

Advanced Image Analysis based system for Automatic Detection of Malarial Parasite in Blood Images

Jigyasha Soni^{#1}

1. Dept. of Electronics & Telecommunication

SSCET Bhilai (India)

jigyasha2006_elex@yahoo.com

Nipun Mishra ^{*2}

2. Dept. of Electronics & Telecommunication

SSCET Bhilai (India)

*mishranipun@gmail.com

Abstract-Malaria is a serious global health problem, and rapid, accurate diagnosis is required to control the disease. An image processing algorithm to automate the diagnosis of malaria in blood images is developed in this project. The image classification system is designed to positively identify malaria parasites present in thin blood smears, and differentiate the species of malaria. Images are acquired using a charge-coupled device camera connected to a light microscope. Morphological and novel threshold selection techniques are used to identify erythrocytes (red blood cells) and possible parasites present on microscopic slides. Image features based on colour, texture and the geometry of the cells and parasites are generated, as well as features that make use of a priori knowledge of the classification problem and mimic features used by human technicians. The first order features provides the basic mathematical ranges for different types of parasites. A two-stage tree classifier distinguishes between true and false positives, and then diagnoses the species (*Plasmodium falciparum*, *P. vivax*, *P. ovale* or *P. malariae*) of the infection. Malaria samples obtained from the various biomedical research facilities are used for training and testing of the system.

Keywords-

component, formatting, species, insert

I. Introduction

Malaria is caused by parasites of the species *Plasmodium* that are spread from person to person through the bites of infected mosquitoes. A parasite is an organism that lives off another organism. Animals can also get malaria, but malaria cannot be passed from humans to animals or from animals to humans.

A. Facts and Figures

Approximately, 40% of the world's population, mostly those living in the world's poorest countries, is at risk of malaria. A child dies of malaria every 30 seconds. Every year, more than 500 million people become severely ill with malaria. [1]. Between 300 million and 500 million people in Africa, India, Southeast Asia, the Middle East, the South Pacific, and Central and South America have the disease. The worldwide annual economic burden of malaria, calculated to include spending on prevention and treatment as well as loss of productivity due to illness, was estimated at US\$ 500 million in 2005[2].

B. Diagnosis of Malaria

The definitive diagnosis of malaria

infection is done by searching for parasites in blood slides (films) through a microscope. In peripheral blood sample visual detection and recognition of Plasmodium spp is possible and efficient via a chemical process called (Giemsa) staining. The staining process slightly colorizes the red blood cells (RBCs) but highlights Plasmodium spp parasites, white blood cells (WBC), and platelets or artifacts. The detection of Plasmodium spp requires detection of the stained objects. However, to prevent false diagnosis the stained objects have to be analyzed further to determine if they are parasites or not. [1]

In the fig.1:, There are four types of human malaria – Plasmodium falciparum, P. vivax, P. malariae, and P. ovale. P. falciparum and P. vivax are the most common. P. falciparum is by far the most deadly type of malaria infection

