33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Ontology-based Malaria Parasite Stage and Species Identification from Peripheral Blood Smear Images

Vishnu V. Makkapati and Raghuveer M. Rao

***Abstract*— The diagnosis and treatment of malaria infection requires detecting the presence of the malaria parasite in the patient as well as identification of the parasite species. We present an image processing-based approach to detect parasites in microscope images of a blood smear and an ontology-based classification of the stage of the parasite for identifying the species of infection. This approach is patterned after the diagnosis approach adopted by a pathologist for visual examination, and hence, is expected to deliver similar results. We formulate several rules based on the morphology of the basic components of a parasite, namely, *chromatin dot(s)* and *cytoplasm*, to identify the parasite stage and species. Numerical results are presented for data taken from various patients. A sensitivity of 88% and a specificity of 95% is reported by evaluation of the scheme on 55 images.**

I. Introduction

Malaria is a highly prevalent disease affecting millions of people in many parts of the world with several thousands of deaths attributed to it every year. Several methods exist to diagnose malaria but microscopic observation of blood smears is considered the *gold* standard. The blood smears are stained to highlight the parasites, with Giemsa and Leishman being the popularly used dyes. The determination of the species of malaria is critical to the treatment of the disease.

The smears are manually observed by a pathologist, and hence, the quality of the diagnosis varies depending on the skill. Inaccurate diagnosis can lead to wrong treatment and can also result in death of the patient. A pathologist typically observes around 100 High Power Fields (HPF) to conclude the presence/absence of infection, and the time taken per patient will vary between 15-20 minutes. This leads to fatigue, especially in peak infection seasons, resulting in wrong diagnosis. To overcome these problems, methods for automatic smear analysis based on image processing have been proposed recently [1], [2].

The method proposed in [3] extracts features such as histogram, Hu moments, and relative shape measurements to detect parasite-infected Red Blood Cells (RBC). A K-Nearest Neighbor classifier is used with an appropriate distance measure to evaluate the performance of individual and concatenated features. It was concluded that the concatenated features do not necessarily yield better performance than the individual features. Color-based features were found to be useful to discriminate between parasite and non-parasite RBCs; however, no attempt was made to classify the species of the parasite.

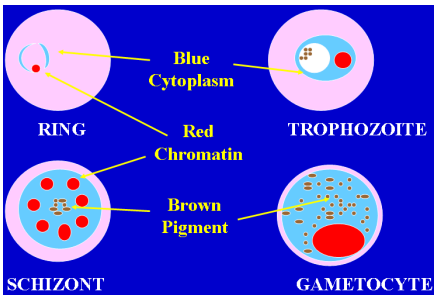


Fig. 1. Recognizing Erythrocytic Stages: Schematic Morphology [5].

The scheme presented in [4] counts the number of malaria parasites in a given image. The input images are enhanced for contrast using histogram equalization. Edge correlation is used to identify the boundary edges of RBCs. A binary mask is used to detect the interior regions of RBCs. A binary filling operation is first performed on the boundary contours of RBCs, which are then morphologically eroded to obtain the inner regions of the RBCs where parasites may be located. However, even here, no method is proposed to identify the species of the parasite.

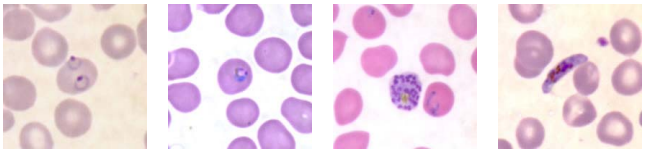
The technique discussed in [6] determines the pattern spectrum using granulometry and quantifies RBC characteristics. This information is then thresholded to identify RBCs and potential parasites. Features based on color, shape, and morphology are extracted from infected RBCs. The features are classified using one neural network to determine valid infected RBCs. Another neural network is used to classify the species of malaria.

The scheme discussed in [7] is based on a multi-stage estimation process with minimal prior knowledge starting from a representation of RBCs. It compensates for imaging variability by using a correction process on the intensity component in HSI space. An elliptic template is designed and used to extract RBCs from an image.

