Automatic Detection of Malaria Parasite based on Microscopic Image Analysis

By

Abebe Bekele

A master thesis

Presented in partial fulfillment of the requirements for the Degree of Master of Science in Biomedical Engineering

Addis Ababa Institute of Technology

Addis Ababa University

Advisor: Masreshaw Demelash (PhD)

Co-advisor: Mr. Geremew Tasew (MSc)

Addis Ababa, Ethiopia, February 2017

Declaration

I, the undersigned, declare that this thesis is my original work. It has never been presented for a degree in any other institution and that all sources of materials used in it have been duly acknowledged.

Name: _____

Signature: _____

Date: _____

This MSc. thesis has been submitted for examination with my approval as an advisor.

_____

Masreshaw Demelash (PhD)

Certificate of Examination

This is to certify that the thesis prepared by Abebe Bekele entitled *'Automatic Detection of Malaria Parasite based on Microscopic Image Analysis'* submitted in partial fulfillment of the requirements for the degree of Master of Science in Biomedical Engineering (Bioinstrumentation and Imaging) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee

Examiner _____Signature _____ Date _____

Examiner _____Signature _____ Date _____

Advisor _____Signature _____ Date _____

_____

Chief of Department or Graduate Program Coordinator

Acknowledgements

First and foremost, I would like to express my gratitude to Dr. Masreshaw Demelash for his recommendations and support with this thesis and for being my supervisor.

I am grateful to my Co-supervisor Mr. Geremew Tasew for sharing with me his expertise in malaria disease and encouraging me during all course of the research.

I would also like to thank Adama Malaria Center for characterizing and permitting free use of their malaria samples.

Finally, I would like to extend my sincere gratitude to the staff of Malaria and other Parasite Research Laboratory at the Ethiopian Public Health Institute (EPHI) for their collaboration during image acquisition.

Abstract

Automatic Detection of Malaria Parasite based on Microscopic Image Analysis

Abebe Bekele

Addis Ababa University, 2017

Malaria is a serious global health problem and its diagnosis is usually done manually by compound light microscopy which is time consuming, tiresome and subjective. To support this manual method, in this master thesis, we designed and developed a system which is able to automatically detect plasmodium parasites from images of blood smears acquired by ourselves using a digital light microscope.

In this method, blood smears taken from patients who were infected with plasmodium parasites were prepared. Digital images were then acquired by the light microscope and saved in the computer. Red blood cells (RBCs) are first segmented by marker control watershed algorithm, where the foreground markers are obtained from circular Hough transform and background markers from distance transform. The plasmodium infected RBCs are then detected in the Hue-Saturation-Intensity (HSI) color space. Thresholding on hue component of HSI color space is used to detect the chromatin dots of the parasite. Plasmodium falciparum and plasmodium vivax, the two dominant plasmodium species which cause the vast deaths in Ethiopia, are differentiated based on the size of infected RBCs.

The performance of the proposed system for RBC segmentation, parasite detection and species differentiations was analyzed by comparing with the gold standard manual method for the total of 91 images of thin blood smears. The result shows that 97% of the RBC counts are similar to the gold standard with 97.5% sensitivity and 84.4% positive predictive value for plasmodium parasite detection at the cellular level. The species differentiations were done for each image with the accuracy level of 91.46%.The result showed the potential of the method for supporting the mass screening of malaria parasite.

Keywords: *Digital Microscope, Plasmodium, Thin Blood Smears, Watershed Algorithm, Circular Hough Transform, Distance Transform, Hue-Saturation-Intensity (HSI), Thresholding*

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

Acronyms

CCD	Charge Coupled Device
CDC	Central Disease Control
CDC DPDx	CDC Parasite Image Library
C-Y	Cyan Yellow
DINF	Double Infection
EPHI	Ethiopian Public Health Institute
FN	False Negative
FP	False Positive
GUI	Graphic User Interface
GUIDE	Graphical User Interface Development Environment
HIS	Hue Intensity Saturation
HSV	Hue Saturation Value
HT	Hough Transform
PCR	Polymerase Chain Reaction

PF	Plasmodium Falciparum
PLT	Platelet
PPV	Positive Predication Value
PV	Plasmodium Vivax
RBC	Red Blood Cell
RGB	Red Green Blue
ROI	Region of Interest
SVM	Support Vector Machine
TP	True Positive
UICODE	User Interface Controls
WBC	White Blood Cell
WHO	World Health Organization

Chapter 1 - Introduction

1.1 Malaria Disease

Malaria is a very serious infectious disease caused by a protozoan parasite of the genus Plasmodium. It is a global health problem causing over 500 million human infections annually, resulting in about 1-2 million deaths, of whom 90% are children in sub Saharan Africa [1]. Figure 1.1 gives an idea how parts of the world have been affected by the disease.

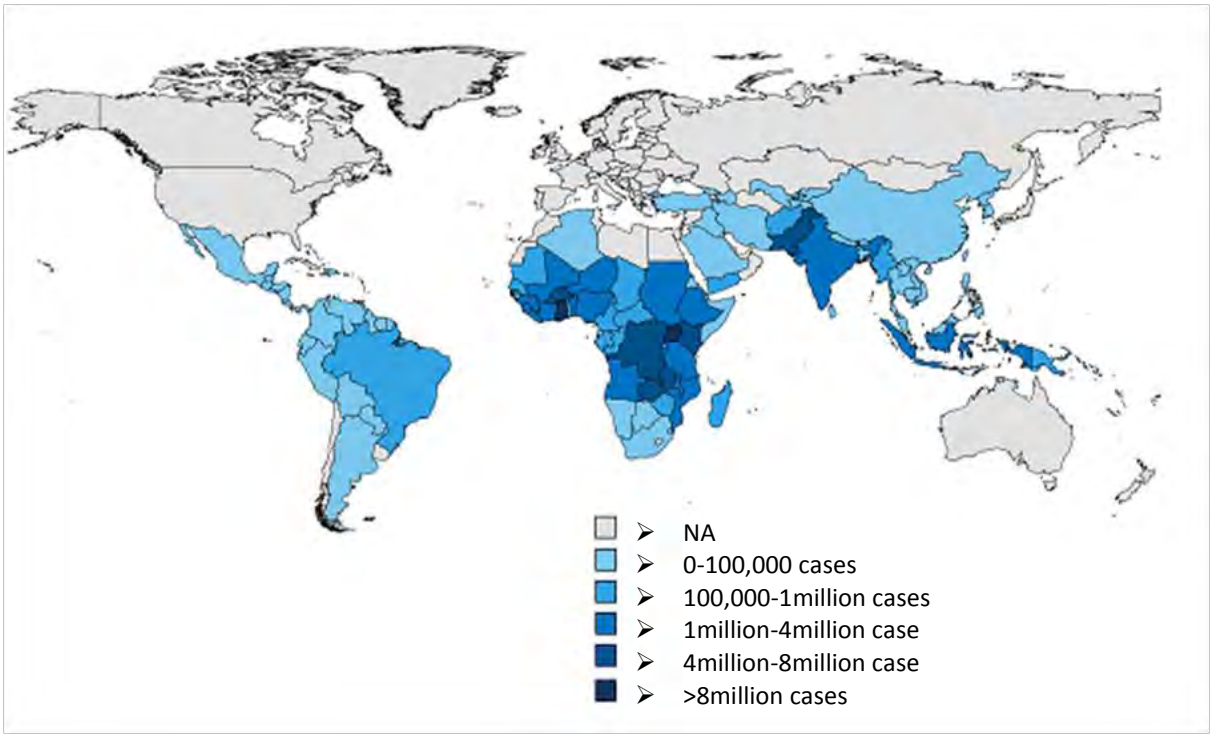


Figure 1-1: Reported Malaria Cases, 2012 [2]

The overall impact of the disease around the world is therefore massive, and it is an enormous burden to developing countries like Ethiopia. It is estimated that about 75% of the landmass of Ethiopia is malarious and 68% of the Ethiopian populations, estimated at

about 54 million in 2010, live in malaria risk areas [3] which resulted in an average of 9.5 million infections per year between 2001 and 2005 with 70,000 deaths each year [4]. The fact that malaria is now moving to the highland regions of Ethiopia is also highly alarming [5].

1.2 Malaria Parasite

Malaria is transmitted by infected female Anopheles mosquitoes which carry Plasmodium sporozoites in their salivary glands. There are four species of Plasmodium parasites that infect human and result in four kinds of malarial fever: Plasmodium Falciparum (*PF*), Plasmodium Vivax (*PV*), Plasmodium Ovale (*PO*), and Plasmodium Malariae (*PM*) [6]. *PF* is most common in tropical and subtropical areas. It causes the most dangerous and malignant form of malaria without relapses and contributes to the majority of deaths associated with the disease. *PV* shows the widest distribution and is characterized by reappearances of symptoms after a latent period of up to five years. With similar characteristics, *PO* appears mainly in tropical Africa. *PM,* is also widely distributed but much less than *PV* or *PF*. In Ethiopia *PF* and *PV* are the most dominant malaria parasites, distributed all over malarious areas of the country and accounting for 60-70% and 30-40% cases respectively [7].

There are three phases of development in the life cycle of most species of plasmodium [8]:

- Exo-erythrocytic stages in the tissues (usually in the liver)
- Erythrocytic schizogony (i.e. protozoan asexual reproduction) in the erythrocytes
- Sexual process (beginning with the development of gametocytes in the host and continuing with the development in the mosquito)

When an infected mosquito bites humans, as shown in Figure 1-2, several hundreds of sporozoites (the protozoan cells that develop in the mosquito's salivary gland and infect new hosts) may be injected directly into the blood stream, where they remain for about 30 minutes and then disappear. Many sporozoites are destroyed by the immune system, but some enter the cells in the liver. Here they multiply rapidly by a process referred to as exo-erythrocytic schizogony. When schizogony is completed, the parasitic cells produced by

asexual reproduction in the liver, termed as merozoites, are released and invade the erythrocytes.

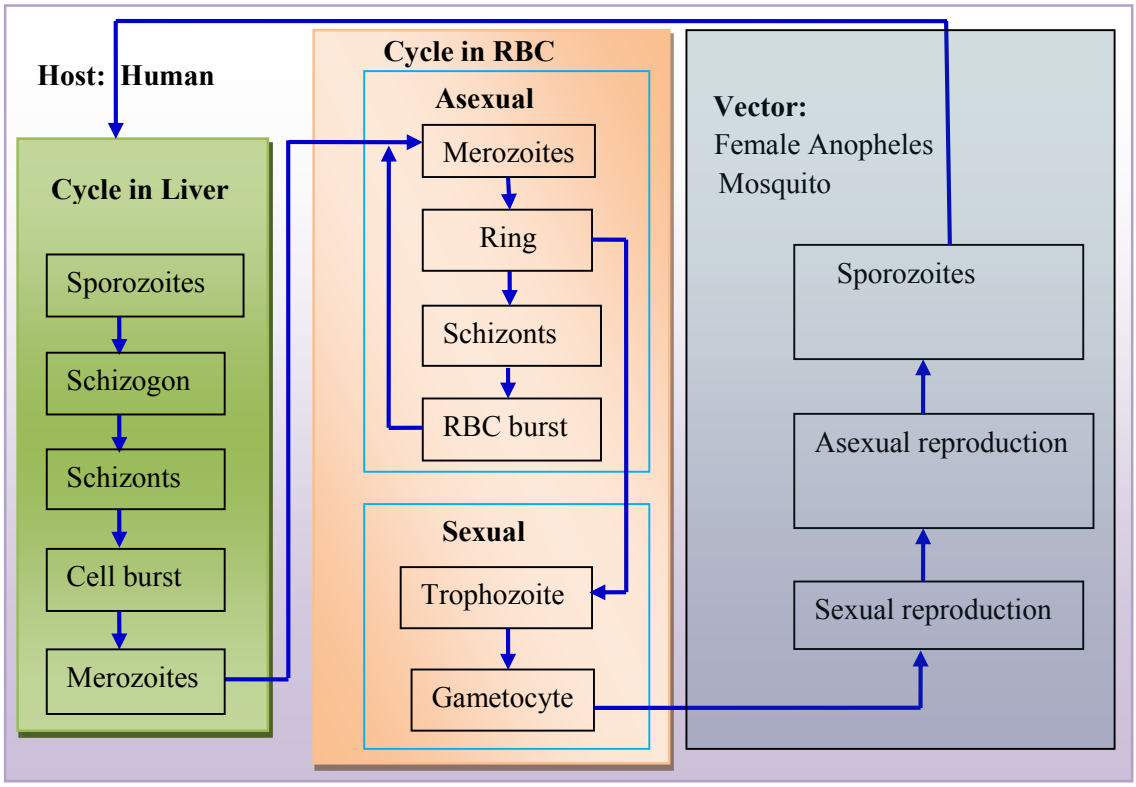


Figure 1-2: Life Cycle of Plasmodium Species

When the released merozoites enter erythrocytes, the erythrocytic cycle begins. This process is referred to as erythrocytic schizogony. Within an erythrocyte, the parasite is first seen using light microscope as a minute speck of chromatin surrounded by scanty protoplasm. The plasmodium gradually becomes ring-shaped and is known as ring or immature trophozoite. It grows at the expense of the erythrocyte and assumes a form differing widely with the species but usually exhibiting active pseudopodia (i.e. projections of the nuclei). Pigment granules appear early in the growth phase and the parasite is known as a mature trophozoite. As the nucleus begins to divide, the parasite is known as a schizont. Dividing nucleus tends to take up peripheral positions and a small portion of cytoplasm gathers around each.

The infected erythrocyte ruptures and releases a number of merozoites which attack new corpuscles and the cycle of erythrocytic schizogony is repeated. The infection about this time enters the phase in which parasites can be detected in blood smears.

1.3 Microscopic Diagnosis of Malaria

In general malaria diagnosis involves identifying malaria parasites or antigens produced in patients' blood. Rapid and effective early malaria diagnosis does not only alleviate suffering of the patients and save lives of patients, but also decreases the rate of malaria transmission to the rest of the community. Microscopic examination of blood film is the most common method of malaria diagnosis. A number of new methods such as polymerase chain reaction (PCR) have been developed in recent years for the diagnosis of malaria. Though PCR is more accurate than microscopic examination, it is expensive, and requires a specialized laboratory [9].

As a result, microscopic examination of blood films remains the gold standard for laboratory diagnosis of malaria because it is economical and reliable method. In this method, diagnosis initially requires observing blood smeared slide under the compound light microscope for detecting the presence or absence of parasites. The practitioners generally observe two types of slides using fast Giemsa stain protocol to highlight the parasites. The two types of blood slides are thick smeared and thin smeared slides, where the violet colored dots within the RBC can be identified as stained parasites.

The probability of detecting parasite in thick blood smeared slides is higher because of the larger volume of blood observed. If the expert suspects the presence of Malaria parasite in thick blood smeared slides then the diagnosis process is followed by examination of a thin smeared blood slide.

The thin blood smear slide is used for the enumeration of the infection. The process of enumeration requires manual count of the number of RBCs in the observed microscopic field followed by manual counting of number of infected RBCs. Apart from enumeration, medical laboratory specialists also need to report the life cycle and the species causing the infection; thus the process of diagnosis is extensively dependent on experts.

The simplest guide to distinguish between the four species of malaria is to look at the effect the parasite has on infected RBSs.

The size of infected RBC, and whether or not staining reveals Schüffner's dots or Maurer's dots (also known as Maurer's clefts) within the cell are the most common features often used to differentiate the species. Table 1 summarizes the rules for species differentiation.

Parameter	P.falciparum	P.vivax	P.ovale	P.malariae
Size of RBC	Not enlarged	Enlarged	Enlarged	Not enlarged
Shape of RBC	Round	Round or oval	Round	Round
Chromatin dots	Large red dots	Small red dots	Small red dots	Few tiny dots
Mature schizont	12-30 merozoites	12-24 merozoites	4-12 merozoites	6-12 merzoites

Table 1: Some rule for species differentiation

1.4 Statement of the Problem

Though the manual microscopic examination is still the gold standard for malaria diagnosis, it is tedious, time consuming and requires special training and considerable expertise. Various studies have aimed to assess the accuracy of the manual microscopic diagnosis of malaria. It was shown that this manual method itself may not be a reliable screening method when it is performed by clinical examiners who have lack of training; especially in the rural areas where malaria is endemic [10]. A pervious field study in Ethiopia showed that the agreement rates among clinical examiners for a set of samples are below expectations (i.e. corrected agreement of the operational and reference reader was less than 0.53) [11]. This means that the result of manual malaria screening technique is subjective, i.e., the result varies among different observers.

In this thesis, a system is developed that can automatically detect malaria parasites whereby the decision is made by a computer using image processing algorithms applied on digital images acquired on a digital microscope.

1.5 Research Objective

General Objective

The whole purpose of this thesis is to develop a system that detects malaria parasites automatically from microscopic images such that malaria diagnosis could be fast, reliable with minimum human reliance, less subjective and is, therefore, more consistent in applying diagnostic criteria.

Specific Objectives

- Acquiring the digital images of locally available plasmodium infected samples
- Designing and developing an automatic system for:-
    - RBCs segmentation
    - Plasmodium detection
    - Identification of plasmodium infected RBCs
    - Determining the number of infected and non-infected RBCs
    - Determining the level of parasite infection
    - Differentiating between *PF* and *PV*
- Comparing the processed method with the gold standard method

1.6 Materials and methods

Malaria samples were collected from Adama malaria center. Thin blood films were prepared from collected samples and stained using Giemsa protocol. All slides had already been examined and verified by expert microscopists who had given a species-specific diagnosis. Digital Images of the samples were acquired using acquisition setup found at the Ethiopian Public Health Institute (EPHI), Parasitology Laboratory.

There are many different paradigms for image analysis in computer vision, which can be utilized to achieve the objectives listed in the previous section. In this thesis, however, marker controlled watershed transform for RBC segmentation and thresholding in the HSI color space for plasmodium detection have been used. The details of materials and methods are presented in chapter 3.

1.7 Relevance of the Research

The introductions of high speed computers and fast CCD cameras have made image processing one of the vital tools used during medical screening and diagnosis. Malaria is one of the serious health problems particularly in developing countries like Ethiopia. However malaria diagnosis is often made through visual inspection of blood samples under a microscope. As rigorous literature search in this thesis showed no local research have been done to automatically detect malaria parasite from locally available microscopic images. This research is intended to alleviate the problems related to manual microscopic diagnosis of malaria which is entirely dependent on the skill, experience and motivation of a human technician. Automated plasmodium detection from microscopic images would allow entire slides to be examined uniformly by a computer with no fatigue. It would also constitute a diagnostic aid for the increasing number of cases of imported malaria in traditionally malaria free areas, where practitioners lack experience on the disease. This method use computer and digital camera in addition to microscope, which incur additional cost, but if once the system is fully developed it could be easily adopted on an Android application to reduce the cost. This is stated as one of the future work in section 6.2.

1.8 Literature Review

In various literatures, different approaches have been reported on image analysis for use in malaria detection. In this section, we present few of the approaches related to microscopic medical image analysis.

Damahe *et al.* [12] presented an approach to detect the RBCs and the parasite in the blood cell images in view of malaria detection. Their approach was applied on the S and V components of the microscopic images represented in the HSV color space. The images were segmented using Zack's thresholding technique, sequential edge linking algorithm and Euclidian distance based clustering. The approach was implemented on Pathology and CDC DPDx blood cell images.

Abdul-nasir *et al.* [13] presented a color image segmentation approach for detection of malaria parasites that has been applied on malaria images of P. vivax species.

In this method, in order to segment RBCs infected with malaria parasites, the images are first enhanced by using partial contrast stretching. Median filter and seeded region growing area extraction algorithms have been applied for smoothing the images and remove any unwanted regions from the image, respectively. Then, an unsupervised segmentation technique, namely k-means clustering, is used to segment the infected cells from the background. Different color components of RGB, HSI and C-Y color models were analyzed to identify color component that can give significant segmentation performance. Savkare and Narote [14] proposed an automatic approach which uses Otsu's thresholding on gray images and green channel of the blood images for cell segmentation, watershed transform for separation of touching cells. Color and statistical features were extracted from segmented cells and SVM binary classifier was used for classification of normal and parasite infected cells.

Marker constrained watershed algorithm for image segmentation which uses only the binary images and distance transform can lead to over segmentation. If the image objects are irregular in shape and the objects are overlapping and touching, that could lead to under segmentation [13]. Rao and Dempster [15] suggested morphological area opening on distance transform for choosing markers to avoid over segmentation and under segmentation. The morphological area opening shrinks the size of the original object and because of this it is mostly applicable for objects of uniform size. Here in our case the sizes of the objects (RBC, platelets and WBC) have non-uniform area which may result in elimination of some of the objects by area opening.

In this thesis, circular Hough transform (CHT) has been used to split RBCs from background and other blood cells (WBCs and platelets). The circular nature of RBCs is suitable to implement CHT efficiently. Once the RBCs are detected we can use the center pixel of the RBCs as a foreground marker to apply marker control watershed algorithm for RBC segmentation. This is important to distinguish and separate connected RBCs.

The distance transform is used as background marker to demarcate the border of RBCs to avoid any distortion of size during segmentation. The size of RBCs are important for subsequent differentiation of plasmodium species.

1.9 Organization of the Thesis

In chapter 1, the extent of effects of malaria disease across the globe, in Africa and Ethiopia has been presented. The developmental stages of the malaria parasite and the means of transmission were also explained. The different diagnosis methods used to detect the malaria parasite and the problem related to the microscopic diagnosis are briefly discussed. The general and specific objectives of the thesis, materials and methods used the relevancy of the research and literature reviews were also presented in this chapter. The remaining parts of the thesis have been organized in the following way.

Chapter 2 discuss the fundamental theories and mathematics tools used for digital image processing. Digital color images and digital monochromic images are briefly discussed. Morphological image processing, in conjunction with the fundamental morphological operations (erosion and dilation) and their derivatives(image reconstruction, labeling, region filling and boundary extraction) which are helpful in the development of the proposed algorithm in this thesis are briefly explained.

