Automated Malaria Parasite Detection in Thin Blood Films:- A Hybrid, Illumination and Colour Constancy Insensitive Morphological Approach

Reni, S., Kale, I. and Morling, R.C.S.

This is a copy of the author's accepted version of a paper subsequently published in the *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS* Kaohsiung, Taiwan 02 Dec 2012, IEEE .

It is available online at:

https://dx.doi.org/10.1109/APCCAS.2012.6419016

Automated Malaria Parasite Detection in Thin Blood Films:- A Hybrid, Illumination and Colour Constancy Insensitive Morphological Approach

S.Kareem,I.Kale,R.C.S Morling
Applied DSP and VLSI Research Group
University of Westminster
London, United Kingdom
s.kareem@my.westminster.ac.uk

***Abstract*—This paper illustrates the automated diagnosis of malaria parasite (Plasmodium species) in microscopic images of Giemsa stained thin blood films. The procedure adapts a morphological approach for blood cell identification and uses the image features such as intensity, histogram, relative size and geometry for further analysis. Two methods of object classification have been described for parasite detection; one based on relative size and morphology and the other based on intensity variation. An analytical study on both methods has been performed further to validate the accuracy of the methods.**

I. INTRODUCTION

Malaria can be fatal if not detected and treated on time. The disease is the second most dangerous after tuberculosis and is endemic in areas of Africa, Asia, South America, and to a lesser extend in areas of Middle East and Europe [1]. Malaria is transmitted by the female anopheles mosquitoes which carry the protozoan parasite of the genus Plasmodium. In peripheral blood sample, definitive diagnosis of malaria is done by visual detection and recognition of the parasite in a stained sample of blood through a microscope. The common staining technique for malaria diagnosis in thin blood films is the Giemsa stain. If examined through a light microscope, the Plasmodium species, white blood cells (WBC) and platelets or artifacts will appear as saturated dark blue-purplish whereas the red blood cells (RBC) are lightly colorized. However the visual effect of staining varies according to the lighting and imaging conditions. But the manual microscopic diagnosis is laborious, time consuming and the accuracy depends on the expertise of the microscopist. Hence automating the process is important to provide an accurate, reliable and objective tool for the complete eradication of the deadly disease.

In this paper an attempt has been made to describe an automated process which utilizes a different approach compared to the existing works done in this field and reported in the open literature. Automated Parasite detection in blood images has been addressed in various studies [2-5]. In a blood film diagnosis scenario one should consider two facts; firstly the image will contain highly stained parasites, WBC and artifacts, very slightly stained RBCs and non-stained plasma. The plasma should be considered as the background of the image and all the other cell/components as the foreground objects.

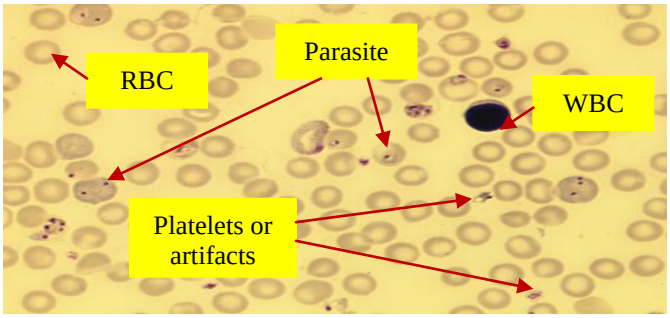


Figure 1. Giemsa stained thin blood film

Secondly the parasites reside in the RBCs, which mean a deeply stained component on a lighter background. The latter is usually ignored in most of the existing studies, focusing mainly on the stained/non-stained classification. Figure 1 will demonstrate how a peripheral blood image will look like after Giemsa staining. The studies [2-4] were based on parasite marker extraction by thresholding using color histograms and hence have the disadvantage of tagging all the platelets and other artifacts as parasites. A computerized automated diagnosis of malaria has been done by [5] which perform minimum area watershed segmentation for blood cell detection. The process first estimates the stained classes using non-parametric histogram method and uses features such as Hu-moments, color auto-correlogram etc for parasite classification. However the method requires complex pre-processing such as color normalization which includes transforming the original image to match a reference image color characteristic and perform further transformation of the foreground regions. In this paper a simpler approach is demonstrated which avoids pre-processing techniques that has to be used to correct the illumination issues which usually occur during image acquisition.

