35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Automated Detection of Malaria in Giemsa-Stained Thin Blood Smears

Mark C. Mushabe, Ronald Dendere and Tania S. Douglas, *Senior Member, IEEE*

***Abstract*— The current gold standard of malaria diagnosis is the manual, microscopy-based analysis of Giemsa-stained blood smears, which is a time-consuming process requiring skilled technicians. This paper presents an algorithm that identifies and counts red blood cells (RBCs) as well as stained parasites in order to perform a parasitaemia calculation. Morphological operations and histogram-based thresholding are used to extract the red blood cells. Boundary curvature calculations and Delaunay triangulation are used to split clumped red blood cells. The stained parasites are classified using a Bayesian classifier with their RGB pixel values as features. The results show 98.5% sensitivity and 97.2% specificity for detecting infected red blood cells.**

I. INTRODUCTION

In 2010 the World Health Organisation reported that malaria caused an estimated 655000 deaths from 216 million cases, most of which occurred in sub-Saharan Africa. The current gold standard of diagnosis is the analysis of blood smears by light microscopy [1]. This is used for parasitaemia determination (i.e. the percentage of infected cells) as well as species and life-cycle identification. This process can be very sensitive and specific but requires a skilled technician. Microscopic analysis is also time-consuming and is prone to inaccuracy and inconsistency [2].

Automated image analysis for parasite detection and parasitaemia calculation could reduce the burden of skilled technicians as well as provide objectivity and repeatability of results [3]. Such automated methods need to have the performance of a skilled technician and cater for a variety of images and problems such as uneven illumination, artefacts in the images and clumped red blood cells. Current research involves a pre-processing stage that optimises the input image for automation, the extraction and segmentation of red blood cells and the detection of stained pixels. These operations can be performed by morphological operations, thresholding and statistical classification using morphology, texture and colour [4, 5, 6].

We use morphological and statistical classification to detect malaria in blood smears by identifying and counting red blood cells and *Plasmodium* parasites. Morphological operations and histogram-based thresholding are used to extract the red blood cells (RBCs) and boundary curvature calculations and Delaunay triangulation are used for splitting clumped RBCs; these methods were proposed by Berge et al for RBC detection, with an absolute error rate of 2.8% with manual detection as the gold standard [4].

Corresponding author: T.S. Douglas (Phone: +27 21 4066541; fax: +27 21 4487226; email: tania@ieee.org)
M.C. Mushabe, R. Dendere and T.S. Douglas are with the MRC/UCT Medical Research Unit and the Biomedical Engineering Programme, Department of Human Biology, University of Cape Town, South Africa

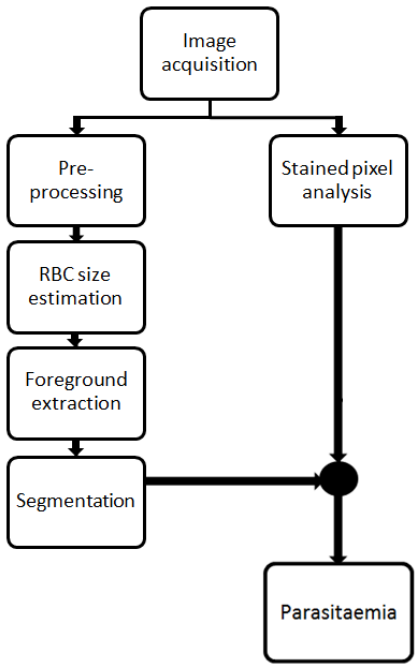

Figure 1: Steps involved in the automated parasitaemia determination algorithm

Parasites in blood smears are stained by Giemsa and colour classification can be used to determine which pixels in an image are stained or not. Pixel classification based on colour was used by Khutlang et al [7] to detect tuberculosis in Ziehl-Neelsen stained sputum smears. Tek et al [6] performed parasite detection and species and life-cycle identification by making use of multi-class classification methods to identify stained pixels. A modified K-nearest neighbour (KNN) classifier was used to detect stained pixels for the purpose of parasite/non-parasite classification. This research also investigated the use of Fisher linear discriminant (FLD) and back propagation neural network (BPNN) classifiers but found the modified KNN classifier to have the best performance with an overall accuracy of 93,3%, 72.4% sensitivity and 97.6% specificity. We use a linear Bayesian classifier to identify stained pixels.

Figure 1 shows the steps taken in the algorithm to identify and count the RBCs and the parasites.

