

# Blood Smear Malaria Parasite Detection

Austin Zheng

Department of Electrical Engineering

Stanford University

austinzheng@gmail.com

## I. INTRODUCTION

Malaria, a blood-borne disease transmitted by mosquitoes, involves the infection of red blood cells in humans and other organisms by protists of the genus *Plasmodium*. Current state of the art for medical diagnosis and research purposes involves drawing a blood sample from a patient or research subject. This blood sample is smeared onto a slide and stained in order to color cell nuclei. Because mature red blood cells do not possess nuclei, the stain only strongly marks malarial parasites. The slide can then be examined under a microscope in order to count the number of infected red blood cells. Figure 1 presents an example micrograph.

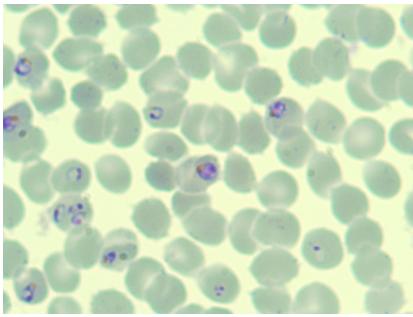


Fig. 1. Example image

A red blood cell is considered infected if at least one parasite can be detected within its interior. White blood cells and free-floating parasites are not considered. The current state of the art involves manual counting by a laboratory technician or other individual, who can distinguish staining artifacts from actual nuclei, white blood cells, and (depending on specific requirements) life cycle and species of malarial parasites [1]. Although manual counting is relatively inexpensive to implement, adequate sensitivity requires proper training and supervision of technicians. This poses problems for both medical care providers in impoverished regions of the world as well as laboratory settings which may benefit from automation of a tedious and time-consuming task [2]. Conceivably, automation of this task could both facilitate laboratory efficiency as well as provide an alternative diagnostic tool in conjunction with mobile phone based microscopy in developing countries [3].

This paper describes a basic image processing pipeline

implemented in MATLAB which detects *Plasmodium falciparum* parasites within a micrograph of a blood smear slide and attempts to determine which red blood cells in an image are infected. Efficacy of this pipeline is demonstrated, and a number of potential improvements and directions for further development are discussed.

## II. PREPROCESSING

Blood smear micrographs (see Figure 1) are first transformed into the hue-saturation-value (HSV) color space using the `rgb2hsv` function. This transforms a standard red-green-blue image (such as those obtained from digital cameras) into a three-dimensional vector whose dimensions correspond to the hue (H), saturation (S), and value (V) of each pixel in the image (Figure 2).

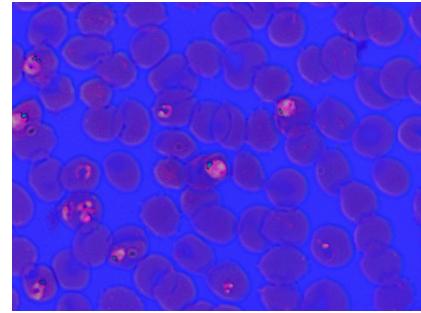


Fig. 2. HSV color space image

Two masks are created from the V component of the image, which is relatively insensitive to the presence or absence of purple stain areas. The V component is subjected to histogram equalization (`histeq`). Histogram equalization tends to ‘spread out’ the range of grayscale values comprising an image and provides a form of normalization that allows both edges and nuclei to be more easily thresholded across differently lit images. Because the interiors of red blood cells are significantly darker than the surrounding background Otsu thresholding is carried out in order to create an area mask which determines whether or not a particular region of the image lies within a red blood cell (Figure 3).

