J Digit Imaging (2012) 25:542–549
DOI 10.1007/s10278-011-9442-6

Image Analysis Approach for Development of a Decision Support System for Detection of Malaria Parasites in Thin Blood Smear Images

Keerthana Prasad · Jan Winter · Udayakrishna M. Bhat · Raviraja V. Acharya · Gopalakrishna K. Prabhu

Published online: 7 December 2011

Abstract This paper describes development of a decision support system for diagnosis of malaria using color image analysis. A hematologist has to study around 100 to 300 microscopic views of Giemsa-stained thin blood smear images to detect malaria parasites, evaluate the extent of infection and to identify the species of the parasite. The proposed algorithm picks up the suspicious regions and detects the parasites in images of all the views. The sub-images representing all these parasites are put together to form a composite image which can be sent over a communication channel to obtain the opinion of a remote expert for accurate diagnosis and treatment. We demonstrate the use of the proposed technique for use as a decision support system by developing an android application which facilitates the communication with a remote expert for the final confirmation on the decision for treatment of malaria. Our algorithm detects around 96% of the parasites with a false positive rate of 20%. The Spearman correlation r was 0.88 with a confidence interval of 0.838 to 0.923, $p<0.0001$.

Keywords Color image analysis · Decision support system · Telemedicine · Malaria diagnosis

K. Prasad (✉) · J. Winter · U. M. Bhat
Manipal Centre for Information Science, Manipal University,
Manipal 576104, India
e-mail: keerthana.prasad@manipal.edu

J. Winter
e-mail: jwinter@gmx.net

U. M. Bhat
e-mail: uadyakrishnam@gmail.com

R. V. Acharya
Department of Medicine, Kasturba Hospital, Manipal University,
Manipal 576104, India
e-mai: rvacharya@manipal.edu

G. K. Prabhu
Department of Biomedical Engineering,
Manipal Institute of Technology,
Manipal 576104, India
e-mail: gk.prabhu@manipal.edu

Introduction

Malaria is a serious disease prevalent in developing countries. The World Health Organization estimates 300–500 million malaria cases and a few millions of deaths per year [1, 2]. It is a leading cause of mortality and morbidity in tropical and subtropical countries. Hence, it is an important health problem which needs to be addressed.

Malaria is caused by protozoan parasites of the *Plasmodium* genus by entering the bloodstream. The four species of malaria parasites infecting humans are *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. *P. falciparum*, which is found worldwide in tropical and subtropical areas, is the only species that can cause severe, potentially fatal malaria. *P. vivax* is found mostly in Asia, Latin America and in some parts of Africa. Because of the population densities especially in Asia, it is probably the most prevalent human malaria parasite. *P. ovale* and *P. malariae* are less frequently encountered.

Many techniques have been developed to diagnose malaria such as flow cytometry, fluorescent microscopy, PCR, etc, but light microscopy remains the gold standard for laboratory confirmation of malaria. Malaria parasites can be identified by examining a drop of the patient's blood under the microscope, spread out as a "blood smear" on a microscope slide. Thick and thin smears of Giemsa-stained peripheral blood samples are used for diagnosis. Thin smears are used to assess

degree of infection and species of malaria parasite. The specimen is stained with Giemsa stain, which expresses the parasites in a distinct manner. Giemsa stain is used to differentiate nuclear and cytoplasmic morphology of platelets, red blood cells, white blood cells and parasites. Giemsa staining solution stains up nucleic acids, and therefore those components containing the nucleic acids, namely parasites, white blood cells and platelets, are highlighted in a bluish purple color. Red blood cells are usually colored in slight pink. The rings of the trophozoites take up pale blue color.