II. METHODS

A. *Image acquisition*

The algorithm was developed and tested using Giemsa-stained thin blood smears infected with *Plasmodium*

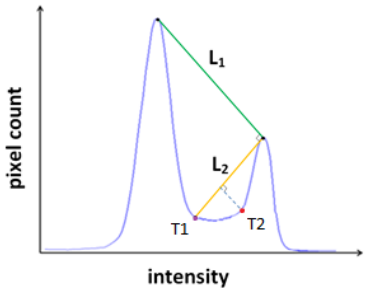

Figure 2: Zack's thresholding algorithm for RBC segmentation

falciparum that were prepared and cultured in a laboratory environment. Images of the slides were captured using a Zeiss Axioskop 2 bright field microscope (motorized 1999 model) with a 12 V, 100 W halogen lamp, a 1.3 numerical aperture, and a 100x oil immersion objective. The images were captured with an Axiocam high resolution colour camera using the Axiovision imaging software (version 4.7) at a resolution of 1296 x 1024.

B. *Preprocessing*

The greyscale version of the RGB image is extracted and filtered with a median filter and morphologically closed using a disk-shaped structuring element. This is to smooth the image and remove small pixel noise. The background is extracted by morphologically closing the image with a disk-shaped structuring element and this is subtracted from the filtered image to correct uneven illumination.

C. *RBC size estimation*

The foreground extraction and segmentaion portions of the algorithm are dependent on the size of the RBCs and an estimation of the average RBC size is required.

Grey scale granulometry is a method that estimates the size of objects in an image using a series of morphological openings with a structuring element of a fixed shape (in the case of RBCs, this shape was circular) that has been used in studies on RBC size estimation [4, 5]. The resultant pattern spectrum shows the distribution of object sizes within the image. A peak represents a large number of objects equal in size to that of the corresponding structuring element.

A filtered, grey scale version of the input image (from the preprocessing stage) is inverted and used for the RBC size estimation.

D. *Foreground extraction*

1) Determining the threshold

The histogram of the illumination corrected image is bimodal, with the first peak representing the background intensities and the second peak, the RBC intensities. Berge et al. [4], extending the work of Le et al. [8] suggested iteratively using Zack's thresholding algorithm to determine the threshold between the two peaks and extract the foreground.

This method finds the point furthest from a line drawn that connects the two peaks (see L1 in Figure 2). Berge et al [4] suggested a second, more accurate threshold, found by using the line connecting the first threshold (T1 in Figure 2) to the line connecting the peaks and finding the point furthest from this second line (see L2 in Figure 2) resulting in the final threshold T2 in Figure 2.

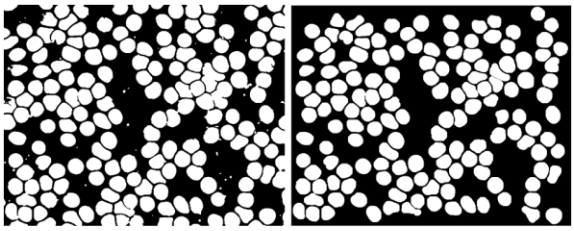

Figure 3: RBC segmentation; left: original binary image; right: after iterative thresholding and border clearing

2) Creation of RBC binary mask

Utilising Zack's thresholding algorithm twice is not as effective in images with large clumps of RBCs. An iterative process is used in which the threshold is adjusted and imposed on the resultant clumps. This process is repeated 3 times or until no clumps larger than 3 RBCs are found. The binary mask is then improved using morphological operations to remove small artefacts, fill unwanted gaps, improve RBC shape and smooth their edges [4].

E. *Segmentation*

Once the foreground has been extracted in some cases there are still clumped RBCs and as such an iterative clump splitting process is performed. The individual cells are stored while the extracted clumps are split [4].

For each clump the boundary is extracted and its curvature is calculated using methods suggested in [10] and used in [4].

The boundary curvature vector for each object is analysed for regional maxima, indicating points of concavity from which potential split lines will be drawn. An iterative process used in [4] for RBC clump splitting and inspired by [11, 12] is applied.

The original individual RBCs and resultant individual RBCs from the clump splitting process are combined into a RBC binary image. Cells that were unnecessarily split are morphologically reconstructed and added back to the RBC binary image (an example of the input and final RBC binary image is shown in Figure 3) [4].

F. *Stained pixel analysis*

1) Feature generation

The original RGB image is used as the input to this stage of the algorithm. The red (R), green (G) and blue (B) values of each pixel are extracted as features in order to classify each pixel into one of two classes: stained or non-stained. Twenty one images were used to extract the data (RGB pixel values) to train and validate the stained pixel classifier: 11952 non-stained pixels and 2093 stained pixels. This data was divided into training and validation data in the ratio 3:2 with the validation data used for parameter tuning and evaluating the classifiers. Figure 4 shows the distribution of the RGB pixel values for the background, RBC and stained pixels.

