IMPROVED RED BLOOD CELL COUNTING IN THIN BLOOD SMEARS

Heidi Berge[1], Dale Taylor[2], Sriram Krishnan[1], Tania S. Douglas[1]

[1]MRC/UCT Medical Imaging Research Unit, Department of Human Biology
[2]Division of Clinical Pharmacology
University of Cape Town, South Africa

ABSTRACT

Quantification of the extent of malaria parasite infection (parasitaemia) continues to rely on time-consuming manual microscopy of Giemsa-stained blood smears. We present an algorithm that counts red blood cells in thin blood smear images, the first step in the determination of malaria parasitaemia. Morphological methods and iterative thresholding are used for red blood cell segmentation, and boundary curvature calculations and Delaunay triangulation for red blood cell clump splitting. Our results compare well with those of published semi-automated methods, with an absolute error of 2.8% between manual and automatic counting of red blood cells.

Index Terms— malaria; erythrocyte; segmentation

1. INTRODUCTION

Malaria parasitaemia is the determination of the parasite burden in the blood, and is useful for prognosis and in objective evaluation of responses to treatment, especially in the presence of drug resistance. Laboratories generally have a poor ability to measure parasite loads [1]. Parasite load in thin blood films is expressed as a percentage of infected erythrocytes (red blood cells - RBCs). Although parasites are readily recognised and counted by trained technologists, estimating the total number of RBCs present in microscope fields is a source of error [2]. Microscopy remains the gold standard for diagnosing malaria and is used for parasitaemia determination and species identification [3].

Automated image analysis for parasite detection and characterisation could reduce the workload on technicians and the dependence on human skill and experience, especially when automation is extended to include image acquisition and microscope control. A number of research groups have used pattern recognition to detect malaria and calculate parasitaemia in thin blood smears.

We concentrate on the segmentation and counting of red blood cells. We use alternatives to or modifications of components of algorithms developed by other researchers, in an attempt to overcome the user interaction required by the best performing available methods. Morphological methods and iterative thresholding are used for RBC segmentation, and boundary curvature calculations and Delaunay triangulation for clump splitting.

2. METHODS

Microscope images of Giemsa-stained thin blood smears were captured using a Zeiss Axioskop 2 microscope (motorised 1999 model) with a 1.3 numerical aperture, a 100W halogen lamp and a 100x oil immersion objective. Images were captured using an Axiocam high resolution colour camera at a resolution setting of 1296 X 1024 pixels with exposure time approximately 794.5 ms.

2.1 Preprocessing

The RGB image is converted to greyscale, filtered using a 5x5 median filter to smooth the image, and morphologically closed using a disk-shaped structuring element of radius 2 pixels to remove pixel noise. Morphologically closing the image using a disk-shaped structuring element of radius 80 pixels produces the background image. This size was determined empirically to be large enough to remove clumps of RBCs. The background image is then subtracted from the original filtered image (Fig. 1) to correct uneven illumination.

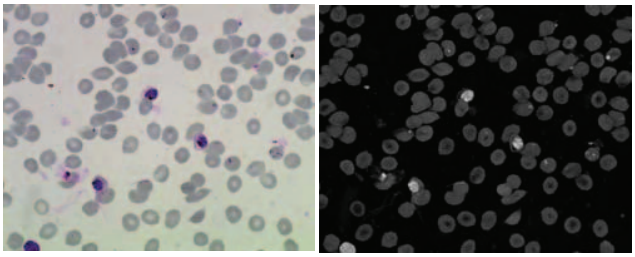

Fig. 1. Pre-processing; left: original image; right: after background subtraction.

2.2 RBC Segmentation

The histogram of our background subtracted image (Fig. 1) has two distinct peaks, the first representing the background intensities and the second representing the RBC intensities. Le et al. [4] suggest Zack's method [5] to determine the

RBC threshold: a line is drawn between the two peaks and the point between the two peaks that is furthest from this line is used as a threshold. We found that this method thresholds unwanted foreground in images with a distinct RBC histogram peak. A more accurate threshold is obtained by repeating this method using a line which joins the original threshold point to the second peak, and finding the furthest point away from it (Fig. 2). An actual threshold of 90% of the calculated threshold is used, to account for the thin central pallor of some RBCs.

