ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

Automatic Malaria Detection from Images of Microscopic Thin Blood Films

Daniel Zemene Mequanint

A Thesis Submitted to the Department of Computer Science in Partial Fulfillment for the Degree of Master of Science in Computer Science

Addis Ababa, Ethiopia

March 2016

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

Daniel Zemene Mequanint

Advisor: *Yaregal Assabe (PhD)*

This is to certify that the thesis prepared by *Daniel Zemene Mequanint*, titled: *Automatic Malaria Detection from Images of Microscopic Thin Blood Films and* submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name Signature Date

Advisor: _____

Examiner: _____

Examiner: _____

ABSTRACT

Malaria is one of the most common infectious diseases, causing wide spread sufferings and deaths in various parts of the world. The accurate and timely diagnosis of malaria infection is an essential requirement to control and cure the disease. Automated method of detecting malaria parasite improves accuracy, saves time, reduces the required human resource and minimizes human errors. In line with this, determination of parasitemia is also a crucial step to measure the amount of Plasmodium parasites in the patient's blood, an indicator for the degree of infection. Dealing with this, many researches have proposed different algorithms to solve the problem. However, accuracy of detecting the presence of parasites and estimating parasitemia under the occurrence of cells with complex structural arrangement is still regarded as a challenging task. Hence, in this research, a better technique to address Plasmodium parasite detection and computing parasitemia on the basis of counting red blood cells (RBCs), also called erythrocytes, that are extracted from microscopic thin blood films is looked for.

After images of microscopic thin blood film are acquired, we apply color median filter as a preprocessing activity to enhance the quality of images. A novel segmentation technique is proposed to isolate background from foreground (RBCs and parasites) and also to separate Plasmodium parasites from the rest of blood cell components. A hybrid of adaptive threshold and color structure tensor threshold is applied to come up with a better segmentation result. We use a total of 11 features derived from geometric, color structure tensor and color characteristics to detect RBCs and parasites. A novel red blood cells arrangement identification technique is proposed to deal with overlapping RBCs, which enables us to compute parasitemia even with the presence of high degree of RBCs overlap.

The proposed algorithms are tested using sample image data collected from different sources. A diagnostic accuracy of 93 % is achieved to detect Plasmodium parasites with sensitivity and specificity values of 94% and 91%, respectively. This was achieved even in the context where poorly illuminated microscopic images and complex red blood cells' arrangement existed. The accuracy of the proposed algorithm to compute parasitemia is found to be better than a system developed based on watershed segmentation technique, which is very popular and frequently used in many research works.

Keywords: *Plasmodium parasites, Color structure tensor, Parasitemia, Watershed segmentation, Red blood cell arrangement*

ACKNOWLEDGMENT

First and foremost, I want to offer this endeavor to our God the Almighty for the wisdom he bestowed upon me, the strength, peace of my mind, and good health in order to finish this research.

I would like to express my sincere gratitude to my advisor Dr. Yaregal Assabie for the continuous support of my study, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. My sincere thanks also goes to staff of Tikur Anbessa Hospital /Pathology Department/, who have dedicated their time for supplying equipment when it was needed. A special thanks goes to Dr. Mahlet Arayaselassie for her kind sharing on how to set up and use microscope while capturing images from thin blood films. I also wish to express my gratitude to the medical staff members from the Ethiopian Health Institution for providing a collection of blood samples' specimen and constructive ideas about the behavior of parasite under microscopic investigation.

