Automated Protocol for Counting Malaria Parasites (*P. falciparum*) from Digital Microscopic Image Based on L*a*b* Colour Model and K-Means Clustering

J. Opoku-Ansah *jopoku-ansah@ucc.edu.gh*
School of Physical Sciences
Department of Physics,
Laser and Fibre Optics Centre (LAFOC),
University of Cape Coast, Cape Coast, Ghana

B. Anderson *banderson@ucc.edu.gh*
School of Physical Sciences
Department of Physics,
Laser and Fibre Optics Centre (LAFOC),
University of Cape Coast, Cape Coast, Ghana.

J. M. Eghan *meghan@ucc.edu.gh*
School of Physical Sciences
Department of Physics,
Laser and Fibre Optics Centre (LAFOC),
University of Cape Coast, Cape Coast, Ghana.

J. N. Boampong *jonboamus@yahoo.com*
School of Biological Sciences
Department of Biomedical and Forensic Sciences,
University of Cape Coast, Cape Coast, Ghana.

P. Osei-Wusu Adueming *adueming21@yahoo.com*
School of Physical Sciences
Department of Physics,
Laser and Fibre Optics Centre (LAFOC),
University of Cape Coast, Cape Coast, Ghana.

C. L. Y. Amuah *camuah@ucc.edu.gh*
School of Physical Sciences
Department of Physics,
Laser and Fibre Optics Centre (LAFOC),
University of Cape Coast, Cape Coast, Ghana.

A. G. Akyea *angelakyea@yahoo.com*
School of Physical Sciences
Department of Physics,
Laser and Fibre Optics Centre (LAFOC),
University of Cape Coast, Cape Coast, Ghana

Abstract

Basis for malaria parasites diagnosis in most hospitals and clinics, especially in developing countries, which is manually done, is strenuous and time-consuming. In this paper, we present an automated protocol for counting malaria parasites (*P. falciparum)* from digital microscopic red blood cells (RBCs) images based on L*a*b* colour model and K-Means clustering algorithm using Matlab. This method is device-independent, perceptually uniform and approximates human vision. An image slide of size 300 x 300 x 3 pixels of RBCs with malaria parasites has been counted in less than 10 seconds using a computer with 64-bit Intel (R) Celeron (R) Central Processing Unit and processing speed of 2.20 GHz. The digital counts have a good correlation

with the manual counts. This automated protocol has the potential of providing fast, accurate and objective detection information for proper clinical management of patients.

Keywords: Malaria Parasites, Image processing, Segmentation, K-means, cluster, L*a*b* colour model.

1. INTRODUCTION

Malaria is a worldwide health problem, causing more than 1 million deaths annually. Its annual economic burden is estimated to be around several hundreds of millions United State dollars.[1-4] In spite of available techniques such as fluorescent microscopy, rapid antigen detection methods and polymerase chain reaction (PCR),[1, 5, 6] manual counting of parasites from blood smears is commonly used for malaria diagnosis,[7-9] in most hospitals and clinics especially in developing countries. This manual counting is strenuous, arduous, and requires an expert microscopist [10-14]. Moreover, the accuracy of the final count depends largely on the expertise of the microscopist [8]. It has been observed that the agreement rates among microscopists for the counts are low [13] and is attributed to biasness. For prompt diagnosis, treatment and control of the malaria , fast, accurate and objective detection of malaria parasites is required to facilitate the process [1].

In this paper we present a new automated protocol for counting malaria (*Plasmodium falciparum)* parasites based on L*a*b* colour model and K-means clustering algorithm trained with digital RGB images captured from Giemsa stained thin blood films.