In this paper we propose digital image processing methods [8] that detect the parasite stage and identify the species by mimicking the steps used by a pathologist for diagnosing malaria from the microscope examination of blood smears. This method, based on ontological description, uses the morphology of the two basic components of a parasite, namely, *chromatin dot(s)* and *cytoplasm*. Several rules are proposed to detect the stage of the parasite first and then use

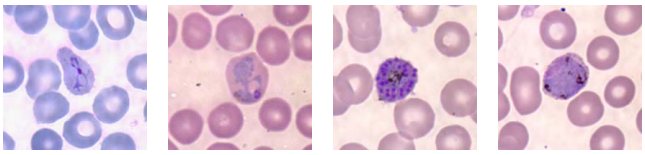
V.V. Makkapati is with Philips Research Asia - Bangalore, Philips Innovation Campus, Philips Electronics India Limited, Manyata Tech Park, Nagavara, Bangalore - 560 045, Karnataka, India vishnu.makkapati@philips.com

R.M. Rao is with the U.S Army Research Laboratory, RDRL-SES-E, 2800 Powder Mill Road, Adelphi, MD 20783-1197, USA raghuveer.rao@arl.army.mil



(a) Ring (b) Trophozoite (c) Schizont (d) Gametocyte

Fig. 2. *P. falciparum*: Stages found in blood [9].



(a) Ring (b) Trophozoite (c) Schizont (d) Gametocyte

Fig. 3. *P. vivax*: Stages found in blood [9].

it for species classification.

II. Background

The stage and species of a parasite can be described based on the appearance of chromatin dots and cytoplasm. The presence of a chromatin dot is the primary cue for detecting infected RBCs. However, several artifacts may also be detected as chromatin dots. Hence, a pathologist uses multiple cues based on chromatin dot(s) and cytoplasm for malaria diagnosis.

Figure 1 shows a schematic of various stages of the parasite and some sample images are given in Fig. 2 and Fig. 3. The trophozoites (early form known as *ring*) are identified using the presence of a chromatin dot and associated cytoplasm in the form of a ring. Schizonts are usually detected based on the dispersion of the chromatin dots. The gametocytes are classified as *macro* or *micro* depending on whether the chromatin is compact or diffused, respectively. The analysis of cytoplasm is usually not considered in schizont and gametocyte stages.

Four species of malaria namely *P. falciparum*, *P. ovale*, *P. vivax*, and *P. malariae* infect human beings. Of these, *P. falciparum* and *P. vivax* account for more than 99% of the infections, and hence, we focus our attention on identifying them. Table I describes the appearance of the parasite and we use it to formulate our ontology-based method to identify the stage and species of a parasite.

III. Proposed Method

We present a technique based on ontological description to identify various stages and species of parasites. This approach attempts to mimic a pathologist, and thus, is expected to yield similar results. Chromatin dots are more visibly distinguishable than cytoplasm and provide important cues about the stage of the parasite. Therefore, we segment chromatin dots and analyze them to identify the parasite stage.

The green component of an image offers maximum separation between the background and RBC pixels. The green component is first segmented to detect the RBCs by choosing an optimal threshold [10]. The pixels belonging to chromatin dots have higher saturation values and are segmented by thresholding the saturation component of the image. The segmented RBCs and chromatin dots are given unique labels. However, several artifacts may also be segmented as chromatin dots. We evaluate whether a chromatin dot is valid by identifying whether it is inside a RBC or not.

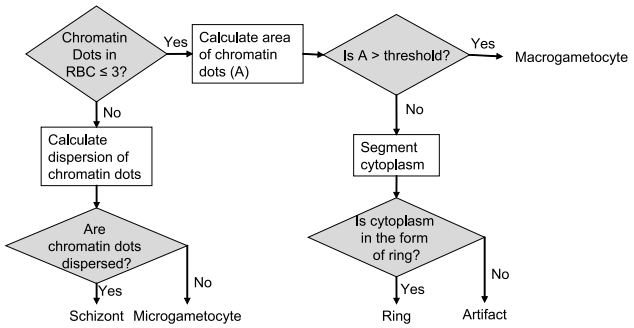


Fig. 4. Flowchart of the parasite stage identification algorithm.

