Image Analysis of Blood Slides for Automatic Malaria Diagnosis

Mahdieh Poostchi, Ilker Ersoy, Emile Gordon, Abhisheka Bansal, Kannappan Palaniappan,
Susan Pierce, Sameer Antani, George Thoma, Stefan Jaeger

***Abstract*— Malaria is a serious global health problem, claiming the lives of 450,000 children per year. A fast and reliable test for diagnosing malaria would be a promising approach to fight this disease. We present an automatic system for diagnosing and quantifying a malaria infection in cultured red blood cells on thin films, using image processing techniques. We measure an average error of 1.8% by comparing the true frequency of infected cells with the automatically computed infection frequency, which encourages applying our technique for malaria diagnosis in the field.**

I. Malaria

Malaria is caused by parasites transmitted through mosquito bites, which infect the red blood cells and lead to symptoms such as seizures and coma in severe cases. The common method for malaria diagnosis is microscopy, during which an expert slide reader visually inspects blood slides for parasites. Given the millions of slides inspected every year all over the globe, this is an extremely laborious, costly and unreliable process. We propose an automatic visual inspection of blood slides by digitized image analysis. Our goal is to run our system on a smartphone that captures slide images through a microscope's eyepiece and returns automatic counts of infected and uninfected cells. Only a few approaches have been published in the literature so far [1, 2].

II. Image Analysis

Our approach is unique in that we apply an accurate n-color level-set cell detection technique with 98% precision on each slide, in the cell segmentation step. This allows us to better identify and split touching cells. We also present a one-to-one matching technique for segmentation evaluation. All cells for a given blood slide are extracted using the segmentation mask. In the next step, a feature vector will be provided for each cell by computing its color features in normalized RGB color space. The collection of feature vectors for infected and uninfected cells form the training input and testing data for a support vector machine (SVM), which is the final processing step to classify and label a cell as either being infected or uninfected.

III. Results and Discussion

We evaluate our method on a set of 70 thin blood slides with about 10,000 red blood cells. Each cell was manually annotated by an expert into either the infected or uninfected category, using our online annotation tool (firefly.cs.missouri.edu). This provides us the ground-truth (GT) infection ratio for each slide, which is computed as the ratio of the number of infected cells over the total number of cells in the slide. We then apply our processing pipeline and a ten-fold cross-validation scheme to train and test the SVM classifier. The accuracy of the SVM in correctly identifying infected cells is 97%, with a sensitivity of 93%, and a specificity of 98%. Fig. 1 shows a comparison of the actual infection ratio and the computed infection ratio based on the classifier output, averaged over ten folds. The average infection ratio error we obtain is about 1.8%. These experiments encourage us to test the system in the field. Future experiments will concentrate on making our approach robust against different staining variations encountered in the field.

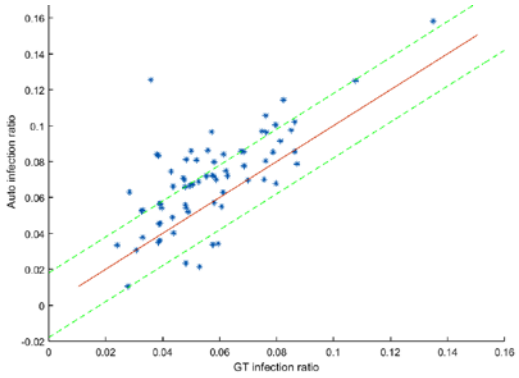

Figure 1. True versus automatic infection ratio using normalized-RGB; true vs auto infection ration average distance is 0.0178.

Acknowledgment

The authors thank Rahul Singh for his help with the Firefly online annotation tool (firefly.cs.missouri.edu).

References

[1] Linder N, Turkki R, Walliander M, et al. "A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears," in PLoS One 9(8), e104855, 2014.

[2] Y. M. Alomari, S. N. H. S. Abdullah, et al. "Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm," Computational and mathematical methods in medicine, Vol. 2014, 2014, pp. 1–17.

*This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). The US Department of Health and Human Services provides support through the HHS Innovation Ventures Fund Program.

M. Poostchi, S. Antani, G. Thoma, S. Jaeger* are with the Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD 20894, USA (e-mail: {mahdieh.poostchi, sameer.antani, george.thoma, stefan.jaeger}@nih.gov. (*phone: 301-435-3198; fax: 301-402-0341).

I. Ersoy, K. Palaniappan are with the Department of Computer Science, University of Missouri-Columbia, Columbia, MO 65211 USA (e-mail: ersoyi@health.missouri.edu, palaniappank@missouri.edu).

E. Gordon and S. Pierce are with the Laboratory of Immunogenetics, A. Bansal is with the Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland 20852, USA (email: emilebrinkgordon@gmail.com, spierce@niaid.nih.gov, abhisheka.bansal@nih.gov).