This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference.

Image Analysis for Malaria Parasite Detection from Microscopic Images of Thick Blood Smear

Ishan R. Dave
S.V. National Institute of Technology, Surat, India
Email: ishandave95@gmail.com

***Abstract*—Malaria is a major health issue and causes millions of deaths a year worldwide. Since the diagnosis of malaria is predominately done using light microscopy method, well-trained microscopists are required. By using thick blood smear, a large amount of blood can be examined quickly and easily. This work deals with the automatic estimation of parasite density in 'parasites per microliter of blood' from the microscopic images of Giemsa-stained thick blood smear. The algorithm is primarily divided into three steps: (1) preprocessing & segmentation, (2) feature extraction and (3) classification. In the preprocessing step, an attempt is made to reduce variations due to various factors like lighting conditions and concentration of staining solution. The image is segmented using adaptive thresholding, followed by several mathematical morphological operations. In the second step, various features based on shape, texture, color and frequency domain are extracted. Using the classification step, the parasite candidate is classified into its correct life stage or classified as leukocytes. The novelty of the algorithm is that it can detect all the life stages (ring, trophozoite, schizont, gametocyte) of parasites and leukocytes unlike detecting only ring life stage in the state-of-the-art algorithms. The discrepancy in the automated parasite count by the proposed algorithm is 7.14%, which is suitable for computer aided diagnosis (CAD) of malaria according to world health organization (WHO) quality control standards.**

***Index Terms*—Medical Image Analysis, Malaria, Computer Aided Diagnosis (CAD), Thick Blood Film.**

I. Introduction

Malaria is one of fatal diseases worldwide and can lead to death rapidly if it is untreated. In 2015, 214 million cases of malaria were estimated, which caused around 438 000 deaths [1]. Malaria is prevalent in the tropical and subtropical regions near the equator like Asia, Sub-Saharan Africa, and Latin America. The disease is caused by Genus Plasmodium parasites and transmitted by the female anopheles mosquito. There are mainly four species of Genus Plasmodium that infect humans are: Plasmodium Vivax, Plasmodium Falciparum, Plasmodium Ovale and Plasmodium Malariae. Among these species, Plasmodium Vivax causes significant health issues worldwide. According to world health organization (WHO), there were total 13.8 million cases were estimated globally in 2015. Since 2010, WHO has suggested malaria diagnostic testing by either light microscopy or rapid diagnostic test (RDT) [1]. RDT is used where microscopy or well-trained microscopist is not available; this method gives instant results but main disadvantage of RDT is that parasite density cannot be estimated. Microscopy method is "Gold Standard" for detecting malaria parasites and estimating parasite density [2].

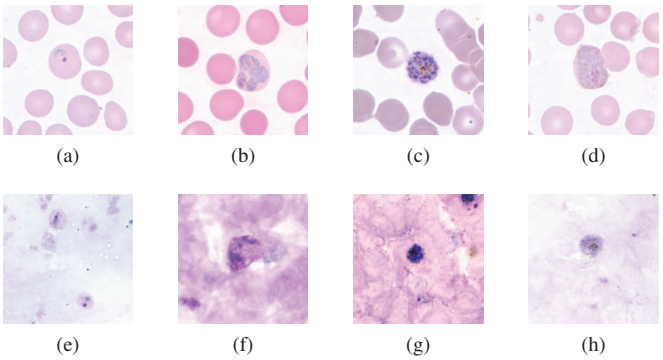


Fig. 1. Plasmodium vivax life stage in thin and thick smear (a) ring form in thin smear (b) trophozoite form in thin smear (c) schizont form in thin smear (d) gametocyte form in thin smear (e) ring form in thick smear (f) trophozoite form in thick smear (g) schizont form in thick smear (h) gametocyte form in thick smear.