A. Parasite Stage Identification

First, we count the number of chromatin dots in a RBC. The presence of at most three chromatin dots in a RBC suggests the presence of a ring or macrogametocyte stage of the parasite. If the number of chromatin dots is more than three, it suggests that the stage is either a schizont or microgametocyte. We choose the appropriate parameters to distinguish the parasite stages in either of these two cases. Figure 4 shows the flowchart of the proposed method.

Multiple rings may be observed in a *P. falciparum* infection but the total area of the chromatin dots in a RBC will be smaller as compared to that of a macrogametocyte. Thus, an area threshold is used to distinguish them. To validate the rings further, we segment the cytoplasm and check if it is in the form of a ring. The dominant hue range in a RBC will indicate the color of the RBC and the remaining pixels will typically belong to the parasite. We divide the entire hue range of 360^o into six segments, where each segment has a width of 60^o and is centered around a color type as given in Table II. Then, we identify the segment with the maximum number of pixels and declare it as the dominant hue segment.

The pixels that do not belong to the dominant hue segment are marked as parasite pixels. The cytoplasm pixels have smaller saturation values as compared to the chromatin dot pixels, and a threshold determined using [10] is used to segment them. The cytoplasm is declared to be valid by verifying whether an associated chromatin dot is present inside it (i.e., if the cytoplasm is in the form of a ring, the distance between the centroids of chromatin dot and cytoplasm should be less than the radius of cytoplasm). Figure 6 shows an RBC with a ring stage parasite and the histogram of saturation values of its parasite pixels. A threshold value of 0.3412 is chosen using [10] to segment the

TABLE I

APPEARANCE OF PARASITE [9]

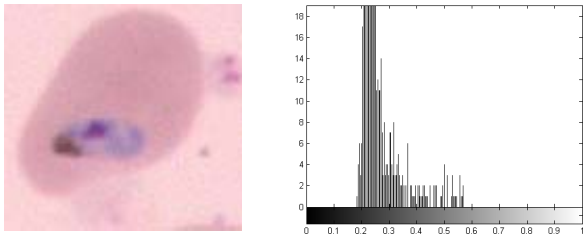
Stage	P. falciparum	P. vivax
Ring	delicate cytoplasm; 1-2 small chromatin dots; occasional applique (accolle) forms	large cytoplasm with occasional pseudopods; large chromatin dot
Trophozoite	seldom seen in peripheral blood; compact cytoplasm; dark pigment	large ameboid cytoplasm; large chromatin; fine, yellowish-brown pigment
Schizont	seldom seen in peripheral blood; mature = 8-24 small merozoites; dark pigment, clumped in one mass	large,may almost fill RBC; mature = 12-24 merozoites; yellowish-brown, coalesced pigment
Gametocyte	crescent or sausage shape; chromatin in a single mass (macrogametocyte) or diffuse (microgametocyte); dark pigment mass	round to oval; compact; may almost fill RBC; chromatin compact, eccentric (macrogametocyte) or diffuse (microgametocyte); scattered brown pigment

Fig. 5. Hue scale [11], [12]

TABLE II

HUE ANGLE AND COLOR TYPE

Hue Angle	Color Type
0^o	red
60^o	yellow
120^o	green
180^o	cyan
240^o	blue
300^o	magenta



(a) (b)

Fig. 6. (a) RBC infected with ring stage parasite. (b) Histogram of saturation values of parasite pixels.

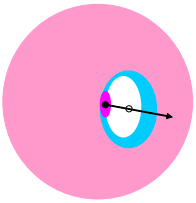


Fig. 7. Schematic showing calculation of thickness of cytoplasm. Solid and hollow dots denote the centroids of chromatin dot and cytoplasm, respectively.

cytoplasm. Figure 8 shows a sample image and its segmented chromatin dots and cytoplasm.