My thanks extend to Minale H/Michael who has spent many hours for proof, experiment, reading and discussion. Thank you Engidawork Demeke and Yosef Wubalem for your constructive comments on my report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

AMD	Automatic Malaria Detection
CDC	Center for Disease Control and Prevention
CST	Color Structure Tensor
DIP	Digital Image Processing
FDREMH	Federal Democratic Republic of Ethiopia Ministry of Health
HSV	Hue-Saturation-Value
KNN	K Nearest Neighbor
MAE	Mean Absolute Error
MIRBC	Malaria Infected Red Blood Cell
PCA	Principal Component Analysis
RBC	Red Blood Cell
RBCs	Red Blood Cells
RDT	Rapid Diagnostic Test
RGB	Red-Green-Blue
SVM	Support Vector Machines
WBC	White Blood Cell
WHO	World Health Organization

CHAPTER ONE: INTRODUCTION

1.1 Background

Blood disorders can cause morphological changes in mature red blood cells. Hence, by investigating blood smears morphologically, we can study and distinguish blood diseases such as malaria. On the other hand, because most of the available methods are manual, expensive, time consuming and dependent on the expert's expertise so for improving detection rate and increasing its accuracy, a method of automated analysis is crucial.

Malaria is a serious infectious disease caused by a peripheral blood parasite of the genus Plasmodium. It is a parasitic disease in tropical and sub-tropical areas that causes large number of deaths. Malaria is caused by a parasite in the blood; the parasites are very small in size and can be seen only under a microscope with high magnification. There are four types of Plasmodium malaria species namely Plasmodium falciparum, Plasmodium vivax, Plasmodium oval and Plasmodium malaria. Each species may happen to be in any of the three stages called trophozoites, schizonts or gametocyte. Malaria remains a major public health problem in many countries of the world. According to World Health Organization (WHO) report 2015 [1], despite the progress in reducing malaria cases and deaths, it is estimated that 214 million cases of malaria occurred worldwide in 2015, leading to 438 000 malaria deaths. The global burden of mortality is dominated by countries in sub-Saharan Africa, accounting for more than 35% of the global total of estimated malaria deaths.

Malaria is also a major public health problem in Ethiopia. It has been consistently reported as one of the three leading causes of morbidity and mortality in the past years [2]. According to the revised version of the 2004 malaria diagnosis and treatment guidelines developed by the Federal Democratic Republic of Ethiopia Ministry of Health (FDREMoH) [2], Plasmodium species have different geographical distributions and different species have different clinical features which require different treatment. Plasmodium falciparum and Plasmodium vivax are the two dominant parasite species with relative frequency of 60% and 40%, respectively. This proportion varies from place to place and from season to season. In malaria epidemic situations, Plasmodium falciparum is the dominant parasite species that causes severe manifestations and almost all malaria deaths happen due to the infection by this parasite. The accurate and timely diagnosis of

malaria infection is essential to control and cure the disease. According to FDREMH, there are three ways of malaria diagnosis [2].

- Sign and symptoms are the most traditional methods commonly used in rural areas where there are scarcities of medical equipment.
- Rapid Diagnostic Test (RDT) also called "dipsticks" uses immuno-chromatographic techniques to detect Plasmodium-specific antigens in a finger prick blood sample.
- Microscopic diagnosis of malaria is based on examination of blood films stained with Giemsa [3]. This method of diagnosis also gives quantitative results that can be used in the evaluation of degree and rate of clearance of parasitemia. Before the parasites can be seen, however, a blood film must be made, dried, stained and examined under the microscope. When the blood analyst examine stained parasites, the diagnosis of malaria is confirmed.

In addition to this, automatic detection is also a type of diagnosis based on image taken from microscopic thin blood films. This diagnostic approach needs right quantification and accurate detection of Plasmodium parasites. A poorly designed automated count for RBCs (Red Blood Cells) fails to compute parasitemia under complex arrangement of RBCs (erythrocytes). A complex arrangement of erythrocytes will be formed due to the dynamic nature of them while they are squeezed to supply oxygen to tissues in different parts of the body. Another reason is the quality of the staining procedure before examining the erythrocytes under microscope which might make them to occur in clustered or overlapped form [3]. Still another factor that creates complex arrangement is poor segmentation process or existence of RBCs around the border of the image which makes RBCs to have broken morphological structure.