2. RELATED WORK

Advances in digital imaging have made it possible for red blood cells (RBCs) infected with malaria Plasmodium falciparum parasites to be imaged and studied. As a result, automatic counting of these parasites has been made possible [15-23]. However, in most of these works, image segmentation was based on red, green and blue (RGB) colour model, which is device-dependent [24] and the colour information is affected by intensity manipulation [25]. Also, the model is not perceptually uniform and does not approximate the human vision. Therefore, features for image segmentation do not appear in a more definite colour scheme in the RGB colour space [26].
S.S. Savkare and S.P. Narote have used Support Vector Machines (SVM) which relies on recognition and classification of cells [23].

3. THE L*A*B* COLOUR MODEL

In contrast to the RGB colour model, the L*a*b* colour model is designed to approximate the human vision using image from RGB colour space via XYZ system. The transformation between RGB colour model and L*a*b* colour model is linear.[26, 27] The aim of the L*a*b* colour model is to eliminate device dependency and bring about perceptual uniformity. The L*a*b* colour model is made up of a luminance, L* and chroma coordinates a* and b*, and are defined as:[28]

$$L^* = \begin{cases} 116.0\left(\frac{Y}{Y_{wp}}\right)^{1/3} - 16 & \frac{Y}{Y_{wp}} > 0.008856 \\ 903.3\left(\frac{Y}{Y_{wp}}\right) & \frac{Y}{Y_{wp}} \le 0.008856 \end{cases}$$

$$a^* = \begin{cases} 500\left[\left(\frac{X}{X_{wp}}\right)^{1/3} - \left(\frac{Y}{Y_{wp}}\right)^{1/3}\right] & \frac{X}{X_{wp}}, \frac{Y}{Y_{wp}} > 0.008856 \\ 7.787\left(\frac{X}{X_{wp}}\right) + \frac{16}{116} & \frac{X}{X_{wp}} \le 0.008856 \end{cases}$$

$$b^* = \begin{cases} 200\left[\left(\frac{Y}{Y_{wp}}\right)^{1/3} - \left(\frac{Z}{Z_{wp}}\right)^{1/3}\right] & \frac{Y}{Y_{wp}}, \frac{Z}{Z_{wp}} > 0.008856 \\ 7.787\left(\frac{Y}{Y_{wp}}\right) + \frac{16}{116} & \frac{Y}{Y_{wp}} \le 0.008856 \end{cases}$$

where X,Y and Z are defined as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{pmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Y_R & Z_G & Z_B \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The columns of the matrix give the tristimulus values of the RGB values in the XYZ coordinate system. The tristimulus values of white point (*wp*), R = G = B = 1, is given as [24];

$$X_{wp} = x_{wp} / y_{wp},$$

$$Y_{wp} = 1$$

and

$$Z_{wp} = z_{wp} / y_{wp}$$

where $x = \frac{X}{X+Y+Z}$, $y = \frac{Y}{X+Y+Z}$ and $z = \frac{Z}{X+Y+Z}$

3.1 The K-Means Clustering Algorithm

K-means clustering algorithm is commonly used to segment multi-dimensional data sets [29, 30]. Its aim is to minimize the sum of squared distances between all points and cluster centers [30]. This process involves initially selecting known *K* cluster centres

$$m_1(1), m_2(1), ..., m_k(1)$$

since the final clusters depend on this initial selected cluster centres *K*. Having prior knowledge of the number of clusters expected enables one to select the cluster centres *K*.

Using the relation

$$n \in Q_j(k) \quad \text{if} \quad \left\| n - m_j(k) \right\| < \left\| n - m_i(k) \right\|$$

for all $i = 1, 2, ..., K$; $i \neq j$; where $Q_j(k)$ denotes the set of samples whose cluster centre is $m_j(k)$ the sample $\{n\}$ is distribute among the *K* clusters. New cluster centres, given as

$$m_j(k+1) = \frac{1}{N_j} \sum_{n \in Q_j(k)} n, \qquad j = 1, 2, ..., K$$

where N_j is the number of samples in $Q_j(k)$, are computed such that the sum of the squared distances from all points $Q_j(k)$ to the new cluster centre is minimized. If $m_j(k+1) = m_j(k)$ for all $j = 1, 2, ..., K$ then the algorithm has converged and the process is terminated.