It is possible to exploit optical artifacts in the V component by conducting thresholding on a band of values (for histogram-equalized images, between around 0.6 and 0.7)

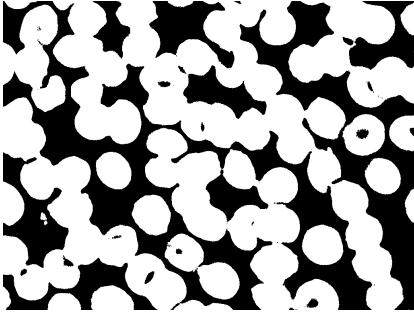


Fig. 3. RBC area mask

to produce a reasonable approximation of the edges of the cells (Figure 4). This outline mask is then subjected to Canny edge detection [4] to create an edge map comprised of single-pixel edges (Figure 5). It is also possible, depending on the characteristics of the map, to perform morphological dilation and erosion on the area mask to create this outline. A particularly complex example is provided in Figures 6 and 7. In this example it is conceivable that morphological erosion could be carried out on the area mask and then applied to the edge mask in order to create a significantly cleaner edge mask devoid of the ‘noise’ beyond the edges of the cells.

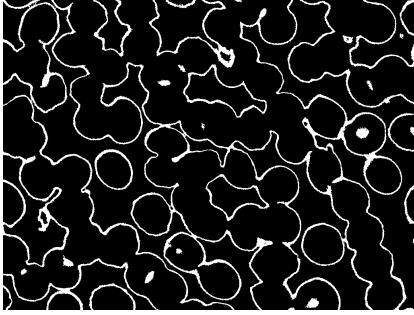


Fig. 4. RBC outline mask



Fig. 5. Edge detection (portion)

### III. PARASITE CANDIDATE DETECTION

The H and S components are used to detect and differentiate regions which have been stained (including candidate parasite

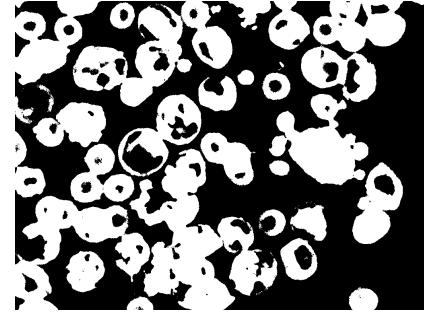


Fig. 6. Poor-quality area mask

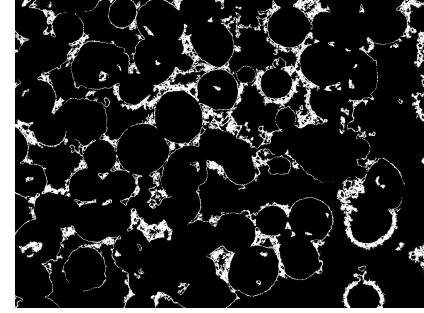


Fig. 7. Edge mask

nuclei). The H and S components are subjected to histogram equalization and then thresholded to produce a mask consisting of all purple regions in the input image, each of which is believed to correspond to either a parasite or part of a parasite (Figure 8). Small region removal is carried out in order to suppress extremely small signals caused by poor image quality or noise from imprecise threshold choice. Once this is complete the remaining regions in the image can be identified and labeled using `bwlabel`.

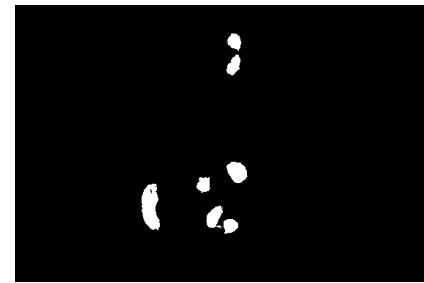


Fig. 8. Candidate region mask (portion)

### IV. CELL SEGMENTATION

Once parasite nuclei candidates have been isolated and identified, it is necessary to examine each one in order to determine whether or not it lies within the confines of a red blood cell, and if so estimate the center and approximate extent of the red blood cell. This task is complicated by certain common image properties – overlapping and deformed

blood cells, low contrast, and blur from poor focusing.

Cell segmentation is accomplished by examining a ‘window’ surrounding each parasite nuclei candidate. This window comprises a 500 pixel by 500 pixel area surrounding the centroid of the nuclei region on the candidate mask. Cell outline and cell area masks corresponding to this window are obtained and used to determine whether blood cells exist within the window region and whether or not the parasite candidate in question lies within a red blood cell.

