website [8] and captured from the Reference Laboratory of Malaria, in Sudan Ministry of Health.

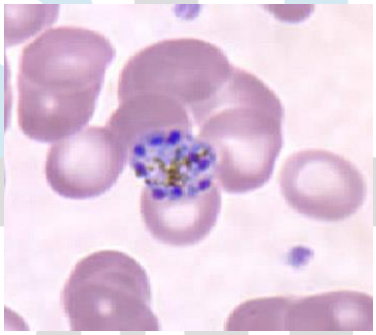

Fig -2: Input Image sample

3.2 Image Pre-Processing:

Pre-processing step incorporates noise reduction, smoothening of image. In this paper, we utilized bilateral filter for smoothening of color image and In range filter is utilized for edge sharpening. This outcome is subtracted from original to enhance the image. The bilateral filter is a non-linear digital filtering technique, used to expel noise from images. In bilateral filtering every pixel is supplanted by a weighted average of its neighbors. This perspective is imperative since it makes it simple to acquire intuition about its behavior, to adjust it to application-specific requirements. It depends just on two parameters that show the size and complexity of the features to preserve. In range filtering is a procedure that cleans up appearances and takes into consideration particular featuring of particular data. After pre-processing, image is send to cell segmentation block to segment cells.

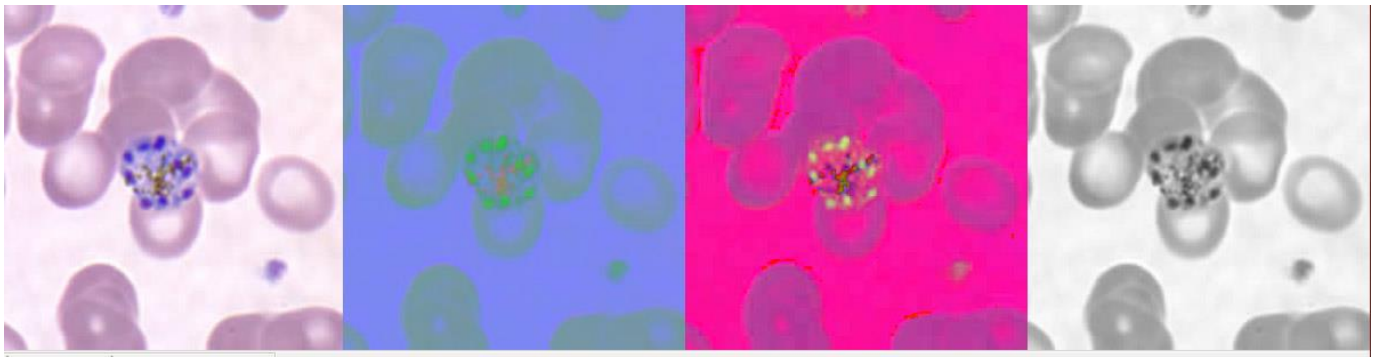

Fig -3: HSV to gray scale conversion of input image

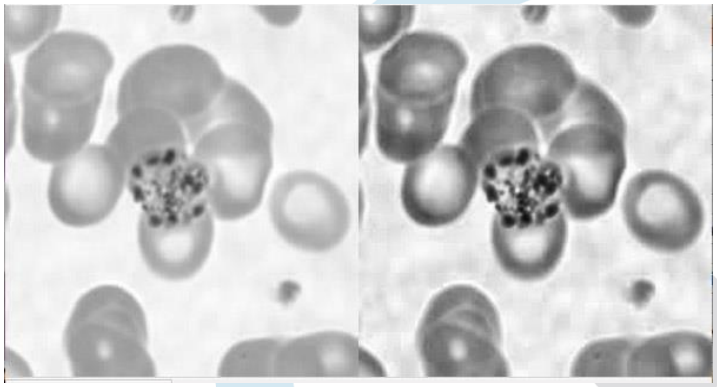

Fig -4: CLAHE

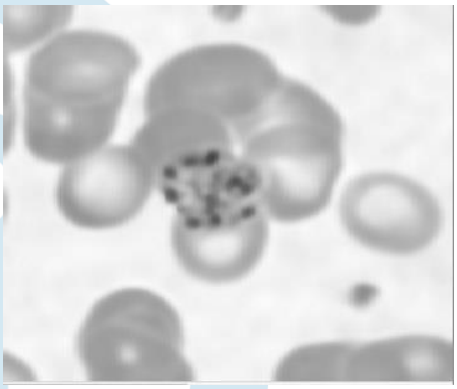

Fig -5: Bilateral Filtering of input image

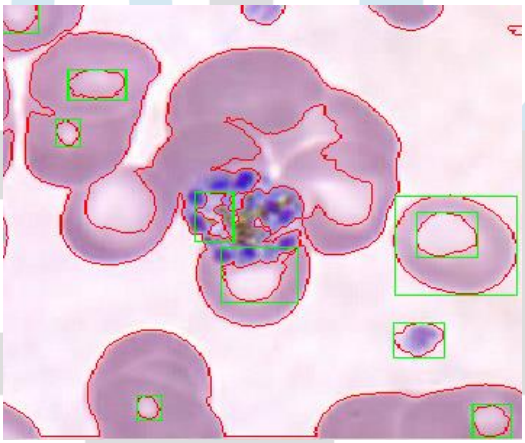

Fig -6: Post-processing of input image

3.3 Image Segmentation

Image segmentation is the way toward partitioning a digital image into different fragments. The objective of segmentation is to simplify and change the representation of an image into something that is more important and less demanding to examine. Image segmentation is ordinarily used to locate objects and boundaries in images. All the more unequivocally, image segmentation is the way toward allotting a label to each pixel in a image to such an extent that pixels with a similar label share certain visual characteristics. The consequence of image segmentation is an arrangement of segments that aggregately cover the whole image, or a set of contours extracted from the image. Every one of the pixels in a locale is comparative concerning some trademark characteristic or computed property, such as color, intensity, or texture. Adjoining districts are essentially unique concerning similar characteristics. At the point when connected to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms. Meaningful segmentation is the initial step from low-level image processing changing a grayscale or color image into at least one images to high-level image description in terms of features, objects, and scenes.