4. MATERIALS AND METHODS

Five digital images of Giemsa stained thin film slides showing *Plasmodium falciparum* (*P. falciparum*) were retrieved from the Public Health Library of the Center for Disease Control and Prevention website [31]. They were retrieved in *portable document format* (PDF) and re-saved as joint *photographic expert group* (JPEG) format. These images were used as training set. Another set of digital images, twenty in number, with similar preparation were acquired from Giemsa stain thin film slides consisting of *P. falciparum* as test samples. These slides were obtained from the Department of Biomedical and Forensic Science, University of Cape Coast (UCC) on May 29, 2012. Images of the slides were captured using Olympus CX41 Microscope, with Olympus Imaging Camera (GMBH) model LC20 and Olympus LCMicro Software for the developed protocol counting using Matlab. The test sample slides were later sent to the Central Regional Hospital, Cape Coast, Ghana for experts counts of malaria parasites.

4.1 Image Transformation From RGB To L*A*B* Colour Space

The RGB image was transformed via XYZ image space to the L*a*b* image space to produce L*, a* and b* components. The pixels belonging to the RBCs and the parasites were identified based on *K-means* clustering using the chromaticity information. The intensities of the colour bands were clustered into different groups using the *K-means* clustering algorithm. The clustering was done by first separating the a* and b* components of the image since all the colour information was contained in these components. The a* and b* components images were combined to an index image, from which objects contained were then separated by colour. The result was a set of three distinct gray level regions in the image, where each region is relatively homogeneous in terms of pixel intensity. These were parasites intensities (cluster 1), RBCs intensities (cluster 2) and background intensities (cluster 3) referring to the colours existing in the original image. Therefore, it was assumed that there were three distinct gray level images referring to the colours existing in the original image to be separated with the *K-means* clustering algorithm. The output variables were cluster indices and the cluster centres. The mean cluster centre was then calculated for all the cluster centres in order to identify the one with the smallest cluster centre value in the image. It has been determined experimentally that the dark blue or purple colour has the smallest cluster centre value [32]. Histogram-based thresholding algorithm,[33] was then applied to cluster 1 to separate the parasites pixels from the background. Binary mask was generated where the value 1 represents the area where parasites were present and 0 corresponded to the background. All pixels outside the given range were not included in the parasites segmentation. Binary morphological operators were then used to ensure that the enclosed regions with distinctly different sizes and shapes from those of RBCs were removed. Smoothing algorithm was used to eliminate irregularities along the contours of RBCs.

4.2 Identification of Parasites Features and Counting

Connected-component labeling was performed to identify the parasites, labeling each of them with the same label. In order to use morphological methods for parasites identification, the shape

and size of the parasites in the image were considered. A disk- shape structural element of radius 2 pixels was chosen and applied onto the filled regions in the image in order to obtain regions where the parasites were located [21]. Morphological opening by the structuring element, followed by dilation of the result, was then applied on the image. The resultant image was then converted to gray scale and a structuring element of ball-shape with radius 6 pixels and height 2 pixels were then applied on the resulted gray scale image. The extent of thickening was controlled by structuring the element shape. The parasites were equated to the dilated image and labeled using a binary labeling system. These labeled pixels representing the malaria parasites were then counted. Figure 1 shows the resulted program for transforming images from RGB colour space to L*a*b* colour space, segmentation in a*b* components, and counting of the parasites.