The parasitemia is measured by counting the number of parasites present in the specimen. The hematologist has to study about 100 to 300 microscopic views of the blood smear to detect the parasites and evaluate the extent of infection. This is a subjective, laborious and time consuming task. It has been reported in [3] that the agreement rates among clinical experts for the diagnosis are surprisingly low. Moreover, the accuracy of diagnosis depends on quality of reagents, microscope and on experience of the laboratorian. Malaria is a seasonal and contagious disease, and hence it needs to be brought under control as early as possible. This calls for rapid and accurate diagnosis. Also, handling huge number of cases in a short time is a challenging task. In many developing countries, microscopy is not reliable since the microscopists are not well trained and are overworked. Conversely, in endemic countries, laboratory technicians are often unfamiliar with malaria and are likely to judge the parasitemia inaccurately. Thus, it is important to automate the diagnosis of malaria parasite detection so that the burst and spread of malaria can be controlled. Automation could play a very important role in the case of mass screening for malaria. It is very important to produce a common standard tool which makes the diagnosis based on a certain criteria and thus eliminates subjective assessment and makes the procedure highly repeatable.

The recent convergence of technologies is opening up new avenues for telemedicine applications. In many parts of the world, more so in developing countries, the rural areas without sufficient healthcare infrastructure or trained professionals are often well covered by mobile phone service providers. There is a lot of scope to develop innovative low cost technologies which can have a good impact on the healthcare systems of the developing countries. There have been efforts to use mobile phone based diagnostic systems [4, 5] for diagnosis of malaria. These studies have focused their attention to acquire images of smears of the specimen which can then be sent over a telecommunication channel for analysis.

Automated malaria parasite detection has been attempted by several groups using various approaches such as color histogram based thresholds, morphological operators, granulometry, shape and texture parameters, etc.

In ref. [6], the authors reported the study of analysis of blood cell images to detect abnormalities in which they binarized the original image using a fuzzy measure and labeled all the cells in the image. These labeled cells were then classified into red blood cells (RBC), white blood cells (WBCs) and platelets using a hierarchical neural network model with the help of their features, size and color.

In the work of Di Ruberto et al. [7, 8], digital images were acquired by scanning the color photographs of stained malarial blood from a microscope. They used morphological granulometry to obtain the size distribution of RBCs. They used HSV color space components for analysis. A simple thresholding of these components and their product generated a marker image for WBC and parasites. In this study, morphological approach was used to classify the parasites into immature trophozoites, mature trophozoites and gametocytes.

In the work of Ross et al. [9], digital images were acquired in JPEG format and were preprocessed. The complemented green component of the true color image was used in this study. The preprocessing and analysis steps followed the work of Di Ruberto et al. [7, 8] to some extent. This was followed by image segmentation, generation of features and classification. The preprocessing of the image included filtration by a median filter, followed by a morphological area closing filter. The overlapping RBCs were distinguished from the single ones with the help of granulometry. The overlapping RBCs were then separated using morphological operators used by Di Ruberto et al. [7, 8]. Geometric, texture features and color features were generated from the obtained RBCs. They used a two stage tree classifier with infection classified as positive or negative at the first node and the species assigned at the second node.

Sio et al. in [10] developed an image analysis based program for accurate determination of parasitemia called MalariaCount. The algorithm consists of four stages namely edge detection, edge linking, clump splitting and parasite detection. Adaptive histogram equalization was used for preprocessing.

Tek et al. in [11] reported a method to detect malaria parasites in which they separated stained and non-stained pixels using a Bayesian pixel classifier and then classified the stained pixels as parasites or non-parasites using a distance weighted K-nearest neighbor classifier. Four features namely color histogram, Hu moments, shape measurements and color autocorrelogram were selected for the second step of classification.

Diaz et al. [12] reported a color segmentation technique for separation of pixels into parasite, red blood cell and background, based on standard supervised classification algorithms. Supervised classification techniques namely KNN, Naive Bayes, SVM and Neural network were evaluated on different color spaces namely RGB, normalized RGB, HSV and YCbCr. For the classification of pixels into one of the three classes, training sample sets were manually extracted by an expert, and each of the pixels of the training sample was labeled accordingly. A classification model was

trained for each of the color spaces. The classified color space was then used as a look-up table.

In the work of Le et al. [13], the authors report a semi-automatic image processing approach to detect malaria parasites. Here, the comparison based approach differentiated solid components in the smears by exploiting the interrelations between different observations and radiometric representation. The algorithm is comprised of six steps namely nucleated components detection, image decomposition, erythrocyte size estimation, leukocytes and malarial gametocytes identification, erythrocyte segmentation and finally parasitemia estimation.