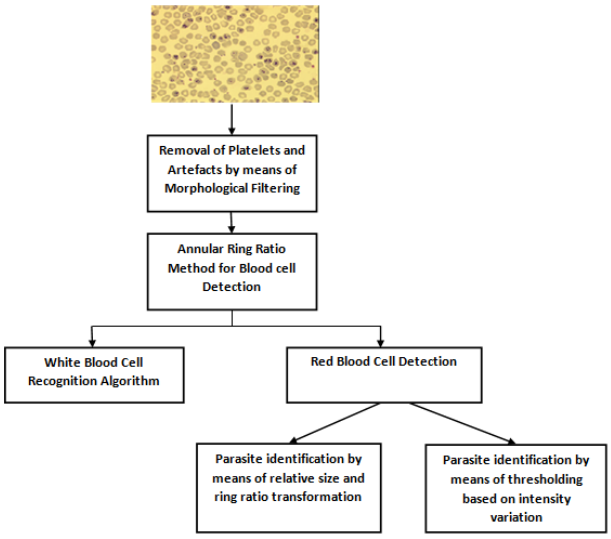


Figure 2. Diagrammatic representation of the diagnostic set up

A. *System overview*

The aim of our system is to recognize different objects present in the image prior to differentiating them as parasites and non-parasites. The foreground region of an infected cell consists of RBCs, WBCs, parasites, platelets and any artifacts or noises from various other imaging factors. The method performs morphological filtering in order to remove the platelets and other noises. To detect the rest of the cells, a novel method called annular ring ratio (ARR) method is proposed [6] which will provide the location information of WBC and RBCs present in the image. Once all the foreground components are detected, an algorithm based on the average pixel intensity values has been proposed to differentiate the white blood cells from the rest of the cells. After separating the WBCs, the algorithm proceeds to detect the parasites present in each RBC using the location information obtained from ARR transform method. In order to find the parasites, we propose two separate methods; one by means of modified ARR method based on relative size of the nucleus of the parasite and other based on thresholding with respect to the intensity of the pixels. The steps involved in the whole processing method have been portrayed diagrammatically in Fig.2.

The paper is organized as follows. Section II describes the method used to differentiate the platelets, WBCs and RBCs in the image. In Section III, the two proposed approaches for parasite detection are explained. Section IV provides the results and Section V draws the conclusion.

II. IDENTIFICATION OF BLOOD CELLS

This section deals with the methods to distinguish the foreground and background regions of a blood image and identify each of them. For analytical purpose, the background is considered to be the plasma which is lighter than the stained foreground region which consists of platelets, RBCs and WBCs. A morphological approach has been executed to differentiate each of the cells.