Chapter 3 presents materials required and methods flowed to do the current research. The method includes sample collection, blood smear preparation, staining, digital image acquisition, RBC segmentation, plasmodium detection and species differentiation.

Chapter 4 briefly describes the design and development of the automated system for RBC segmentation, plasmodium detection and species differentiation. The flow chart of the algorithm used for RBC segmentation, plasmodium detection and species differentiation accompanied by the processed images are described in detail.

Chapter 5 presents the results and discusses the proposed method by comparing it against the gold standard method used for RBC segmentation, plasmodium detection and species differentiation. The comparison is made for *PF* infected images, *PV* infected images and non-infected images used as controls.

Chapter 6 illustrates the conclusion and future prospective of the work.

Chapter 2 - Morphological Image Processing

2.1 Digital Images

A continuous grayscale image is defined as a two dimensional function f(x, y). The spatial coordinates are represented by the values of x and y. The intensity of the image at coordinates (x, y) is the output f(x, y). A digital image is formed by sampling the function f(x, y) at discrete values of x and y and through quantization of the output intensity f.

2.1.1 Digital Monochrome Images

For a monochrome image, the intensity ranges between black and white called grayscale. In a digital image equipped with an 8-bit camera, it is common to define the intensities as integers where black is intensity 0 and white is intensity 255. Subsequent to sampling and quantization, the digital image is represented as a matrix of integer values containing M rows and N columns; we say the digital image is of size MxN. The elements of the matrix are known as pixels. The discrete coordinates x and y are integer valued and indicate row and column positions respectively. It is common to define the origin of the digital image as (0, 0). However the image processing toolbox in MATLAB defines the origin to be (1, 1); we adopt this convention. For example, the next pixel below the origin will be at (2, 1) and the pixel in the last row and last column is at position (M, N). It is important to note that the coordinate (x, y) in the digital image is the position of pixels and not the position of sampled points in the continuous image. The matrix representation of a digital image f(x, y) is shown below:

$$f(x,y) = \begin{pmatrix} f(1,1) & f(1,2) & \dots & f(1,N) \\ f(2,1) & f(2,2) & \dots & f(2,N) \\ \vdots & \vdots & \vdots & \vdots \\ f(M,1) & f(M,2) & \dots & f(M,N) \end{pmatrix} \qquad 2.1$$

2.1.2 Digital Color Images

Color images are composed of individual monochrome images. The RGB color system uses three 2D component images r(x, y), g(x, y) and b(x, y) representing the red, green and blue values respectively. The color image can be thought of as MxNx3 array where the red, green and blue component images are stacked along the 3rd dimension.

An illustration of RGB color representation is shown in Figure 2-1. The color of a pixel (x, y) is characterized by a 3D vector p(x, y) = [r(x, y), g(x, y), b(x, y)]. When all three values are identical the pixel is displayed as grey.

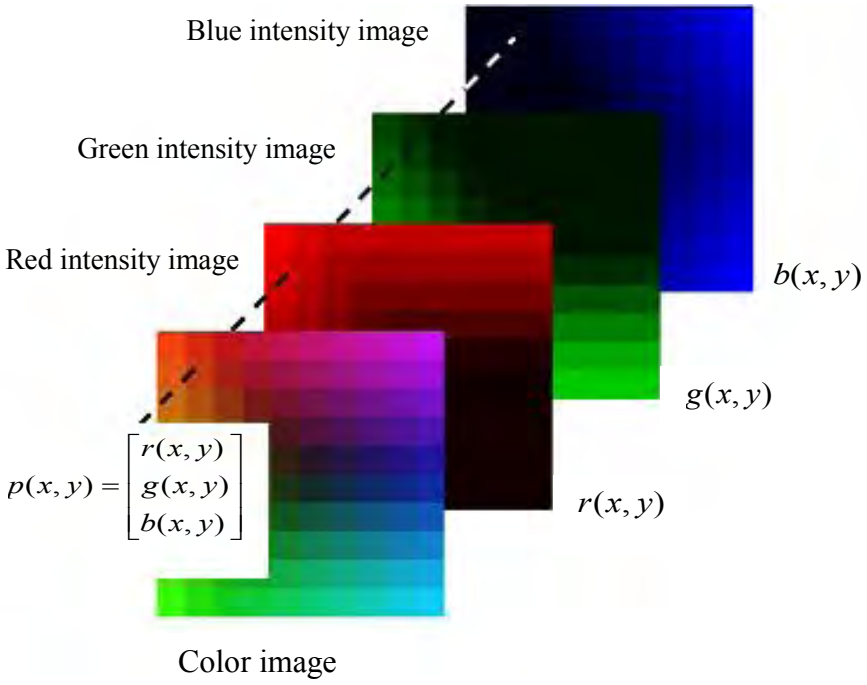


Figure 2-1 : Digital color Image representation using the RGB color system

2.2 Morphological Image Processing

The word morphology commonly denotes a branch of biology that deals with the form and structure of animals and plants. We use the same term here in the context of mathematical

morphology. It was developed about forty years ago and is a powerful tool for image processing [16]. Unlike other tools such as Fourier method, morphological image processing is directly related to the image shape [16], and it has many applications. Erosion and dilation are the two primitive operations which are fundamentals to understand morphological operations and many other morphological algorithms [17].

Erosion operation is a morphological operation for reducing the foreground area. The effect of this operation is shrunk foreground. The foreground is reduced from its outer edge to inside its area. If there is a hole inside the foreground area, the hole enlarges. It uses a structuring element and it is done with a convolution operation between the image and the structuring element. This operation is done for binary images. The erosion process will set the foreground pixel to be background if there is part of the structuring element that reaches the background while the center of the structuring element reaches the foreground edge. Figure 2.2 gives an illustration of erosion process with the disk-shaped structural element.

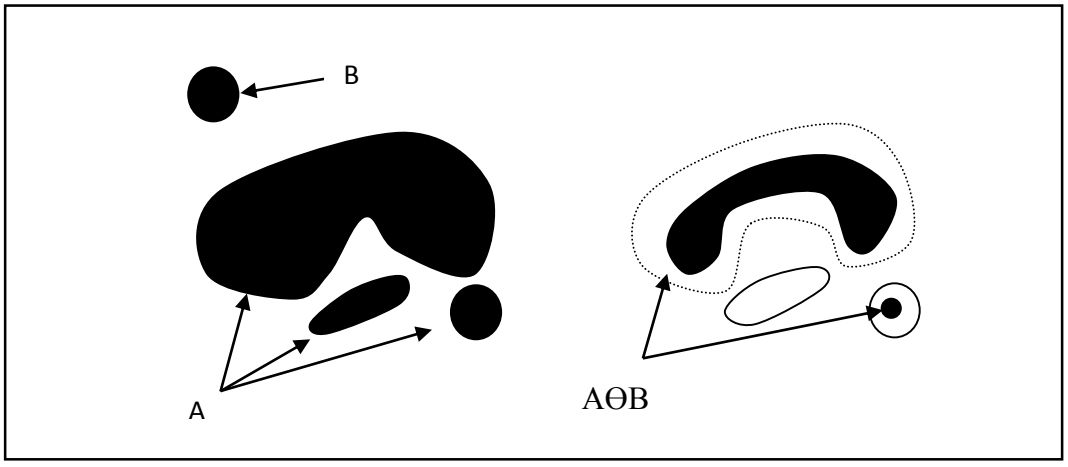


Figure 2-2: Erosion of a set A by a disk-shape structure element B

The erosion of A by B, denoted AΘB, is defined as

$$A\Theta B = \{z|\ (B)_z \subseteq A\} \qquad 2.2$$

Where, the notation $B \subseteq A$ means B is a subset of A. This equation says that the erosion of A by B is the set of all points z such that B, translated by z, is contained in A.

Dilation operation is the reverse operation of erosion. While erosion is for reducing the foreground, dilation is for enlarging it. The foreground is stretched from its outer boundary. If there is a hole inside the foreground, the hole shrinks. Like erosion, dilation operation uses structural element. The structural element is used in convolution with the image.

The dilation process will set the background pixel to be foreground if there is part of the structuring element that reaches the foreground when the center of the structuring element is still in the background area. Figure 2.3 gives an illustration of dilation process with the disk-shaped structural element.

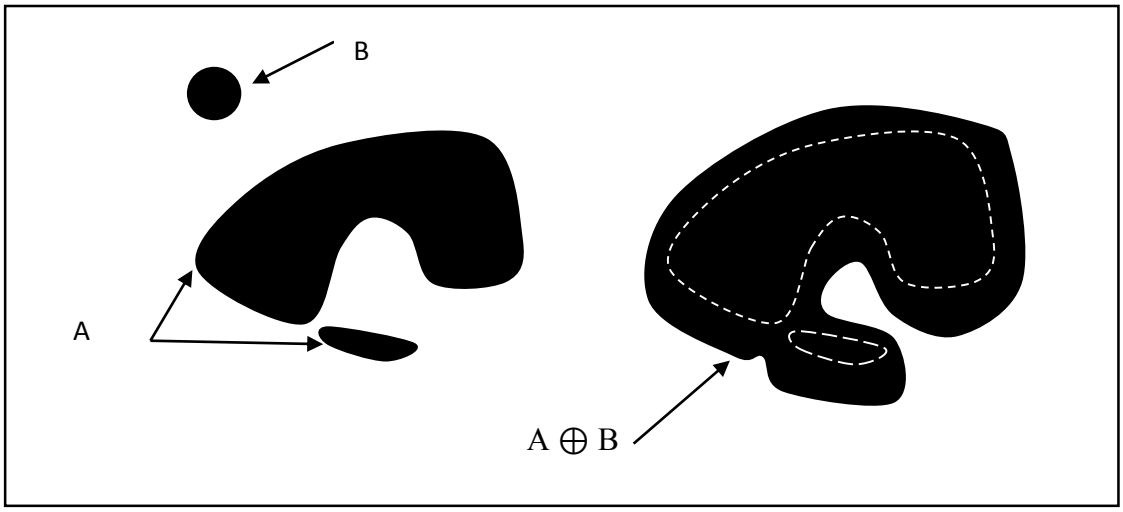


Figure 2-3: Dilation of a set A by a disk-shaped structural element B

The dilation of A by B, denoted $A \oplus B$, is defined as

$$A \oplus B = \{z | (-B)_z \cap A \neq \emptyset\} \qquad 2.3$$

Where Ø is the empty set, B is the structural element and -B is the reflection of B. In words, the dilation of A by B is the set of all points such that the reflection of structuring element B translated overlap with A at least at one element.

It is a convention in image processing to let the first operand is the image and the second operand the structuring element, which usually is much smaller than the image. The next

parts of this section are heading for image reconstruction, labeling, region filling and boundary extraction which are helpful in the development of the proposed algorithm.

2.2.1 Morphological Reconstruction of Binary Images

Reconstruction is a morphological transformation involving two images and a structuring element (instead of a single image and structuring element). One image, the marker, is the starting point for the transformation. The other image, the mask, constrains the transformation. Figure 2-4 shows the morphological reconstruction of a binary image.

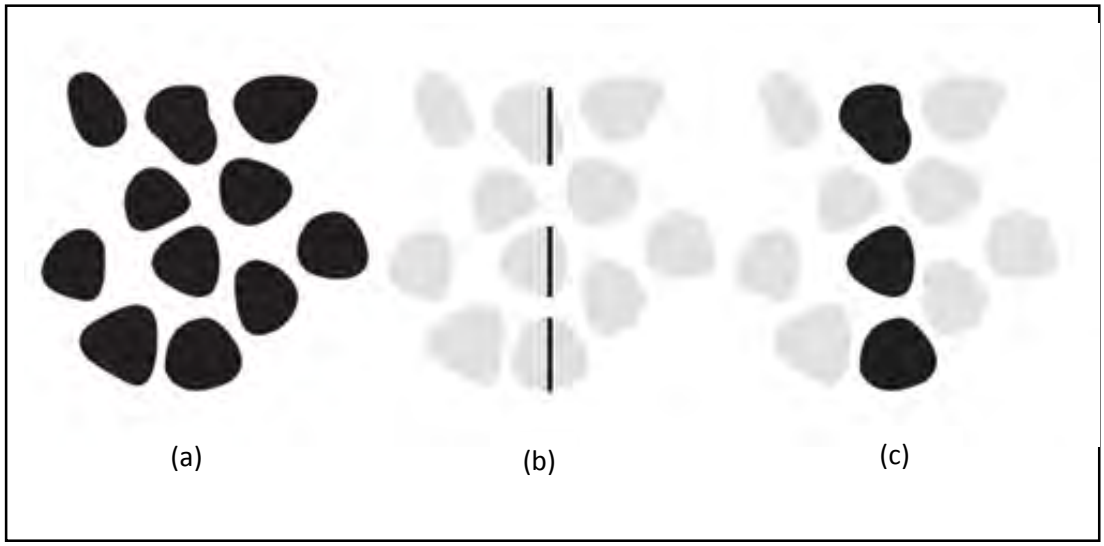


Figure 2-4: Reconstruction from markers
(a) Input image (b) Marker image (c) Reconstructed image

If G is the mask and F is the marker, the reconstruction of G from F, denoted RG (F), is defined by the following iterative procedure:

1. Initialize h_1 to be the marker image F
2. Create the structuring element B=ones
3. Repeat: $h_{k+1} = (h_k \oplus B) \bigcap G, until\ h_{k+1} = h_k$
4. $R_G(F) = h_{k+1}$; Marker F must be a subset of G: F ⊆ G

2.2.2 Labeling Connected Components

The concepts discussed so far are applicable mostly to all foreground (or all background) individual pixels and their immediate neighbors. In this section we consider the important "middle ground" between individual foreground pixels and the set of all foreground pixels. This leads to the notion of connected components, also referred to as objects in the following discussion.

A connected component is just defined in terms of a path, and the definition of a path in turn depends on the type of adjacency used. This implies that the nature of a connected component depends on which form of adjacency we choose; 4 and 8 are the most commonly used adjacencies. Figure 2-5 illustrates the effect that adjacency can have on determining the number of connected components in an image.

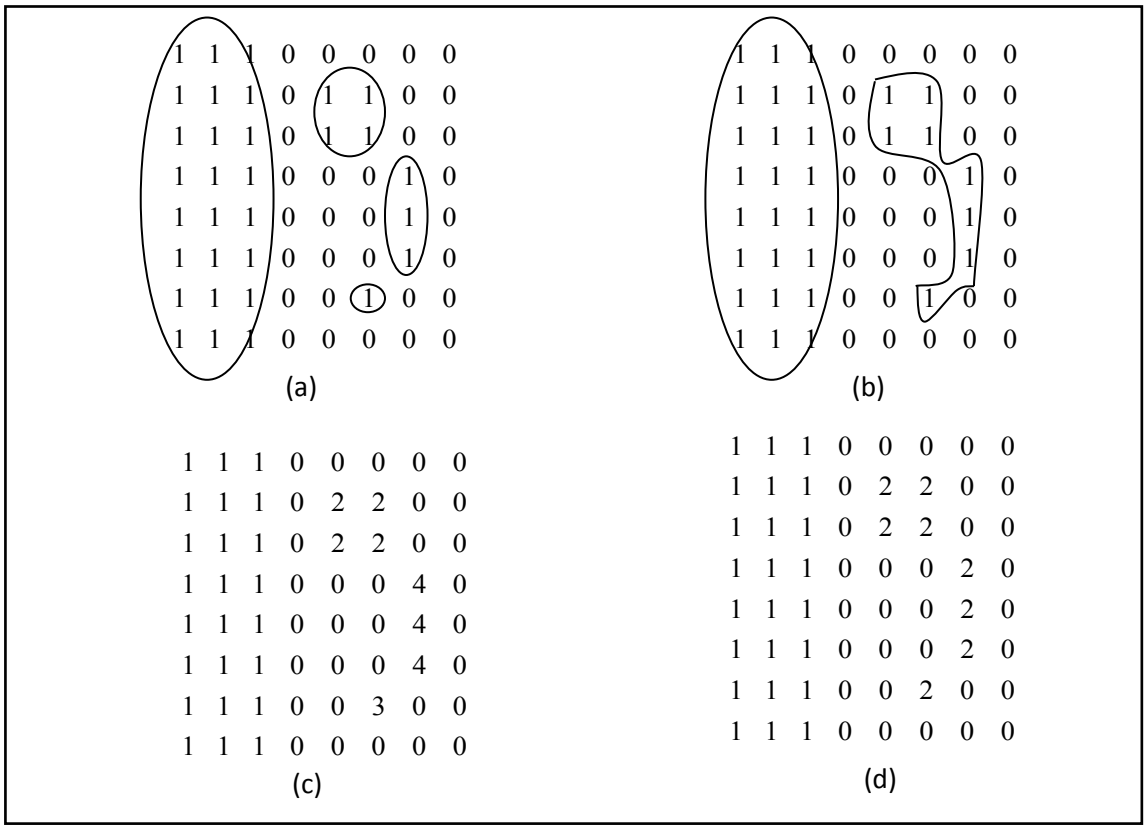


Figure 2-5: Connected components

(a) 4-connected components; (b) 8-connected components; (c) Labeled Matrix obtained using 4-connectivity; (d) Labeled matrix obtained using 8-connectivity

Figure 2-5(a) shows a small binary image with 4 connected components by using 4 adjacency. Figure 2-5(b) shows that choosing 8-adjacency reduces the number of connected components to two. Figure 2-5(c) indicate the labeling for 4 connected components and Figure 2-5(d) shows labeling of 8 connected components.

2.2.3 Region Filling

Hole filling operations are widely used in medical image processing [18]. Almost all the image processing operations of medical scans produce a binary form of original image in any stage. The binary forms of medical images are normally produced by simple segmentation techniques such as thresholding. They contain foreground objects surrounded by background regions.

Sometimes a set of background regions lying completely within the foreground regions appear due to imperfection in the binary conversion identified by the optimal thresholding. It is known as holes within the foreground objects or region of interest (ROI). For example, let I denotes a binary image and suppose that we choose the marker image F to be 0 everywhere except on the image border, where it is set to 1 - I:

$$F(x,y) = \begin{cases} 1 - I(x,y) \; if \; (x,y) \; is \; on \; the \; border \; of \; I \\ 0 \; otherwise \end{cases} \qquad 2.4$$

Only dark pixels of *I(x, y)* touching the border have a value of 1 in *F(x, y)*.The binary image with all regions (holes) filled is given by:

$$H = [R_{I^c}(F)]^c \qquad 2.5$$

Where R indicates reconstruction, c indicates complement and F the marker.

For example dilation based propagation can be used to fill interior holes of segmented objects in a thresholded image. Figure 2-6 shows an example of such a procedure. Starting from the binary segmented image of the object shown in Figure 2-6(a), one inverts this image to create a mask. Then the border of the image is used as the marker of a propagation (reconstruction) inward toward the mask.

This generates the image shown in Figure 2-6(b). Inverting this propagated image produces the desired result, which contains the object with all interior holes filled (Figure 2-6(c)).

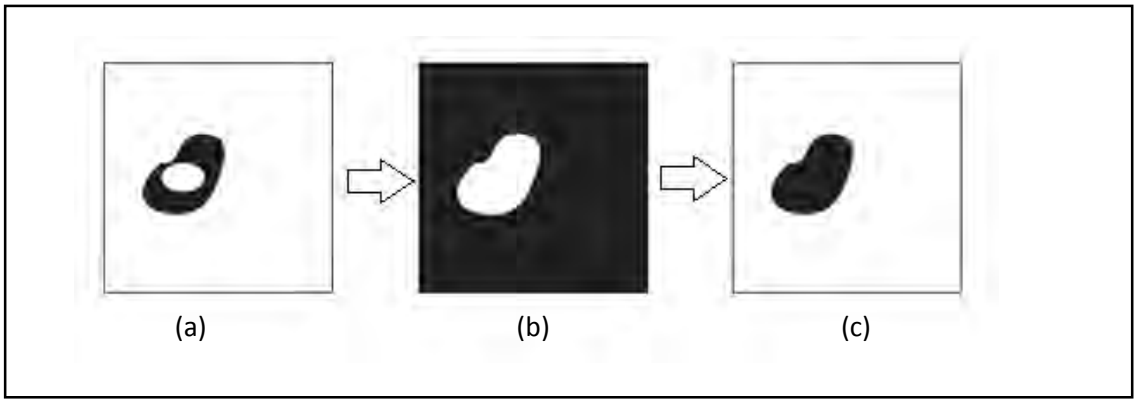


Figure 2-6: Filling interior holes of segmented objects using binary morphological operations

2.2.4 Border Object Rem0oval

Another useful procedure is the removal of border touching objects. In quantitative microscopy the objects that are connected to the image border are partially obscured and usually not suitable for subsequent analysis.

Suppose we define the marker image, *F*, as

$$F(x,y) = \begin{cases} G(x,y) \; if \; (x,y) \; is \; on \; the \; border \; of \; G \\ 0 \; otherwise \end{cases} \qquad 2.6$$

G was the original image. Then, using G as the mask, the reconstruction equals:-

$$H = R_G(F) \qquad 2.7$$

H contains only the objects touching the border *G-H. G-H* shows image that do not touch the borders.

Figure 2.7 illustrates the procedure how to eliminate border touching objects. Here the binary thresholded image (Figure 2-7(a)) is used as the mask, and the border of the image is used as the marker (equation 2.7). The border clearing algorithm first computes the morphological reconstruction which simply extracts the objects touching the border as

shown in Figure 2-7(b).Then obtains the new image with no objects touching the borders by equation G-H as shown in Figure 2-7(c).