In Ref.[14], the authors provide an overview of computer vision studies of malaria diagnosis and provide critique of these works. They also provide a perspective of the future work for realization of automated microscopy diagnosis of malaria.

There is a need for a decision support system which aids a hematologist in malaria parasite detection in thin blood smear images. The management of review of all microscopic views of the sample is quite a laborious task for the hematologist. A decision support system is highly desirable in a laboratory setup of a hospital for regular use and also for screening programs where a large number of cases need to be managed in a short period of time. In this paper, we attempt to develop an algorithm to detect malaria parasites which can be used to develop a decision support system.

Methods

Blood specimen collected from patients is spread as thin blood smear, stained with Giemsa stain and examined with a 100× oil immersion objective. This study uses a total of 200 images out of which 100 images represented *P. vivax* infection, another 50 images represented *P. falciparum* infection and the other 50 images represented joint manifestation of *P. vivax* and *P. falciparum* infection. Out of the 200 images, 20 images were used to manually study the typical sizes of RBC, WBC, schizonts and the threshold values suitable for the segmentation. These images were carefully acquired to obtain uniform staining and illumination. All the procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation of the university.

Image Analysis Algorithm

The task of detection of malaria parasites is divided into subproblems namely detection of ROIs and RBCs (segmentation), analysis of the RBCs (feature extraction) and detection of infected RBCs (classification).

As the blood smear is stained and the parasites have a certain color, the basic approach for detecting ROIs is a color segmentation. The following subsections describe the subalgorithms which are used to detect malaria parasites. A block diagram of the algorithm can be seen in Fig. 1.

Background Elimination

After observing the gray-level histogram, it is obvious that a threshold segmentation is suitable to extract the background (plasma) from the image, because there are two main peaks corresponding to the background and foreground. Therefore, first the mean of the gray-level values is calculated, and the maxima above and below are detected from the histogram of the input image. The mean of those two maxima is the threshold value for the background extraction. The threshold is obtained for each image individually.

RBC Detection

The hematologist needs to know how many RBCs are infected. Therefore, a counting algorithm had to be developed. The knowledge of the size and of the position of the RBCs is also needed for separating agglomerated RBCs. This task is solved by the morphological erosion with a disk as structuring element on the binary representation of the background extracted image [8]. The operation is repeated several times within a predefined interval of the radius of the disk size, which equals to the estimated RBC radius range. For this study, the suitable range of the radius was calculated from the training set as 17 to 58. At the beginning the smallest radius size is used, while each step increases the disk radius by one. When the disk has approximately the size of the RBC, the result of erosion on the RBC is only a few pixels. When the disk gets bigger than the RBC, the result of the erosion is null. On the whole image each connected component object is observed, and when it disappears in the eroded image the position and the size are saved.

ROI Detection

A connected component object of the image is considered as a ROI, if it contains chromatin stain and the blue ring part. Therefore, color segmentation is applied on the HSI representation of the image to extract the blue ring part and the chromatin of the parasites in two stages. In the first segmentation stage, the blue ring part is extracted and in the second stage, the chromatin dots, schizonts and gametocytes are extracted. Along with these, white blood cells and platelets too get selected. Hence, to eliminate WBCs and platelets, the result of the first segmentation stage is referred. If the selected region contains any stained pixels from the first color segmentation, it is retained as an ROI.

To get rid of small noise objects like platelets or stains, a pre-classification concerning the connected component object size is applied. The ROIs have to be in a certain range of