The dispersion and total area of the chromatin dots is more in schizonts than in microgametocytes. To reduce complexity we use the centroid of a chromatin dot as its representative to calculate the dispersion. We use measures of dispersion such as range and mean deviation to analyze the chromatin dots. These parameters are thresholded to distinguish schizonts from microgametocytes.

B. Parasite Species Identification

We analyze each stage of the parasite independently to determine the species (Table I). Schizonts are rarely spotted in *P. falciparum* (Fig. 2(c)) and their presence will suggest *P. vivax* infection (Fig. 3(c)). The gametocytes of *P. falciparum* are crescent shaped (Fig. 2(d)), while those of *P. vivax* are round (Fig. 3(d)). The eccentricity e of an RBC infected by a gametocyte is determined as

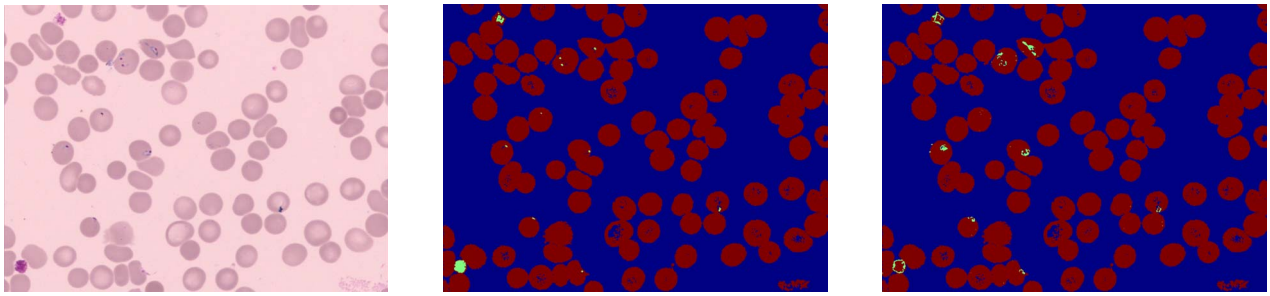
$$e = \sqrt{1 - \frac{b^2}{a^2}} \tag{1}$$

where a and b denote the lengths of the semi-major axis and semi-minor axis, respectively. The eccentricity of *P. falciparum* gametocytes will be more as compared to *P. vivax* and we choose an appropriate threshold to distinguish them.

We conducted experiments to identify species from the ring stage. The rings of *P. vivax* are usually thicker (Fig. 3(a)) than those of *P. falciparum* (Fig. 2(a)). Multiple rings in a RBC suggest the presence of a *P. falciparum* infection. In the case of a single ring being found, we threshold the area and thickness of the cytoplasm. Figure 7 shows the schematic for calculation of the thickness of cytoplasm. We determine the thickness of the cytoplasm in the direction of the line joining the centroids of the chromatin dot and cytoplasm. If P_c and P_y denote the centroids of the chromatin dot and cytoplasm, respectively, then the equation of this line can be written as

$$P(t) = tP_y + (1-t)P_c \tag{2}$$

where $P_c = P(0)$ and $P_y = P(1)$. The points $P(t), t > 1$ will correspond to the points on this line away from the centroid of the cytoplasm. The parameter t is incremented in steps of $\frac{1}{d(P_c,P_y)}$ until a pixel belonging to the cytoplasm is hit where $d(P_c, P_y)$ is the Euclidean distance between points P_c and P_y. The parameter t is further incremented until a pixel not belonging to the cytoplasm is hit. The distance between these pixels is taken as the thickness of the cytoplasm. However, we found it difficult to detect species from rings due to fragmented cytoplasm (Fig. 8(c)) and the presence of mature trophozoite stages of the parasite (Fig.



(a) Sample Image (b) Chromatin Dots (c) Cytoplasm

Fig. 8. Sample image and its segmented chromatin dots and cytoplasm where blue, red, and green colors indicate the pixels corresponding to background, RBCs and parasites respectively.