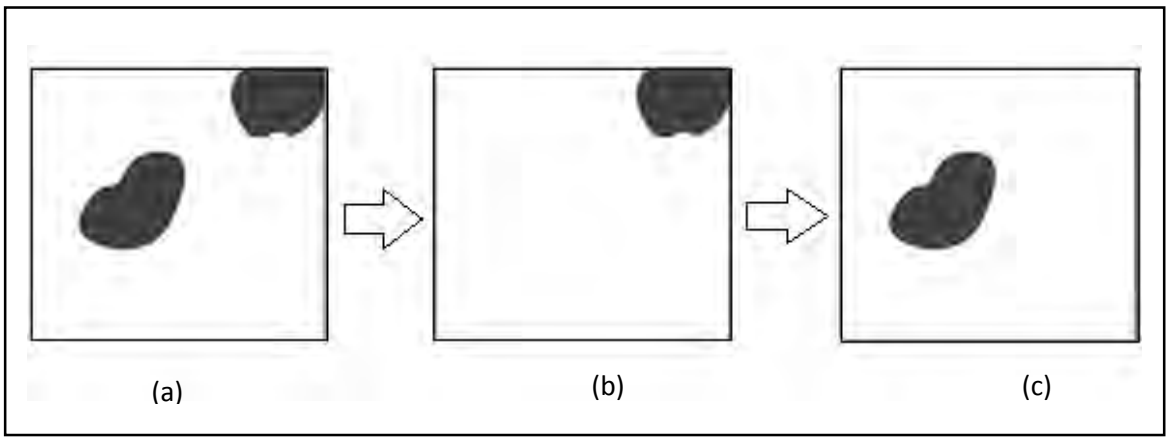


Figure 2-7: Removing border objects using binary morphological operations

Chapter 3 -Materials and Methods

This section explains the materials and methods used for automatic detection of malaria parasite. It begins by explaining blood film preparation, the staining technique applied and the image acquisition setup used to accomplish the intended objective of the thesis work. Finally, the details of image processing technique for RBC segmentation and plasmodium detection are discussed.

Malaria parasite detection using microscopic image analysis has similar approach as that of the manual microscopic diagnosis. It consists of the following procedure as indicated in Figure 3-1 which begins with data collection and ends with result verification.

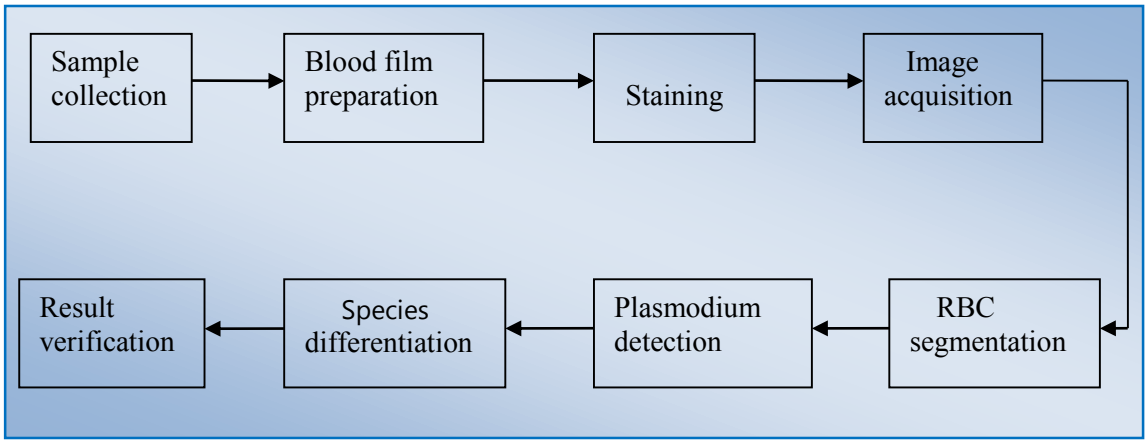


Figure 3-1: Block diagram of the research methodology

3.1 Blood Film Preparation

To establish the diagnosis of malaria, a blood film must be prepared from fresh blood obtained by pricking a patient's finger with a sterile, non-reusable lancet. Blood films are made by placing a drop of blood on one end of a slide, and using a spreader slide to disperse the blood over the slide's length. Two types of blood films preparation, thick and thin, are used during microscopic diagnosis of malaria. Both thick and thin films should be prepared and examined in all cases of suspected malaria.

3.1.1 Thick Blood Film

Thick blood film consists of a thick layer of lysed erythrocytes. The blood elements (including parasites, if any) are more concentrated (~30x) than in an equal area of a thin smear, allowing a greater volume of blood to be examined. Because a larger volume (~6 μl) of blood is examined in the thick film, it is much better than the thin film for detection of low levels of parasitemia and reappearance of circulating parasites during infection, recrudescence or relapse. Thick film is therefore the most suitable method for the rapid detection of the parasite. However, thick film method does not permit an optimal review of parasite morphology for species identification [3].

If the thick smear is positive for malaria parasites, the thin smear should be used for species identification. Thus, the thick films are performed to detect and quantify (parasite density) malaria parasites in routine malaria microscopic diagnosis.

3.1.2 Thin Blood Film

Thin blood film consists of blood spread in a layer such that the thickness decreases progressively toward the feathered edge. In the feathered edge, the RBCs should be in a single layer, not touching one another. Thin blood smear should be fixed with methanol so that the parasites are found intact inside the RBCs.

In thin film examination, the morphological identification of the parasite to the species level is much easier and provides greater specificity than the thick film examination. However, low-density infections can be missed and require more time to read. Thin blood film is used to assist in the identification of the malaria species after the parasites have been seen in the thick film examination. In this thesis, since we are going to differentiate species, the thin blood films were used. This blood films were prepared and stained at Adama Malaria Center and brought to the Ethiopian Public Health Institute (EPHI) for image acquisition using a digital microscope.

3.2 Staining

Microscope examination of blood is usually performed on stained samples. Staining is a chemical process which highlights the object of interest in the sample; usually it is specific to the application. Staining greatly helps to enhance the image contrast. The common staining technique used for malaria diagnosis in peripheral blood thin blood films is the Giemsa-stain; however, in thick films Field's staining technique is preferred due to its speed.

It is appropriate first to clarify the terms stained pixels and stained objects before embarking on the technical discussions. A stained pixel is a pixel in the image which has distinctive, saturated color compared to the color of the RBCs. A stained object can be roughly defined as a connected stained pixel group and the body that includes the group (if any). In the case of an infected RBC, the stained object is ideally the whole RBC region which includes the stained pixels. In the case of a white blood cell (WBC), it is ideally the WBC nucleus and surrounding cytoplasm. For a platelet, it is the only connected stained pixel group (since they do not have cytoplasm); artefacts appear randomly and sometimes it can be on top of the RBCs.

Figures 3-2 shows few examples of images that contain stained pixels. At this stage of the discussion, the categorization of the stained objects is not important; however to visualize the above definitions, the stained objects on the images are marked such that: Plasmodium, WBCs, artefacts, and platelets are indicated with the letters (P), (W), (A), and (PLT), respectively. Note that, for the plasmodium (P) the parasite shape and maturity is not always the same, hence the infected RBC appearances vary from one to another.

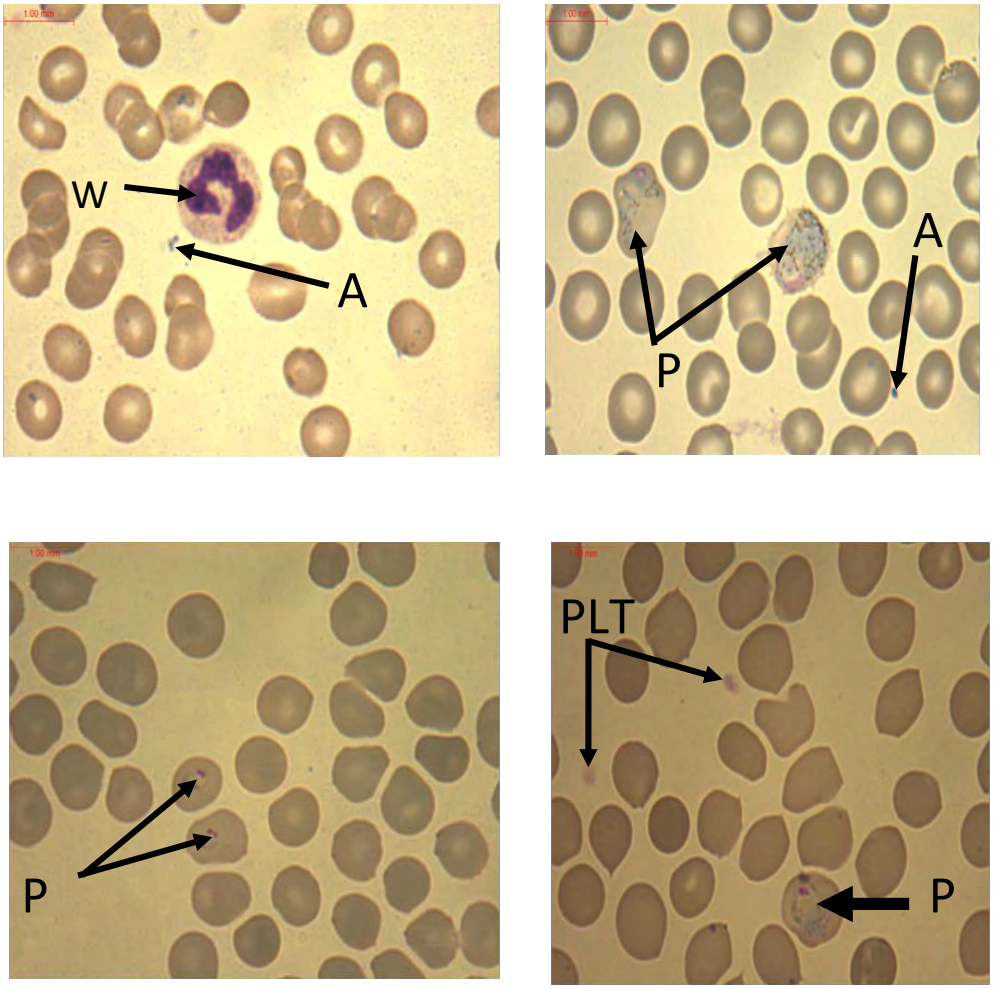


Figure 3-2: Stained objects examples
(p) Parasite, (W) White blood cell, (PLT) Platelet, (A) Artifacts

3.3 Image Acquisition Setup

The overall scheme of the image acquisition system is shown in Figure 3-3. It consists of a microscope, a digital camera and a computer. The camera is connected to the computer which enables remote capturing and visualization of the slide fields. The link is provided by a USB interface and Lica software drivers [19]. In this system, there are no motors on the microscope and no automatic control is used for the slide positioning table or focus. Hence, there is no control link between the microscope and the computer.

Figure 3-3: Image acquisition setup

The microscope used in this study is Leica DM LS2. It has an overall magnification of up to 1000x (with 100x objective and 10x eyepiece). It is equipped with Köhler illumination system. Köhler illumination is an important standard to setup microscopes to provide uniform illumination of light on the sample [20]. The importance of uniform illumination is explained in more detail in section 4.1.

The images were acquired as shown in Figure 3-3 by using setup found at the Ethiopian Public Health Institute (EPHI), inside their malaria and other parasite research laboratory. The samples obtained mostly had low number of plasmodium parasites in early stages (ring trophozoites) of their life cycle. Such features are often hard to detect. In addition some samples did not have any parasites (negative controls).

The samples were stained using a Giemsa protocol to highlight the parasite(s) and were initially examined by a laboratory technologist who had expertise in malaria diagnosis. Slide images were acquired using a Leica DFC320 Digital Camera. In total 91 images were acquired. 9 negative control cases were included in a blinded fashion to check the specificity of the method.

3.4 RBC Segmentation

Image segmentation is an important and, perhaps, the most difficult task in image processing. And it is based on three principal concepts: edge detection, threshold, and

region growing. In this work, we implement the concept of watershed transformation which embodies those three principles [21].

RBC segmentation is one of the most important steps in the development of automatic detection of malaria. It is a crucial step for the subsequent tasks of plasmodium detection, species identification and parasitemia (level of parasite infection) determination. RBC segmentation involves proper exclusion of artifacts which are found outside of the RBC. In humans, mature RBCs also called erythrocyte cells are flexible biconcave disks that lack a cell nucleus and most organelles [22] as shown in Figure 3-4.

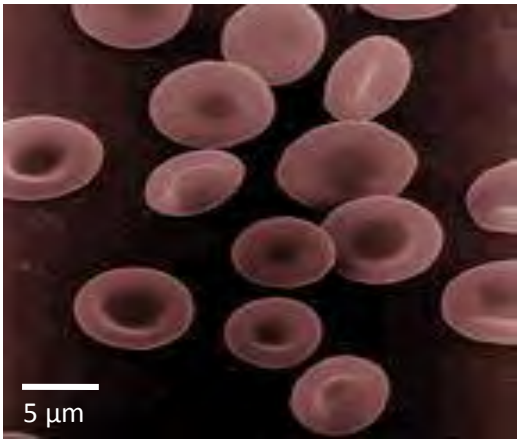


Figure 3-4: Bi-concave shape of human RBC [22]

This shape allows for more surface area to bind oxygen and carbon dioxide during gas exchange. The bi-concave disc shape also makes the cell more flexible so that it can get through the smaller blood vessels. This nature of RBC together with the artifacts makes RBC segmentation not easy.

3.4.1 Watershed Transform

In grayscale mathematical morphology, the watershed transform, originally proposed by Digabel and Lantuéjoul and later improved by Beucher and Lantuéjoul, is the method of choice for image segmentation [23].The watershed transform can be classified as a region-based segmentation approach. Because the regions in the image characterized by small variations in gray levels have small gradient values. Thus in practice, we often see

watershed segmentation applied to gradient of an image, rather than to the image itself. The aim of the watershed transform is to search for regions of high intensity gradients (watersheds) that divide neighbored local minima (basins). The intuitive idea underlying this method comes from geography [24].

Figure 3-5 shows illustration of immersion process of watershed transforms (C_1, & C_2 are catchment basins).

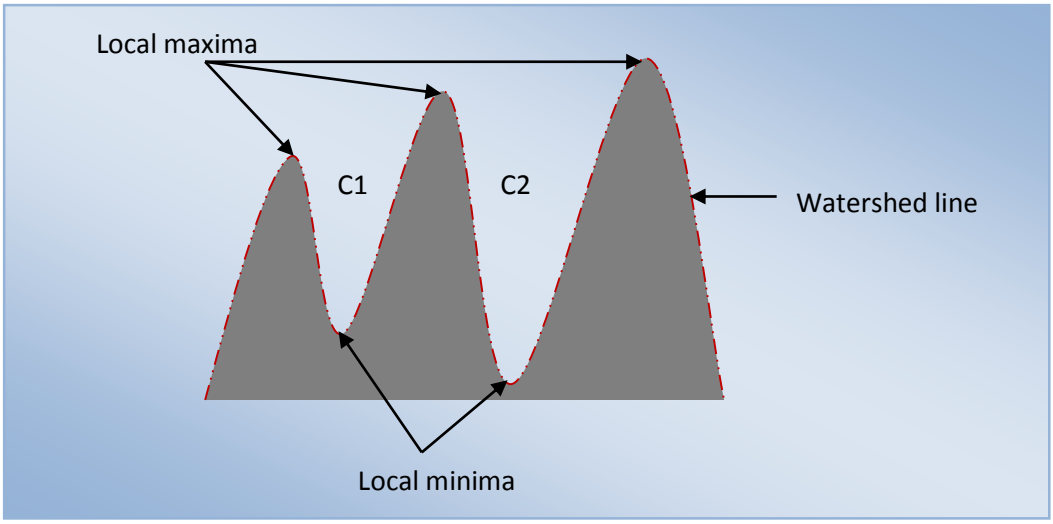


Figure 3-5: Illustration of watershed transform

The idea of watershed can be viewed as a landscape immersed in a lake; catchment basins will be filled up with water starting at each local minimum [25]. Dams must be built where the water coming from different catchment basins may be meeting in order to avoid the merging of catchment basins. The watershed lines are defined by the catchment basins divided by the dam at the highest level where the water can reach in the landscape. As a result, watershed lines can separate individual catchment basins in the landscape.

Figure 3-6(a) shows the original image. Figure 3-6 (b) shows RGB labeled binary image.

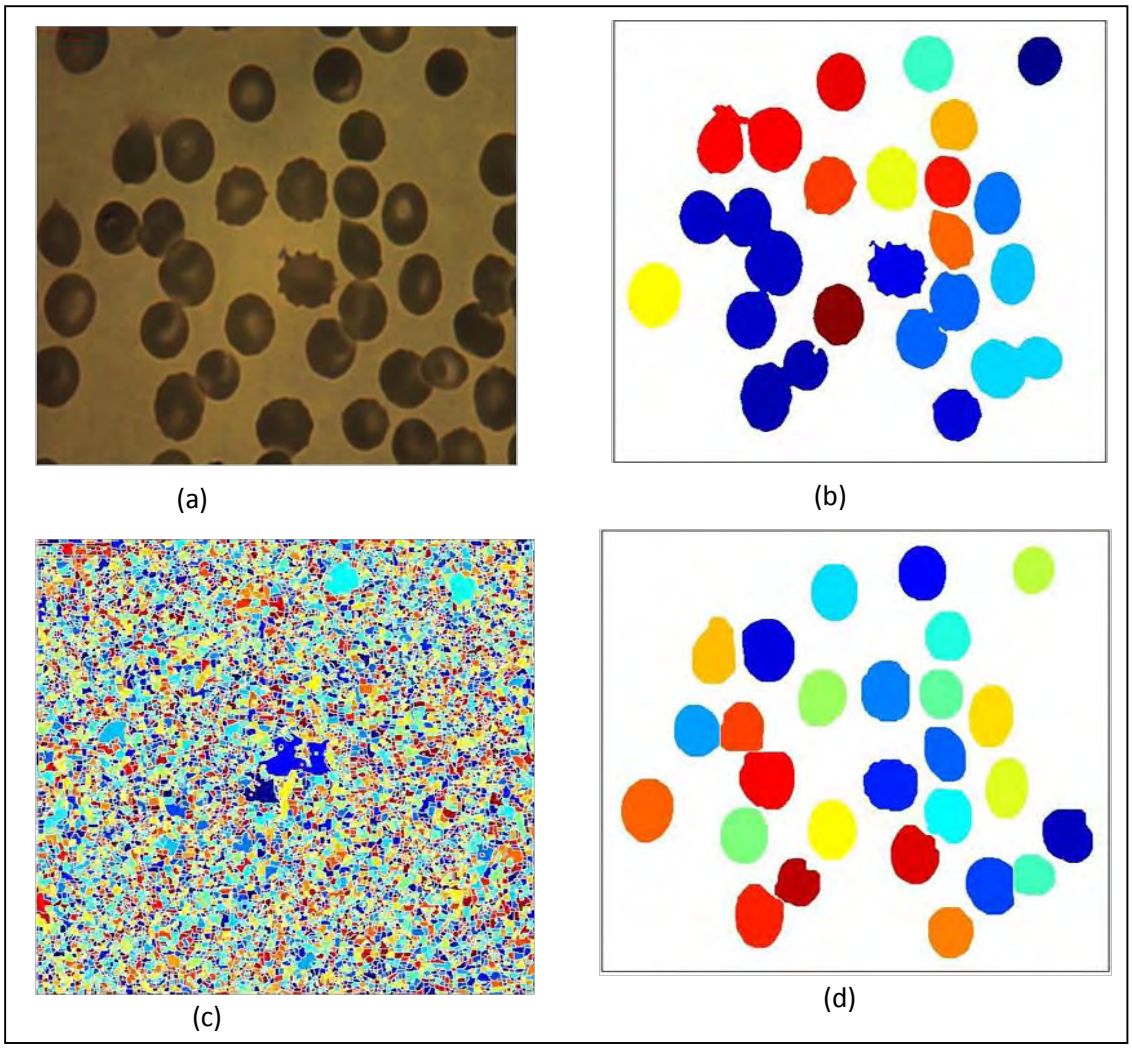


Figure 3-6: Comparison of watershed and marker controlled watershed segmentation

(a) Original image (b) RGB labeled binary image (c) Watershed transforms of gradient image (d) Proposed segmentation method

In this method we are able to count only 19 RBCs because touching objects are counted as single object that are indicated by the same color. Figure 3-6 (d) shows RGB labeled image following marker controlled watershed segmentation. In this method we are able to count 27 RBCs because the touching objects are disconnected and counted as independent objects that is indicated by different color. Figure 3-6 (c) shows direct application of watershed transform on gradient intensity image which leads to over segmentation.

This is due to local minima as a result of noise and biconcave shape of RBC (with a thin central disc nature). To overcome the problem of over segmentation and to separate touching RBCs in an image, we use marker controlled watershed transform where the markers are obtained from circular Hough transform and background markers from distance transform.

3.4.2 Markers Controlled Watershed Segmentation

Several approaches exist to remedy the over-segmentation problem, such as region merging algorithm, modified gradient algorithms, marker-controlled methods, multi-scale segmentation, and hierarchical segmentation. The main advantage of the marker controlled watershed method over other previously developed remedies in segmentation method is that it allows segmentation of particular objects and thus is ideal for counting applications [26].

Separating touching objects in an image is one of the most difficult image processing operations. The marker controlled watershed transform is often applied to resolve this problem. The watershed transform finds catchment basins and watershed ridge lines in an image by treating it as a surface where light pixels are high and dark pixels are low. Segmentation using the watershed transforms works well if one can identify or mark foreground objects and background locations [27]. By following this basic statement, we obtained the foreground markers from circular Hough transform and the background markers from distance transform.

3.4.2.1 Circular Hough Transform

The Hough Transform (HT) is a standard method for shape recognition in digital images. It was first applied to the recognition of straight lines and later extended to circles, ellipses and arbitrarily shaped objects [28].

As RBCs are geometrically similar to circles, we used circular HT to locate the center of each RBC. HT can be used to determine the parameters of a circle when a number of points

that fall on the perimeter are known. As shown in Figure 3-7 the center points are represented as red cells in the parameter space drawing.