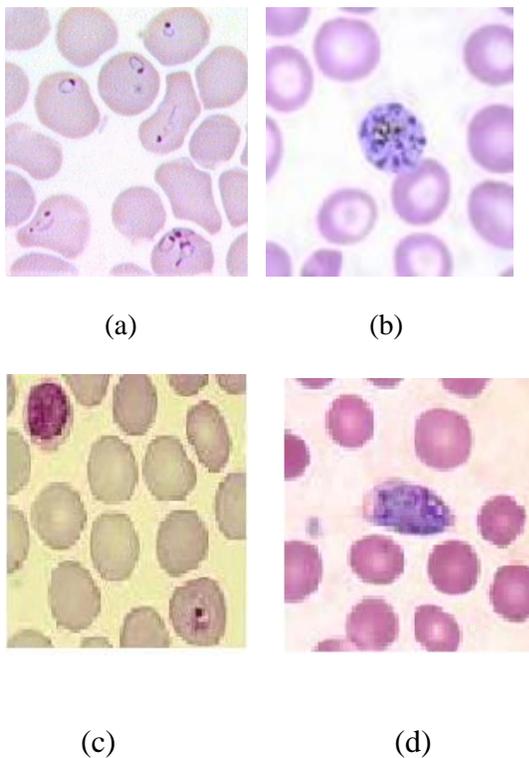


Fig1. a) Plasmodium Falciparum
 b) P. Vivax
 c) P. Malariae

d) P. Ovale

C. Goals

The biggest detraction of microscopy, namely its dependence on the skill, experience and motivation of a human technician, is to be removed. Used with an automated digital microscope, which would allow entire slides to be examined, it would allow the system to make diagnoses with a high degree of certainty. It would also constitute a diagnostic aid for the increasing number of cases of imported malaria in traditionally malaria-free areas, where practitioners lack experience of the disease.

D. Objectives

The objective of the project is to develop a fully automated image classification system to positively identify malaria parasites present in thin blood smears, and differentiate the species. The algorithm generated will be helpful in the area where the expert in microscopic analysis may not be available. The effort of the algorithm is to detect presence of parasite at any stage. One of the parasites grows in body for 7 to 8 days without any symptoms. So if this algorithm is incorporated in routine tests, the presence of malarial parasite can be detected

II. Proposed Algorithm

The design follows the same steps as that of a pattern recognition problem. But the best part of the algorithm is the usage of the most appropriate algorithm for each stage. The test algorithms illustrated above give an insight about the algorithm to be used for each stage. The process is given below.

- 1.) Image Acquisition and database collection
- 2.) Image Analysis

- 3.) Image Segmentation
- 4.) Feature Generation
- 5.) Classification of Parasite and result verification

A. Image acquisition and database collection

Oil immersion views (10x1000), of Giemsa stained blood films were captured using a binocular microscope mounted with a digital camera. Captured images were 460 pixels X 307 pixels bitmap images. Fig 2. shows the sample slide of *P. falciparum*. The database consists of 110 images.

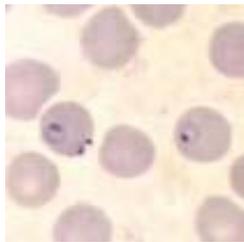


Fig2. Plasmodium Falciparum

B. Image analysis:

For a real time system using time varying image sequences, speed is an important criterion to be considered. Also there has to be a compromise between maximizing signal extraction and minimizing output noise: the so-called “Uncertainty Principle” of edge detection. I have implemented a new approach to low-level image processing - SUSAN (Smallest Univalued Segment assimilating Nucleus) Principle [16], which performs Edge and Corner Detection and Structure Preserving Noise Reduction. Canny edge detector, which has become one of the most widely used edge finding algorithms, is found to be ten times slower than this SUSAN approach. The results are stable for Canny but the edge

connectivity at junction is poor and corners are rounded. The fact that SUSAN edge and corner enhancement uses no image derivative, explains why the performance in the presence of noise is good. The integrating effect of the principal together with its non-linear response gives strong noise rejection.

Algorithm: The following steps are performed at each image pixel:

- Place a circular mask around the pixel in question.
- Calculate the number of pixels within the circular mask which have similar brightness to the nucleus. These define the USAN.
- Subtract USAN size from geometric threshold to produce edge strength image.
- Use moment calculations applied to the USAN to find edge direction.
- Apply non-maximum suppression thinning and sub-pixel estimation, if required.
- Hence SUSAN allows image edges, corners and junctions to be accurately and quickly found and a related method for reducing noise whilst preserving image structure. The localization of the features is independent of the mask size used and noise suppression is shown to be good.