A. *Removing the platelets and artefacts.*

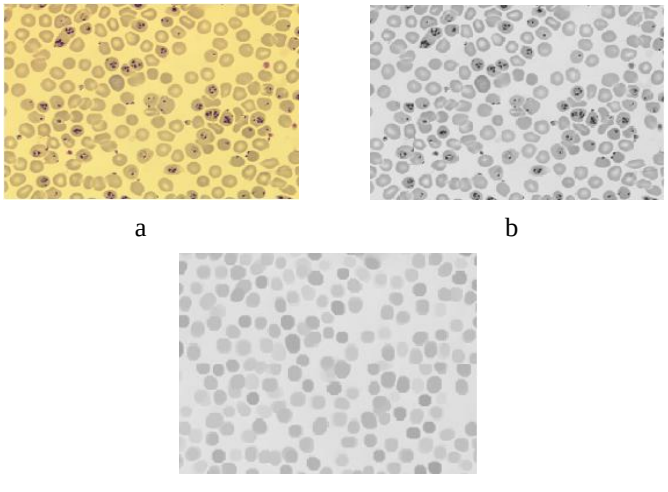
As explained in section I the platelets and artifacts are highlighted by staining. The artifacts are present due to various factors such as chemical processing, noise from the microscope lighting or from the scanner. Any stain/non-stain classification tool tends to consider them as either the parasite or white blood cell. Two fundamentals factors are considered to distinguish them from other stained components in the image. Firstly the size of the platelet or artifact is comparatively smaller than other cell components. Secondly the parasites reside within the RBC whereas platelets/ artifacts are found outside the RBCs. Since the morphological operations are well suited for biomedical image analysis, a novel morphological filtering based on the size of RBC has been proposed for platelets and/or artifacts elimination [6]. In this technique the color image is converted to gray scale to speed up the processing. It then undergoes morphological dilation followed by morphological erosion. The dilation is performed by a concentric ring structuring element and erosion by disk shaped structuring element. The radius of the structuring element depends on the radius of the RBC, so that any component smaller than the RBC will be removed. Let A be the concentric ring structuring element and B be the disk shaped structuring element. First the image is dilated using A;

$$I_\delta = I \oplus A \qquad (1)$$

Followed by erosion with B;

$$I\varepsilon = I_\delta \ominus B \qquad (2)$$

Where I is the gray scale image. This sequence of operation will potentially remove all the platelets and noises in the image. An annular ring structuring element was chosen in order to avoid the lighter centre patch of the RBC attenuating the intensity of the filtered (closed) grayscale image (Iε) as would occur if a conventional disk is used. Fig 3 shows the result of the morphological operation in a given image.



a b c

Figure 3. Morphological filtering using dilation followed by erosion.(a)Original Image (b)Gray level image (c) Iε

B. WBC and RBC Detection Algorithm

As demonstrated in Fig 3 the closed image (Iε) has all the foreground blood cells with uniform intensity distribution. These cells are detected using the ARR transform method which calculates the ratio of outer intensity to the inner intensity by means of an annular ring structuring element. The peak detection algorithm applied on a ratio transformed image will provide the location of the centroid of each cell. The next task is to differentiate the WBC and RBCs. A WBC detection algorithm has been explained in [7] which estimate the presence of the WBC if the mean pixel intensity and variance falls within a threshold range. This method is effective for post treatment malaria diagnosis, where the parasitemia which is a measure of blood infection is comparatively low. However, for a heavily infected blood image, there is a risk of false positive prediction of WBC. Hence a modified WBC detection algorithm has been introduced in this paper.

A closed image (Iε) of a blood film with WBC is shown in Fig 4(b). Since the WBC radius is greater than normal RBCs the closing operation will not remove it and will be detected by ARR method. The WBC detection algorithm is developed such that the mean intensity of the region connected to each pixel of the closed image at the location provided by the ARR transform is calculated. Any region with an average intensity less than a selected threshold will be diagnosed as a WBC. Let (x, y) be the coordinates of each pixel. The region of Iε surrounding the coordinate (x, y) will be considered as WBC if

$$\mu\ (A(x, y) < \mathrm{WBC}_{\mathrm{thresh}} \qquad (3)$$

Where A(x, y) is the area surrounding each coordinates of the image Iε and $\mu\ (A(x, y))$ is the mean (average) intensity of the area A(x, y). $\mathrm{WBC}_{\mathrm{thresh}}$ is the threshold value manually evaluated by experimenting on a set of different images. Once the WBCs are detected the area surrounding the rest of the located pixel coordinates are considered to be RBCs. Fig 4(a-d) demonstrates the 4 different steps involved in the algorithm.