This research presents automatic detection and counting of red blood cells in poorly illuminated microscopic blood cell images with the presence of complex red blood cells' arrangement. The proposed work individuates the red blood cells from the other blood cells in the blood cell images after separating foreground object from background and then compute parasitemia through isolating overlapped cells. Detection of Plasmodium parasites and treating cells lying around the border of the image are also presented.

The shape of red blood cell is described as a biconcave disc. It is the commonest cell in thin blood films [3]. There are about 5, 000, 000 in each microliter (μl) of blood. With Giemsa staining [3], the red cell appears as a pale-greyish to light-pink disc measuring about 7.5 μm in diameter. Figure 1.1 [3] demonstrates the possible morphological structure under healthy environment and components of infected RBC that can be seen when it is examined under microscope.

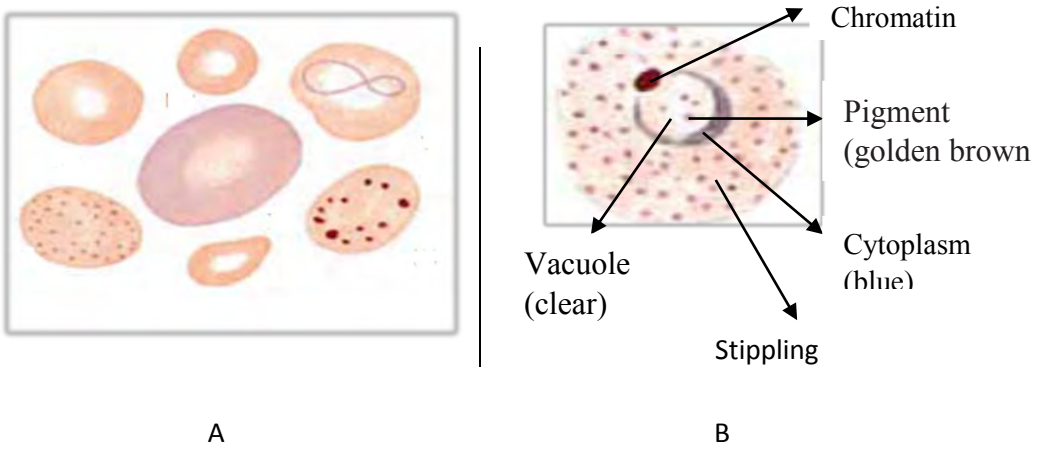


Figure 1.1: (A) Normal RBC (B) Components of malaria infected RBC in Giemsa stained thin blood films

Figure 1.1[3] shows the possible cases of normal RBC (A) and the different components of plasmodium parasite after infection is made (B). Some cells may appear larger than normal red blood cells (normocytic) and in some rare cases may even appear oval as shown in Figure 1.1 (A). In peripheral blood sample, visual detection and recognition of Plasmodium parasites is possible via a chemical process called Giemsa Staining. Giemsa stain colors each part of a malaria parasite differently. With good staining, it is easy to distinguish the parts shown in the diagram. Under normal circumstances the staining process slightly colorize the red blood cells but highlight Plasmodium parasite, WBC (White Blood Cell), and platelets [3].

Figure 1.2 [3] shows blood cell components when examined under microscope as a result of malaria infected RBCs that can be observed with a unique blue dominant color feature with ring shape. However, the red blood cells look pale color though it is affected by illumination.