C. Image segmentation:

Techniques have been proposed earlier that make use of thresholding or morphology to segment an image. In this section we have presented a technique that takes benefit of morphological operation and thresholding at appropriate positions in the whole process to maximize the productivity of the algorithm. In order to use morphological methods for image segmentation, the shape and size of the objects in the image must be known. The most commonly used morphological procedure for estimating size distribution of image components is the Granulometry. The size and eccentricity of the erythrocytes are also required for the calculation of some feature values (as

these can be indicative of infection). The shape of the objects (circular erythrocytes) is known a priori, but the image must be analyzed to determine the size distribution of objects in the image and to find the average eccentricity of erythrocytes present. A pattern spectrum showing the size distribution of objects in a sample can be calculated using Granulometry [15].

The next stage of the process identifies and segments potential parasites and erythrocytes from the image background. To extract the infected erythrocytes, it is first necessary to identify them from the combination of parasites and erythrocytes in the image, and then segment them from the background.

This algorithm relies primarily on thresholding. The key to successfully segmenting an image using thresholding is threshold selection. The histogram of the complemented, green component of the sample image (fig.3) is a bimodal distribution typical of all the images considered.

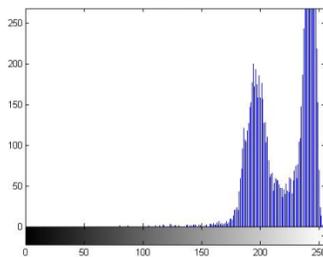


Fig. 3: - Bimodal histogram of Plasmodium Falciparum

The threshold level is automatically selected using a modified version of the method that maximizes the separability of the resultant classes of the grey-level histogram [17]. The principle mode is due to the grayscale intensities of the image background, and the second mode is due to those of the erythrocytes in the image.

Two threshold levels need to be determined from the histogram: one for erythrocytes, and one for parasites. The first threshold is selected to separate the erythrocytes from the background of the image. This essentially means separating the two modes of the image histogram. The first threshold is selected to separate the erythrocytes from the background of the image. This essentially means separating the two modes of the image histogram.

The resulting thresholded binary mask of erythrocytes then has all holes with an equivalent radius less than an empirically determined average erythrocyte radius removed. A morphological opening using a disk-shaped SE with a radius 40% of the mean erythrocyte radius is applied to smooth the objects in the image, and any objects with an equivalent radius of less than half the mean erythrocyte radius are removed. The problem with this binary image of erythrocytes is that clusters of cells are not separated. The next step is to select the second threshold to find parasites present in the image. A global threshold level, taking the threshold as the first local minimum in the histogram after the mode due to erythrocytes, is not sensitive enough. This is a common problem experienced with global threshold selection, caused by inconsistent intensities in the image. The solution is to find local threshold levels. The erythrocytes, having already been identified, provide excellent image regions in which to find these, especially since valid parasites are only found inside erythrocytes. The threshold is then found by taking the first minimum after the principal mode of the histogram incorporating only the erythrocytes.

While this method has greater sensitivity, it is at the expense of a reduced specificity. There are also cases, particularly with *P. ovale*, where the global threshold is able to detect parasites that are missed by the local thresholds. This is due to colorization of the infected cells, which shifts

the principle mode of the local histograms of the affected cells. For this reason, both local and global thresholds are used, and the union of the two binary images is used as the parasite marker image.

Invalid objects in the marker image (objects detected with the global threshold that lie outside any erythrocyte) are removed by taking the intersection of the parasite marker image with the binary mask of erythrocytes. The erythrocyte mask is dilated first, to ensure that 'blister' forms of the parasites, that appear to bulge out of the edge of the cells, are not removed.

Other artifacts in the blood containing nucleic acid, particularly white blood cell nuclei, are also detected by this thresholding. They are removed by excluding all objects greater than an empirically determined size (chosen to exclude all objects greater than the largest trophozoite that one would expect to find.)

The infected cells are identified by morphologically reconstructing the erythrocyte mask with the valid parasite marker. Binary reconstruction simply involves extracting the connected components of an image (the mask) that are marked by another image (the marker), where cells are clustered together, if an infected cell forms part of the group, then the entire aggregation is reconstructed.