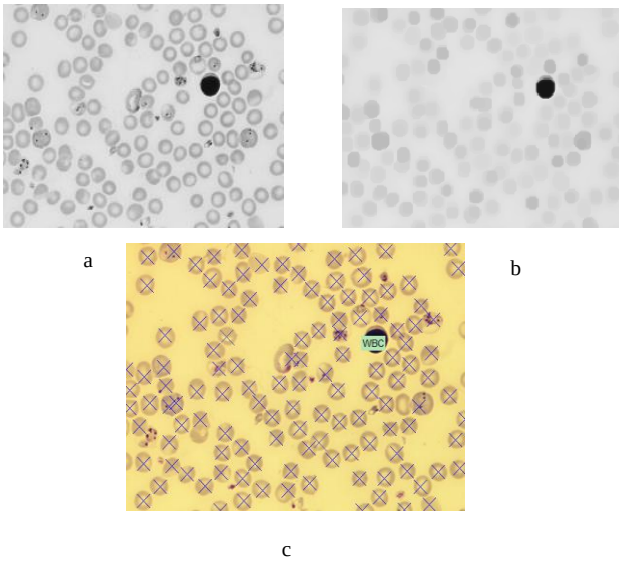


Figure 4. Blood cell detection.(a)Gray scale image (b) Iε (c) RBC and WBC differentiated and marked

III. INFECTED CELL IDENTIFICATION

The infected cells are RBCs having stained nuclei of malaria parasite. Depending on the maturity of the parasite RBCs will contain 1-8 nuclei within them. Smaller size, circular shape and saturated color are the distinctive features of these nuclei. Two classification methods based on these features are described in the following section; one based on morphology and the other based on intensity variation.

A. Identification based on size and morphology

Morphologically, an infected cell is a connected group of pixel in an RBC region, with circular stained group of pixels. The method proposed in this section utilizes modified ARR method to detect these circular objects. Hence in a lighter background of RBC, defined by its area, the peak detection algorithm picks up the centroid of the nuclei/chromatin dots of the parasite. The radius of the annular ring structuring element depends on the radius of the nuclei which is normally 2-3 pixels. Since the centroid location of the RBC is already known (see section II), the ring ratio will be performed in the neighborhood of the RBC center. This will eliminate any risk of false detection of platelets /artifacts that are present outside the RBC radius as parasitic nuclei.

B. Identification based on Intensity variation

Infected cells are highlighted as purplish dark object in the image even though the visual extent of staining varies according to the lighting and imaging conditions. The color features are very well differentiated in Hue-Saturation-Value color space that in gray level. Hence to detect bright object within an RBC, the hue component of the original image is used. The method performs image binarisation using thresholding of the hue image and search for the localized maxima. The search is defined in terms of RBC radius surrounding the already located centroid. Following Fig 5 shows the Hue component and its binary image used in the process.

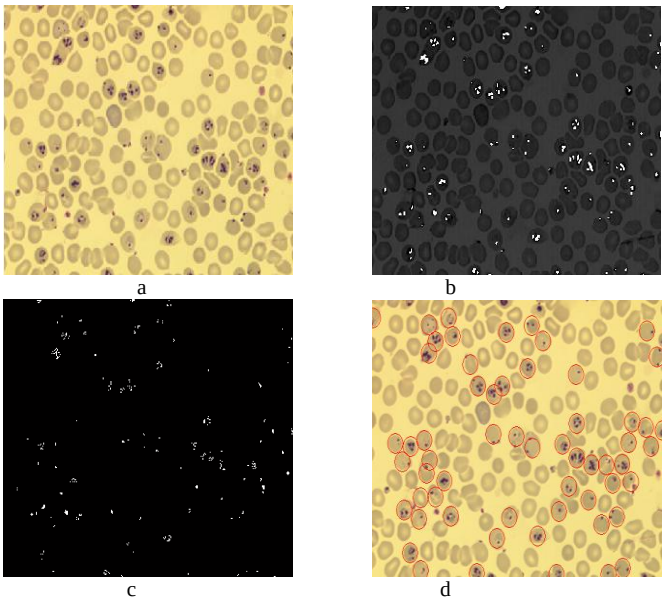


Figure 5. Infected cell detection (a) Original image (b) Hue component (c) Binary image (d) Infected cell identification

IV. EXPERIMENTAL RESULTS

The images used for experimental analysis were obtained from National Institute of Medical Research (NIMR), UK. The method was experimented in MATLAB environment on a set of images with a range of magnifications and apparent variations in stain color as well as lighting and other imaging conditions. The resolution of the image is modified and reduced to 512 x 381 pixels TIFF image to speed up the performance.