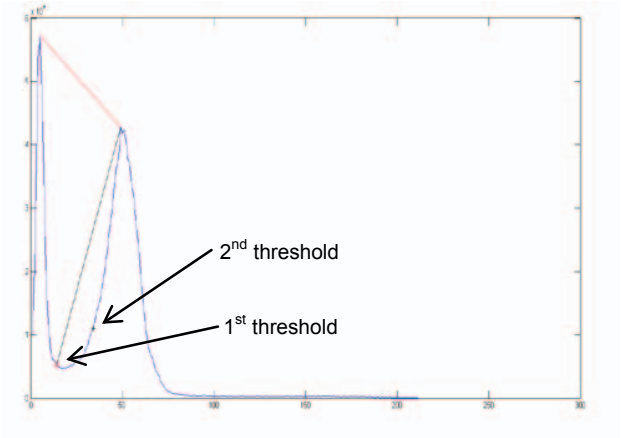

Fig. 2. Threshold for RBC segmentation in a clear image. The thresholds found by the first and the second applications of Zack's method are indicated.

The binary image resulting from the above thresholding is adequately segmented in cases where RBCs are sparse. However, in images with high RBC concentrations, large clumps may result. RBC areas range from 3900-5400 pixels, therefore an area threshold at 15000 pixels is applied to include clumps of 3 RBCs or more in the output image. The binary conversion is repeated 3 times, with a 10% increase in the threshold each time. An example is shown in Fig. 3.

Objects of area less than 1000 pixels are considered too small to be RBCs and are removed from the resulting thresholded image. Poorly segmented objects of area less than 3000 pixels (e.g. RBCs having an inner pallor in the original image) are closed to improve their shape and larger segmented objects are opened to smooth their edges. The smaller and larger objects are combined and their holes filled, to produce the final RBC image.

RBC segmentation serves as preparation for the clump splitting stage, which separates overlapping and touching RBCs.

2.3 RBC Size Estimation

Granulometry, using a series of morphological openings with a circular structuring element, has been used in various studies for RBC size estimation. A peak in the pattern spectrum results if the image contains a high number of objects of size equal to that of the corresponding structuring element. This method is limited by its assumption of a circular shape for RBCs and its computationally intensive passes through the entire image.

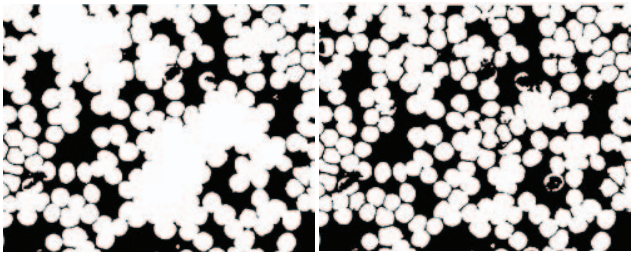

Fig. 3. RBC segmentation; left: original thresholded image; right: after iterative thresholding.

RBC size estimates are based here on area rather than radius, because of the varied shape of RBCs. The areas are sorted in ascending order and the derivative of this area curve is examined: peaks indicate large changes in area. The largest acceptable size of a single RBC is assumed to be located where the area curve first starts to increase significantly.

RBC areas were empirically found to range from approximately 3900 – 5400 pixels, and derivative increases beyond 400 pixels typically represented clumped objects. Therefore, the algorithm initially locates the first peak in the derivative greater than 400 pixels and beyond the median object size. If, however, the area is less than the minimum RBC size found empirically, or greater than the maximum, the algorithm correspondingly increases or decreases the threshold by 50%. If the area curve is flat (no clumps), the size estimate is that of the largest object in the image.

2.4 RBC Clump Splitting

Our clump splitting process is most successful when applied iteratively. The first step extracts all the objects larger than the maximum RBC area. Objects larger than 105% of the typical RBC area are targeted to eliminate abnormally large single cells. The single cells are then stored while the larger clumps are split.

2.4.1 Boundary Extraction and Curvature Calculation

The boundary curvature is calculated as suggested in [6] for each object with respect to the derivatives of the x and y coordinates on the boundary. Gaussian first and second derivatives are convolved with the x and y boundaries to obtain the respective boundary derivatives. The Gaussian parameters are: $\sigma = 3$, kernel size = ceil(8.5σ) and $\varepsilon = 0.01$. Where the denominator vector of the above curvature calculation is equal to zero, it is replaced by ε [7].