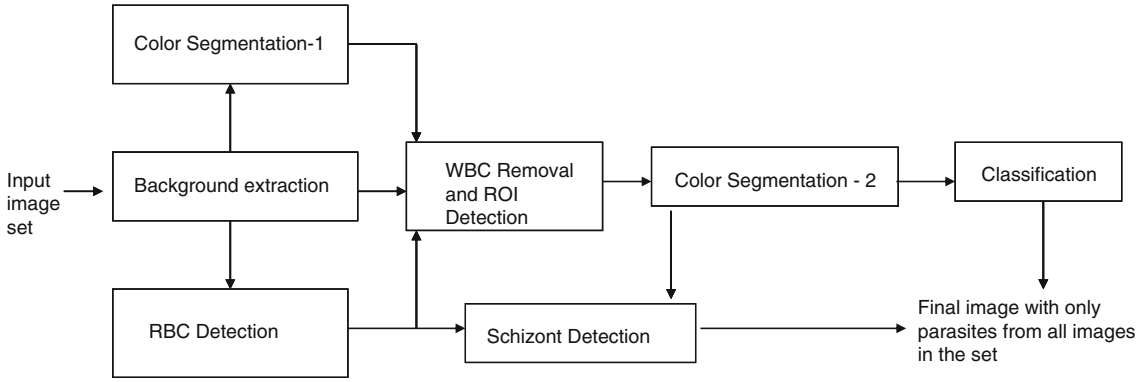

Fig. 1 A block diagram of the algorithm

total number of pixels, which suits the approximated RBC size. Each region containing less than a minimum number of pixels is removed from the image. Each region containing more than maximum number of pixels is analyzed further. This is due to the fact that RBCs often are clustered and not separated from each other after the background extraction. Therefore, the additional information of the RBC detection algorithm is used. For this study, the suitable value for the minimum size was found to be 2,000, and the maximum size was found to be 14,000. As in the step above, each detected RBC of that clustered region is being analyzed whether it contains any stained chromatin dot pixel or not. Therefore, only the areas of that subregion which contain detected RBCs with any stained chromatin dot pixel inside will be kept in the image; the rest is eliminated.

WBC Removal

As white blood cells (WBCs) have similar color as that of the chromatin dot, they are also selected as a ROI. This might result in false positive classifications further on. To prevent this, WBCs are removed and not treated as a ROI if the percentage of stained chromatin dot pixels is above a certain threshold. This threshold was found to be 45% for this study through trial and error method. Such connected component object is then not kept as a ROI. This works robustly and does not lead to any false negatives, because the percentage of chromatin dot stained pixels is lower for all the stages of life cycle of the parasites.

Schizont Detection

It is found that some of the schizonts get selected into ROI in the first step of color segmentation. But some are found to have got missed if they do not contain any pixels of chromatin color. Hence, a schizont detection step becomes necessary. This makes use of the detected RBC regions and the output of the second color segmentation step. The detected RBC region is used as a mask on the output of the second step of color segmentation. If, in a detected RBC region, the percentage of pixels in the output of second color segmentation step crosses a minimum value of 9%, it is detected as a schizont.

Classification

The problem of classification is to find suitable descriptors in order to classify each detected ROI into its relevant class. Two classes are used: infected RBCs and normal RBCs. Three descriptors have been chosen out of many: the percentage of chromatin dot stained pixels, the percentage of ring part stained pixels and the standard deviation of the value channel of the HSV representation of each ROI are taken.

A 2D classification was used with the classification parameters being the two percentages of stained. Standard deviation has been considered as an additional parameter. A training set of 20 images was examined manually, and in a supervised process the threshold values for these parameters were extracted [15]. If any of the two descriptors is above its threshold, then the ROI was classified as an infected RBC. A confidence score is obtained based on how many of the parameters are above the minimum value. Thus, each region classified as an infected region is assigned a score which represents the confidence factor for the classification.

Generation of Composite Image

Once the suspected regions have been identified in each of the views in the image set of a particular specimen, they are all put together to generate a composite image.