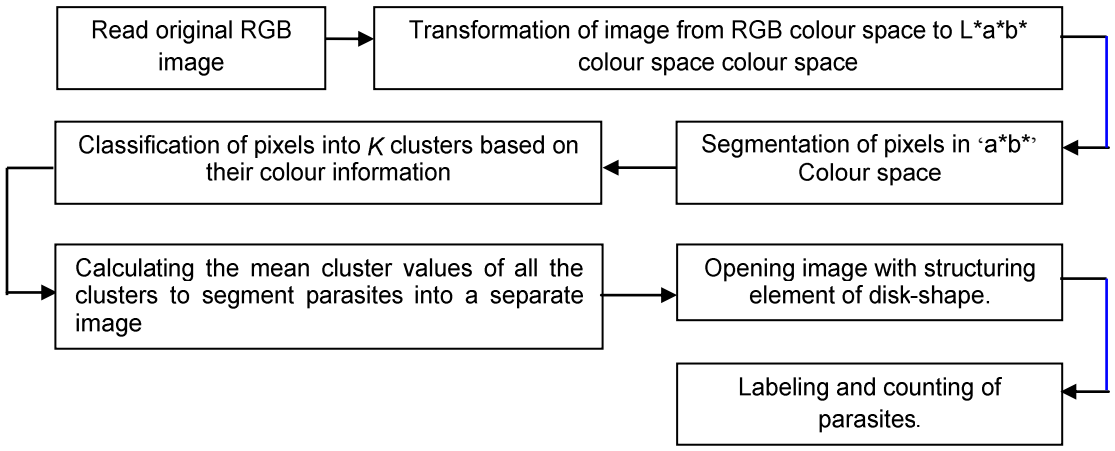
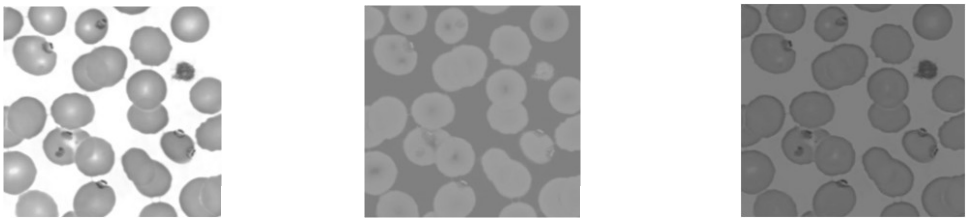


FIGURE 1: Structured program for transformation of images from RGB colour space to L*a*b* colour space, segmentation and clustering of chroma components, a*b*, and counting of malaria *Plasmodium falciparum* parasites.

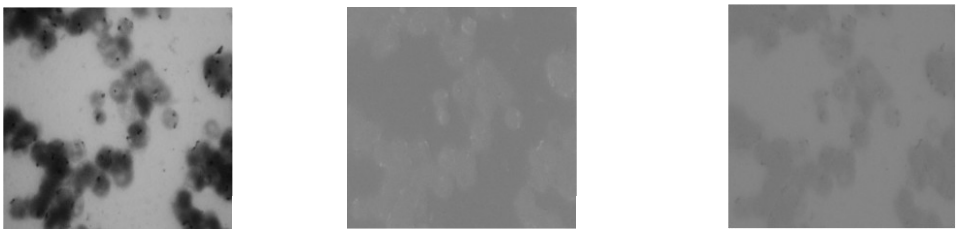
5. RESULTS AND DISCUSSIONS

Figures 2 and 3 show the results of the image transformation from the RGB via XYZ image space to the L*a*b* image space of the images retrieved from the Website and the captured images respectively. In the L*a*b* colour space, colours of pixels which appear similar are located close to each other and produce a range of colours whose form depend on the primaries and the white point.



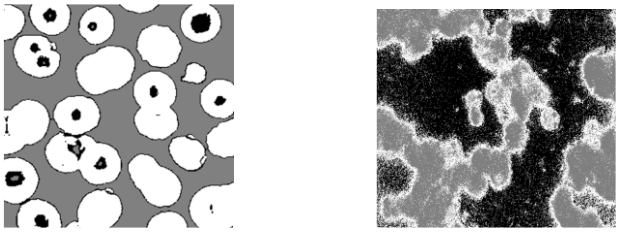
(a) (b) (c)

FIGURE 2: Processed images: (a) L* component (b) a* component and (c) b* component of images retrieved from the Website.