Six ring-shaped masks of progressively larger radius are ‘slid’ across the outline mask in order to detect circular regions, including incomplete regions corresponding to cells which are overlapping, partially outside the window region, or misshapen. This results in a number of line segments which are then labeled and categorized using `regionprops`. Line segments whose length is computed to be greater than 20% of the average circumference of the mask are considered potential candidates for the edge of a red blood cell; line segments are weighed according to the proportion of their length to the average circumference in order to place greater emphasis on longer, continuous line segments. The weighed sum of all candidate line segment lengths is then calculated; if this value exceeds a threshold based on the circumference of the mask a circle (and by extension, a red blood cell) is considered to have been detected and the current coordinates are stored. The RBC area mask for the window is also checked; proposed circles which do not lie within a dark green region are considered spurious and rejected. Figure 9 demonstrates an application of this algorithm to a portion of the edge map shown in Figure 5.

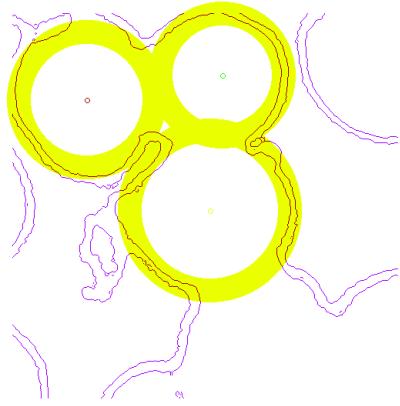


Fig. 9. Sample window region, with estimated cell locations and radii

A number of parameters affect the effectiveness of the circle detection algorithm. An increased difference between the inner and outer radii of the mask allows for improved detection of ovoid and deformed cells, but also results in increased sensitivity to spurious line segments and ‘premature detection’ (fitting a smaller circle to an outline where a

larger circle might have more accurately reflected the size of the red blood cell). The line segment length sum threshold also reflects a trade-off between detecting circles with poorly-defined edges (for instance, circles which overlap) and spurious fitting. Finally, weighting of candidate line segments reflects a balance between ‘rewarding’ longer line segments as being more likely to signify an actual cell edge while still being capable of detecting cells whose edges are comprised of several shorter line segments.

Once cell segmentation has been carried out for all nuclei candidate regions the cell radius and center coordinate information can be used to both count the number of infected red blood cells and plot candidate cell locations onto the original image for further examination. A post-processing step not currently implemented calculates the distance of candidate locations and attempts to consolidate close-by cell candidates into clusters; each cluster is then ‘consolidated’ into a single cell for counting and analysis purposes.

## V. RESULTS

The image processing pipeline was applied to five test images. Candidate nuclei within 250 pixels of the edge of the image were intentionally ignored due to time constraints and limitations with the test harness. These images were chosen due to their lack of visual noise (red blood cells are evenly colored, no background artifacts) in order to carry out a baseline performance evaluation and identify strengths and weaknesses in the pipeline.

Test images 2 and 3 exhibit the highest performance. Test image 2 (Figure 11) demonstrates the identification of four red blood cells containing parasites and estimates a boundary for each of the cells. These boundaries fall into four distinct clusters, which could be correlated through further analysis (consolidating cells with very close centers together). In test image 3 (Figure 12), seven cells containing parasite nuclei are identified, and the borders of the predicted red blood cells are clearly delineated.

Test images 1 (Figure 10) and 4 (Figure 13) demonstrates several limitations of the system as it currently exists. Test image 1 is dominated by parasites in the schizont stage, which appear as round purple clusters containing many dark purple spots. Each of the spots within the schizont is currently detected as a separate candidate and analyzed as such. The three schizonts in the middle of the image are clearly detected, but a large number of spurious detections characterize both these three schizonts and the fourth one to the upper right. These spurious detections are caused when cell segmentation is carried out on some of the schizont spots near the edge of the cluster. Test image 4 exhibits similar issues with schizont detection, exhibiting two high accuracy detections of schizonts and a large number of close-by

spurious signals.