In this paper, superpixel calculation is utilized in which superpixels aggregate pixels into perceptually important atomic regions that can be utilized to supplant the rigid structure of the pixel grid in images. Thusly, image primitives and redundancy can be decreased incredibly. It is more advantageous and powerful to compute image features based on regions than pixels.

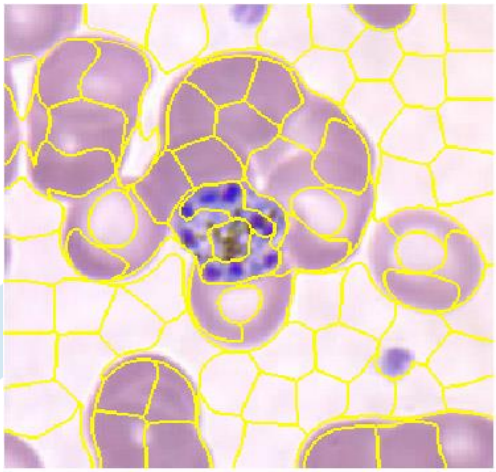

Fig -7: Superpixel segmentation of input image

3.4 Feature Extraction

Feature extraction changes the higher dimensional information to the less measurements with the goal that it diminishes the measure of resource required to depict the resource information. It diminishes the many-sided quality in breaking down the lager information which requires huge measure of memory and computation power. It builds the combination of variables such that it diminishes the resource usage and furthermore describes information with adequate precision. It likewise checks for the excess information and changed in to reduced features. The important data from the information is discovered utilizing feature selection and after that got utilizing feature extraction, utilizing this information we can perform any operation this will be same as the input data. In image data, we get change and get the required information these information are called features. In this paper, feature extraction is done with the help of Local Binary Pattern.

The original LBP operator marks the pixels of a image with decimal numbers, called Local Binary Patterns or LBP codes, which encode the nearby structure around every pixel. It continues in this manner, as showed in Fig.8: Each pixel is contrasted and its eight neighbors in a 3x3 neighborhood by subtracting the center pixel value. The subsequent entirely negative values are encoded with 0 and the others with 1. A binary number is acquired by concatenating all these binary codes a clockwise direction beginning from the top-left one and its relating decimal value is utilized for labeling. The inferred binary numbers are referred to as Local Binary Patterns or LBP codes.

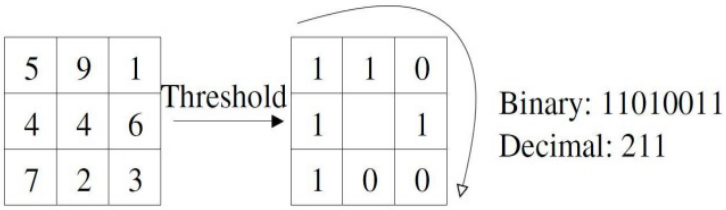

Fig -8:An example of the LBP operator

One limitation of the essential LBP operator is that its small 3x3 neighborhood can not catch prevailing features with largescale structures. To manage the surface at various scales, the operator was later generalized to utilize neighborhoods of various sizes .A local neighborhood is characterized as an arrangement of sampling focuses uniformly centered on a circle which is focused at the pixel to be labeled, and the sampling point that don't fall inside the pixels are interjected utilizing bilinear interpolation, consequently taking into account any radius and any number of sampling points in the area. Fig. 9 demonstrates a few cases of the extended LBP operator.