(a) (b) (c)

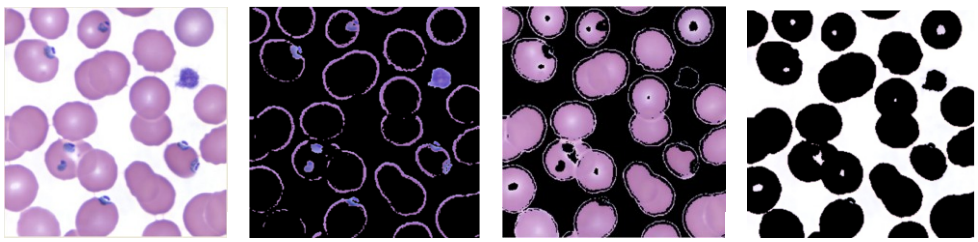
FIGURE 3: Processed images: (a) L* component (b) a* component and (c) b* component of image captured.



(a) (b)

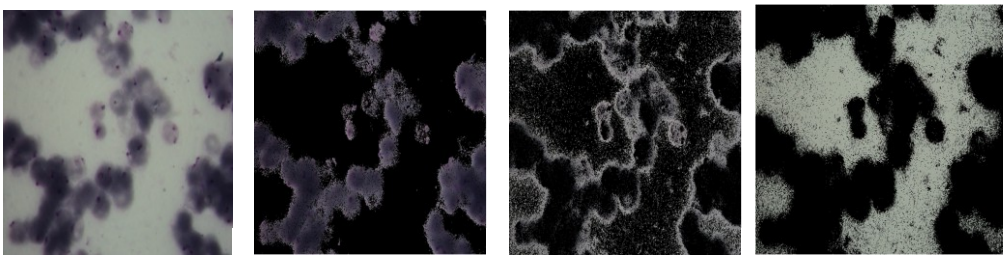
FIGURE 4: Combined a*b* colour component images: (a) images retrieved from the Website and (b) images captured

The combined a*b* components image shows that colour distributions are smooth, with pixel intensities close to zero. The chrominance distribution shows that there are very few pixels which are on the luminance axis, and this is presented in Figure 4. Figures 5a and 6a show one of the original images retrieved from the Website and a captured image respectively. The light shade of pink colour shows the colour of the RBCs and the dark shade of purple colour shows the parasites in the RBCs. The *K* cluster images depict the colours of the RBCs and the parasites existing in the original images but different background, as shown in Figures 5b-d and 6b-d. In Figure 5b-d, cluster 1 show light shade of purple colour for the parasites, cluster 2 show light shade of pink for the RBCs and cluster 3 show black colour for the background.



(a) (b) (c) (d)

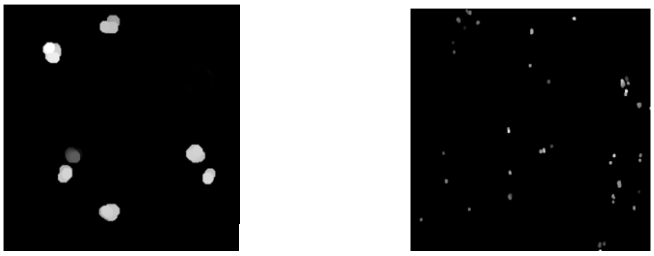
FIGURE 5: Original image and the *K (=3)* clustered images of the a*b* colour space showing the colours existing in the original image: (a) the original image from the website (b) parasites colour (shade of purple) image, (c) red blood cells colour (shade of pink) image and (d) background colour (white) image.