There are two types of blood films prepared for microscopic diagnosis: thin and thick [2]. A thin blood smear is used to identify species of the parasite, as the appearance of the parasite is retained. A thick blood smear is used to detect the presence of parasite and parasite density. A thick blood smear is more than ten times sensitive than a thin blood smear, as it allows examination of a larger volume of blood [2]. The appearance of the parasite is distorted in thick blood smear, so it is hard to identify species and life stages of parasite [3]. In the routine examination of blood films a person needs to count all the life stages of parasites and leukocytes from at least 100 microscopic fields (oil immersion, $100\times$ objective, $10\times$ eyepiece) in order to estimate parasite density in 'parasites per microliter of blood', which can be found from following Eq. (1) [2].

$$\frac{\text{Number of parasites counted} \times 8000}{\text{Number of leukocytes}} = \text{parasites per microliter.} \quad (1)$$

It is very exhaustive, error-prone and time-consuming process. It takes about 10 minutes of expert to examine 100 oil immersion fields [2]. The results in this examination may vary due to the human discrepancy. Parasite of a species is further classified according to its life stages: ring, trophozoite, schizont, gametocyte. In P. vivax, a range of trophozoite and schizont stages is mostly seen [2]. All forms of P. Vivax Parasite are shown in Fig. 1.

978-1-5090-4442-9/17/$31.00

An overview of the approaches to the computer aided diagnosis (CAD) is provided in the following. Tek et al. [4] proposed an approach for detection and classification of malaria parasites and species from thin blood films. Features based on color, local granulometry and shape were extracted from the region of interest (ROI) and k nearest neighbor (kNN) classifier was used in order to classify from extracted features. Frean [5] proposed a novel method to detect malaria parasites from thick blood smear based on open access software ImageJ. Frean used parasite size frequency distributions as statistical parameters in particle counting algorithms of the digital image analysis program in order to estimate parasite density, but the method doesn't give desired output for parasite density lower than six parasites per image. Kaewkamnerd et al. [6] proposed an algorithm to classify species (P. Vivax and P. Falci.) from the thick blood smear with classification success rate of approximately 60%, but the algorithm needs further validation as it was tested on 20 images only. Arco et al. [7] proposed an algorithm based on contrast limited adaptive histogram equalization (CLAHE), adaptive thresholding and morphological process in order to detect malaria parasites from gray scale image of a thick blood smear. The limitation of [7] is, it can detect only ring life stages of parasites based on size only; artifacts or platelets of the size of parasites can be misclassified as a parasite. Elter et al. [8] proposed a two-stage algorithm in order to detect ring form of P. Falciparum from thick smear images. But, in [8] only ring life stage of parasites is detected and other three forms of parasites are neglected, which may lead to wrong diagnosis. Although thick blood smear is more than ten times sensitive than thin blood smear for diagnosis of malaria [2], there is no published article that covers computer aided detection of all the life stages of malaria parasites from the thick blood smear. The proposed method detects all the forms of parasites of P. Vivax and leukocytes, which can be directly used to estimate parasite density from Eq. (1).

The organization of rest of this paper is as follows. Description of material and database is given in Section II. The methods used in preprocessing, are explained in Section III. Feature extraction from the parasite candidates is described in Section IV. Classifiacation & results are explained in Section V. The work is concluded with a discussion in Section VI. Fig. 2 shows block diagram of the proposed algorithm.