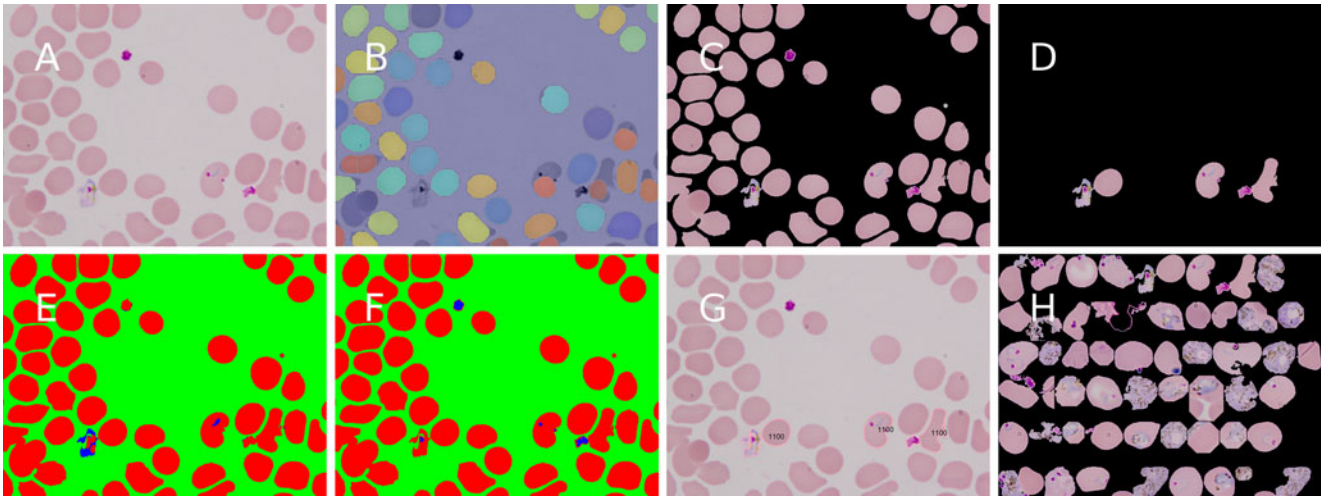

Fig. 2 Results of various steps of the algorithm. **a** Original image. **b** The color coded output of RBC detection step. **c** Output of background extraction. **d** ROI detection. **e** and **f** The *blue color* in the image represents the detection of the ring part and chromatin content respectively. **g** The classification scores on the detected infected RBC. **h** The composite image generated from all the views of the particular specimen

This image can be sent to a remote expert for confirmation on the parasite detection and also for identification of the species of the parasite through a telecommunication channel.

Confirmation from Remote Expert

Since we intend to make this application to be useful in remote areas without internet connectivity too, we design an android application which facilitates the viewing of the composite image, received by the remote expert through the multimedia messaging service (MMS) using a mobile service provider from the laboratory where the specimen image analysis is carried out.

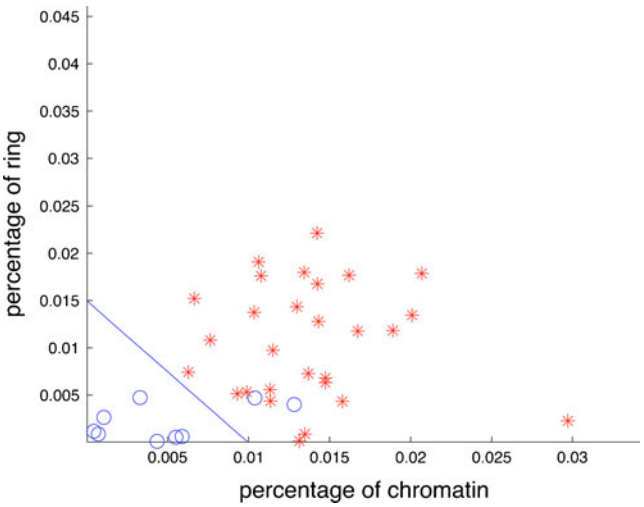

Fig. 3 Classification parameters from the training set

Results

The outputs of different steps of the algorithm and the classification parameters used are shown in Figs. 2 and 3, respectively.

This automated analysis was correlated with the manual evaluations, and the correlation was found to be highly significant. The statistical analysis was done with the help of freely available software named GraphPad. The Spearman correlation coefficient r was 0.888 with a confidence interval of 0.838 to 0.923, $p<0.0001$.

The classification of regions with the use of two percentages of pixels representing the chromatin and ring as parameters is shown in Fig. 3. Blue circle represents normal RBC, and red star represents malaria parasite.

The details of parasites actually present, the number of parasites detected by the algorithm and the number of parasites which were missed by the algorithm are tabulated in Table 1.