(a) (b) (c) (d)

FIGURE 6: Original image and the *K (=3)* clustered images of the a*b* colour space showing the colours existing in the original image: (a) the original image captured (b) parasites colour (light shade of purple) image, (c) red blood cells colour (light shade of pink) image and (d) background colour (light shade of gray) image.

as dark blue or purple colour [16, 23]. The histogram-based thresholding process is base on the fact that pixels values belonging to the parasites differ from the values of the RBCs and the background. The morphological opening followed by dilation was done to remove completely regions of parasite that were not contained in the structuring element. This smoothened parasite contours, broke thin connections and removed thin protrusions. Image dilation grows and thickens the parasites identified and count them in the images. Sample of the images are presented in Figure 7.



(a) (b)

FIGURE 7: Protocol count images of: (a) image retrieved from Website and (b) captured image.

The score plot of the parasites counted from the twenty images, using the Manual counts and the new Protocol counts, are presented in Figure 8. The plot reveals 98.87 % positive correlation between the Manual counts and the new Protocol counts in the images. This indicates the potential of using the Protocol counting as a fast, accurate and objective way of counting malaria parasites. Comparing with literature image processing in L*a*b colour space is unique since any RGB image can be used. Figure 9 shows a Graphical User Interface (GUI) for the segmentation and counting of the malaria parasites from the RGB images after transforming into L*a*b* colour space. The GUI developed can be easily used by an expert microscopist.

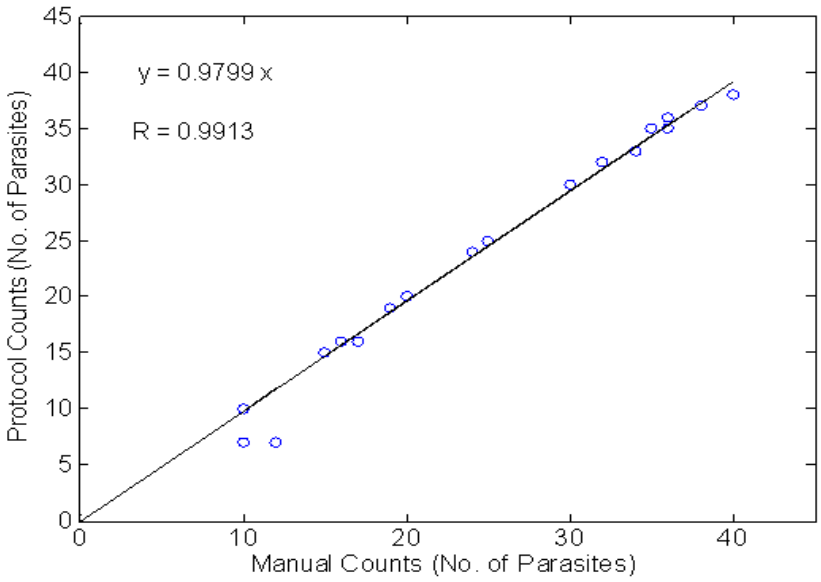


FIGURE 8: Scatter plot of the number of parasites counted using manual counts, by a specialist at the Central Regional Hospital, Cape Coast, Ghana, and the Protocol counts using the developed Matlab codes.