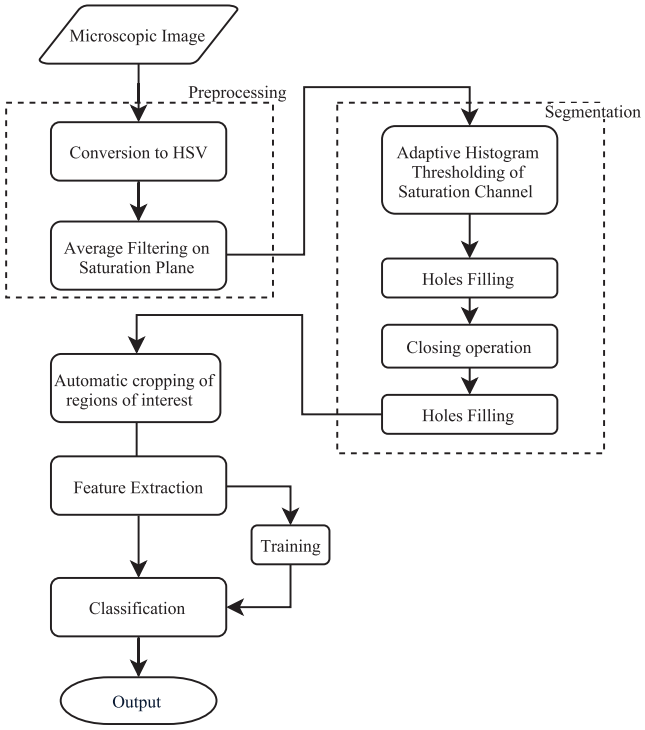


Fig. 2. Block diagram of the proposed algorithm.

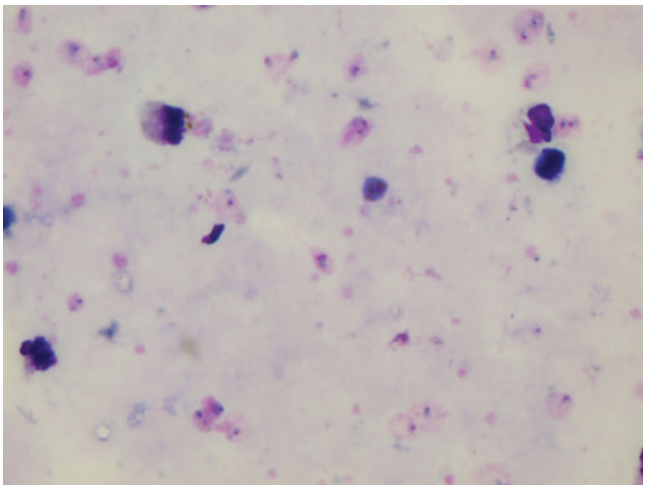


Fig. 3. Microscopic image of thick blood film.

II. Material

We acquire total 87 images of microscopic fields of 5 different Giemsa-stained thick blood films collected from distinct patients, with variation in staining solution concentration and parasite density. The images are captured from Olympus E-330 Digital Camera, attached on Olympus CX41 Microscope at $100\times$ oil immersion objective. The resolution of each image is 3136×2352 pixels and image is used in compressed jpeg form to reduce memory requirements. The dataset is randomly split into two subsets: 43 images are used for the training set and 44 images are used for the testing set. A sample captured microscopic image is shown in Fig. 3.

III. Preprocessing and Segmentation

In this section, the image is segmented into foreground (stained particles i.e. parasites of all life stages, leukocytes, platelets and many other artifacts) and background (non-stained particles i.e. liaised red blood cells). A large number of false positive (stained objects) may come as regions of interest, which will be eliminated in next sections. The captured images are converted in HSV colorspace to reduce

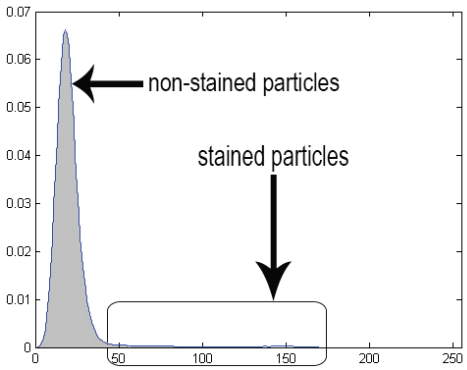


Fig. 4. Histogram of saturation channel.

(a) Output of adaptive thresholding. (b) Filled. (c) Closed + filled. (d) Filled + closed + filled.