The screenshots of the mobile application are shown in Figs. 4, 5 and 6. The composite image is displayed as an

Table 1 Details of parasite detection

	Plasmodium vivax	*Plasmodium falciparum*	Mixed Infection	Total
Parasites present	129	104	56	289
Parasites detected	121	100	47	268
Parasites missed	8	4	9	21

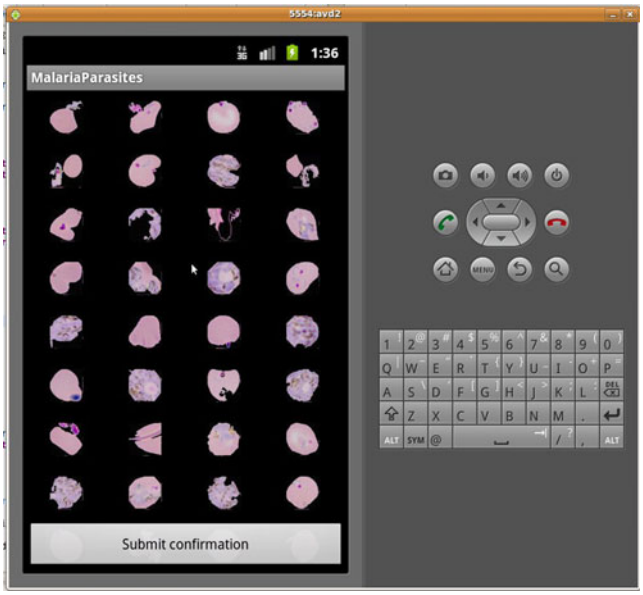

Fig. 4 Screen shot of the android application showing the subimages

array of subimages, each subimage representing the affected RBC as detected by the image analysis program, shown in Fig. 4.

The expert can click on the subimage to view it individually, and in case he finds it as a false positive case, he can indicate so by selecting the radiobutton 'No' at the bottom of the screen, as shown in Fig. 5. This can be repeated for all suspicious subimages, as required by the expert. Once he is done with this step, he clicks on the 'Submit confirmation' button which takes him to a screen with the prescription option, shown in Fig. 6.