To separate these clusters so that the infected cell can be isolated and extracted, a **modification** of the morphological technique used in Di Ruberto et al. [20] is used. A morphological opening filter, using a disk-shaped SE with radius equal to the mean erythrocyte radius less the standard deviation, is applied to the grayscale, morphologically filtered green component of the image to remove any objects smaller than an erythrocyte. The morphological gradient—the difference between

a dilation and erosion of the image—is then calculated using a diamond-shaped SE with unity length.

The segmentation method is applied to each object in the reconstructed binary image of erythrocytes individually. Those objects that do not exceed the area of a circle with radius equal to the mean erythrocyte radius plus the standard deviation are regarded as being single cells, and are unmodified.

Unlike the method in Di Ruberto et al. [20], where the morphological gradients are used to generate marker images for the watershed algorithm, the objects deemed to be overlapping erythrocytes are segmented as follows. First, the intersection of the morphological gradient image and the dilated cell cluster is taken. This image is then transformed to a binary image by thresholding any value greater than zero. A series of morphological operations, namely a closing operation, thinning, and spur-removal are then applied to generate a contour of the segmented erythrocytes. The contours are filled, and the segmented mask is again reconstructed with the valid parasite marker image to result in a segmented mask of infected cells.

The erythrocytes that have been identified as possibly infected are then extracted from the image and passed to the next stage of the algorithm for feature generation. The binary mask of the erythrocyte, as well as a binary mask (obtained by local threshold selection based on the image histogram as detailed above) of parasite-like objects present in the cell, is also passed to the next stage.

III. Feature Generation and Classification

A. Feature Generation:

Two sets of features are used for development. The first set will be based on image

characteristics that have been used previously in biological cell classifiers, which include geometric features (shape and size), colour attributes and grey-level textures.

It will be advantageous to apply expert, a priori knowledge to a classification problem. This will be done with the second set of features, where measures of parasite and infected erythrocyte morphology that are commonly used by technicians for manual microscopic diagnosis are used. It's desirable to focus on these features, because it is already known that they are able to differentiate between species of malaria.

B. Feature Classification

The final classification of an erythrocyte as infected with malaria or not, and if so, the species of the parasite, falls to the classifier. The classifier is a two-stage tree classifier, with an infection classified as positive or negative at the first node, and the species assigned at the second node.

The design of a tree classifier has the following steps: the design of a tree structure (which has already been assigned), the selection of features to be used at every node, and the choice of decision rule at each node [22]. The same type of classifier is used at both nodes.

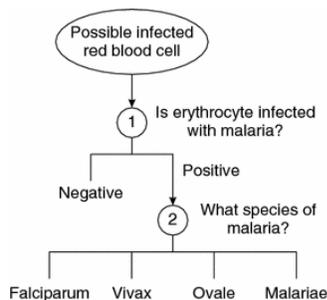


Fig. 4: - Structure of the tree classifier

The features selected for the first classifier are those that describe the colour and texture of the possible parasites. The features used by microscopists to differentiate malaria species are selected for the second classifier. The training goal is to minimize squared errors, and training is stopped when the error of a validation set increased. This is done to avoid overtraining.

IV. Results

The performance and accuracy of the algorithm are analyzed using two measures: **sensitivity**, the ability of the algorithm to detect a parasite present; and **positive predictive value (PPV)**, the success of the algorithm at excluding non-infected cells. These values are expressed in terms of true positives (TP), false positives (FP) and false negatives (FN):

$$Sensitivity = \frac{TP}{TP + FN}$$

$$PPV = \frac{TP}{TP + FP}$$

The algorithm has been tested on various malaria parasites. The results are summarized in the table 1 various stages of Image Processing.