Fig. 5. Mathematical morphological operations.

variations due to brightness and concentration of stain solution, which highly affects in RGB colorspace. The stained and non-stained particles have a significant difference in their saturation channel. The saturation channel of image is filtered through an average filter of size 3×3 in order to reduce high variations due to noise. Histogram of the saturation channel is shown in Fig. 4. Since most of the non-stained particle has low saturation channel values, the segmentation is done using adaptive histogram thresholding technique to segment stained particles from the saturation channel. As shown in Fig. 4, stained-particles' part covers very small area on X-axis of the histogram. By finding cumulative distribution function (CDF), one can note that stained-particles have CDF values higher than 0.95. The threshold level is taken where $\mathrm{CDF} = 0.95$. Using this threshold level saturation channel image is binarized as shown in Eq. (2).

$$\text{Pixel value} = \begin{cases} \mathbf{1}, & \text{if saturation value} > T \\ \mathbf{0}, & \text{if saturation value} < T \end{cases} \quad \text{where, } \mathrm{CDF}(x = T) = 0.95. \tag{2}$$

The output of adaptive histogram thresholding is shown in Fig. 5(a). Mathematical morphological operations are used to extract the desired components from the segmented binary image. Holes are filled in order to eliminate small background areas inside the foreground contours (Fig. 5(b)). It is followed by closing operation to fill the small gaps between components of the same particle (Fig. 5(c)). Closing operation is performed using a disk shaped structuring element of radius 3. Holes are filled again to eliminate background pixels inside foreground contours (Fig. 5(d)). The results of morphological operations are shown in Fig. 5. The output after morphological operations is shown in Fig. 6.

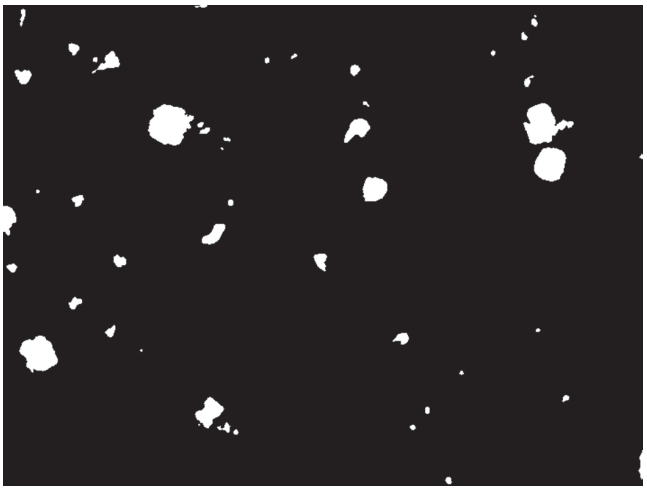
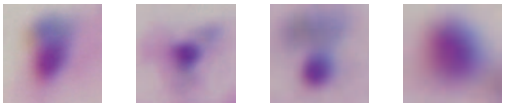
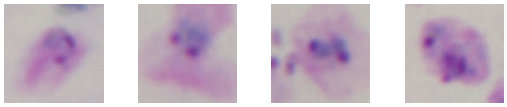


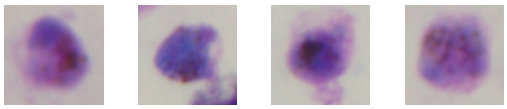
Fig. 6. Output of preprocessing and segmentation step.



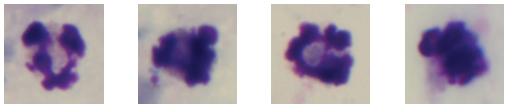
(a) Class-1: ring life stage.



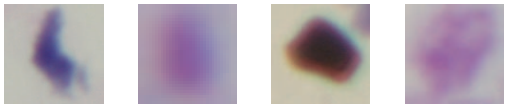
(b) Class-2: Early trophozoite life stage.



(c) Class-3: Late trophozoite life stage.



(d) Class-4: Leukocyte (WBC).



(e) Class-5: Non-parasite.

Fig. 7. Cropped regions of interest.