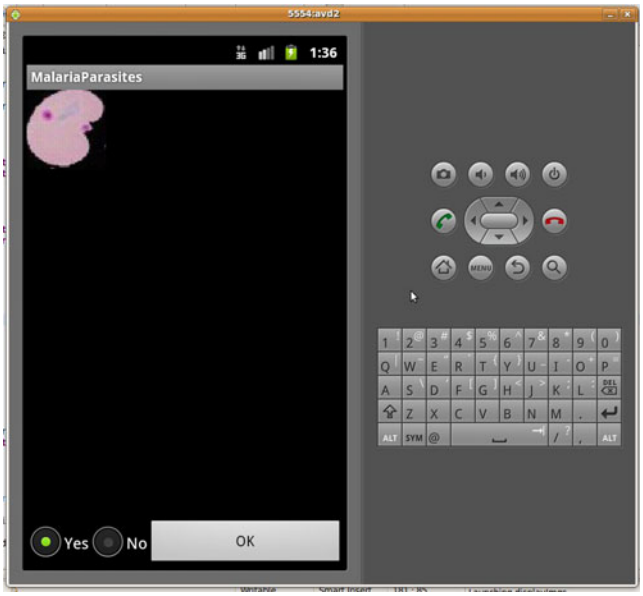

Fig. 5 Screen shot of the android application showing the selected subimage

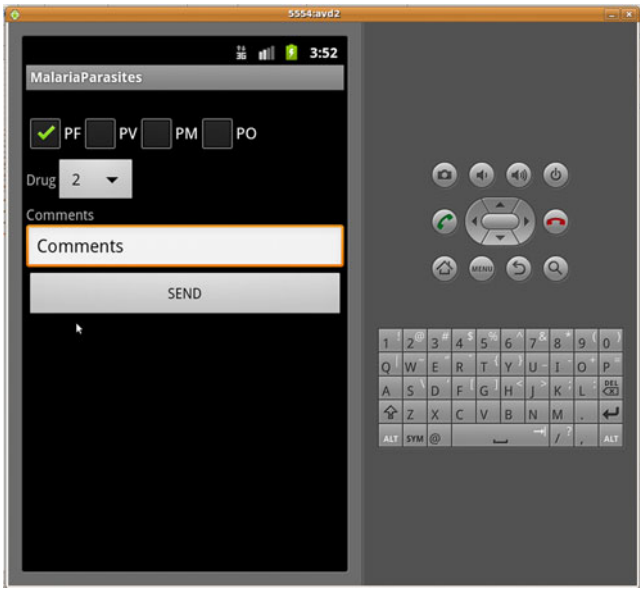

Fig. 6 Screen shot of the android application showing the prescription form for entering the final diagnosis and comments of the expert

Also, the species of the parasite can be selected by the expert. The 'send' button generates a text message which indicates the false positive cases, species of the parasite, suggested drug and dosage and any other comments, and sends it to the laboratory. Another application on the receiver's side can read the SMS from the expert, parse it and generate a prescription for the patient.

Discussion

The previous studies on blood smear analysis used traditional image processing techniques like thresholding [9], edge detection [10, 16], region growing [17], morphological operations and granulometry [7–9, 18] for the study of parasitemia. Some studies have focused on estimation of RBCs in case of overlapping cells [10, 16, 19]. Authors in [13] have developed a comparison based approach to handle the variations that may occur during sample preparation and image acquisition phase. The practicality, robustness, accuracy of these solutions and their applicability for diagnosis are questioned in [14]. Many of these efforts to automate malaria parasite detection fail to efficiently differentiate between parasites, artifacts and healthy blood components [8, 16, 20]. There are a few studies which differentiate between an infected RBC and a healthy RBC [9, 11]. Those few studies which address this problem [9, 11] fail to provide results comparable to manual analysis.

Malaria being a contagious disease, the primary goal of automation of the diagnosis is to provide quick and efficient treatment which can control the spread of malaria. Decision

on treatment depends mainly on the correct identification of the species of the parasite, parasitemia and the dominant life stage of parasites in the specimen. Identification of the species of the parasite requires a lot of heuristic judgment, and any mistake in this step can have a very serious impact. A technique which attempts automation of this step has to be extremely robust and highly accurate.

Our approach attempts to address this issue. Malaria is prevalent in developing countries, and the microscopists are either overworked or not adequately trained. Mass screening requires reliable and fast technique to diagnose malaria. Moreover, facilities for diagnosis and treatment of malaria are not available in all rural or primary health centers. Studies on mobile phone based microscopy have been reported in [4, 5] to diagnose malaria. But it is not very practical to send images of all the views of the specimen to a remote expert over a telecommunication channel for diagnosis. An image analysis algorithm which can detect parasites with very high sensitivity and a facility for teleconsultation with an expert for the final decision on the treatment can handle malaria efficiently. Hence, we developed an algorithm which picks up all abnormal regions from all the views of the specimen and puts all of them into a composite image which can be sent over a telecommunication channel for an expert opinion. The identification of species and decision on treatment can be obtained by an expert. This facilitates accurate diagnosis and treatment.

The algorithm evaluates the proportion of colors present in a trophozoite, schizont and gametocyte and only if the region of interest satisfies the color contents specific to these, it is identified as a malaria parasite. Depending on how far the color proportions match, a color score is given to each of the abnormal regions selected. The score value for each region classified as infected indicates the confidence factor for the decision of classification. This technique reduces possibility of false positive detection and thus is more robust in handling WBCs, platelets and other staining artifacts.

We used MATLAB on a Pentium machine with 1 GB RAM. The computation of each image took around 2 min. In our approach, we used a gray-level threshold technique for background elimination. Alternately, the green channel could also be used to threshold and obtain the background as reported in [14]. We used morphological operations for RBC detection step. This is a time consuming task. This could also be achieved by using the initial selection of a typical RBC by the hematologist, which could provide the information of the approximate size of the RBC. This information could be used to set a threshold for the size of connected component. These steps can be adopted for achieving better computational efficiency.