More than 200 images were manually evaluated in order to set up the threshold values to detect the peaks in ARR method, image binarisation as well as the $WBC_{thresh.}$ Also histogram evaluations of these images were carried out to finalize the values. The information regarding the size distribution of different components in the image is extremely crucial for the proposed methods. Area granulometry was applied on these entire images to acquire the cell size (radius) information [6].

A performance measurement of the two parasite detection methods were conducted on a set of 200 images carrying around 40000 RBCs. The accuracy of the system is defined to be a measure of True positive (TP-positive result for a positive sample), True negative (TN-negative result for a negative sample), False positive (FP-positive result for a negative sample) and False negative (FN-negative result for a positive sample) and is given as:

$$Acc = TP + TN / (TP + TN + FP + FN) \quad (4)$$

The outcomes of the performance measurement estimate the Sensitivity, Specificity, Positive prediction value (PPV) and Negative prediction value (NPV). Following table provides a comparison of the performance measurement of the two parasite identification techniques.

TABLE I. PERFORMANCE OUTCOME

Outcome	**Method used**	
	I. Parasite Identification using size and morphology	***II.Parasite Identification by means of Intensity variation***
Sensitivity	85%	90%
Specificity	89%	86%
PPV	90%	87%
NPV	84%	90%
Overall Accuracy	87%	88%

These are the experimental outcomes of the analysis of the parasite classification routine. However, it would be appropriate to validate the results with experts in the medical field. The overall accuracy is a measure of miscalculations for a given set of samples. Even though Method II is slightly more accurate than Method I, other factors has to be considered in terms of hardware complexity. Method I work on modified annular ring ratio on gray scale images whereas Method II looks for local maxima of a binary image converted from the hue component of the original image. Our experiments shows that hue component of a stained blood image is invariant to highlighting and shading and holds more information compared to the gray level. But method I perform faster than method II in MATLAB environment. More work has to be carried out to evaluate the complexity of each method in terms of hardware implementation.

V. CONCLUSION

An automated malaria diagnostic system in peripheral Giemsa stained blood sample images is addressed in this paper. The core of the system is the ARR method which is insensitive to illumination and color constancy. The system can be used as an effective Red blood cell and white blood cell detection tool for any blood film analysis outside the spectrum. The results show an accuracy rate above 85%, with a sensitivity of 90%. The future work in the area aims to provide a better and controlled diagnosis performance compared to manual laboratory analysis in a real time diagnosis scenario.

REFERENCES

[1] WHO: Global report on antimalarial efficacy and drug resistance: 2000-2010

[2] K.N.R.M.Rao,,"Application of Mathematical Morphology to Biomedical Image Processing," PhD thesis , 2004

[3] C.Di Ruberto, Andrew G. Dempster, S.Khan, and B.Jarra,"Automatic thresholding of infected blood images using granulometry and regional extrema,"ICPR 2000,pages3445-3448

[4] C.Di Ruberto, Andrew G. Dempster, S.Khan, and B.Jarra,"Analysis of infected blood cell images using morphological operators,"Image and vision computing,vol 20,no.2,pp.133-146,February 2002

[5] Boray Tek, Andrew G. Dempster,Izzet Kale."Computer Vision for microscopy diagnosis of malaria".Malaria Journal 2009..

[6] S.Kareem,R.C.S Morling, I Kale" A Novel Method to Count the Red Blood Cells in Thin Blood Films",IEEE International Symposium on Circuits and Systems (ISCAS), 2011, Pages 1021-1024

[7] S.Kareem, I Kale, R.C.S Morling," Automated P.falciparum Detection system for Post-treatment Malaria Diagnosis using Modified Annular Ring Ratio Method",IEEE 14th International Conference on Computer Modelling and Simulation , March 2012