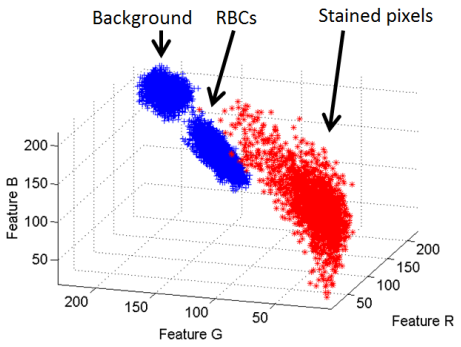

Figure 4: Distribution of RGB pixel values in the training data

2) Pixel classification

Using the Pattern Recognition Toolbox in MATLAB and following from the research by Tek et al, a K-nearest neighbour (KNN) classifier was investigated. This classifier assigns to a query pixel the most frequent class of its K nearest neighbours based on Euclidean distance. This is a simple and intuitive approach that can be easliy visualised in the 3-dimensional RGB space. From the training data a K value of 3 was found to minimise the classification error of the KNN classifier [6]. A linear Bayesian normal classifier (with the class frequencies used as the prior probabilities) was also investigated following from the pixel classification work of Khutlang et al [7].

The linear Bayesian normal classifier was chosen due to its reduced computational requirements and similar performance to the KNN classifier. This classifier is based on Bayes' rule; it minimises the probability of error in assigning a class to a query object and assigns a class which has the highest probability at that position [6].

3) Creation of stained pixel marker

Once each pixel in the input image is classified into the stained or non-stained class these labels are used to create a binary marker image (shown in Figure 5). This binary marker image is then refined by morphological operations to remove small pixel noise and fill holes within parasites.

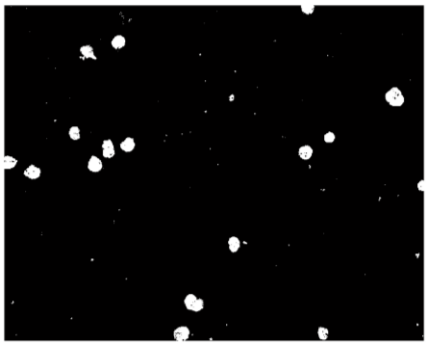

Figure 5: Stained pixel marker

G. *Parasitaemia calculation*

Objects touching the image border in the RBC and stained pixel binary images are removed as per WHO protocols [2].

The RBC binary image and the stained pixel binary image are analysed together to eliminate parasites found outside a red blood cell (not considered in parasitaemia calculations). A final count of the RBCs in the RBC binary marker image and the number of infected RBCs is performed. In determining the parasitaemia level multiple parasites found in a single RBC are counted as a single infected RBC.

III. RESULTS

Twenty one images were used to train and test the stained pixel classifier. The performance of the KNN and linear Bayesian normal classifiers was tested on the validation data and is shown in TABLE 1.

TABLE 1: PERFORMANCE OF STAINED PIXEL CLASSIFIERS

	Sensitivity	**error[a]**
KNN (k = 3)	99.8	0.0012
Linear Bayesian	99.5	0.0055

a. classification error based on error counting, weighted by the class priors

Seven slides containing thin smears of blood infected with *Plasmodium falciparum* were used as the testing set for the entire algorithm (a total of 42 images). Figure 6 shows the final output image with the RBC and parasite counts. The results of identifying infected RBCs (i.e. performing an accurate RBC and parasite count to determine parasitaemia) are shown in Table 2. These results are compared with those of Ross et al who used morphological operations and threshold selection techniques along with a two-stage tree classifier with features based on colour, texture and cell geometry and with the parasite/non-parasite results of Tek et al who used morphological operations and a modified KNN classifier [5, 6].

TABLE 2: EVALUATION OF THE INFECTED RBC DETECTION ALGORITHM

	Sensitivity	**Specificity**	**% error[b]**
This study	98.5	97.2	1.6
Ross et al [5]	85.3		
Tek et al [6]	83.2	96.8	

b. (|manual_count– algorithm_count|/manual_count)*100

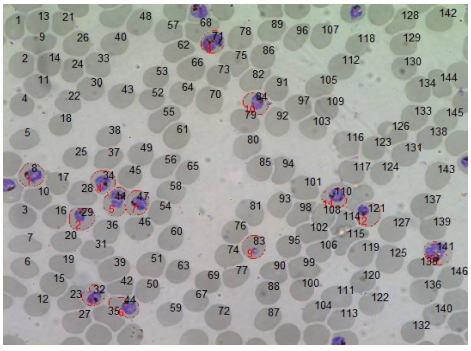

Figure 6: Resultant output image with RBC and parasite count

IV. DISCUSSION

This algorithm identifies red blood cells and malarial parasites in microscope images in thin blood smears. The algorithm also performs a parasitaemia calculation. The algorithm incorporates methods suggested by research on automated malaria detection as well as methods used in the analysis of cellular components in microscope images.