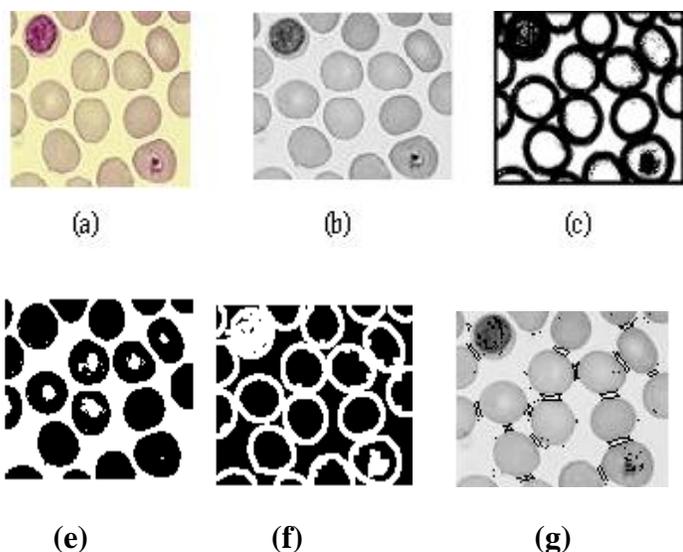


Fig. 5:(a)Original image (b)Gray scale image
 (d)Susan output (e)binary segmented image
 (f)Dilated gradient mask (g)final detected cell

Results for Sensitivity and Positive predictive value-

The test results of 10 blood images consisting of 199 Red blood cells are included in a table. The values are tabulated below and are compared with manual counting

Test images	Algorithm 1		Algorithm 2		Manual counting	
	RBC	Parasites	RBC	Parasites	RBC	Parasites
1	12	2	12	2	12	2
2	12	4	12	4	12	4
3	27	2	27	2	27	2
4	13	1	13	1	13	1
5	11	1	11	1	11	1
6	13	1	13	1	13	1
7	24	1	24	1	24	1
8	11	1	11	1	11	1
9	51	7	51	7	51	7
10	25	3	25	3	25	3

Table 1

Table no.2 gives the results of first order features of different parasite affected blood cells

S.No.	Image	Variance	Energy	Entropy
1.	Simple RBC	3.1408	0.1928	1.7969
2.	P.Falciparum	1.0069	0.3703	1.2036
3.	P.Malarie	1.3944	0.2608	1.4989
4.	P.Oval	3.5076	0.2206	1.6735
5.	P.Vivax	1.1738	0.2742	1.3599

Table2

Table 3 gives the summary of all the algorithms tested and also shows the comparison between previous work done in this field and my work

S.NO	PREVIOUS WORK BY	SENSITIVITY	PPV	AVERAGE EXECUTION TIME
1	Jean Philippe thiran	NA	NA	1 MINUTE
2	Ci. Di.Ruberto	NA	NA	NA
3	Silvia Halim	low	NA	LOW
4	Selena w.s.sio	NA	28-81%	30 Seconds to process a single image
5	F.Borey Tek	74%	88%	LOW
6	Our obervation	99%	100%	5 SEC.Average

Table 3

IV Conclusion

The proposed automated parasite detection algorithm has any advantages compared to other diagnostic techniques. It avoids the problems associated with rapid methods, such as being species-specific and having high per-test costs, while retaining many of the traditional advantages of microscopy, viz. species differentiation, determination of parasite density, explicit

diagnosis and low per-test costs.

Apart from overcoming the limitations of conventional methods of parasite detection, the proposed algorithm is optimized to overcome limitations of image processing algorithms used in the past. Among the tested test algorithms, ‘SUSAN edge detection technique’ gave good localization of edges but formed a thick border making cell separation difficult. ‘Otsu’s algorithm’ gave accurate separation of RBCs where as local and global thresholding segmented the parasites. Granulometry provides the size distribution of object in image.. Dilation and erosion are two morphological operations they are used for bridging gaps and eliminating irrelevant detail from a binary image . The first order features provide the mathematical ranges for simple RBC and parasite affected RBC. Results prove that the algorithm developed in this project has better execution speed than all the previous work done in this field, Its also provides best sensitivity than F.Borey Tek and best positive predictive value than Selena W.S.Sio and F.Borey Tek, and is applicable to many other blood cell abnormalities other than malaria in contrast to the algorithm developed by Jean Phillipe. This is because the percentage of pathological differences in various diseases has very less effect on this robust algorithm. The algorithm detects the species of parasite and the with a sensitivity of 99% and a positive predictive value of 100%.