2.4.2 Finding Points of Maximum Curvature

Regional maxima are identified in the curvature curve for each object. These maxima indicate points of concavity (the

points of interest) between which split lines will be drawn. If the clump has area less than twice the maximum RBC area (2 cells in clump) only two concavity points need to be found.

Pairs are found first and split right away, as follows. The first concavity point is taken as the point of maximum curvature in the object boundary, and the second point of concavity is taken as the local maximum approximately opposite the first point, as measured by a boundary distance between the two points of approximately half the total number of pixels in the object boundary.

Larger clumps are analyzed in more detail as the concavity curve of a large clump may have many peaks, yet only a fraction of them will form adequate split pairs. A first step is to find only those pixels with concavity larger than some threshold percentage of the clump boundary length in pixels. It was found empirically that the largest 5-10% of all concavity points in a clump will include the points of interest, therefore a threshold of 8% is used. The vector of concavity points is dilated with a linear structuring element of size 20 elements so as to find only the local maxima, rather than their neighbours. A large clump with its concavity points is shown in Fig. 4.

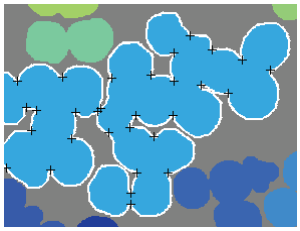

Fig. 4. Concavity points for a large RBC clump.

2.4.3 Selecting the Best Split Lines

The above concavity points can be joined in a variety of ways. Delaunay triangulation is used to draw possible split lines between concavity points, as it is able to provide meaningful edge information without covering the full space of possible split lines [6].

Most cells are expected to have a diameter smaller than the maximum diameter and a split line between two cells should always be shorter than the diameter of the cell itself. Therefore, split lines longer than 95% of the maximum RBC diameter are eliminated. Any edge crossing the background is eliminated.

The concavity-concavity (CC) and concavity-line (CL) alignment are measured [8,9]: the CC alignment has a minimum value when two concavities are directly opposite each other, and a maximum value when they are facing the same direction; the CL alignment is the maximum of the differences in orientation between each of two concavity regions and a candidate split line; small values are desired for both alignments. Good initial CC and CL alignment thresholds were found to be 65 and 35 degrees, respectively. All clumped RBCs are not split, but relaxing the thresholds may result in unwanted lines. An iterative clump split is therefore applied. After each splitting operation, those objects larger than 1.2 times the maximum RBC area are again subjected to the splitting process. The 20% increase in size accommodates those cells that have been split into shapes that include some pixels from their neighbours. The clump splitting algorithm is run a maximum of 5 times, after which remaining clumps are disregarded, and neither their RBCs nor their parasites counted. On each of the first 3 iterations, the alignment thresholds are increased by 5 degrees, and for the last two the increase is 15 degrees.

The split pieces and original single cells are combined in a final image. Any objects smaller than 40% of the maximum RBC area are extracted, closed with a disk-shaped structuring element of radius 15 pixels, and then added back to the decomposed image, in case a clump was split unnecessarily into smaller pieces,. Lastly, any cells touching the image border are removed. The division lines are drawn using Bresenham's line algorithm [10], which finds the shortest path between two pixels.

2.5 Final Counting

The RBC image is examined for unwanted splits. Objects smaller than a third of the maximum RBC area are removed from the decomposed RBC image; the remaining objects are counted to provide the total number of RBCs in the image. An example of an RBC counting result is shown in Fig. 5.

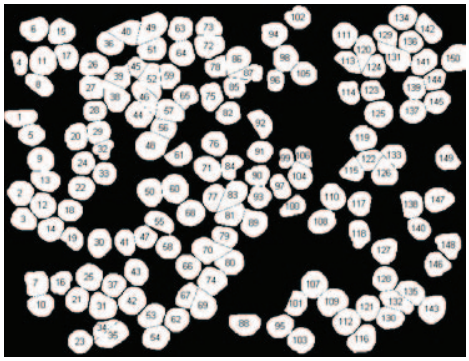

Fig. 5. Result after 4 iterations of clump splitting, closing and counting.