IV. Feature Extraction

8-Connected components of the output of image segmentation (Fig. 6), are labeled. Regions of interest are cropped from the original image according to size and location of labels. Each ROI is square with centroid at centroid of the label. Fig. 7 shows ROIs of different size. Different kinds of features based on shape, texture and color are extracted

TABLE I
EXTRACTED FEATURES.

Feature	Details	Type	Input image	No. of features
Central moments	Mean, variance, skewness, kurtosis	Statistical moment features	Grey scale	4
Hu's invariant moments [9]	Moments up to 7th order	Statistical moment feature	Grey scale	7
Zernike's moment [10]	Moments up to 12th order	Statistical moment feature	Grey scale	49
Chen's geometrical statistical features [11]	Features for NOC1, NOC0, IRGL1, IRGL0	Texture (Geometry) based	Grey scale	16
Haralick's co-occurrence matrix [12]	Features from GLCM matrix with offset of 0°, 45°, 90°, 135°	Texture (Intensity) based	Grey scale	24
Hue features	Features form sampled hue histogram	Color features	Hue plane	22
Area feature	Area of output after high pass filtering	Frequency domain	Grey scale	1

from the ROIs. Extracted features are shown in Table I. Total 123 features are extracted from each ROI to classify it in correct class. Area feature is obtained from frequency domain after passing gray scale ROI to a high pass filter (gaussian, standard deviation = 1), followed by global threshold to get binary image. Some stained objects (non-parasite) of low-frequency components are removed successfully using this feature.

V. CLASIFICATION AND RESULTS

Regions of interest are classified in five classes: ring form, early trophozoite (ET) form, late trophozoite (LT) form, leukocyte and non-parasite (artifact, platelets and other stained particles) as shown in Fig. 7. Since it is very hard to discriminate schizont and gametocyte (shown in Fig. 1(g) and 1(h)), both will be considered under a single class 'Late Trophozoite'. Cubic support vector machine (SVM) classifier is used for classification. The classifier is trained using all 123 features of 2323 ROIs (extracted from 43 training data images) and tested on 2290 ROIs (extracted from 44 testing data images). There is no overlap between training dataset and testing dataset. Table II shows 5-class confusion matrix after classification of 2290 testing ROIs. In the confusion matrix, by considering ring, ET and LT under a same class of 'parasite', one can find the parasite count by the proposed algorithm and manual parasite count. The discrepancy of parasite count by the proposed algorithm is 7.14%, which is in the limits by WHO quality control [13]. The performance measurement of classification is done using sensitivity and specificity. Sensitivity and specificity can be found from true positives (TP), false positives (FP), true negative (TN) and false negative (FN) from the following Eqs. (3) and (4), respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

TABLE II
5-CLASS CONFUSION MATRIX, USING ALL 123 FEATURES FOR TRAINING.

	Ring	ET	LT	WBC	Artifact
Ring	201	5	0	0	24
ET	8	22	1	0	13
LT	0	1	40	1	6
WBC	0	0	5	62	4
Artifact	46	11	5	1	1834

TABLE III
ACCURACY IN PARASITE COUNT USING DIFFERENT NUMBER OF FEATURES FOR TRAINING.

No. of features used for training	Sensitivity	Specificity
31 (25% of 123)	0.6770	0.9360
62 (50% of 123)	0.7764	0.9492
92 (75% of 123)	0.7840	0.9619
123 (100% of 123)	0.8634	0.9660