V-REFERENCE

1. World Health Organization. What is malaria?
Facts sheet no 94.

<http://www.who.int/mediacentre/factsheets/fs094/en/>.

2. Foster S, Phillips M, *Economics and its contribution to the fight against malaria*. Ann Trop Med Parasitol 92:391–398, 1998.

3. 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.

4. Bloland PB (2001) *Drug resistance in malaria*, WHO/CDS/CSR/DRS/ 2001.4. World Health Organization, Switzerland, 2001.

5. Gilles H.M. The differential diagnosis of malaria. *Malaria. Principles and practice of malariology* (Wernsdorfer W.H., McGregor I eds), 769-779, 1998.

6. F. Castelli, G.Carosi, *Diagnosis of malaria*, chapter 9, Institute of Infectious and Tropical Diseases, University of Brescia (Italy).

7. Baird J.K., Purnomo, Jones T.R. *Diagnosis of malaria in the field by fluorescence microscopy of QBC® capillary tubes*. Transactions of the Royal Society of Tropical Medicine and Hygiene; 86: 3-5, 1992.

8. Anthony Moody, *Rapid Diagnostic Tests for Malaria Parasites*, Clinical Microbiology Reviews, 0893-8512/02/\$04.00_0 DOI: 10.1128/CMR.15.1.66–78.2002, p. 66–78, Jan. 2002.

9. Brown A.E., Kain K.C., Pipithkul J., Webster H.K. Demonstration by the polymerase chain reaction of mixed Plasmodium falciparum and P. vivax infections undetected by conventional microscopy. Transactions of the Royal Society of Tropical Medicine and Hygiene; 86: 609-612, 1992.

10. Jean-Philippe Thiran, Benoit Macq, *Morphological Feature Extraction for the*

Classification of Digital Images of Cancerous Tissues. IEEE Transaction on Biomedical Engineering, Vol. 43, no. 10, October 1996.

11. C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra. Automatic thresholding of infected blood images using granulometry and regional extrema. In ICPR, pages 3445–3448, 2000.

12. Silvia Halim, Timo R. Bretschneider, Yikun Li, *Estimating Malaria Parasitaemia from Blood Smear Images.* 1-4244-0342-1/06/\$20.00 ©IEEE, ICARCV 2006.

13. Selena W.S. Sio, *Malaria Count: An image analysis-based program for the accurate determination of parasitaemia,* Laboratory of Molecular and Cellular Parasitology, Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore. May 2006.

14. F. Boray Tek, Andrew G. Dempster and Izzet Kale, *Malaria Parasite Detection in Peripheral Blood Images,* Applied DSP & VLSI Research Group, London, UK, Dec 2006.

15. Rafeal C. Gonzalez, Richard E. Woods, *Digital Image Processing,* 2nd Edition, Prentice Hall, 2006.

16. S. M. Smith, J. M. Bardy, *SUSAN—A New Approach to Low Level Image Processing,* International Journal of Computer Vision, Volume 23, and Issue 1 Pages: 45 – 78, may 1997.

17 N. Otsu, A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics,* 9(1):62.66, 1979.

18 J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram. *Graphical Models and Image Processing,* 29:273.285, 1985.

19 T.W. Ridler and S. Calvard, Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man and Cybernetics,* SMC-8:630.632, 1978.

20 Di Ruberto C, Dempster A, Khan S, Jarra B, *Analysis of infected blood cell images using morphological operators.* Image Vis Comput 20(2):133–146, 2002.

21 M. K. Hu, Visual pattern recognition by moment invariants," IRE Trans. Information Theory, vol. 8, pp. 179{187, 1962

22 Mui JK, Fu K-S, Automated classification of nucleated blood cells using a binary tree classifier. *IEEE Trans Pattern Anal Machine Intell* 2(5):429–443, 1980