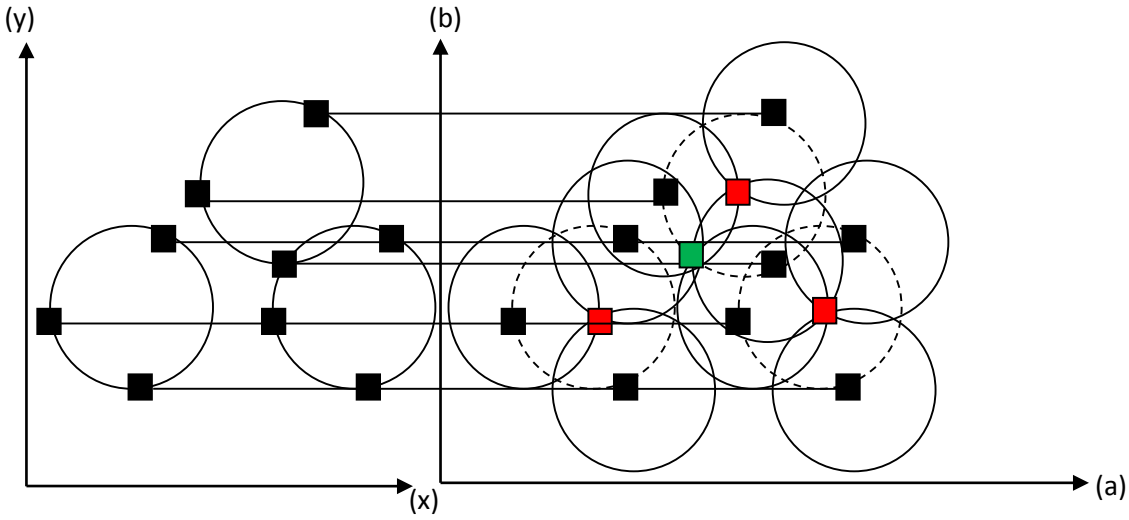


Figure 3-7: Geometric space (left) and parameter space (right)

There may still exist some unwanted overlap of circles which can cause spurious centers, shown by the green cell in the Figure3-7. These spurious circles can be removed by matching the parameter space with geometric space based on the "maximum score" (vote strategy) so it's pretty insensible to outliers i.e. only point with maximum score should be considered as the center of the circle.

$$x = a + r\cos(\theta) \quad 3.1$$

$$y = b + r\sin(\theta) \quad 3.2$$

The parameter space of Figure 3-7 shows that the locus of points (a, b) in the parameter space fall on a circle of radius r centered at (x, y). The true center point will be common to all parameter circles, and can be found with a Hough accumulation array.

3.4.2.2 Distance Transform

The distance transform of a binary image is the distance from every pixel of the object component which is black pixels to the nearest white pixel. In binary images there are only two gray levels 0 and 1 where 0 stands for black and 1 for white. Only one catchment basin

will appear in the topography of a binary image surface only when two black blobs are connected together.

The three most common ways of measuring distance are Euclidian distance, City-block distance and Chessboard distance.

Euclidean Distance: It is by far, the most commonly used measure of distance and is given by equation 3.3.

$$D_e = \sqrt{((i-k)^2 + (j-l)^2)} \qquad 3.3$$

Where the two points have spatial indices (i, j) and (k, l), respectively

City-Block Distance: City-block distance is an approximation to the Euclidean measure that is computationally faster. It is also called Manhattan distance or the absolute value metric. It is written as [29].

$$D_m = |i-k| + |j-l| \qquad 3.4$$

Chessboard Distance: The chessboard distance measure is the maximum separation in either the x or y direction between the two points. Also known as the maximum value metric, it is written as [29]

$$D_c = Max(|i-k|, |j-l|) \qquad 3.5$$

Chen *et al* [30] evaluated and compared the performances of watershed segmentation for binary images with different distance transforms including Euclidean, City-block and Chessboard and they found that Chessboard distance transform can achieve better watershed segmentation results than Euclidean distance transform and City-block distance transform.

In the proposed method, since no any significant difference were observed in each of the distance transforms, the common Euclidian distance transform were used to get the background markers.

3.5 Plasmodium Detection

The most important part of computerized diagnosis of malaria is the detection of plasmodium parasites. Plasmodium detection involves determining whether or not a person's blood is infected with malaria parasites. If the result is found positive a further classification is required to specify the species of malaria parasite, which is presented in section 3.6. For plasmodium detection, we used the Hue-Saturation-Intensity (HSI) color model.

The HSI color system separates color information of an image from its intensity information [31, 32]. The RGB color model is defined with respect to a unit circle. However the HSI space representation is through a double cone, as shown in Figure 3-8.

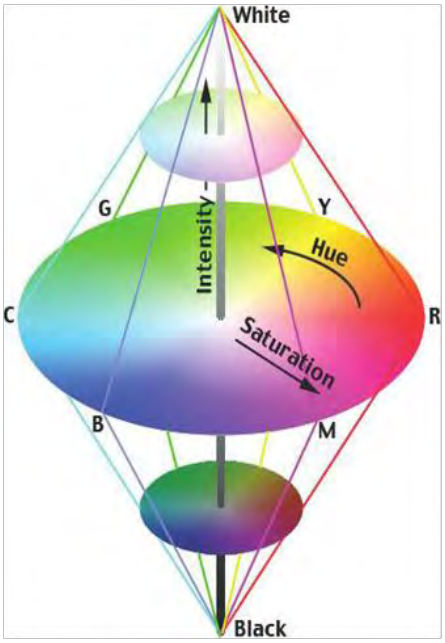


Figure 3-8: HSI color space representation [32]

The center of this double cone is a circumference divided into angles of equal magnitude. For this reason the value of **Hue (H)**, which describes the color by its wavelength, covers angles whose amplitude varies between 0 º and 360 º. The letters R, Y, G, C, B and M refer

to Red (0 º), Yellow (60 º), Green (120 º), Cyan (180 º), Blue (240 º) and Magenta (300 º) respectively.

The radial distance from intensity axis represents the **Saturation (S)** found in every color and takes values from 0 (at periphery) to 1(at the axis), indicating how the color is diluted with white light. When the saturation component is close to 0, colors only reflect a change between black and white. **Hue (H)** is a color attribute that describe a pure color; perhaps saturation gives a measure of the degree to which a pure color is diluted by white light [32]. Finally, the axis through the two cones corresponds to the **Intensity (I)** brightness component. This has a normalized value from 0 (black) to 1 (white) and indicates the amount of light in a color. Removing a small circumference of the figure formed by two cones, colors close to an intensity of 1 are lighter than those close to zero.

Mathematical HSI color space is calculated from RGB color space:

$$Hue = \begin{cases} \theta \; if \;\; B \le G \\ 360 - \theta \; if \;\; B \succ G \end{cases} \qquad 3.6$$

Where, $\theta = \cos^{-1}\left\{ \dfrac{[\frac{1}{2}(R-G)+(R-B)]}{[(R-G)^2+(R-G)(G-B)]^{1/2}} \right\}$

$$Saturation = 1 - (\frac{3}{R+G+B}) \min(R,G,B) \qquad 3.7$$

$$Intensity = \frac{1}{3}(R+G+B) \qquad 3.8$$

To segment objects with different colors, we may apply the segmentation algorithm to the hue component only. For example we may set thresholding on the range of the hue that separates different objects easily [31].

3.5.1 Thresholding of Grayscale Images

Thresholding on the hue components of the HSI color space is used to detect the chromatin dots of the malaria parasite. The chromatin dots are the DNA complex of malaria parasite structure seen in all developmental stages of the plasmodium [33]. Chromatin (part of the parasite nucleus) is usually round in shape and stains red [7].

Binary images can be obtained from grayscale images by thresholding operations. A thresholding operation chooses the pixels that make the object of interest as the foreground pixels and the rest as background pixels. Given the distribution of gray-tone in a given image, certain gray-tone value can be chosen as threshold value that separates the pixels into groups. In a simple case, a single threshold value T is chosen. All the pixels whose gray tone value is greater than or equal to T become foreground pixels and all the rest become background pixels. This threshold operation is called threshold above (equation 3.9). There are many variants including threshold below, which makes the pixels with value less than T the foreground (equation 3.10); threshold inside, which is given a lower threshold and upper threshold and selects pixels whose value are between the two as foreground (equation 3.11); and threshold outside, which is the opposite of threshold inside.

$$F_{T[i,j]} = \begin{cases} 1 \; if \;\; F[i,j] \geq T \\ 0 \; otherwise \end{cases} \qquad 3.9$$

$$F_{T[i,j]} = \begin{cases} 1 \; if \;\; F[i,j] < T \\ 0 \; otherwise \end{cases} \qquad 3.10$$

$$F_{T[i,j]} = \begin{cases} 1 \; if \;\; T_1 \leq F[i,j] \leq T_2 \\ 0 \; otherwise \end{cases} \qquad 3.11$$

Where, T, T1 and T2 are the threshold points.

The main question associated with these simple forms of thresholding is how to choose the threshold value T.

Although there exist different thresholding schemes in the literature (including the adaptive thresholding scheme), simple histogram based approach is used in this thesis as it was found to adequately do the intended purpose.

3.6 Species Differentiation

Once the plasmodium parasite is detected, the next task that the clinician requires to decide is to know what type of species it is; this is because the treatment of the malaria disease varies with the type of species identified. In the manual microscopic malaria diagnosis method this is done by looking at the size of infected RBCs. RBCs infected with *PV* are larger in size than that of *PF* infected RBCs. Hence in our computerized diagnosis of malaria we follow a similar approach to differentiate the species.

This is done through measuring the area occupied by single RBC. Normal and *PF* infected RBCs reside in the same range of area. The area occupied by *PV* infected RBCs are larger than the maximum limit of boundaries for *PF* infected and normal RBCs population. The maximum limit of boundaries for normal and *PF* infected RBCs are set by taking the uppermost area of the population.

The natural size range of normal RBCs is always identical but the image size may vary depending on the magnification of the microscope and the pixel numbers of the camera used. Due to this reason the threshold value in this thesis work may vary in case different microscope or camera with changed magnifications are used for image acquisition. Image resizing can help to solve this problem by creating image with similar pixels number even if different magnifying microscopes or cameras are used.

Chapter 4 – Development of the Automated System

The block diagram of the algorithm developed for the automatic detection of malaria is shown in Figure 4-1. It includes image acquisition, background subtraction image pre–processing, RBC segmentation, plasmodium detection and species differentiation. The detailed block diagram is given in Figure 4-11.

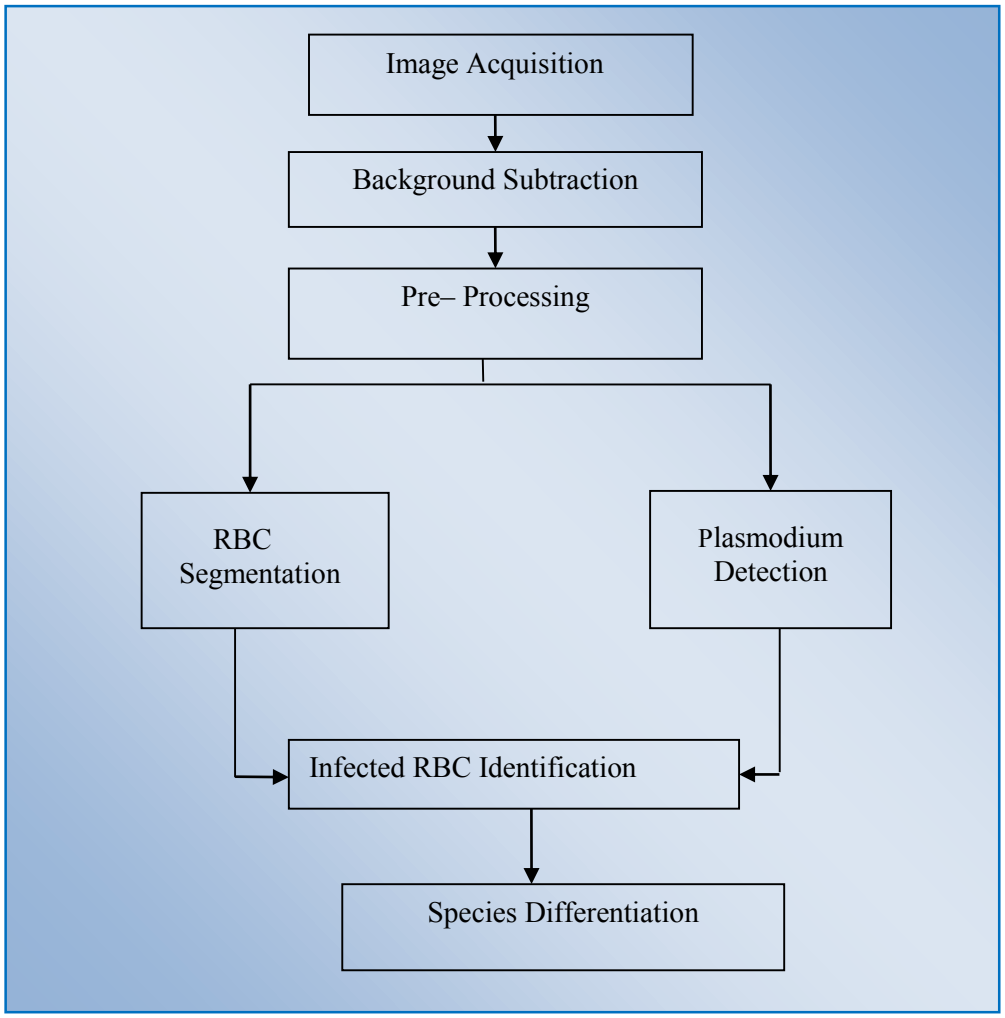


Figure 4-1: Block diagram of proposed algorithm

4.1 Image Acquisition

For developing an automatic malaria diagnosis system using digital image analysis, acquiring a good quality image is the first and crucial step. Among other things, a non-uniform illumination of the samples affects the quality of acquired image. Hence for this technique, selecting a digital microscope that has a uniform illumination system is important to ensure the quality of the image.

A common uniform illumination standard in a microscope setup is Köhler illumination system named after its inventor August Köhler [34]. In illumination (shown in Figure 4-2) the light source is carefully aligned to form collimated light that will uniformly illuminate the specimen. This type of illumination requires the use of an adjustable conjugate stops (the aperture stop and the field stop). Aperture stop is conjugated with light source, whereas the field stop is conjugate with the specimen. Conjugating the field stop with the specimen ensures to have uniform illumination on the specimen. This is often neglected by microscopists since the human vision system is adaptive to local illumination changes. For automatic malaria diagnosis the non-uniform illumination will otherwise produce serious problems at subsequent stages of the process.

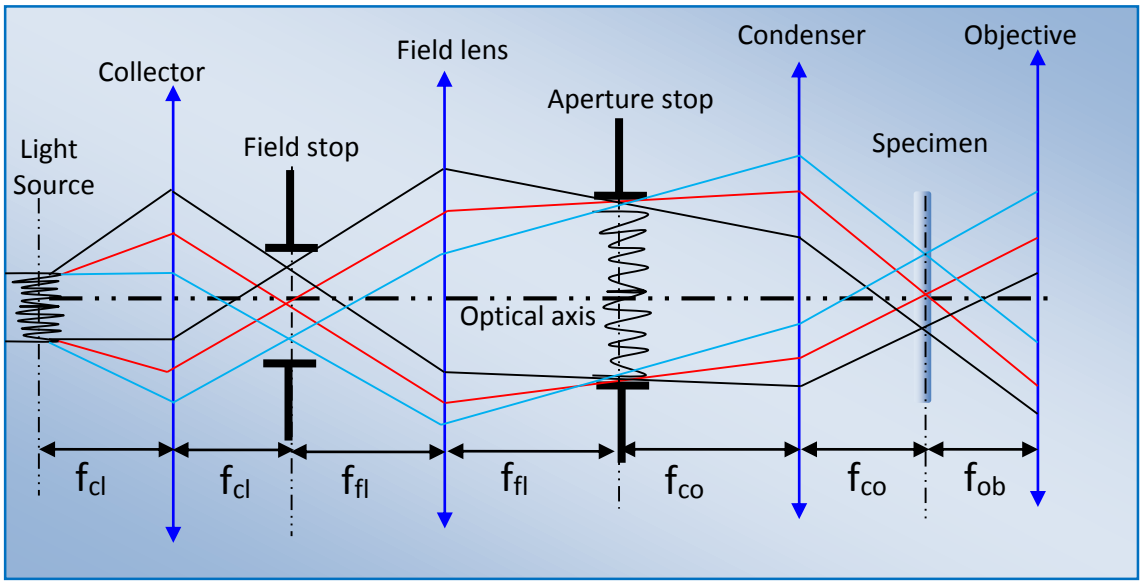


Figure 4-2: Schematic diagram of Köhler illumination

4.2 Background Illumination Correction

Even if a uniform illumination has been employed in our setup, background illumination correction is still important to get rid-off background illumination coming from room light. Background illumination correction can also be applied while acquiring images (a priori correction) [35]. A priori correction can be done using additional images captured at the time of image acquisition. For this purpose we have acquired two images in addition to the specimen image: bright field image with no sample and dark field image with no light.

$$Corrected\ image = Scaling\ \ factor \times \left(\frac{Specimen\ image - Dark\ \ field\ \ image}{Bright\ \ field\ \ image - Dark\ \ field\ \ image} \right) \qquad 4.1$$

Scaling factor is the maximum possible value of grayscale image. Figure 4-3 illustrates the importance of background illumination corrections.

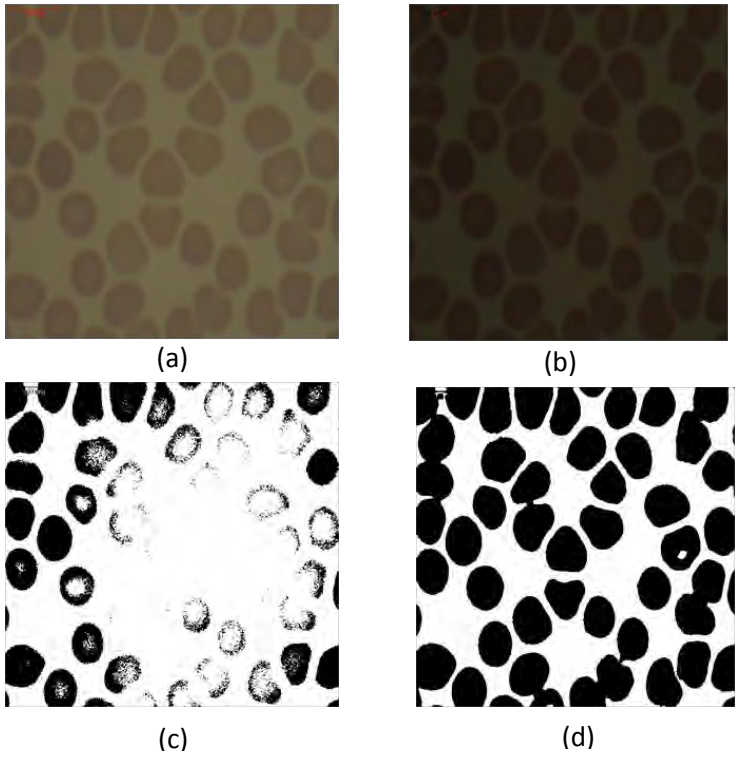


Figure 4-3: Image showing the significance of background correction

(a) Original image, (b) Background corrected RGB image, (c) Binary image before correction, (d) Binary image after correction

Image (a) shows the original RGB image with no background illumination corrections. Image (b) is after the application of background illumination correction which is slightly darker than the original image. Image(c) and (d) are the binary images after binary transform of the original image (a) and illumination corrected image (b) respectively. The negative impact of non-uniform illumination (coming from room light) on image processing is clearly observed in (c) when we compare it with the corrected image (d). This tells us that doing background illumination correction is very important task for image segmentations.

4.3 Pre-processing

Having acquired JPG format image using image acquisition set up as stated in section 3.4, the next task after background illumination correction is image pre-processing. The purpose of image pre-processing is to remove unwanted objects and noise from the image so that it becomes ready for the subsequent image segmentation process.

The image pre-processing therefore enhances the quality of the input image. To decrease the computational time, the original size of the images (2088 by 1550 pixels) is reduced to 500 by 500 pixels. A median filter was used to reduce the salt and pepper noises, and the average filtering was employed to minimize random noises. The optimal image size was carefully selected by testing different sizes not to compromise the performance of plasmodium detection of the automatic system. Image resizing is also important to have uniform image sizes or number of pixels of different samples which allows use of different microscopes or CCD cameras with varied magnifications for image acquisition.

4.4 Algorithms for RBC Segmentation

The detailed algorithm for RBC segmentation is shown in Figure 4-5. Following pre–processing, the image is converted to binary format where holes due to biconcave shape of RBC are filled. Due to the biconcave shape of the RBC, the central pallor is assigned the same features as the background (Figure 4-5(b). In the hole filling process we need to remove the holes in the binary image to ensure accurate RBC count (Figure 4-5(c)).

The circular Hough transform is used not only to get the markers for watershed transform but also used to distinguish RBC from other blood cells (Figure 4- 6(e)).

The minimum and maximum RBC radius values for effective RBC segmentation were found to be 14 and 40 pixels respectively.