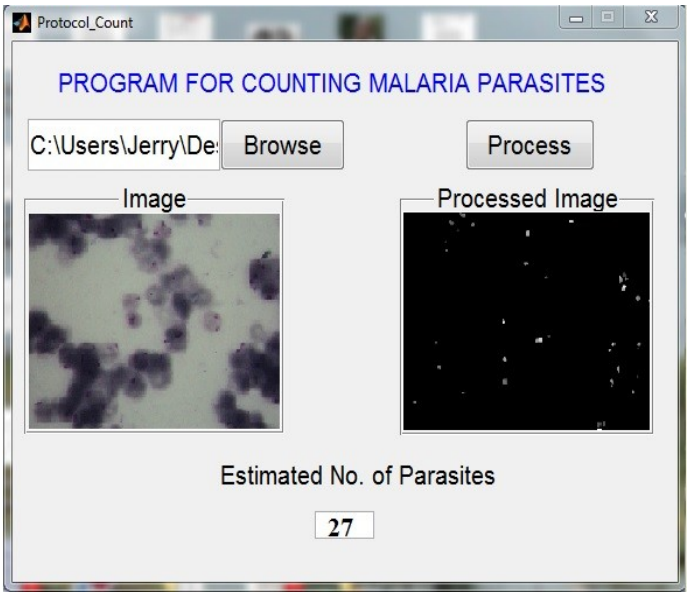


FIGURE 9: Graphical User Interface (GUI) for transformation of images from RGB colour space to L*a*b* colour space, segmentation and clustering of chroma coordinates, a*b*, and counting of malaria *Plasmodium falciparum* parasites.

6. CONCLUSION

Using K-means clustering algorithm based on chromaticity, a*b* space, we have develop Matlab codes to count *P. falciparum* parasites from RGB digital images obtained from thin film blood smear stained with Giemsa. The algorithm was found to be dependent on the stain colour. The Protocol counts provided approximately 99 % correlation with the manual count from a manual count specialist. The computer processing time for executing the developed codes on an image of size 300 x 300 x 3 pixels was found to be less than 10.0 s using a 64-bit Intel (R) Celeron (R) CPU with processing speed of 2.20 GHz.

Acknowledgments
The authors are grateful to the Office of External Activities (OEA) and the Associate Scheme of Abdus Salam ICTP Trieste, Italy for their financial support during their stay at the centre and the use of the library and computing facilities to write this paper. We are also grateful to International Programme for Physical Sciences (IPPS), International Sciences Programme (ISP), Uppsala University, Sweden and Japan International Cooperation Agency (JICA) for funding and provision of microscopes. We are also grateful to the Central Regional Hospital at Cape Coast for allowing us to use their personnel and laboratory facility. Also to Justice Kwabena Sarfo (PhD), of the Biochemistry Department University of Cape Coast for reading through the manuscript text.

7. REFERENCES

[1] M.T. Makler, et al., "*A review of practical techniques for the diagnosis of malaria* " Annals of Tropical Medicine & Parasitology 92, pp. 419-433, 1998.

[2] R.S. Phillips. "Current status of malaria and potential for control" Clin. Microbiol. Rev.14, pp. 208-226, 2001.

[3] C. Shiff. "Integrated approach for malaria control " Clin. Microbiol. Rev. 15, pp. 278-293, 2002.

[4] E. Korenromp et al., World Malaria Report 2005. In Tech rep. World Health Organization, Geneva, 2005.

[5] H. Noedl, C. Wongsrichanalai and W.H. Wernsdorfer. "Malaria drug sensitivity testing: new assays, new perspectives" Trends Parasitol 19, pp. 175-181, 2003.

[6] C.E. Contreras et al., "Stage specific activity of potential antimalarial compounds measured in vitro by flow cytometry in comparison to optical microscopy and hypoxanthine uptake," presented at Mem. Inst. Oswaldo Cruz, Rio de Janeiro 99, Brazil, 2004.

[7] W.H.O. Basic malaria microscopy Part I Learner's Guide, World Health Organization, 1991, pp. 7-10.

[8] M.D. Pammenter. "Techniques for the diagnosis of malari." S. Afr. Med J. 74, pp. 55-57, 1988.

[9] P.B. Bloland. "Drug resistance in malaria," presented at WHO/CDS/CSR/DRS/ 2001.4, World Health Organization, Switzerland, 2001.

[10] S.C.J. Oaks et al., "Malaria: obstacles and opportunities, A report of the committee for the study on malaria prevention and control: Status review and alternative strategies," National Academy Press, Washington, DC, 1991.