Test image 5 (Figure 14) poses significant complications for detection, as a large number of red blood cells are clumped together and edges between adjacent cells are indistinct to nonexistent. There are three strong infected cell correlations, one strong correlation with a spurious stain (due to insufficient candidate processing), and a number of highly imprecise detections which could not properly characterize the cell within which a parasite resided. Finally, of particular note are the two candidates near the lower right section of the image. Both these candidates lie within a group of cells which touch in multiple places and have no visible internal edges to aid with cell segmentation. These candidates demonstrate the weakness of an approach (such as this one) which exclusively tries to fit parasite candidates into blood cells without additional heuristics.

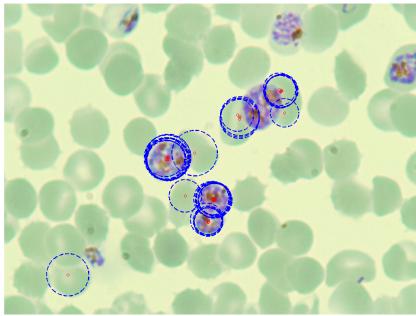


Fig. 10. Image 1, categorized

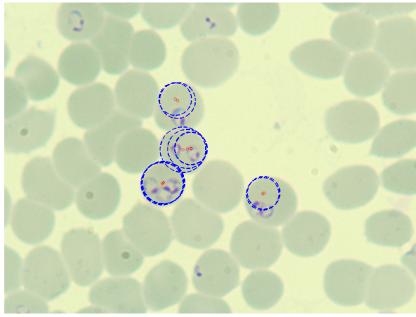


Fig. 11. Image 2, categorized

## VI. FURTHER WORK

### A. Automatic selection of parameters

The image processing pipeline currently depends on a number of hand-selected parameters, most of which are used for thresholding HSV images or the circle detection algorithm. These values were chosen based on empirical performance in order to produce acceptable results with the set of test images, and may or may not be valid for a different batch of

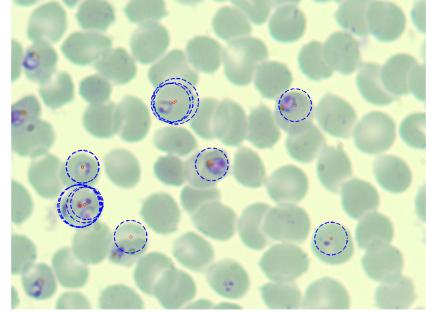


Fig. 12. Image 3, categorized

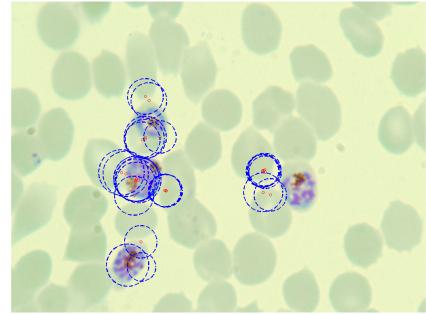


Fig. 13. Image 4, categorized

blood smear slides taken under differing lighting conditions. Further work should be conducted into determining algorithms to adaptively adjust these values with minimal user input.

### B. Parasite candidate mask processing

Currently no processing is carried out on the parasite candidate mask except for small region removal in order to suppress image noise. A significant opportunity exists to carry out analysis on the parasite regions and perform consolidation in order to increase accuracy and reduce noise. In particular, parasites in the schizont life stage exhibit a large number of dark round stains clustered into a roughly circular or ellipsoid area colored a lighter purple, while ring stage parasites are characterized by the presence of a faint purple ring surrounding the nucleus. It should be feasible to use two different candidate masks, one more selective than the other, in order to combine related candidate regions into single candidates. This would significantly reduce redundant work (in particular, circle detection on candidate regions which are part of the same malaria parasite) and allow for different cell detection strategies to be applied based on the type of the parasite.