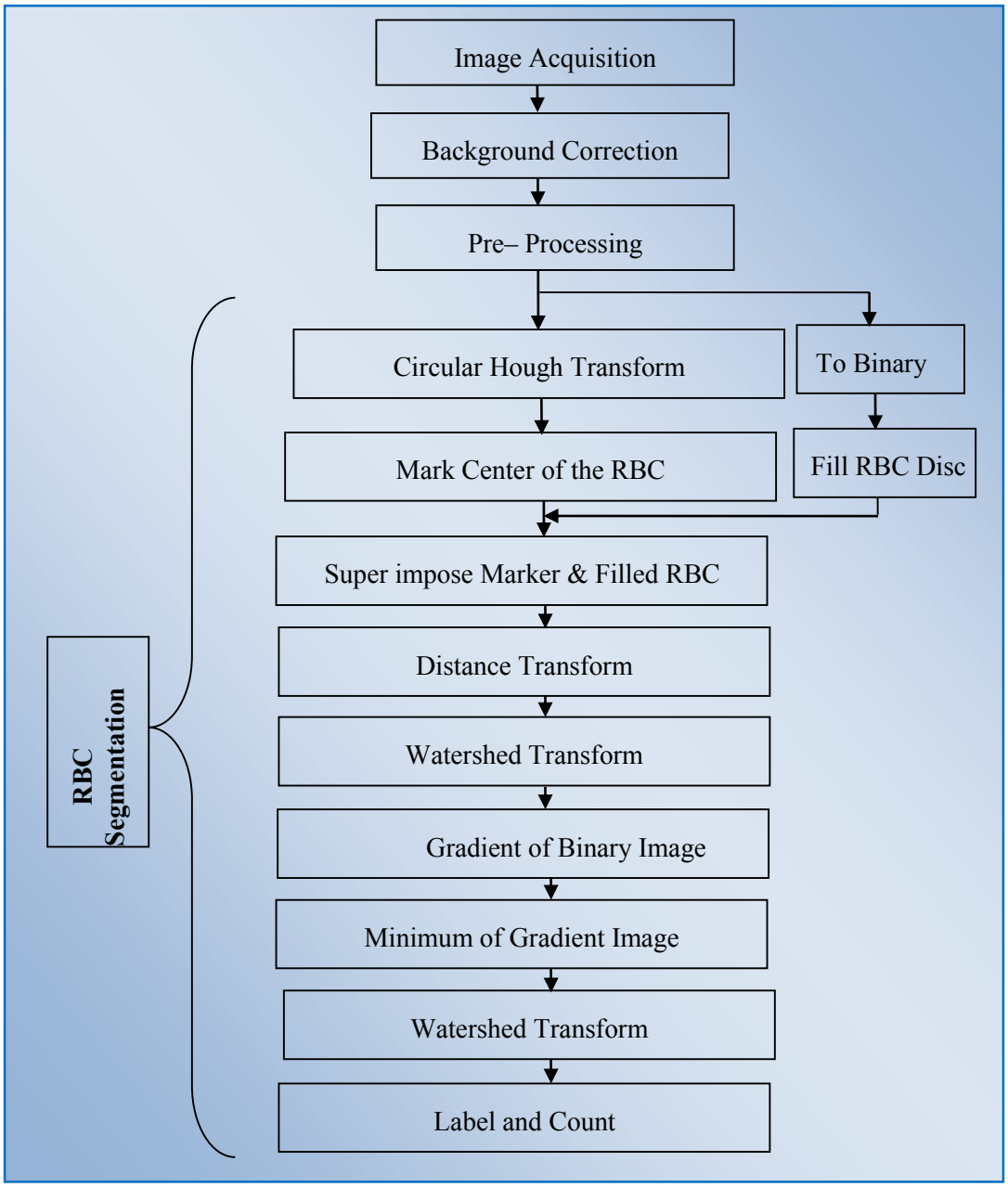
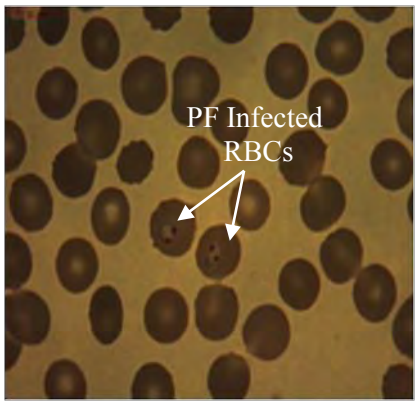
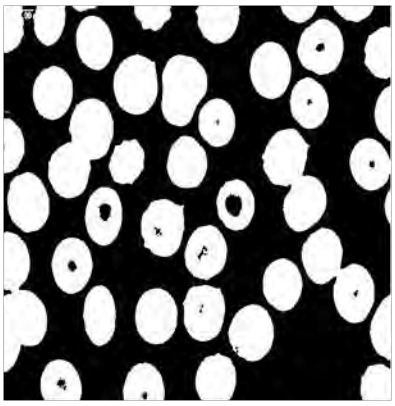


Figure 4-4: Algorithm flow chart for RBC segmentation

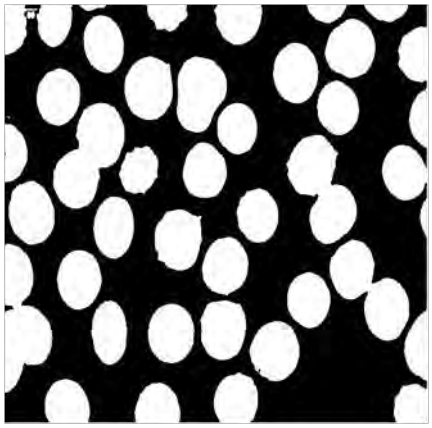
To merge the markers and hole filled binary image, the markers (single pixel at the center of RBC) are superimposed over the binary image. Then by applying watershed transform on distance transform of superimposed images we obtain ridge line of the segmentation (external markers) (Figure 4-5(g)). Next, from calculated gradient magnitude of the binary image, the local minima are suppressed by modifying the intensity of gradient image using morphological reconstruction. Consequently, the image has only regional minima wherever the markers are located. Then by applying watershed transform on modified gradient image we obtain well segmented RBC image (Figure 4-5(i)). Objects connected to the image border are also cleaned because of incomplete information. Finally the segmented images are labeled to count the number of RBCs.



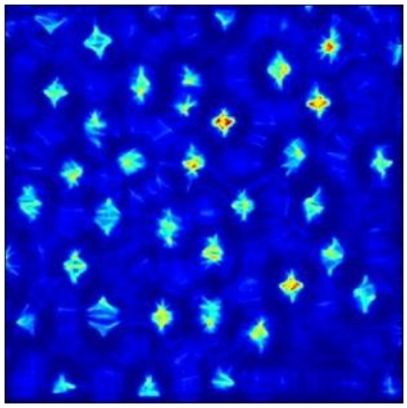
(a)



(b)



(c)



(d)

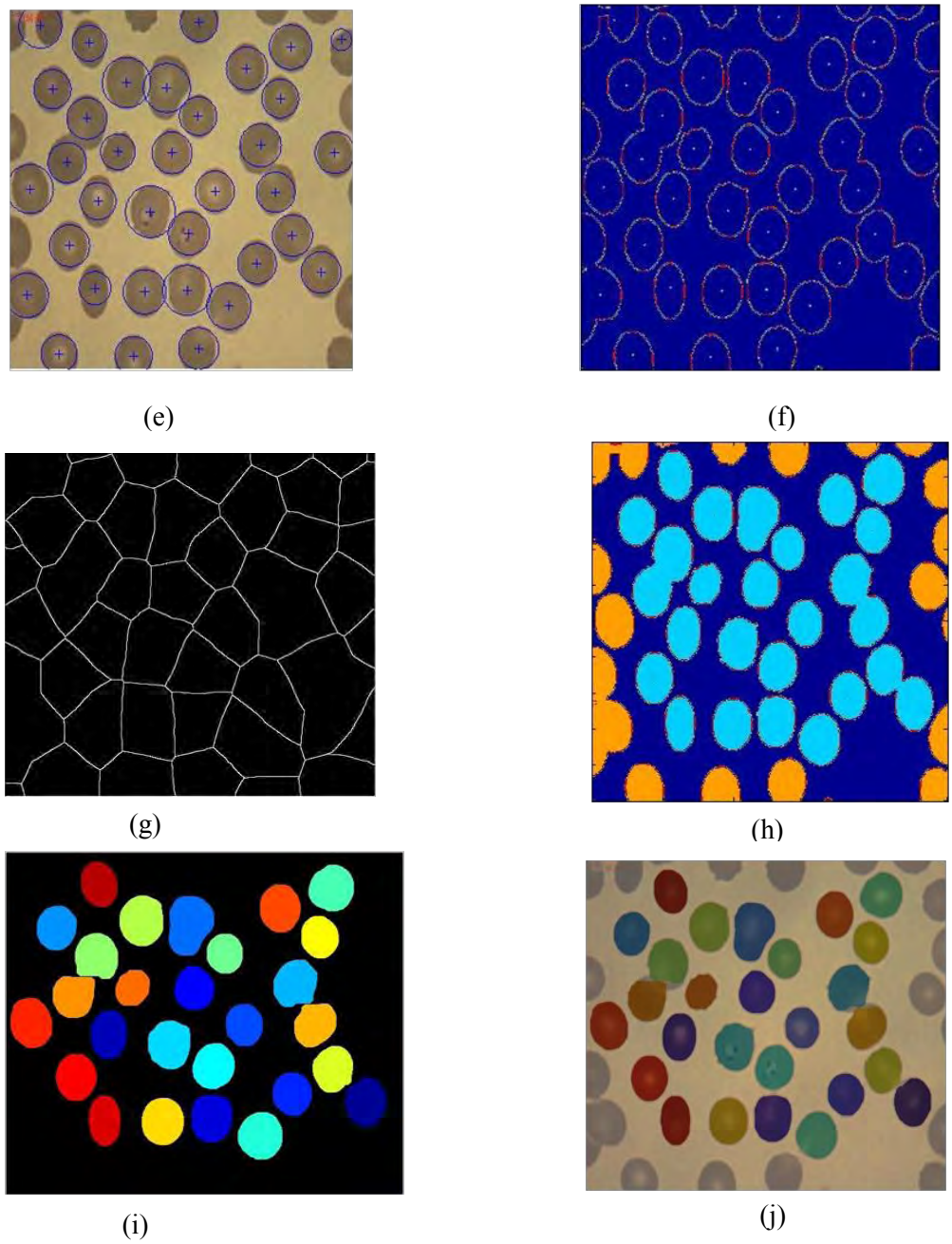


Figure 4-5: Images showing RBC segmentation

(a) Original RGB image, (b) Binary image depict holes, (c) Binary image with hole filled, (d) accumulator array for circular Hough transform, (e) Detected RBC with circular Hough transform, (f) Gradient image superimposed on foreground markers, (g) Watershed ridge line by applying watershed on distance transforms, (h) image showing clearing of the border, (i) Marker controlled Watershed Segmented &RGB labeled RBC, (j) Segmented RBC superimposed on original image

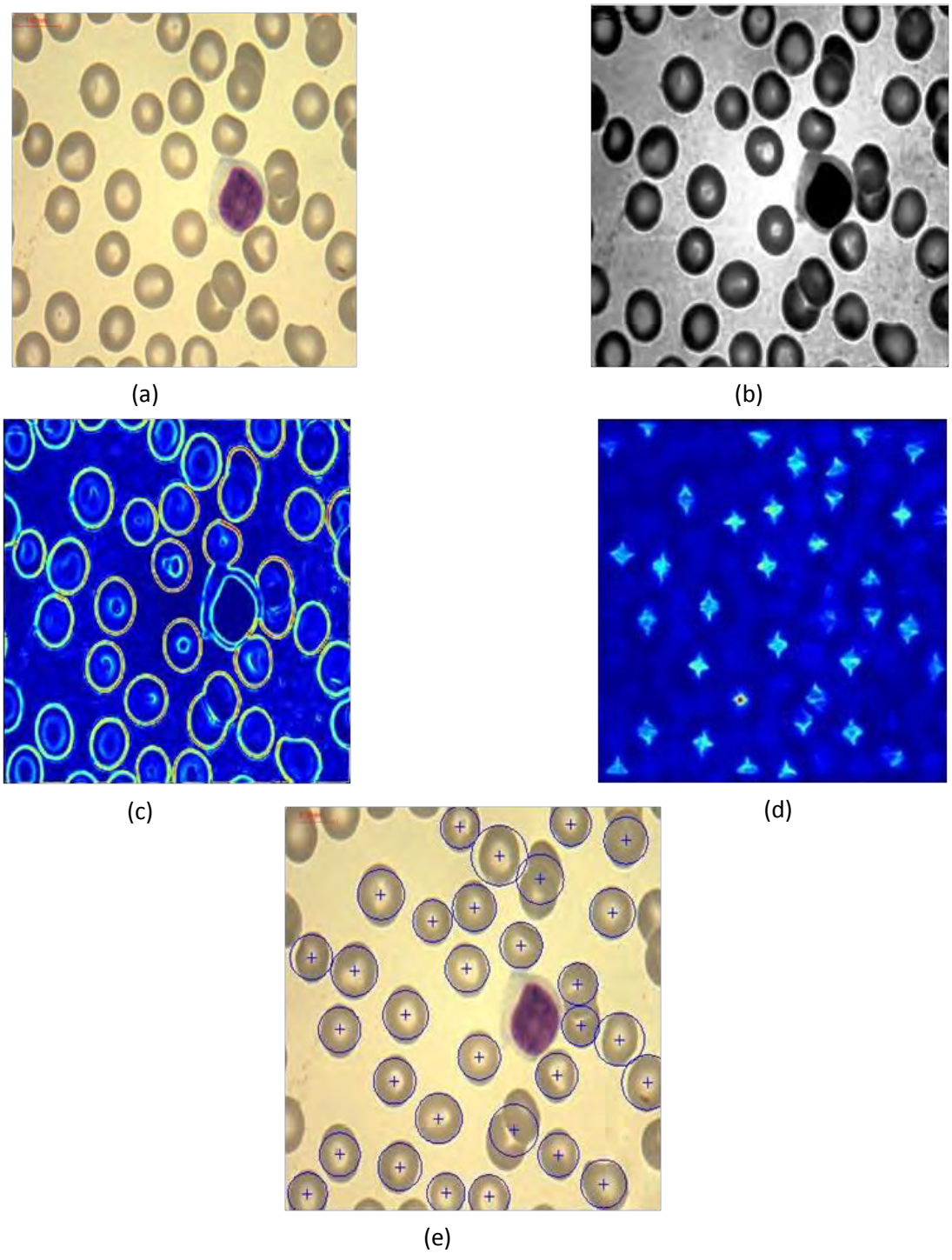


Figure 4-6: Circular Hough transforms:

(a) Original Image (b) green channel of RGB (c) Gradient transform (d) Accumulator from CHT (e) Original image with circle detected

4.5 Algorithm for Plasmodium Detection

As shown if Figure 4-7, RBC segmentation and plasmodium detection are done in parallel. Hence, as shown in Figure 4-4 background correction and image pre-processing are similar with the RBC segmentation.

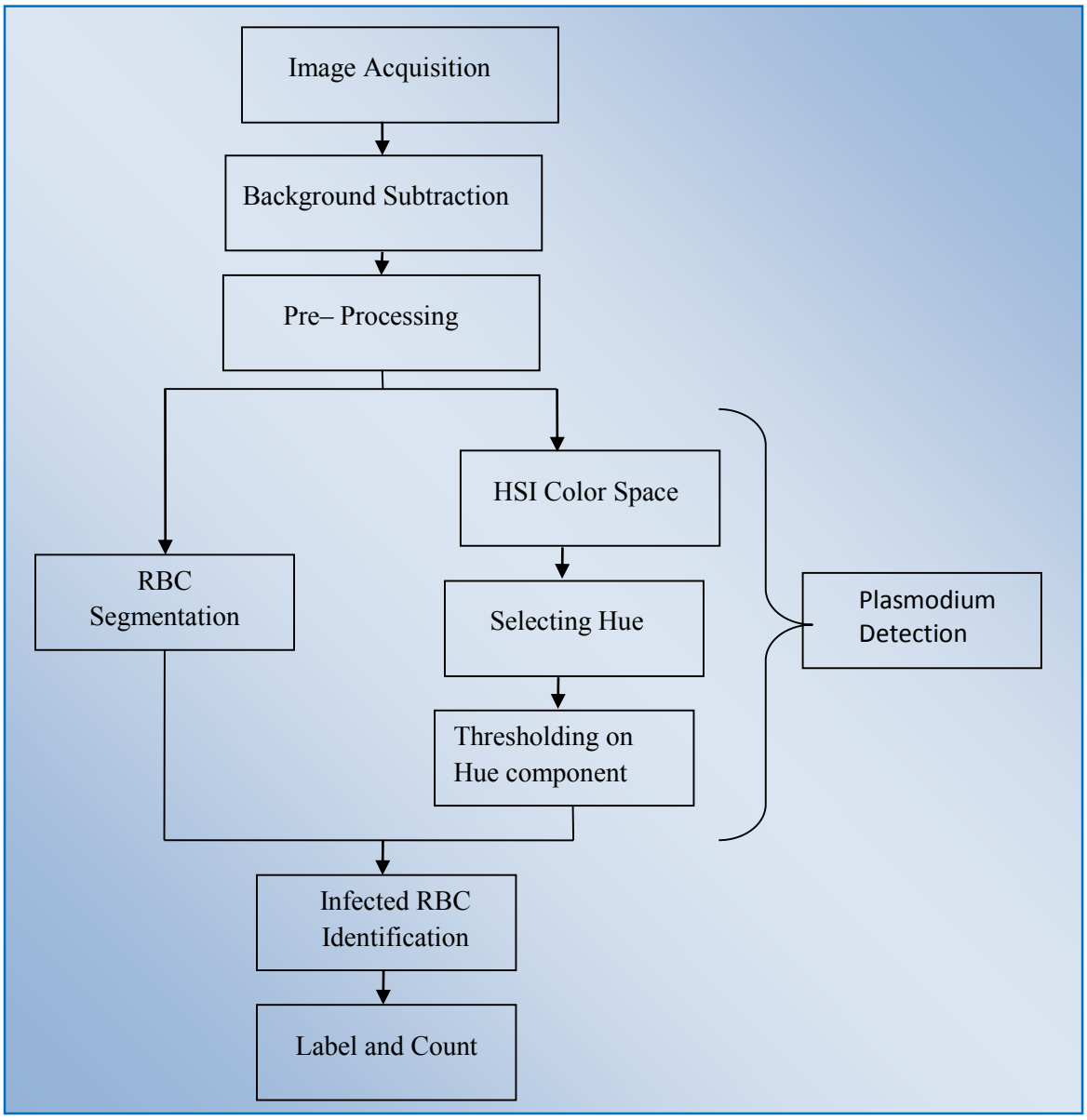


Figure 4-7: Algorithm flow chart for plasmodium detection

For infected RBC segmentation, two independent and parallel processes are required

- RBC segmentation by using marker controlled watershed transform (section 3.4)
- Plasmodium detection using thresholding on Hue component of HSI color space (section 3.5)

These two independent processes are then recombined using binary morphological reconstruction (section 2.2.1). The optimal thresholding point on the Hue component of HSI color space was chosen to detect the chromatin dots of the parasite.

For further clarity, the above algorithm is illustrated below in Figure 4-8 with processed image at each step. Figure 4-8(a) is the original image infected with P.vivax (*PV)*. After background correction and image pre-processing, this image is converted to HSI color model (Figure 4-8(b)). As observed from the image the chromatin dots of the parasite is clearly seen as red dots in this model than in its counter RGB color model. In HSI space Hue (H) component contains the color information of the image and shows the chromatin dots as a white spot (Figure 4-8(c)). Applying thresholding on this gray scale image; only the white mark left as 1 while all other pixel turn to 0 (Figure 4-8(d)).This way, plasmodium is detected.

Once the plasmodium is detected, we applied binary morphological reconstruction (section 2.2.1) to relate segmented RBCs with detected plasmodium spots for identification of infected RBCs. This way, we can select only plasmodium infected RBCs from the all segmented RBCs population (Figure 4-8(e)). Accurate segmentations of infected RBCs are important for the subsequent species detection which is depending on the number of pixels that a single infected RBC occupies. To check if infected RBCs are correctly segmented superimposition on the original image is done (Figure 4-8(f)).