[11] D. Payne "Use and limitations of light microscopy for diagnosing malaria at the primary health care level" Bulletin of the World Health Organization 66, pp. 621-626, 1988.

[12] R.E. Coleman et al., "Comparison of field and expert laboratory microscopy for active surveillance for asymptomatic Plasmodium falciparum and Plasmodium vivax in Western Thailand" Am J Trop Med Hyg. 67, pp. 141-144, 2002.

[13] K. Mitiku et al., "The reliability of blood film examination for malaria at the peripheral health unit" Ethiop J. Health Dev 17, pp. 197-204, 2003.

[14] I. Bates, et al., "Improving the accuracy of malaria-related laboratory tests in Ghana" Malar J., 01 Nov. 2004.

[15] C. Di Ruberto et al., "Morphological Image Processing For Evaluating Malaria Disease" in Proc. Int Workshop on Visual Form, Capri, Italy, 2001.

[16] C. Di Ruberto et al., "Analysis of infected blood cell images using morphological operators" Image and Vis. Comput. 20, pp. 133-146, 2002.

[17] F.B. Tek et al., "Malaria Parasite Detection In Peripheral Blood Images" in Proc Med Image Underst. and Anal Conf, Manchester, UK , 2006.

[18] F.B. Tek, et al., "Malaria Parasite Detection In Peripheral Blood Images," in Proc Br Mach Vis Conf, Edinburgh, UK, 2006.

[19] S. Halim, et al. "Estimating Malaria Parasitaemia From Blood Smear Images," in Proc. IEEE Int. Conf. Control Autom. Robot and Vis, Singapore, 2006.

[20] N.E. Ross et al., "Automated image processing method for the diagnosis and classification of malaria on thin blood smears" Medical and Biological Engineering and Computing 44, pp. 427-436, 2006.

[21] S.W.S. Sio et al., "Malaria Count: An image analysis-based program for the accurate determination of parasitemia" Journal of Microbiological Methods 68, Science Direct, pp. 11-18, 2007.

[22] F.B. Tek et al. "Computer vision for microscopy diagnosis of Malaria" Malaria Journal, 13 July 2009.

[23] S.S. Savkare, S.P. Narote. "Automatic Detection of Malaria Parasites for Estimating Parasitemia" International Journal of Computer Science and Security (IJCSS) 5, pp. 310-315, 2011.

[24] A. Hanbury, J. Serra. "Mathematical Morphology in the L*a*b* Colour Space" Technical report N-36/01/M, Centre of Mathematical Morphology, France, 2001.

[25] G. Mittal et al., "An efficient video enhancement method using L*a*b* analysis" IEEE Computer Society 79, pp. 66-70, 2006.

[26] K. León et al., "Color measurement in L*a*b* units from RGB digital images" Food Research International 39, pp. 1084-1091, 2006.

[27] D. Mery, F. Pedreschi. "Segmentation of colour food images using a robust Algorithm" Journal of Food Engineering 66, pp. 353-360, 2005.

[28] R.W.G. Hunt. "Measurement of Colour Appearance" J. Opt. Soc. Am. 55, pp. 1540-1551, 1965.

[29] L. Kaufman and P.J. Rousseeuw. Finding groups in data: An introduction to cluster analysis, John Wiley & Sons Inc., Hoboken New Jersey, 2005, pp. 155-160.

[30] W.L. Martinez and A.R. Martinez. Exploratory Data Analysis with MATLAB®, Computer Science and Data Analysis Series, Chapman & Hall/CRC, 2005, pp. 135-139.

[31] Center for Disease Control and Prevention (CDC) (2010) Retrieved from http://phil.cdc.gov/phil/home.asp on November 12 2010.

[32] Mathworks Inc., R2010a MATLAB 7.10.0, 2010.

[33] N. Otsu. "A Threshold Selection Method from Gray-Level Histograms" IEEE Transactions on Systems, Man and Cybernetics, 9 (1), pp. 62-66, 1979.