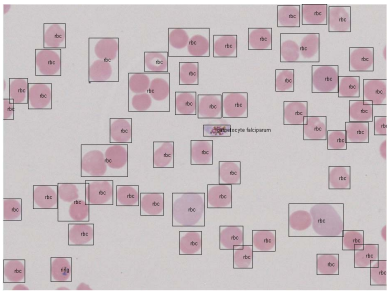


Fig. 9. Classification of a sample *P. falciparum* image.

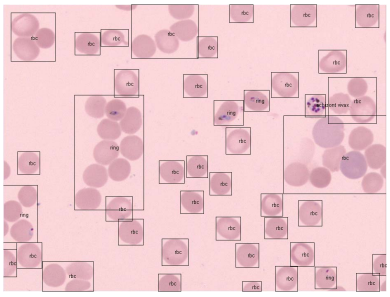


Fig. 10. Classification of a sample *P. vivax* image.

2(b) and Fig. 3(b)).

IV. Numerical Results

We evaluated our scheme using 55 images taken from Lieshman-stained smears of malaria infected patients. There was noticeable variation in the color of the images due to the different pH values of the buffer used. We reported a sensitivity of 88% and a specificity of 95% for parasite stage identification. We were able to identify the species accurately from schizont and gametocyte stages but species identification from rings was found to be difficult. Figs. 9 and 10 show that rings are detected correctly and that classification of species from schizonts and gametocytes is accurate.

V. Conclusion

We proposed a digital image processing method that mimics a pathologist to identify the stage and species of a malaria parasite. This method is based on the ontological description of the morphology of the chromatin dot(s) and cytoplasm of a parasite. Numerical results obtained using various data sets have revealed that the method performs well for stage identification. Species identification from the schizont and gametocyte stages was shown to possess a high degree of accuracy. The approach has significant potential in assisting health personnel in the rapid diagnosis of malaria.

VI. Acknowledgements

The authors would like to thank clinical partners Prof. Raviraj V. Acharya and Dr. Chethan Manohar of Kasturba Medical College, Manipal University, India for providing the data.

References

[1] V. Makkapati and R. M. Rao, "Segmentation of malaria parasites in peripheral blood smear images," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, April 2009, pp. 1361 – 1364.

[2] V. V. Makkapati and S. Naik, "Clump splitting based on detection of dominant points from contours," in *Proc. IEEE Conference on Automation Science and Engineering*, Bangalore, India, August 2009, pp. 197 – 201.

[3] F. B. Tek, A. G. Dempster, and I. Kale, "Malaria parasite detection in peripheral blood images," in *Proc. British Machine Vision Conference*. Edinburgh: British Machine Vision Association, September 2006, p. CDROM.

[4] S. W. Sio, W. Sun, S. Kumar, W. Z. Bin, S. S. Tan, S. H. Ong, H. Kikuchi, Y. Oshima, and K. S. Tan, "Malariacount: An image analysis-based program for the accurate determination of parasitemia," *Journal of Microbiological Methods*, vol. 68, no. 1, pp. 11 – 18, January 2007.

[5] "Malaria diagnostics," http://chsr.aua.am/malaria/eng/docs /learning/presentations/armenia%202002Diagnostics.2.ppt.

[6] N. E. Ross, D. M. Rubin, and C. J. Pritchard, "Automated diagnosis of malaria and other infections," EP Patent 1 787 156A1, May 23, 2007.

[7] S. Halim, T. Bretschneider, Y. Li, P. Preiser, and C. Kuss, "Estimating malaria parasitaemia from blood smear images," in *Proc. International Conference on Control, Automation, Robotics and Vision*. IEEE, 2006, pp. 648 – 653.

[8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Prentice Hall, 2002.

[9] "Dpdx - malaria," http://www.dpd.cdc.gov/dpdx/HTML/ Malaria.htm.

[10] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62 – 66, 1979.

[11] A. Munsell, *A Color Notation*. Baltimore, MD: Munsell Color Corporation, 1946.

[12] "Hue," http://en.wikipedia.org/wiki/Hue.