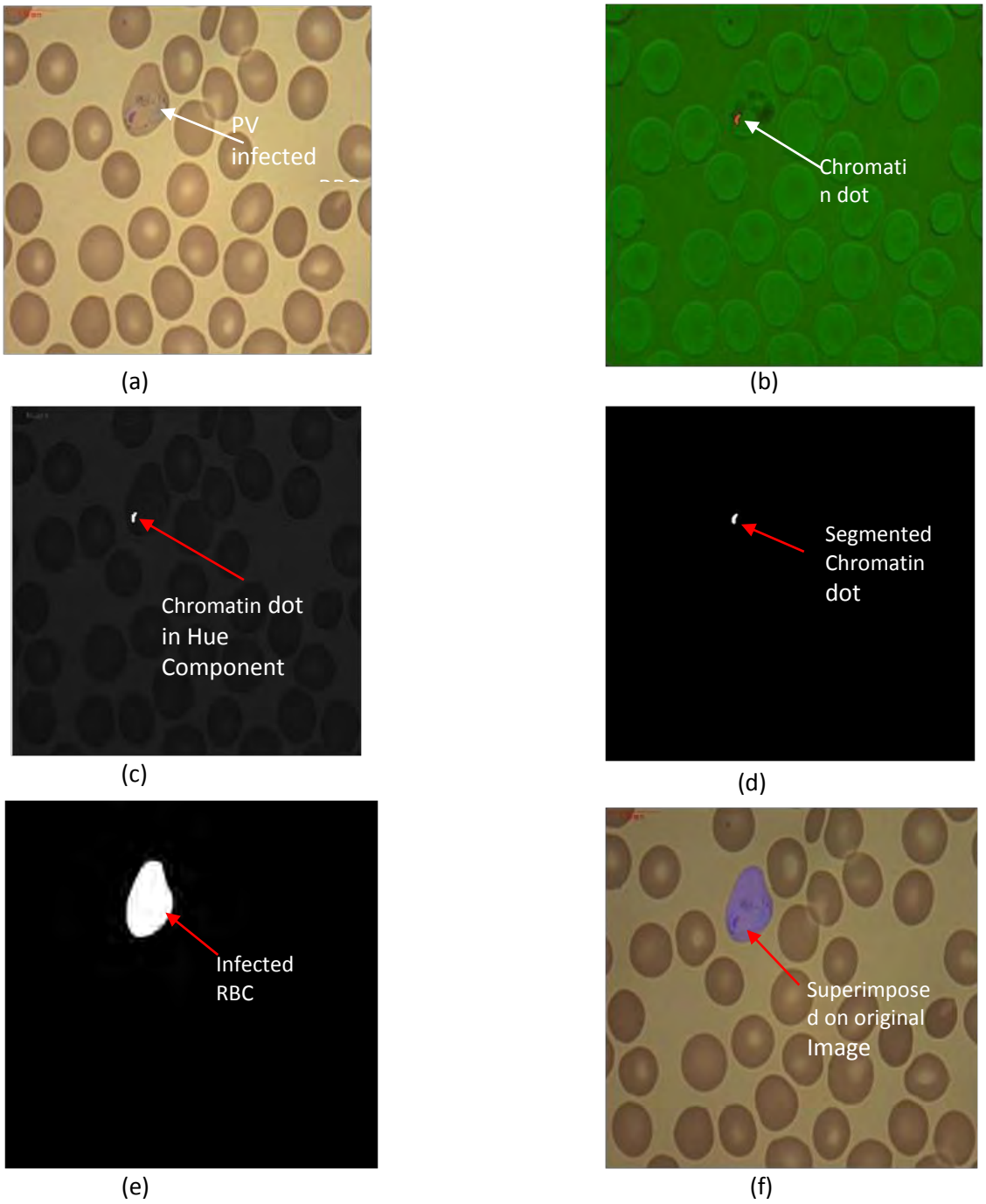


Figure 4-8: Images showing Plasmodium segmentation

(a) Original image, b) HSI colour transform with chromatin dots detected, (c) Hue components of HSI colour model with detected chromatin dot, (d) detected chromatin dot by thresholding, (e) Infected RBC, (f) Infected RBC superimposed on original image

4.6 Algorithm for Species Differentiation

Specie differentiation is the third important task in malaria diagnosis because the medical treatment can differ from one species to the other. As stated in section 1.3 table 1, the two common malaria species found in Ethiopia, *PF* and *PV* are easily distinguished by the area of infected host cells (RBCs). *PV* infected RBCs cover greater area in the digital image than *PF* infected RBCs. The Algorithm flow chart for differentiation of *PF* from *PV* is shown in Figure 4-9.

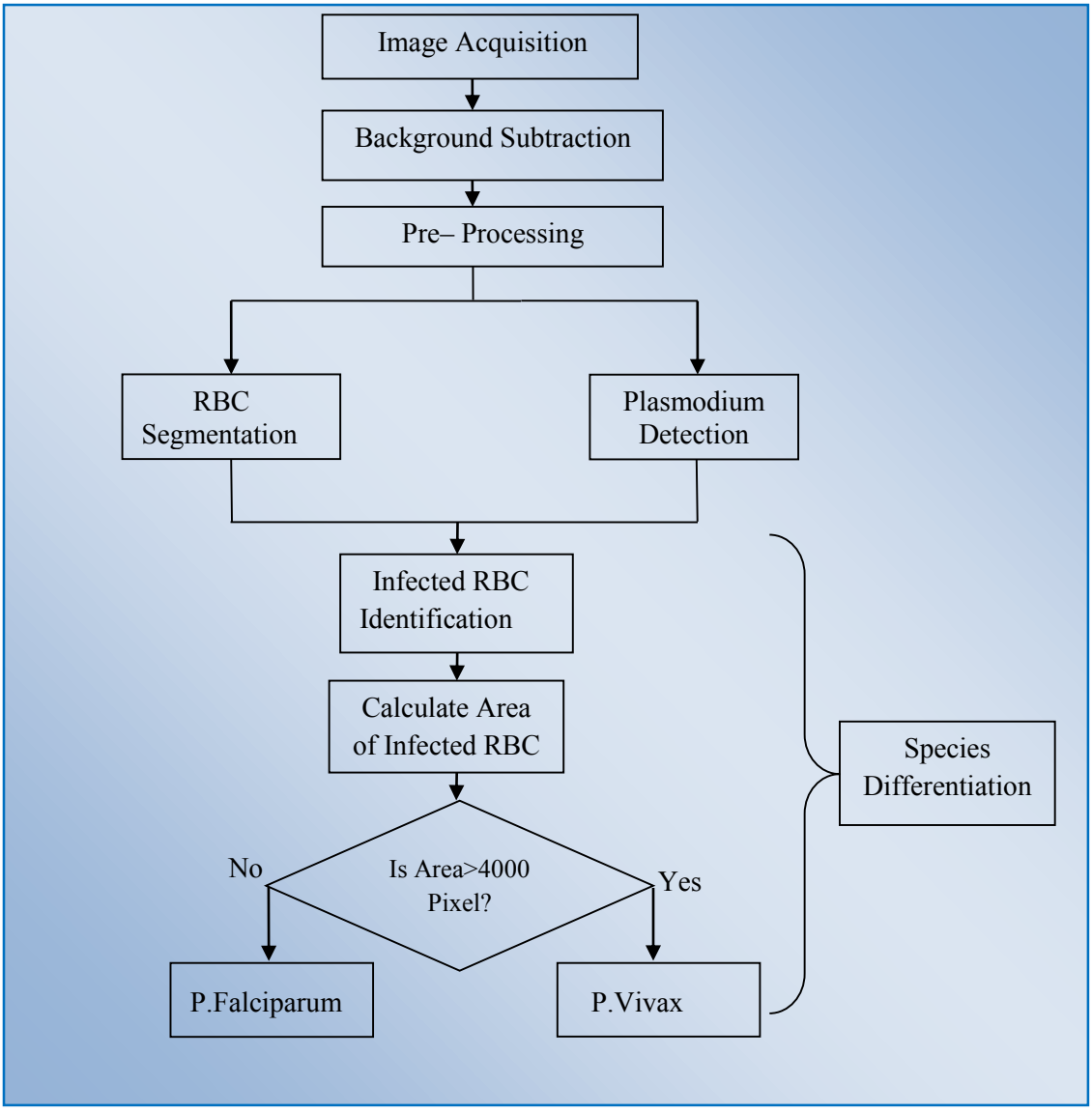


Figure 4-9: Algorithm flow Chart for P. falciparum and P. vivax differentiation

Figure 4-10(a) and Figure 4-10(b) demonstrate area histogram of segmented RBC population of images presented in Figure 4-5(a) and Figure 4-8(a) respectively. From both histograms, we observe that most of the normal RBCs occupy 2500 to 3500 pixels. In Figure 4-10(a), we observe the maximum area of 3917 pixels due to overlap of the two RBCS (Figure 4-5(a)). The two RBCs which are infected with *PF* cover 2809 and 3108 pixels, these values are lower than the threshold value of 4000 pixels. In Figure 4-10(b) the maximum value is 4685 pixels indicating *PV* infection. Here, the comparison is only made between infected RBC where we observe a significant difference.

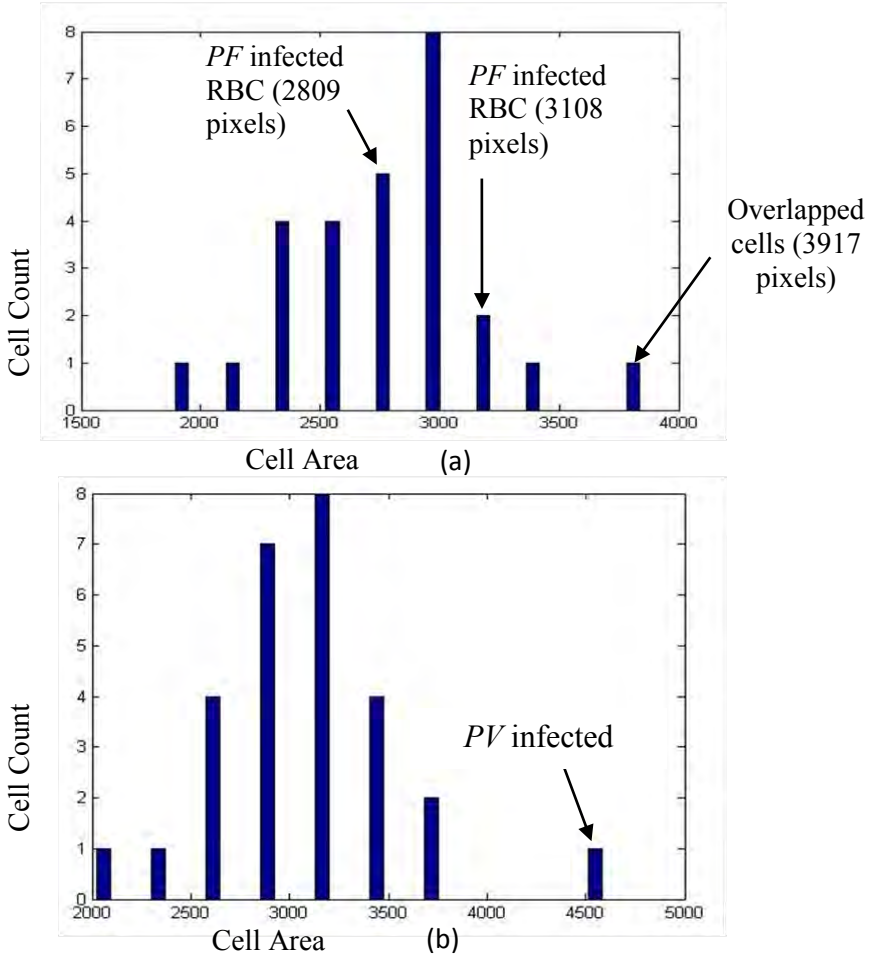


Figure 4-10: Area histogram of plasmodium infected RBC
(a) PF infected RBC; (b) PV infected RBC

As a summary, the overall detailed algorithm developed and implemented for automatic malaria detection is presented in Figure 4-11.

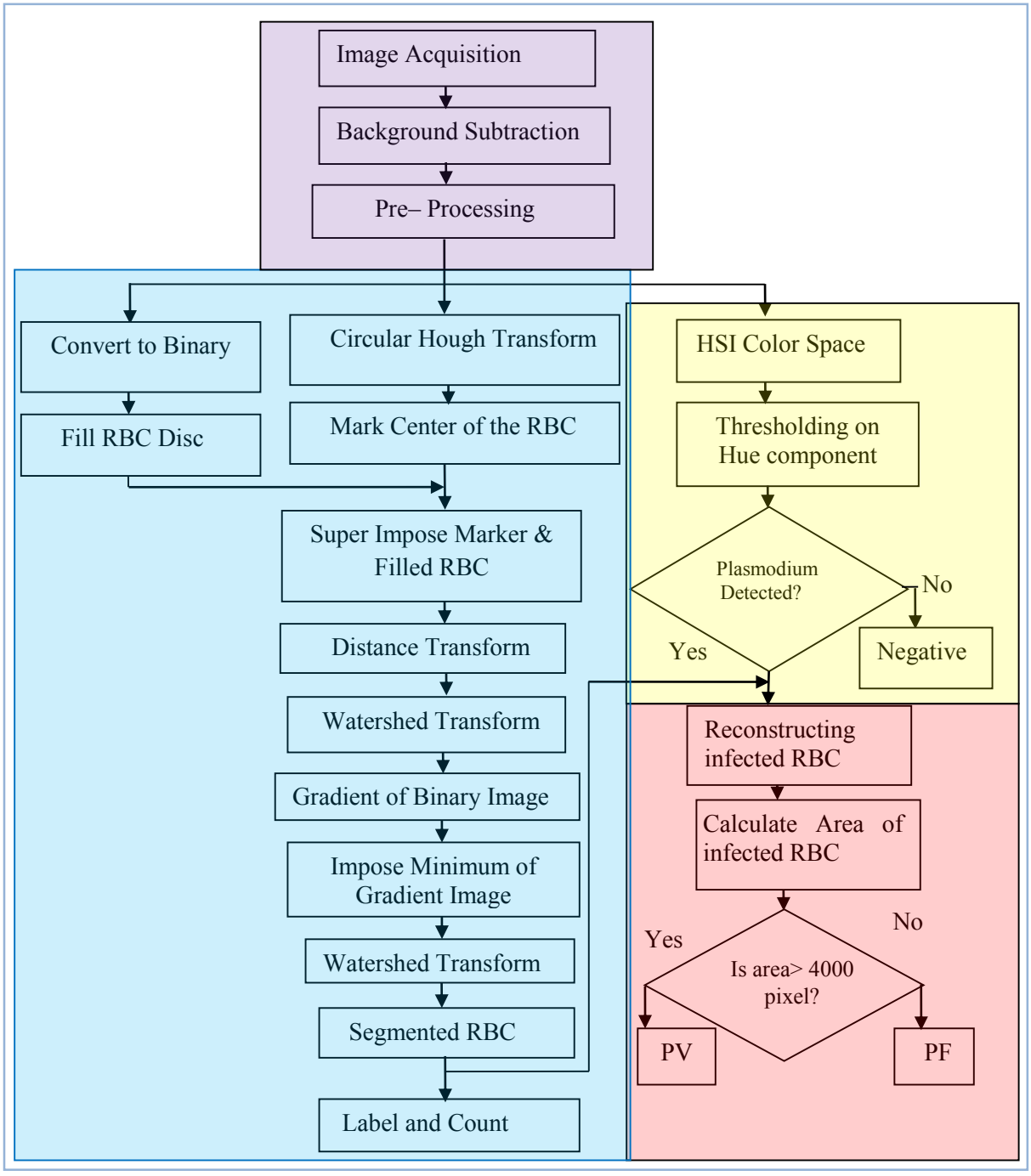


Figure 4-11: Proposed algorithm for RBC segmentation, plasmodium detection and species differentiation

RBC segmentation Plasmodium detection Species differentiation

Chapter 5- Results and Discussion

The proposed method, after being developed, was tested on variety of images with and without malaria parasites to check for the sensitivity and specificity of the algorithm. A total of 91 digital images of thin blood films were acquired using acquisition setup found at the Ethiopian Public Health Institute (EPHI), Parasitology Laboratory. Malaria samples were collected from Adama Malaria Center. Nine thin blood films from different patients were prepared and stained using Giemsa protocol. From these 9 patients 4 were infected with *PF* and 4 were infected with *PV* while the other 1 was not infected by malaria parasite. All thin blood films had already been examined and verified manually by malaria experts who had given a species-specific diagnosis.

From the total of 91 images, 82 of them are infected with plasmodium and the remaining 9 are non-infected. From 82 infected images, 41 are infected with *PF* and the remaining 41 are infected with *PV*.

The following tables (table 2, table 3 and table 4) show the comparison of manual and proposed automatic method for RBC count, parasite count, species differentiation and level of parasite (parasitemia) in each image where only cells within the image boundaries (field of view) are considered. Table 2 shows the images which are free of plasmodium infection. In table 3, we have presented images infected with *PF,* whereas table 4 shows comparison of the images infected with *PV*.

Image S. N	RBC Count			Parasite Count			Species Differentiation			Parasitemia (%)
	Software	Manual	Difference	Software	Manual	Difference	Software	Manual	Difference	
1	24	24	0	0	0	0	–	–	–	–
2	24	25	1	0	0	0	–	–	–	–
3	30	30	0	0	0	0	–	–	–	–
4	18	19	1	0	0	0	–	–	–	–
5	23	24	1	0	0	0	–	–	–	–
6	31	31	0	0	0	0	–	–	–	–
7	24	25	1	0	0	0	–	–	–	–
8	27	27	0	0	0	0	–	–	–	–
9	27	27	0	0	0	0	–	–	–	–
Total	228	232	4	0	0	0	–	–	–	–

Table 2 : Comparison of the automated and manual methods for non-infected images

Image S.N	RBC Count			Parasite Count			Species Detection			Parasitemia (%)
	Software	Manual	Difference	Software	Manual	Difference	Software	Manual	Difference	
1	31	32	1	1	1	0	PF	PF	0	3.22
2	26	27	1	2	2	0	PF	PF	0	7.69
3	25	27	2	1	1	0	PF	PF	0	4.00
4	27	27	0	2	1	1	PF	PF	0	7.40
5	25	27	2	1	1	0	PF	PF	0	4.00
6	16	17	1	1	1	0	PF	PF	0	6.25
7	29	32	3	1	1	0	PF	PF	0	3.44
8	26	28	2	2	2	0	PF	PF	0	7.69
9	25	28	3	1	1	0	PF	PF	0	4.00
10	30	34	4	2	2	0	PF	PF	0	6.66
11	26	26	0	2	1	1	PF	PF	0	7.69
12	18	19	1	1	1	0	PF	PF	0	5.55
13	21	22	1	1	1	0	PF	PF	0	4.76
14	26	26	0	1	1	0	PF	PF	0	3.84
15	28	29	1	1	1	0	PV	PF	1	3.57
16	29	30	1	2	1	1	PF	PF	0	6.89
17	22	23	1	1	2	-1	PF	PF	0	4.54
18	29	30	1	2	2	0	PF	PF	0	6.89
19	29	29	0	2	1	1	PF	PF	0	6.89
20	26	26	0	1	1	0	PF	PF	0	3.84
21	25	25	0	1	1	0	PF	PF	0	4.00
22	30	31	1	1	1	0	PF	PF	0	3.33
23	28	29	1	1	1	0	PF	PF	0	3.57
24	22	23	1	2	1	1	PF	PF	0	9.00
25	29	31	2	1	1	0	PF	PF	0	3.44
26	30	30	0	2	1	1	PF	PF	0	6.66
27	32	34	2	2	1	1	PF	PF	0	6.25
28	26	27	1	2	2	0	PF	PF	0	7.77
29	25	25	0	2	1	1	PF	PF	0	8.00
30	33	34	1	1	1	0	PF	PF	0	3.00
31	27	29	2	1	1	0	PF	PF	0	3.70
32	27	27	0	1	1	0	PF	PF	0	3.70
33	30	30	0	1	1	0	PF	PF	0	3.33
34	24	24	0	1	1	0	PF	PF	0	4.16
35	21	22	1	1	1	0	PF	PF	0	4.76
36	25	27	2	1	1	0	PF	PF	0	4.00
37	27	27	0	2	2	0	PF	PF	0	7.4
38	22	22	0	2	1	1	PF	PF	0	9.09
39	25	26	1	1	1	0	PF	PF	0	4.00
40	25	27	2	1	1	0	PF	PF	0	4.00
41	26	27	1	1	1	0	PF	PF	0	3.80
Total	1073	1116	43	56	48	8	–	–	1	5.26(Average)

Table 3: Comparison of automatic and manual methods for PF infected images

Image S.N	RBC Count			Parasite Count			Species Detection			Parasitemia (%)
	Software	Manual	Difference	Software	Manual	Difference	Software	Manual	Difference	
1	28	28	0	1	1	0	PV	PV	0	3.57
2	27	29	2	1	1	0	PV	PV	0	3.70
3	26	27	1	2	1	1	DINF	PV	1	7.69
4	27	27	0	1	1	0	PV	PV	0	3.70
5	20	20	0	1	1	0	PV	PV	0	5.00
6	27	27	0	1	1	0	PV	PV	0	3.70
7	22	23	1	1	1	0	PF	PV	1	4.54
8	28	29	1	1	1	0	PV	PV	0	3.57
9	29	30	1	1	1	0	PV	PV	0	3.44
10	28	28	0	1	1	0	PV	PV	0	3.57
11	22	22	0	1	1	0	PV	PV	0	4.54
12	28	28	0	1	1	0	PV	PV	0	3.57
13	30	29	-1	1	1	0	PV	PV	0	3.33
14	35	34	-1	2	1	1	DINF	PV	1	5.71
15	27	28	1	1	2	-1	PV	PV	0	3.70
16	30	32	2	1	1	0	PV	PV	0	3.33
17	27	27	0	1	1	0	PV	PV	0	3.70
18	25	25	0	1	1	0	PV	PV	0	4.00
19	26	26	0	2	1	1	DINF	PV	1	7.69
20	23	23	0	1	1	0	PV	PV	0	4.34
21	25	25	0	1	1	0	PV	PV	0	4.00
22	24	24	0	2	2	0	PV	PV	0	8.33
23	30	31	1	1	1	0	PV	PV	0	3.33
24	26	26	0	1	1	0	PV	PV	0	3.84
25	28	28	0	1	1	0	PV	PV	0	3.57
26	24	25	1	2	2	0	PV	PV	0	8.33
27	27	27	0	1	1	0	PV	PV	0	3.70
28	23	23	0	1	1	0	PV	PV	0	4.34
29	23	24	1	1	1	0	PV	PV	0	4.34
30	35	35	0	1	1	0	PV	PV	0	2.85
31	26	26	0	1	1	0	PV	PV	0	3.84
32	22	22	0	1	1	0	PV	PV	0	4.54
33	28	28	0	1	1	0	PV	PV	0	3.57
34	30	30	0	1	1	0	PV	PV	0	3.33
35	26	27	1	1	1	0	PV	PV	0	3.84
36	25	25	0	2	2	0	PV	PV	0	8.00
37	26	26	0	1	1	0	PV	PV	0	3.84
38	32	32	0	3	2	1	DINF	PV	1	9.37
39	26	27	1	1	1	0	PV	PV	0	3.84
40	29	29	0	2	1	1	DINF	PV	1	6.89
41	27	27	0	1	1	0	PV	PV	0	3.70
Total	1097	1109	12	50	46	4	–	–	6	4.58(Average)

Table 4 : Comparison of the automated and manual methods for PV infected images

As shown in Figure 5-1, the manual RBC counts of both *PF* and *PV* infected images are compared with the automatic RBC counts.