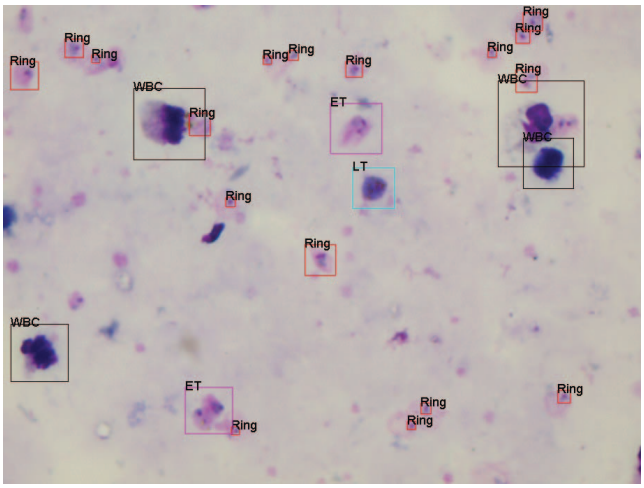


Fig. 8. Final output image annotated by the proposed algorithm.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{4}$$

By reducing a number of features using principal component analysis (PCA), the accuracy in the parasite count decreases. Table III shows classification accuracy for a different number of features used for training using principal component analysis (PCA). The algorithm performs at its best when all 123 features are selected for training. Final output image, annotated by the proposed algorithm is displayed in Fig. 8.

VI. DISCUSSION

In this paper, a novel method is developed that detects all life stages of P. Vivax and leukocytes from the microscopic images of thick blood smear. The parasite count by the proposed algorithm is with a discrepancy of 7.14%, which is acceptable by the WHO quality control limits. By using the proposed method, parasite density can be estimated using all life stages (ring, trophozoite, schizont, gametocyte) of malaria

parasites, unlike state-of-the-art algorithms, that can estimate parasite count only from ring life stage.

It is worth to note that, accuracy may be improved by using more images for training, by improving slide quality and improving image acquisition method. This work can be extended for detecting different Plasmodium species and their different life stages.

Acknowledgements: Author would like to acknowledge the help supported by:

1. Dr. Nitin Khanna, Department of Electrical Engineering, IIT-GN
2. Team of Summer Research Internship Program (SRIP-16), IIT-GN
3. Dr. Pranav Desai, Desai Metropolis, Surat
4. Dr. Sunil Patel, Sarvoday Lab, Gandhinagar
5. Dr. Hansa Goswami, BJMC, Ahmedabad
6. Mr. Sharad Joshi, Department of Electrical Engineering, IIT-GN.

REFERENCES

[1] W. H. Organization, *World malaria report 2015*. World Health Organization, 2015.

[2] W. H. Organization and C. for Disease Control, *Basic Malaria Microscopy: Tutor's guide*. World Health Organization, 2010.

[3] D. Warhurst and J. Williams, "Acp broadsheet no 148. july 1996. laboratory diagnosis of malaria," *Journal of clinical pathology*, vol. 49, no. 7, p. 533, 1996.

[4] F. B. Tek, A. G. Dempster, and İ. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 21–32, 2010.

[5] J. A. Frean, "Reliable enumeration of malaria parasites in thick blood films using digital image analysis," *Malaria Journal*, vol. 8, no. 1, p. 1, 2009.

[6] S. Kaewkamnerd, A. Intarapanich, M. Pannarat, S. Chaotheing, C. Uthaipibull, and S. Tongsima, "Detection and classification device for malaria parasites in thick-blood films," in *Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2011 IEEE 6th International Conference on*, vol. 1. IEEE, 2011, pp. 435–438.

[7] J. Arco, J. M. Górriz, J. Ramírez, I. Álvarez, and C. G. Puntonet, "Digital image analysis for automatic enumeration of malaria parasites using morphological operations," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3041–3047, 2015.

[8] M. Elter, E. Haßlmeyer, and T. Zerfaß, "Detection of malaria parasites in thick blood films," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 5140–5144.

[9] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[10] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[11] Y. Q. Chen, M. S. Nixon, and D. W. Thomas, "Statistical geometrical features for texture classification," *Pattern Recognition*, vol. 28, no. 4, pp. 537–552, 1995.

[12] R. M. Haralick, K. Shanmugam et al., "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 6, pp. 610–621, 1973.

[13] W. H. Organization et al., "Informal consultation on quality control of malaria microscopy: Who headquarters, geneva, 3 Mar. 2006," 2006.