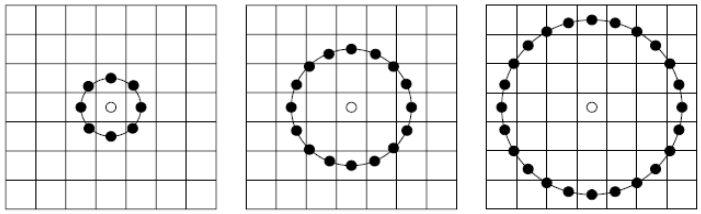

Fig - 9: Examples of the extended LBP operator

In this work, the input image is partitioned into small regions, from which LBP histograms are extracted, and the neighborhood histograms are additionally connected into a spatially upgraded feature vector. In addition, some variations even increment the feature vector length drastically, for example, Extended LBP, VLBP and Gabor Wavelets based LBP. The determined LBP-based feature vector furnishes an over-total representation with repetitive information [11], which could be decreased to be more minimized and discriminative. Moreover, when assembling real-time systems, it is likewise desired to have LBP-based representation with decreased feature length. For every one of the reasons, the issue of LBP include choice has as of late been addressed in many literatures.

Rule-based based Strategy Uniform example is a viable control to choose LBP features, and it has been generally adopted in existing work. There are likewise different rules which could be utilized. For example, Lahdenoja et al. [12] proposed a symmetry level plan for uniform patterns to additionally lessen the length of LBP feature vectors. The symmetry level is turn invariant as per the definition. The most symmetric pattern contains a similar number of ones and zeros, showing a symmetric edge, while the examples with the least symmetry level are the ones comprising of only ones or zeros. It is asserted that the patterns with high symmetry level happen all the more much of the time in the images with more discriminative power [12].

Boosting LBP Features Boosting learning [13] gives a compelling method to highlight selection. In [11], by shifting and scaling a sub-window over face image, numerous more sub- regions are acquired to remove nearby LBP histograms, the separation between the relating histograms of two images is used as the discriminative feature.

3.5 SVM Classifier

The SVM is an intense solution for the classification issues. In this paper, it has been utilized for the acknowledgment and characterization of cells. The fundamental favorable position of the SVM network utilized as a classifier is its great generalization capacity and to a great degree intense learning technique, prompting the worldwide least of the characterized error function. Direct SVM is a straight discriminant classifier working on the principle of most extreme edge between two classes. The decision capacity of the N dimensional input vector x for K-dimensional component space (K>N) is characterized as D(x) = wT (x) +b using function (x). Where (x) = [1(x), 2(x), . . . , K(x)], w as the weight vector of network w = [w1 , w2, ..., wk]T , and b as the bias weight [14].

The learning of the SVM network working in the arrangement mode is gone for the maximization of the division edge between two classes. Basic order calculation is recommended that arranges indicates by assigning them the closer of two parallel planes. Standard support vector machines (SVMs), which assign focuses to one of two half spaces. SVM classifier is utilized for grouping of normal and contaminated cells. Results are appeared in Figure 10.

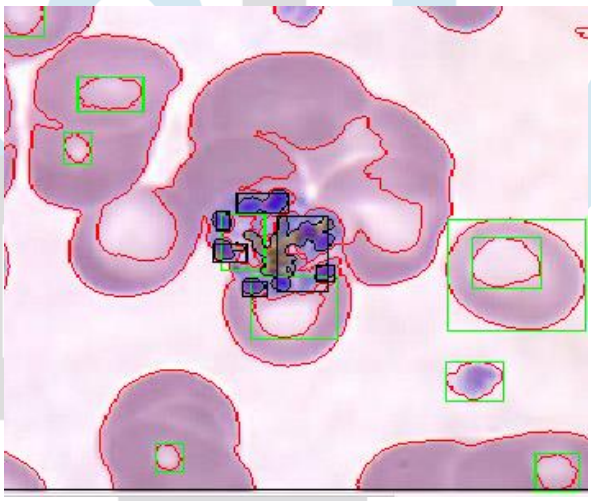

Fig -10: Detected parasite infected cells

IV. RESULT

The described techniques for feature extraction produce an extremely rich group of parameters. Image handled through programmed system segments RBCs from input image, isolate overlapping cells, counts add up to number of erythrocytes and SVM binary classifier distinguish contaminated cells. At long last framework gives number of ordinary cells and infected cell, 15 images handled through programmed framework. Table 1 condenses consequence of manual and automatic parasitemia for 15 images.