The proposed method is designed to keep the false negative rate at a minimum. The false negative rate of this algorithm is about 7% and is tolerable. The false positive rate is about 20%. However, some refinements could improve the results of the algorithm, such as an improvement of the RBC detection sub-algorithm or a further observation for better threshold values for color segmentation and classification. For this work, we have used a carefully acquired set of images from a single laboratory. To make it more robust to color tone variations, a color normalization technique could also be included. As suggested in [14], a count of WBC could also be maintained to obtain an accurate value for parasitemia.

Conclusion

In this paper, we demonstrate the use of color image analysis to develop a decision support system for malaria parasite detection. A prototype of a mobile application has been developed to facilitate communication between a rural primary health centre which may not even have internet connectivity and a remote expert. This enables quick, efficient and easy way to manage malaria by facilitating accurate diagnosis and hence the correct treatment at the right time.

References

1. Phillips RS: Current status of malaria and potential for control. Clin Microbiol Rev 14:208–226, 2001
2. Schiff C: Integrated approach for malaria control. Clin Microbiol Rev 15:278–293, 2002
3. Mitiku K, Mengistu G, Gelaw B: The reliability of blood film examination for malaria at the peripheral health unit. Ethiop J Health Dev 17(3):197–204, 2003
4. Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA: Mobile phone based clinical microscopy for global health applications. PLoS One 4(7):e6320, 2009
5. Mody RM, Dooley DP, Hospenthal DR, Horvath LL, Moran KA, Muntz RW: The remote diagnosis of malaria using telemedicine or e-mailed images. Mil Med 171(12):1167–71, 2006
6. Kim KS, Kim PK, Song JJ, Park YC: Analyzing blood cell image to distinguish its abnormalities. In: Proceedings of the Eighth ACM International Conference on Multimedia 395–397, 2000.
7. Di Ruberto C, Dempster A, Khan S, and Jarra B: Automatic thresholding of infected blood images using granulometry and regional extrema. Proc. of ICPR 3445–3448, 2000.
8. Di Ruberto C, Dempster A, Khan S, Jarra B: Analysis of infected blood cell images using morphological operators. Image and Vision Computing 20(2):133–146, 2002
9. Ross NE, Pitchard CJ, Rubin DM, Duse AG: Automated image processing method for the diagnosis and classification of malaria on thin blood smears. Med Boil Eng Comput 44:427–436, 2006
10. Sio S, Sun W, Kumar S, Bin W: Malariacount: an image analysis based program for accurate determination of parasitemia. Journal of microbiological methods 68:11–18, 2007
11. Tek FB, Dempster A, Kale I: Malaria parasite detection in peripheral blood images. In: Proc Br Mach Vis Conf. Edinburgh UK, 2006.

12. Diaz G, Gonzalez F, Romero E: Automatic clump splitting for cell quantification in microscopical images. Journal of Biomedical Informatics 42(2):296–307, 2007
13. Le M, Bretschneider T, Kuss C, Preiser P: A novel semiautomatic image processing approach to determine *Plasmodium falciparum* parasitemia in Giemsa stained thin blood smears. BMC Cell Biology 9:15, 2008
14. Tek F, Dempster A, Kale I: Computer vision for microscopy diagnosis of malaria. Malaria Journal 170:1362–1369, 2009
15. Han J, Kamber M: Data Mining—Concepts and Techniques. Elsevier, San Francisco, 2006
16. Halim S, Bretschneider T, Li Y, Preiser P, Kuss C: Estimating malaria parasitemia from blood smear images. Proceedings of the IEEE International Conference on Control, Automation, Robotics and Vision 648–653, 2006.
17. Theerapattanakul J, Plodpai J, Pintavirooj C: An efficient method for segmentation step of automated white blood cell classification. Proceedings of the IEEE TENCON 191–194, 2004.
18. Wermser D, Haussmann G, Liedke CE: Segmentation of blood smears by hierarchical thresholding. Computer Vision, Graphics, and Image Processing 25(2):151–168, 1984
19. Won CS, Nam JY, Choe Y: Extraction of leukocyte in a cell image with touching red blood cells. Proceedings of the SPIE Conference on Image Processing 399–406, 2005.
20. Rao KNRM: Application of Mathematical Morphology to Biomedical Image Processing, PhD thesis, U. Westminster, 2004.