The method of RBC detection proposed in [4] and used in this study is less dependent on user defined thresholds and intensity levels than the best performing published methods [8, 13]. The algorithm performs well on images of varying colour, RBC density and level of cluster overlap, providing aequate input to the classification step. Red blood cells that were not thresholded adequately (thus not part of the RBC binary image) and those belonging to clumps too big to be decomposed completely were not counted.

The RBC size estimation step uses a filtered, grey scale version of the input image and is not dependent on the accuracy of the foreground extraction. It does not significantly improve the segmentation process compared to the area estimation used by Berge et al [4] which is less computationally intensive, but it provides a more accurate estimation of the RBC size as determined empirically by the author using the training data.

The classification of stained pixels using the original RGB image is successful and not dependent on the shape of the histogram or a user defined threshold as with the intensity based thresholding methods of [4] and [5]. The method is also not dependent on the morphology of the stained objects as in [11]. The method presented here uses the colour of the Giemsa stain as its discriminating factor, mimicking a human technician. It is also robust against varying illumination. The training data showed significant separability between the stained and non-stained pixels as well as separability between the background and red blood cell pixels.

Our algorithm, as with several others published for similar applications, was tested on a data set of limited variety with most images coming from the same source (laboratory, microscope and camera). Further testing should be performed with images from multiple sources and full blood samples should be tested as well in order cater for different in the image acquisition equipment and to better simulate samples collected in the field, respectively.

REFERENCES

[1] World Health Organisation, "World Malaria Report 2011," Geneva: WHO Press, 2011.

[2] World Health Organisation, "Basic malaria microscopy part 1. Learner's Guide," Geneva: WHO Press, 2010.

[3] J. Frean, "Improving quantitation of malaria parasite burden with digital image analysis," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 102, pp. 1062-1063, Feb. 2008.

[4] H. Berge, D. Taylor, S. Krishnan, T. Douglas, "Improved red blood cell counting in thin blood smears," *International Symposium on Biomedical Imaging*, pp. 204-207, 2011.

[5] N.E. Ross, C.J. Pritchard, D.M. Rubin, and A.G. Duse, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical & Biological Engineering and Computing*, vol. 44, pp. 427-436, 2006.

[6] F.B. Tek, A.G. Dempster, and I. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Computer Vision and Image Understanding*, vol. 114, pp. 21-32, 2010

[7] R. Khutlang, S. Krishnan, R. Dendere, A. Whitelaw, K. Veropoulos, G. Learmonth and T.S. Douglas, "Classification of Mycobacterium tuberculosis in images of ZN-stained sputum smears," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14(4), pp. 949-957, 2010

[8] M.T. Le, T.R. Bretschneide, C. Kuss, and P.R. Preiser, "A novel semi-automated image processing approach to determine Plasmodium falciparum parasitaemia in Giemsa-stained thin blood smears," *BMC Cell Biology 9*, no. 15, 2008.

[9] O. Demirkaya, M.H. Asyali, and P.K. Sahoom, "Image Processing with MATLAB: Applications in Medicine and Biology," *CRC Press, Boca Raton*, 2008.

[10] Q. Wen, H. Chang, and B. Parvin, "A delaunay triangulation approach for segmenting clumps of nuclei," in *Proceedings of the Sixth IEEE International Symposium on Biomedical Engineering*, Boston, Massachusetts, 2009, pp. 9-12.

[11] S.W.S. Sio, W.L. Sim, S. Kumar, W.Z. Bin, S.S.Tan, S.H. Ong, H. Kikuchi, Y. Oshima, K.S.W. Tan, "MalariaCount: An image analysis-based program for the accurate determination of parasitaemia," *Journal of Microbiological Methods*, vol. 68, pp. 11-18, 2007

[12] S. Kumar, S.H. Ong., S. Ranganath, T.C. Ong, and F.T. Chew, "A rule-based approach for robust clump splitting," *Pattern Recognition*, vol. 39, no. 6, pp. 1088-1098, June 2006.

[13] G. Diaz, F.A. Gonzalez, E. Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images," *Journal of Biomedical Informatics*, vol. 42, pp. 296-307, 2009.

[14] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra,"Analysis of infected blood cell images using morphological operators," *Image and Vision Computing*, vol. 20, issue 2, pp. 133-146, 2002