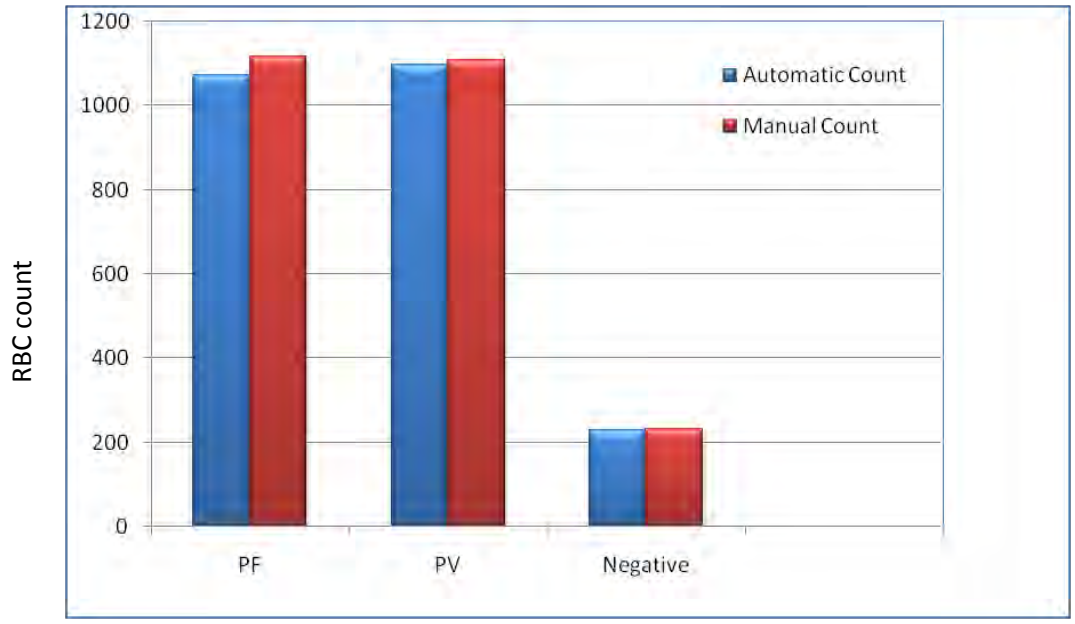


Figure 5-1: Comparison of manual and automatic counts for RBC

From the above histogram, we see that the automatic and manual RBC counts for the total of 41 *PF* infected images are 1073 and 1116 respectively. Similarly the automatic and manual counts for *PV* infected are 1097, 1109 respectively. And 228, 232 for non-infected images using automatic and manual techniques respectively. The total automatic and manual counts of RBCs (infected and non-infected ones) are 2398 and 2457 respectively.

In general automatic count of RBC are less in number than manual count with the count ratio (Equation 5.1) of 0.97 (97%).

$$Count\ ratio = \frac{Automatic\ count}{Manual\ count} \qquad 5.1$$

This is due to the overlap of RBCs where the algorithm is unable to distinguish between double cells. This shows that automatic diagnosis requires carefully prepared thin blood

smears. In this experiment, we have used thin blood films which were prepared mainly for manual diagnosis.

There are only two cases where the automatic counts are one cell more than that of the manual counts as indicated in table 3 of image number 13 and 14 where the differences are indicated by negative sign. These cases are negligible compared with the overall RBC count result.

Figure 5-2 shows the comparison of automatic and manual plasmodium counts. For *PF* infected images, total number of *PF* counted manually and automatically were 48 and 56 respectively. The automatic method counts nine false *PF* and misses one true *PF*, i.e.38 parasites were correctly counted.

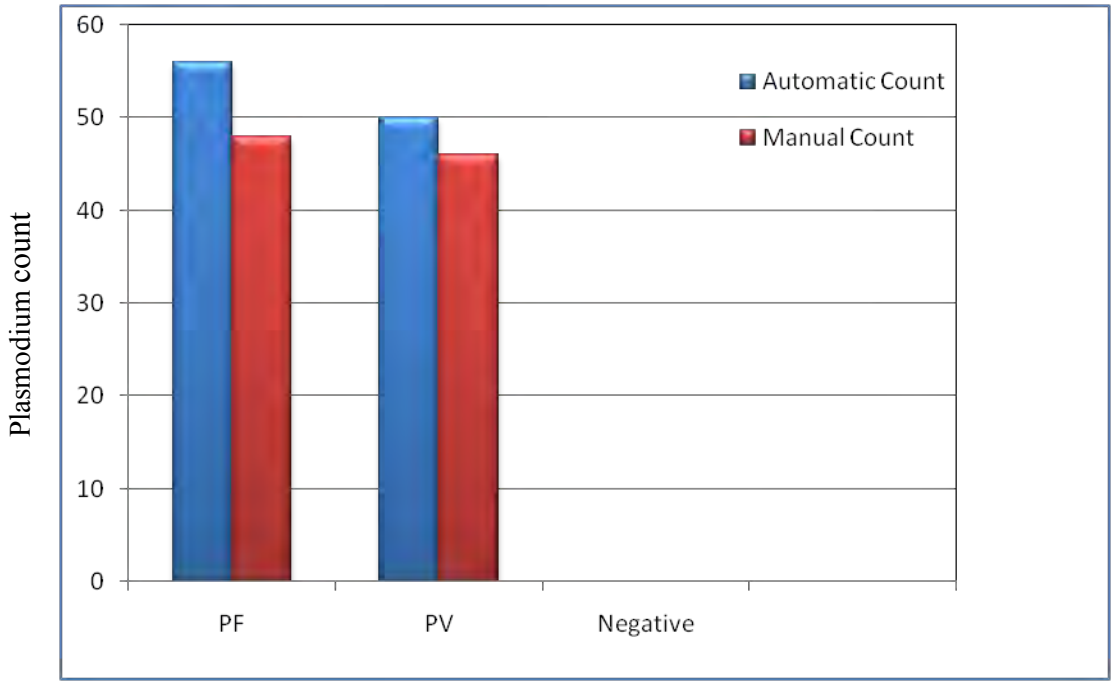


Figure 5-2: Comparison of manual and automatic methods for plasmodium count

For *PV* infected images, total number of parasites counted manually and automatically were 46 and 50 respectively. The automatic method counts five false *PV* and misses one true *PV* i.e.40 parasites were correctly counted.

Chromatin dot is found in all developmental stage of the plasmodium. To categorize the infection as *PV*, the system check whether the size of the infected RBC is greater than the non-infected and *PF* infected images in addition to the chromatin dot detection. It is because of this dual check in the case of *PV* infected images that it has low error when compared to *PF* infected images. There is no any variation observed between the automatic and manual plasmodium count in the case of non-infected images.

For species differentiation the comparison of the automatic and the manual methods is done for the total of 82 plasmodium infected images. As shown in Figure 5-3, from the total of 82 images, 75 of the species differentiation were made correctly, with the accuracy level of 91.46%.

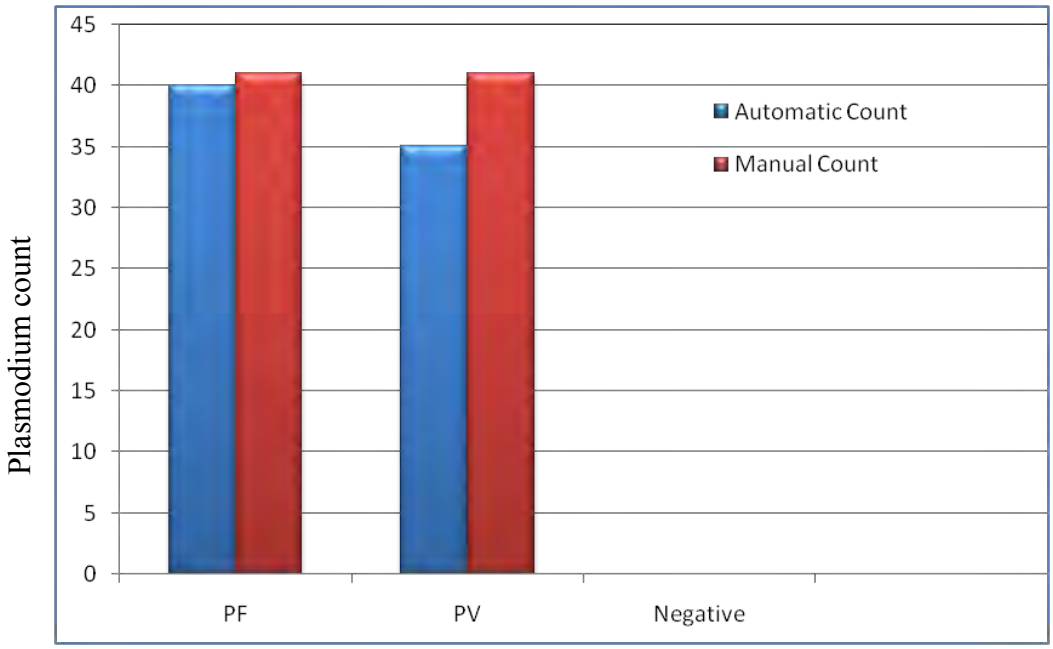


Figure 5-3: Comparison of manual and automatic methods for species differentiation

There are three possible result of species differentiation for each image: *PF*, *PV* and double infection (DIF). As shown in Figure 5-3 for *PF* infected images, only one *PF* is wrongly detected as *PV*. This is due to overlap of RBC which results in enlargement of the size.

There are six wrongly detected plasmodium for *PV* infected images where five of them are detected as double infection and the other one as *PF*. Double infection is possible when

both *PF* and *PV* are detected. The possibility of detecting wrong double infection is high for *PV* images, because normal RBCs are detected as *PF* infected. This could be due to the fact that the size of normal RBCs could be as big as *PF* infected RBCs.

To measure performance and accuracy of the method, we used two parameters: sensitivity and positive predictive value (*PPV*) [38]. Sensitivity is defined as the ability of the algorithm to correctly detect parasite(s). *PPV* is defined as the precision rate of the algorithm.

Mathematically, sensitivity and PPV are expressed as:-

$$sensitivity = \frac{TP}{TP+FN} \quad 5.2$$

$$PPV = \frac{TP}{TP+FP} \quad 5.3$$

Where, *TP*=True positive, *FN*= False negative and *FP*=False positive

The results of all the cases are summarized below, the values of false positive, false negative and true positive are reported in Table 5.

Total parasite case	94
True Positive	78
False Positive	14
False Negative	2
Sensitivity (%)	97.5
Positive predictive value (%)	84.8

Table 5: Summary of all the cases for parasite detection

Chapter 6 - Conclusion and Perspective

6.1 Conclusion

These days, thanks to the advancement of information communication technology, computerized diagnosis of specific diseases based on digital image processing is becoming an open area of study and research that has gained immense popularity in the medical science field. In this thesis work, an automatic detection of malaria parasite was implemented by processing digital images of thin blood smears acquired by a digital microscope. In this research, local data was used which is often missing in most image processing researches due to the unavailability of characterized specimens and well-equipped laboratory. In this thesis, the local images of patients were acquired at the Ethiopian Public Health Institute (EPHI) laboratory using the available digital microscope. The samples used in this research are collected from local malaria patients and characterized at Adama Malaria Control Center. Due to limited access to malaria samples available in other center, only samples collected at Adama Malaria Center were used. Moreover, a single digital microscope was used to acquire all the image samples. This could be mentioned as one of the limitations in this thesis work.

To develop the algorithm of automatic malaria detection, three main categories were identified. These are RBC segmentation, plasmodium detection and species differentiation. RBC segmentation is important to count the number of RBCs which in turn enables us to determine the level of parasite infection. Well segmented RBCs are also improving the sensitivity and specificity of the plasmodium detection and species differentiation.

Segmentation of RBC was done by watershed transform which is fast, robust and widely used in image processing and analysis, but it suffers from over-segmentation [27]. In this thesis, we made some improvements to this algorithm by selecting foreground and background markers in order to solve this problem. The foreground markers are obtained from circular Hough Transform (HT) and the background markers from distance transform. As stated in subsection 3.4.2.1, RBCs are geometrically similar to circle. Hence we used circular HT

to locate the center of the RBC which is used as foreground markers. The watershed transforms segment images as an object whenever it encounter regional minimum which may be due to noise and artifacts. This can lead to over segmentation. The markers at the center of blood cells can reduce the regional minimum only to the region where blood cells are located. The single pixel at the center of blood cell, as marker, indicates the location of the cell. The boundary of the cell was determined by background marker obtained from distance transform (subsection 3.4.2.2).

Plasmodium detection is the most important part in computerized diagnosis of malaria. As stated in subsection 3.5, the HSI color model is used for plasmodium detection. In manual microscopic malaria diagnosis, staining is used to identify parasite from the cells, background and other artefacts. A stained pixel is a pixel in the image which has distinctive, saturated color compared to the color of the red blood cells (RBCs). In the similar way the HSI color model decouples this color information from intensity information which enables the detection of plasmodium.

Species differentiation is the next important part in malaria diagnosis because the treatment for malaria disease is species specific. As stated in section1.2 the most common species in Ethiopia are *PF* and *PV*. Because of this, next task following plasmodium detection is to differentiate these two species. Once RBCs are correctly segmented, it is not difficult to distinguish between these two parasites; because *PV* infected RBCs are larger in size than non-infected and *PF* infected RBCs.

The results of the proposed algorithm in this thesis showed 97% of accuracy for RBCs count and 97.5% sensitivity level and 84.4% positive predictive value for plasmodium detection at cellular level. The species detections were done for each image with the accuracy level of 91.46%. These results show the potential of digital image analysis to be efficiently applied in automatic malaria diagnosis; which is not tedious, not time consuming, not requiring special training and considerable expertise when compared to the manual microscopic examination. The chromatin dots of plasmodium are clearly seen as red dots on hue components in the HIS color space. The sensitivity of plasmodium detection is higher when compared with the positive predication value (PPV). The positive prediction value (PPV) is

lower because the algorithm confuse some artifacts with the chromatin dots of the plasmodium, which could be removed by following good staining technique.

6.2 Perspective

As we could observe from the results, some of the blood cells overlap due to improper thin blood film preparation and some false positives because of artefacts. This leads to the development of protocols of thin blood film preparation and staining for automatic detection.

As indicated in subsection 3.1.1, thick blood film is suitable for the rapid detection of the parasite, although it does not permit an optimal review of parasite morphology for species identification. In the proposed method since we deal with species differentiation thin blood film is used. We suppose that, as it has been done for thin blood film, developing automated system for the thick blood film and integrating it with the proposed method can produce a good result. To decrease the human involvement it is also better to have automatic image acquisition set up. This can be done by using steeper motors to slide the position of the microscope stage to focus the camera on a different field of view in the sample, and acquire the image from the microscope directly to the MATLAB.

The proposed method was developed by using MATLAB software running on a windows platform. Adopting this method on an android application can make it to run on smart phones. The opportunity of using the smart phone as digital microscope to acquire magnified image is an encouraging tool to develop a smart phone based malaria diagnosis system. The smart phone microscope can be realized by coupling magnifying lenses and the camera of the smart phone together. The smart phone based malaria diagnosis system uses a smart phone based camera instead of CCD camera to acquire the digital images of magnified blood smears and directly process the images by the system installed on the smart phone and display the result. In addition to the benefit obtained from this thesis work, it will decrease cost, improve portability and accessibility because smart phones are widely used throughout the world, even in developing countries like Ethiopia where malaria disease is prevalent.

References

[1] N.Tangpukdee,C.Duangdee,P Wilairatana and S.Krudsood, "Malaria Diagnosis: A Brief Review," *Korean J Parasitol, vol. 47*, June 2009.

[2] Kaiser Family Foundation, htpp://kft.org/globaldata/, based on WHO World Malaria Report 2013; December 2013.

[3] Y.Legesse,A.Tegegn T.Belachew, K.Tushune, "Knowledge, Attitude and Practice about Malaria Transmission and Its Preventive Measures among Households in Urban Areas of Assosa Zone, Western Ethiopia" *Ethiop.J.Health Dev.*, 2007.

[4] A. Adugna, "Malaria in Ethiopia," www.EthioDemographyAndHealth.Org.

[5] BBC News, "Malaria spreading to new altitudes", http://www.bbc.com/news/health-26470755

[6] J.D. Smyth, "Introduction to Animal Parasitology," *Cambridge University Press*, Cambridge, 1994

[7] Ethiopian Federal Ministry of HEALTH, "Manual For the Laboratory Diagnosis of Malaria," *vol. 1*, Ethiopian Health and Nutrition Research Institute (EHNRI), September, 2012

[8] W.Ceusters and B.Smith, "Malaria Diagnosis and the Plasmodium Life Cycle: The BFO Perspective," *New York State Center of Excellence in Bioinformatics and Life Sciences*, Nov 2009.

[9] S. C. Parija, "PCR for diagnosis of malaria," Indian J Med Res, July 2010.

[10] J.Somasekar, B.E.Reddy, E.K.Reddy, "An Image Processing Approach for Accurate Determination of Parasitemia in Peripheral Blood Smear Images" *IJCA Special Issue,* 2011.

[11] K. Mitiku,G. Mengistu, and B. Gelaw "The reliability of blood film examination for malaria at the peripheral health unit," *Ethiopian Journal of Health Development, vol. 17*, pp. 197-204, 2003.

[12] L. B. Damahe, R. K. Krishna, N. J. Janwe, Dr. Thakur Nileshsingh V,"Segmentation Based Approach to Detect Parasites and RBCs in Blood Cell Images," *Journal Of Computer Science And Applications vol. 4*, No. 2, July, 2011.

[13] A.S.Abdul-nasir,M.Y.Mashor,Z.Mohamed,"Colour Image Segmentation Approach for Detection of Malaria Parasites

Using Various Colour Models and k-Means Clustering," *WSEAS transactions on biology and biomedicine, vol. 10,* January 2013.

[14] S.S.Savkare, S. P. Narote, "Automatic Detection of Malaria Parasites for Estimating Parasitemia," *International Journal of Computer Science and Security (IJCSS),* vol. 5, 2011.

[15] K.Rao and A. Dempster, "Modification on distance transform to avoid over-segmentation and under segmentation," *Proceedings of International Symposium on Video/Image Process and Multimed Commun Zadar,* Croatia, 2002.

[16] D. Stenning, V. Kashyap, T. C. M. Lee, D. A. van Dyk, and C. A. Young,"Morphological Image Analysis and Its Application to Sunspot Classification", 2010.

[17] Sagar B Tambe,Deepak Kulhare, M.D.Nirmal,Gopal Prajapati,"Image Processing (IP) Through Erosion and Dilation Methods," *International Journal of Emerging Technology and Advanced Engineering*, *vol.* 3.

[18] M. M. Hasan and P. K. Mishra,"Improving Morphology Operation for 2D Hole Filling Algorithm," *International Journal of Image Processing (IJIP)*, *vol. 6*, 2012.

[19] Leica Microsystems, "Leica Application Suite," www.leica-microsystems.com, June 2014

[20] M. W. Davidson and M. Abramowitz,"OPTICAL MICROSCOPY."

[21] J.Choudhary, N.S. Bagri,V. K.Saxena,"Survey of Different Segmentation Method for Low Quality Fingerprint Images," *International journal of engineering science & advanced technology*, vol. 2, 2012.

[22] Saylor Academy,"Red Blood Cell," Http://www.saylor.org/content/epubtest/RedBloodCell.pdf, 10, July, 2014.

[23] J. B. T. M.Roerdink and A.Meijster,"The Watershed Transform: Defnitions, Algorithms and Parallelization Strategies," *IOS Press*, 2001.

[24] J. Serra, "Image Analysis and Mathematical Morphology" *Academic Press,New York,* 1982

[25] M.Kaur, G.Jindal"Medical Image Segmentation using Marker Controlled Watershed Transformation," *International Journal of Computer Science and Technology,* vol. 2, 2011.

[26] H.Tulsani, S.Saxena,N.Yadav, "Segmentation using Morphological Watershed Transformation for Counting Blood Cells," *International Journal of Computer Applications & Information Technology*, vol. 2, Apr-May 2013.

[27] A. Bala, "An Improved Watershed Image Segmentation Technique using MATLAB," *International Journal of Scientific & Engineering Research* vol. 3, 2012.

[28] D.Ioannoua, W.Hudab, A.F.Lainec, "Circle recognition through a 2D Hough Transform and radius histogramming," *Image and Vision Computing17 (1999) 15–26*.

[29] CA Glasbey and GW Horgan,"Image Analysis for Biological Sciences," *John Wiley &Sons*, 1995.

[30] Q.Chen, X.Yang, E.M.Petriu,"Watershed Segmentation for Binary Images with Different Distance Transforms,"*IEEE,* 2004.

[31] H. D. Cheng, X_ H_ Jiang, Y. Sun and Jing Li Wang, "*Color Image Segmentation: Advances & Prospects,*"Utah State University, Dept. of Computer Science.

[32] J.P.Gasparri, A.Bouchet, G.Abras, V. Ballariny , J. I. Pastore, "Medical Image Segmentation using the HSI color space and Fuzzy Mathematical Morphology" *XVIII Congreso Argentino de Bioingeniería SABI 2011 – VII Jornadas de Ingeniería Clínica*.

[33] V.V.Makkapati,R.M.Rao,"Segmentation of Malaria Parasite in Peripheral Blood Smear Images,"*IEEE,* 2009.

[34] M. Spencer, "Fundamentals of Light Microscopy. UK," Cambridge University Press, 1982

[35] L. G., "How to correct background illumination in bright field microscopy," http://imagejdocu.tudor.lu/doku.php?id=howto:working:how_tocorrect_background_illumination_in_brightfield_microscopy, 15 June, 2014

[36] MathWorks· "Introduction to Creating UIs" http://www.apmath.spbu.ru/ru/staff/smirnovmn/files/buildgui.pdf, *2015*

[37] A. B. M. Nasiruzzaman,"Using MATLAB to Develop Standalone Graphical User Interface (GUI) Software Packages for Educational Purposes," Matlab - Modelling, Programming and Simulations, Emilson Pereira Leite (Ed.), 2010.

[38] I.Bates, V.Bekoe, and A.Asamoa-Adu "Improving the accuracy of malaria- related laboratory tests in ghana," *Malaria Journal, vol.* 3, 2004.