Parasitemia is the quantitative content of parasites in the blood.[15] It is used as a measurement of parasite load in the organism and an indication of the degree of an active parasite infection. Figure 11. Shows graphical comparison of manual and automatic parasitemia count.

Image	Manual Parasitemia	Automatic Parasitemia
1	25.00	25.00
2	13.33	6.00
3	11.11	11.00
4	12.50	12.50
5	6.67	7.14
6	16.67	25.00
7	3.03	3.00
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.

Comparison of Manual and Automatic Parasitemia

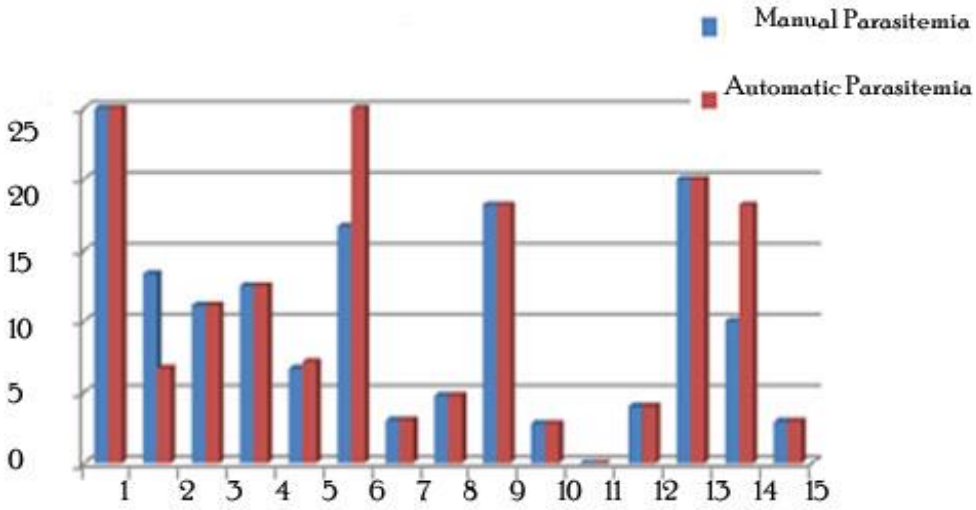

Fig -11**:** Graphical comparision of Manual and Automatic Parasitemia

The measure of performance of the method is evaluated by precision and accuracy. Sensitivity is defined as the probability (percentage) that patients with the infection will have a positive result using the test under evaluation Accuracy refers to the closeness of a measured value to a standard or known value. Precision refers to the closeness of two or more measurements to each other. The values for sensitivity and accuracy are expressed in terms of true positives (TP), false positive (FP), false negative (FN) and true negative (TN) as defined below in expressions.

$$Sensitivity=TP/TP+FN$$

$$Specificity=TN/TN+FP$$

$$Accuracy = TP/ (TP+TN)*100\%$$

V. CONCLUSION

The proposed automated technique for segmentation and grouping of cell is basic. An approach is proposed to identify red platelets with successive order into parasite contaminated and normal cells for estimation of parasitemia. The extraction of red platelets

accomplishes a solid execution and the actual classification of infected cells. Shape based and features are generated for characterization. This approach prompts the high specialization of classifier and results in a general increment in exactness.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] Diaz, G., Gonzalez, F.A. and Eduardo, R. (2009) A semiautomatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Journal of Biomedical Informatics, 42, 296-307. http://dx.doi.org/10.1016/j.jbi.2008.11.005 [Citation Time(s):1]

[7] Di Ruberto, R.C., Dempster, A., Khan, S. and Jarra, B. (2002) Analysis of infectedblood cell images using morphological operators. Image and Vision Computing, 20, 133-146. [Citation Time(s):3]

[8] CDC, "CDC - Malaria - About Malaria - Biology - Malaria Parasites," *USA Government*, 2012. [Online]. Available: http://www.cdc.gov/malaria/about/biology/parasites.html.

[9] 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.

[10] 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.

[11] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, "Boosting Local Binary Pattern-based face recognition," in *Proc. Advances in Biometric Person Authentication*, 2004, pp. 179-186.

[12] O. Lahdenoja, M. Laiho, and A. Paasio, "Reducing the feature vector length in local binary pattern based face recognition," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2005, pp. II: 914-917.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. I: 511-518.

[14] 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.

[15] "parasitemia" at *Dorland's Medical Dictionary.*