3. RESULTS

Eight slides containing thin smears of blood infected with *Plasmodium falciparum* were analysed; 6-7 images were obtained of each slide (49 images in total). The results of running the algorithm as well as results from published studies are shown in Table 1.

Table 1. Evaluation of RBC detection; manual detection is the gold standard. For the current study, the overall results for the dataset are given, with the range for the different image sets in brackets.

	Precision	Recall	% Error[1]
This study	99.5 (98.3-99.9)	96.7 (89.1-100)	2.8 (0.1-9.4)
Le et al. [4]	87.5-99.1	80.1-99.5	
Diaz et al. [11]			1.25±5.43 to 1.78±6.64

[1] |manual - algorithm|/manual*100

4. DISCUSSION

Our algorithm identifies and counts red blood cells in microscope images of thin blood smears. Components of methods suggested by others for malaria detection in blood smears are combined, but our algorithm also incorporates methods from other applications of cellular analysis in microscope images. Iterative analysis is used to improve results.

Evaluation of our algorithm shows results comparable with or better than those of other studies. The best performing published methods [4,11] require user interaction and have limited ability to split RBC clumps. Our algorithm has a greater degree of automation and performs well on images from slides of varying colour, clarity and RBC density and level of overlap.

Cells were not counted if they were too faint and inadequately thresholded, or if the clump they belonged too was too big to be decomposed fully. As only those RBCs that are counted contribute to parasitaemia, the influence of these factors may be limited in practice by examining larger numbers of images per slide.

Segmentation algorithms that rely less on intensity should be explored to avoid dependence of the results on illumination. As with other methods developed for the same application, our algorithm relies on empirically derived thresholds, and results will vary for different image acquisition parameters. Our algorithm avoids the assumption of circular RBC shape, and performs well in detecting RBCs having different shapes.

The algorithm presented here and most of those that have been published for the same application were tested on data with limited variety: samples containing only RBCs and prepared in the same laboratory, and mages captured using the same microscope. Algorithms should be tested on images captured from full blood samples and using different hardware.

The algorithm presented here may be used as a step in a method for automated parasitaemia calculation, which would include RBC counting, parasite identification, and determination of the parasite-RBC ratio.

Automated determination of parasitaemia could ultimately form part of fully automated microscopy systems that also include automated focusing, image acquisition and slide handling.

5. REFERENCES

[1] Frean J.: Improving quantitation of malaria parasite burden with digital image analysis. Transactions of the Royal Society of Tropical Medicine and Hygiene 102, 1062—1063 (2008)

[2] Gering E., Atkinson C.T.: A rapid method for counting nucleated erythrocytes on stained blood smears by digital image analysis. Journal of Parasitology 90, 879--881 (2004)

[3] Tek F.B., Dempster A.G., Kale I.: Computer vision for microscopy diagnosis of malaria. Malaria Journal 8, 1--14 (2009)

[4] Le M.T., Bretschneider T.R., Kuss C., Preiser P.R.: A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears. BMC Cell Biology 9, doi:10.1186/1471-2121-9-15 (2008)

[5] Zack G.W., Rogers W.E., Latt S.A.: Automatic-measurement of sister chromatid exchange frequency. Journal of Histochemistry & Cytochemistry 25, 741--753 (1977)

[6] Wen Q., Chang H., Parvin B.A.: Delaunay triangulation approach for segmenting clumps of nuclei. IEEE International Symposium on Biomedical Imaging, Boston (2009)

[7] Demirkaya O., Asyali M.H., Sahoo P.K.: Image Processing with MATLAB. CRC Press, Boca Raton (2009)

[8] Kumar S., Ong S.H., Ranganath S., Ong T.C., Chew F.T.: A rule-based approach for robust clump splitting. Pattern Recognition 39, 1088--1098 (2006)

[9] Sio S.W.S., Sun W.L., Kumar S., Bin W.Z., Tan S.S., Ong S.H., Kikuchi H., Oshima Y., Tan K.S.W.: MalariaCount: An image analysis-based program for the accurate determination of parasitemia. Journal of Microbiological Methods 68, 11--18 (2007)

[10] Bresenham J.E.: Algorithm for computer control of a digital plotter. IBM Systems Journal 4, 25-30 (1965)

[11] Diaz G., Gonzalez F.A., Romero E.: A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Journal of Biomedical Informatics 42, 296--307 (2009)