Appendices

Appendix A – Graphical User Interface

Graphical user interface (GUI) is software that works at the point of contact (interface) between a computer and its user. It employs graphics elements such as dialog boxes, icons, menus, scroll bars, etc. instead of text characters to let the user give commands to the computer or to manipulate what is on the screen [36].All programs running under a GUI use a consistent set of graphical elements so that once the user learns a particular interface, he or she can use all programs without learning additional or new commands. In this thesis, MATLAB has been employed to develop the GUI.

There are two methods in GUI in MATLAB, using GUI development environment (GUIDE) and coding from the MATLAB editor [37]. Due to its simplicity, we used the first approach to develop the GUI. To develop the GUI by using the first approach, we followed the following 5 steps.

Step 1 – Deciding on a Design

With analysis of the problems, the design layout of the GUI is designed as shown in Figure A-1. This is the paper work where we decide on how to implement the GUI.

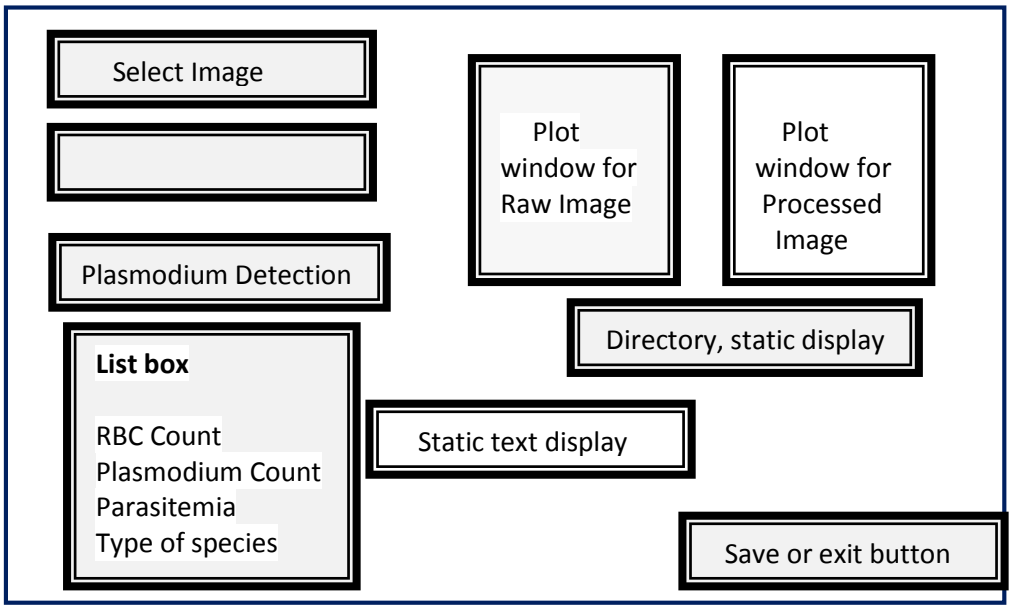


Figure A-1: The design layout of the GUI

Step 2- Use GUIDE to Lay Out the Window

This is the step where we start implementing GUI development on MATLAB. Typing guide command on the MATLAB command window and press enter key and then select "Blank GUI" by clicking "OK" at the bottom of the window opens the GUIDE design window(as shown in Figure A-2)

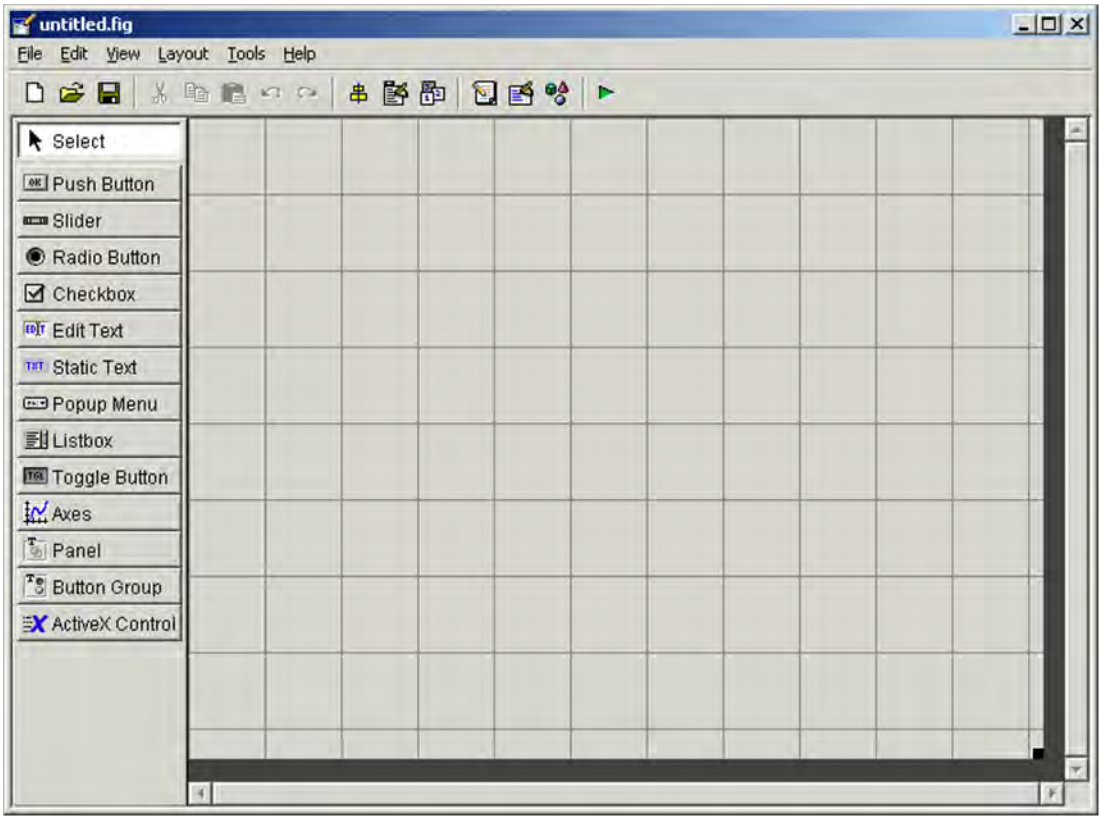


Figure A-2: GUIDE design window [36]

The guide command opens untitled figure which contains all the GUI tools needed to create and lay out the GUI components. GUI components include:- Push Button, Slider, Radio Button, Check Box, Edit Text, Static Text, Pop-up Menu, Listbox, Toggle Button, Table, Axes, Panel, Button Group ActiveX Control. Depending on our design at step 1, we select the Push Button, Listbox, Static Text and Axes.
To organize the GUI, we select the components from the left panel and place them in a required position.

Push button (Style property set to "pushbutton") is perhaps the most prevalent user interface controls (uicontrol) style. It is used primarily to indicate that a desired action should immediately take place. Since push buttons represent actions, they are usually labeled with a verb that describes the action that will take place if the user clicks on the button. Push buttons have a 3-dimensional look that makes it appear as if they are being pressed when the user clicks on the object. In our case we use four pushbuttons; one to select image from storage device, one for RBC segmentation, one for plasmodium detection and the other one to save the processed image or to exit from the GUI.

List boxes are graphical objects that display many lines of text and allow a user to select one or more of those lines. If there are more lines of text than can fit in the list box, a scroll bar will be created to allow the user to scroll up and down within the list box. The lines of text that the user can select among specified cell array of strings, and the value property indicates which of the strings are currently selected. A list box is created by creating a uicontrol whose style property is Listbox. A list box may be added to a GUI by using the list box tool in the layout editor. List boxes can be used to select a single item from a selection of possible choices. We have one list box for RBC count, plasmodium count, to determine the level of parasite and for species differentiation.

The static text style (style property set to text) of uicontrol is available for creating labels, status messages, or other information relevant to the user. The text graphics objects (i.e., those objects created with the text command) cannot be placed on top of frames. For frame objects and to create labels, static text style is used. Static text does not perform any action if the user clicks on any part of the object. In addition, the user cannot edit the information that is displayed. We use two static texts, one to show the source from where the original image is loaded for analysis and the other is used to display the number of RBCs found in the image, the number of plasmodium found, the level of parasite infection and the type of species found in the image when the respective list is selected.

Axes are used to contain the MATLAB plot. Two axes are used to display the original image and the segmented image.

To set the properties of the selected GUI components, click on the button in the layout area and then select property inspector from the toolbar or right click on the button and select inspect properties from the popup menu. The Property Inspector window shown lists of every property available for the component and allows us to set each value. For example, we may set many properties such as color, size, and font and text alignment.

Step 3-Assign Tags Using the Property Inspector

This property is very useful when programming GUIs in a method that requires MATLAB to search for the handle of a specific uicontrol object. The tag property can contain a string row vector of the choice. It is usually assigned a descriptive name that uniquely identifies a particular uicontrol object from all the other uicontrol objects. To assign tag for each component, double click on the component and set the tag properties to descriptive identifier from the property inspector. This name will be needed by the callback function to locate and update the text field.

Step 4 – Save the GUI

Activating the GUI by using "run" menu after we complete with the alignment of the components and setting the value in the property inspect, the GUIDE prompt us to save the file. After saving the file, GUIDE generate two files automatically, a fig-file ((*.fig) and an application m-file (*.m).

- The fig-file contains a complete description of the GUI figure, i.e. whatever uicontrols and axes are needed for the graphical interface.
- The application m-file contains the functions required to launch and control the GUI. The application m-file does not contain the code that lays out the GUI components; this information is saved in the fig-file. The GUI is controlled largely by means of callback functions, which are sub functions in the application m-file.

Step 5 -Write the Code for the Callbacks

M-file is automatically generated MATLAB code to launch and control GUI. GUIDE generates a callback function prototype for GUI components we create. Callback functions are written to determine what action is taken when we interact with a GUI component, e.g. click a button. We can start the GUI by typing the file name in the command window.

Figure A-3 shows a snapshot of the GUI developed in this thesis by following the above steps.

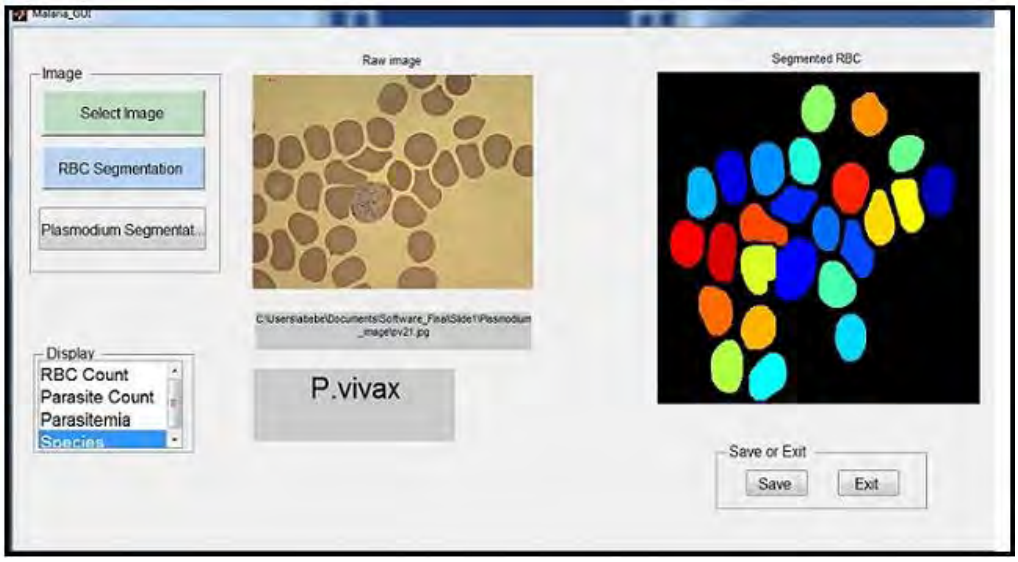


Figure A-3: Snapshot of developed GUI

Appendix B – Portion of the MATLAB Code and Typical Outputs

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% AUTOMATED DETECTION of MALARIA PARASITE
%% BASED on MICROSCOPIC IMAGE ANALYSIS
%% By Abebe Bekele
%%   2015
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clearall;
closeall ;
clc;
```

Reading image for analysis & median and average filtering

```
T=imread('image0881.jpg');
T=imresize(T,[500 500]);
orginal=T;
figure;imshow(T);title('Orginal Image ');
T=im2double(T);
fs=fspecial('average');
Red1=T(:,:,1);
Red1=medfilt2(Red1);
Red1=imfilter(Red1,fs,'replicate');
Green1=T(:,:,2);
Green1=medfilt2(Green1);
Green1=imfilter(Green1,fs,'replicate');
Blue1=T(:,:,3);
Blue1=medfilt2(Blue1);
Blue1=imfilter(Blue1,fs,'replicate');
```

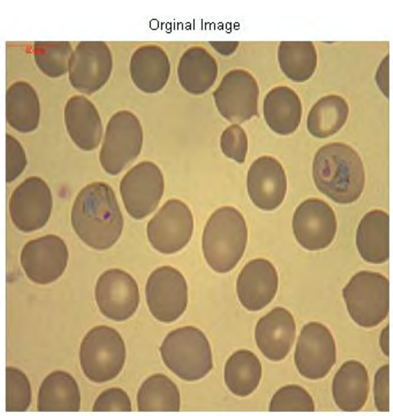


Figure B-1: Original image

Reading Dark Field image & apply median and average filtering

```
Dark_field=imread('image0914.jpg');
Dark_field=imresize(Dark_field,[500,500]);
Dark_field=im2double(Dark_field);
Red2=Dark_field(:,:,1);
Red2=imfilter(Red2,fs,'replicate');
Green2=Dark_field(:,:,2);
Green2=medfilt2(Green2);
Green2=imfilter(Green2,fs,'replicate');
Blue2=Dark_field(:,:,3);
Blue2=medfilt2(Blue2);
Blue2=imfilter(Blue2,fs,'replicate');
```

Reading Bright Field image & apply median and average filtering

```
Bright_field=imread('image0913.jpg');
Bright_field=imresize(Bright_field,[500 500]);
Bright_field=im2double(Bright_field);
Red3=Bright_field(:,:,1);
Red3=medfilt2(Red3);
Red3=imfilter(Red3,fs,'replicate');
Green3=Bright_field(:,:,2);
Green3=medfilt2(Green3);
Green3=imfilter(Green3,fs,'replicate');
Blue3=Bright_field(:,:,3);
Blue3=medfilt2(Blue3);
Blue3=imfilter(Blue3,fs,'replicate');
```

Background illumination Correction

```
RED=Red1.*((Red1-Red2)./(sqrt(Red3-Red2)));
R=(Red1-Red2);
GREEN=Green1.*((Green1-Green2)./sqrt((Green3-Green2)));
G=(Green1-Green2);
BLUE=Blue1.*((Blue1-Blue2)./(sqrt(Blue3-Blue2)));
B=(Blue1-Blue2);
RGB=cat(3,R,G,B);
Corrected_Image = cat(3,RED,GREEN,BLUE);
Corrected_Image=im2uint8(Corrected_Image);
T=Corrected_Image;
T2=RGB;
T2=imadjust(T2,[0 1],[0 1],1.5);
figure;imshow(T2);title('Background illumination Corrected image');
```

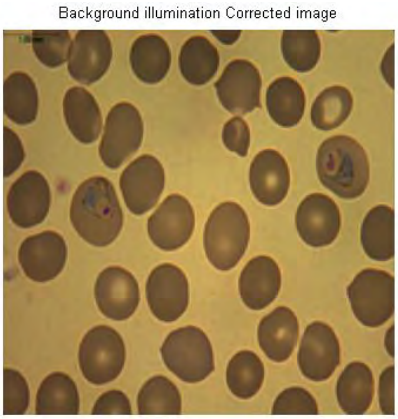


Figure B-2: Background illumination corrected image

RGB to HSI Conversion

```
HSI=rgb2hsi(T2);
figure;imshow(HSI);title('HSI Colour image');
```

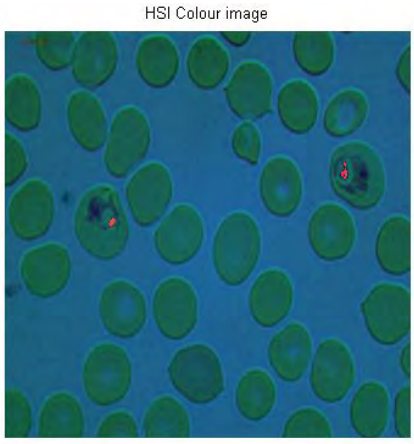


Figure B-1: HSI color image

Median filtering of HSI image

```
H=HSI(:,:,1);
Si=HSI(:,:,2);
I=HSI(:,:,3);
H=medfilt2(H);
Si=medfilt2(Si);
I=imfilter(I,fs,'replicate');
I=medfilt2(I);
I=imadjust(I);
I=histeq(I);
sr=strel('disk',2);
HSI_ALL=cat(3,H,Si,I);
```

```
im=HSI_ALL;
HSi=im;
H1=HSi(:,:,1);
S1=HSi(:,:,2);
I1=HSi(:,:,3);
figure;imshow(H1);title('Hue components of HSI colour Model');
```

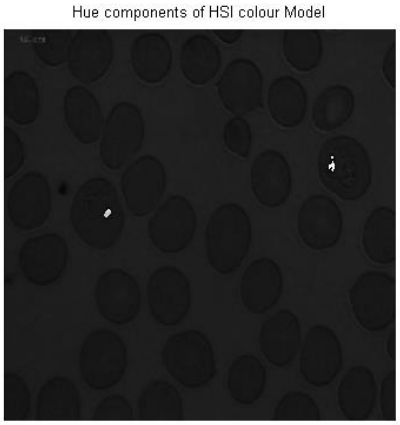


Figure B-4: Hue component of HSI color image

RBC segmentation

```
Seg= segment(T);
 Total=Seg;
figure;imshow(Total);title('Image of Segmented RBC');
 [L,N]=bwlabel(Total);
disp(N);
text([],[],'Total Number of RBC');
sz=size(T2);
```

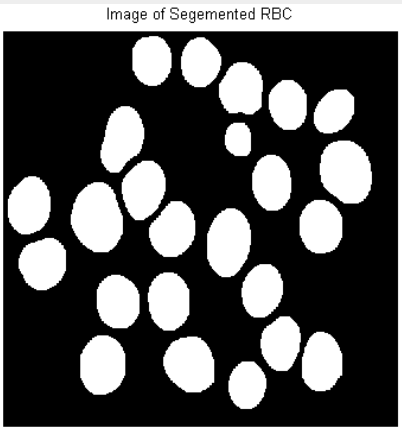


Figure B-5: RBC segmented image

Detection of parasite chromatin dots

```
fori=1:sz(1)
for j=1:sz(2)
if ((HSi(i,j,1)>0.5&&HSi(i,j,1)<1))
H(i,j)=1;
else
H(i,j)=0;
end
end
end
H=logical(H);
figure;imshow(H);title('Detected chromatin dots of the parasite');
```

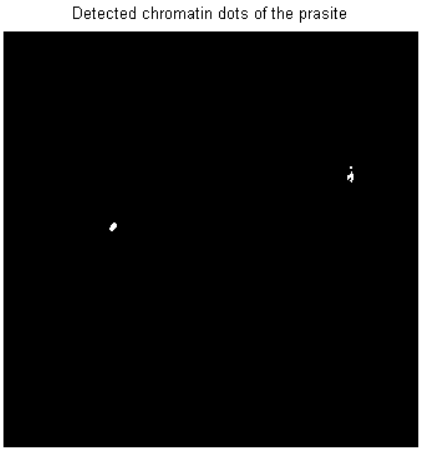


Figure B-6: Detected chromatin dots

Detection of parasite infected RBC

```
Seg=logical(Seg);
Parasite=imreconstruct(H,Seg);
[L2,N2]=bwlabel(Parasite);
figure;imshow(L2);title('Detected parasite infected RBC');
text([],[],'total number of Parasites');
Parastemia=((N2/N)*100);
```

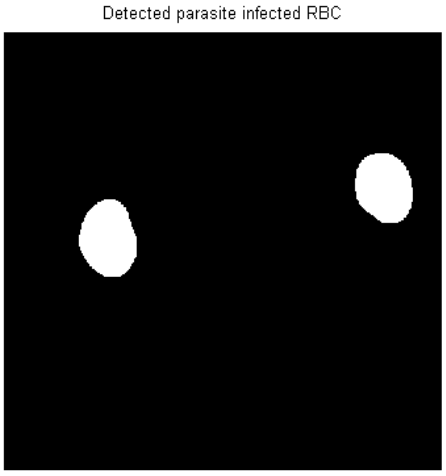


Figure B-7: Parasite infected RBCs

Species Detection

```
cc = bwconncomp(Parasite);
stats =regionprops(cc, 'Area');
index =([stats.Area]>4000);
output_image = ismember(labelmatrix(cc),find(index));
[L3,N3]=bwlabel(output_image);
if N2>0&N3>0&N2==N3
figure;imshow(L3);title('P.Vivax');
elseif N2>0&N3==0
figure; imshow(L2);title('P.Falciparum');
elseif N2>0&N3>0&N2~=N3
figure; imshow(L2);title('Double Infection');
else
disp('Negative')
end
end
end
```

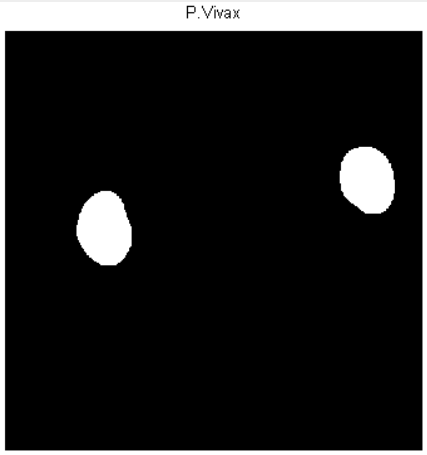


Figure B-8: Species differentiation