### C. Circle detection and heuristics

The circle detection algorithm currently uses only a very rudimentary set of heuristics in order to determine whether the line segments contained within the ring-shaped mask comprise the outline of a cell. There is significant opportunity for improving the methods used to discern valid and invalid line segments. An example may be estimating the curvature

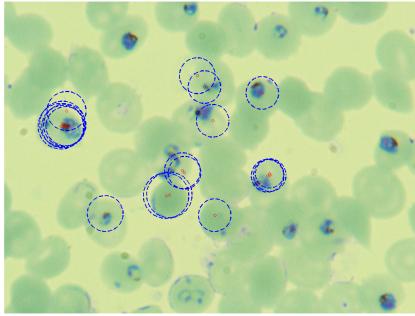


Fig. 14. Image 5, categorized

of a particular line segment by calculating the distance of various points along its length from the mask center, or attempting to detect the sharp ‘points’ in an outline formed when two circles overlap.

Furthermore, in some cases it may be impractical or impossible to extract an outline for a particular group of adjacent cells due to poor edge differentiation. In this case alternative heuristics may be necessary in order to identify or count infected red blood cells in the ‘blob’. At the very least, the pipeline should be able to identify such trouble spots and flag them for human examination.

#### D. Preprocessing

The images used for testing the processing pipeline were chosen due to low levels of visual noise. Visual noise includes faint ‘background’ red blood cells which cannot easily be distinguished from the background through Otsu thresholding alone, as well as dark regions within red blood cells caused by their characteristic shape. Visual noise results in large portions of the area and outline masks which degrades performance. It may be necessary to further characterize and mitigate this noise in order to ensure robust performance.

#### E. Performance

At the present time processing a single blood smear image requires anywhere from thirty seconds to multiple minutes until completion. This is due mostly to the sliding mask circle detection algorithm used to attempt to fit circles to the outline mask. Improvements in parasite mask processing and circle detection, described above, as well as refactoring of existing MATLAB code should greatly decrease the amount of time required to process a single image.

## VII. CONCLUSION

A preliminary image processing pipeline for detecting malarial parasites within red blood cells was developed, and its capabilities and limitations were characterized. Although the pipeline exhibits basic functionality, further work is necessary in order to improve robustness to the point where it can be usefully applied to a wide variety of images. A

number of areas of potential improvement were identified, and will provide guidance as development continues. The current pipeline provides a useful framework which can be refined and extended in order to improve accuracy, tolerance of image noise, and further capabilities.

## ACKNOWLEDGMENT

The author would like to thank Hao Li of Bogoyo Lab at the Stanford University School of Medicine for providing the impetus for this project and the blood smear images. The author also wishes to express his gratitude and appreciation to his project advisor Vijay Chandrasekhar and course teaching assistant David Chen for their support and advice, as well as Professor Bernd Girod for teaching EE 368.

The author would also like to acknowledge the use of MATLAB code by Peter Bone (“Draw a circle in a matrix/image”) and Zhenhai Wang (“Draw a circle.”), obtained from MathWork’s MATLAB Central web site, for drawing circles into matrices and upon graphs.

## REFERENCES

- [1] D.C. Warhurst, J.E. Williams, *Laboratory diagnosis of malaria*, ACP Broadsheet No 148, 1996.
- [2] P. Guerin, P. Olliaro, F. Nosten, P. Druilhe, R. Laxminarayan, F. Binka, W. Kilama, N. Ford, N. J. White, *Malaria: current status of control, diagnosis, treatment, and a proposed agenda for research and development*, The Lancet, Infectious Diseases vol 2, p. 566, 2002.
- [3] Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA (2009) Mobile Phone Based Clinical Microscopy for Global Health Applications. PLoS ONE 4(7): e6320. doi:10.1371/journal.pone.0006320
- [4] T. Nattkemper, W. Schubert, T. Hermann, H. Ritter, *A hybrid system for cell detection in digital micrographs*, 2004.