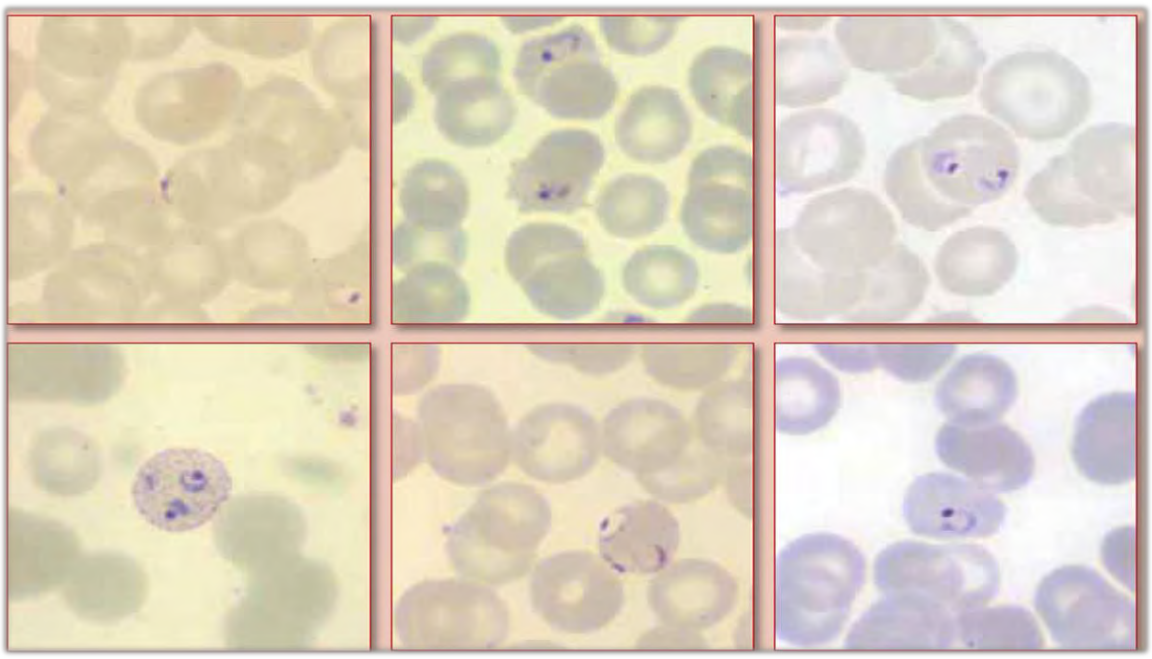


Figure 1.2: Microscopic observation of RBCs and plasmodium parasites in thin blood films

1.2 Motivation

Manual evaluation of blood films is time consuming, error-prone. Moreover, the accuracy of the final diagnosis ultimately depends on the skills and experience of the technician and the time spent studying each slide. The implementation of automated malaria detection using images extracted from thin blood films would result in accurate measurement in malaria diagnosis and improves the delay in treating patients and also reduces the time the physicians spent on the diagnosis. The motivation behind this thesis work is therefore, to implement automatic malaria detection system that enables accurate diagnosis in detecting and estimating parasitemia. There is a need that these systems should be accurate and robust in terms of detecting parasite and computing parasitemia.

1.3 Statement of the Problem

Automatic malaria detection through accurate parasitemia computation is a challenging task. Several research works [4, 5, 6, 7, 8, 9, 10, 11] have been devised to tackle this problem and a promising result was found though challenged back by the following key tasks which the proposed system will take into consideration.

- Lack of effective counting of RBCs and Plasmodium parasites with the presence of complex arrangement of red blood cell such as overlap and broken cells that lay around the border of image, and
- Accurate detection of parasite under poorly illuminated microscopic image as the variability and artifacts in microscopic images of blood samples pose significant challenges.

Due to the above limitations, proper approximation parasitemia is difficult and results in poor automatic examination. Therefore, the hypothesis made is that it is possible to come up with better approach in automatic malaria detection using digital image processing and analysis techniques. The outcome of the proposed study is believed to improve detection and counting of RBCs through image analysis.

1.4 Objectives

General Objective

The general objective of this research is to develop automatic malaria detection system that enables the detection of malaria infected red blood cells from Giemsa stained microscopic thin blood films and to build a model for parasitemia estimation.

Specific objective

In order to achieve the general objective, the following specific objectives are addressed.

- Review of related literature with the aim of better understanding of the techniques, methods and ideas relevant to the problem domain.
- Collection of sample microscopic image data from Giemsa stained thin blood films
- Designing an algorithm for counting and detecting malaria infected RBCs in a given image taken from microscopic Giemsa stained thin blood films.
- Developing a prototype for automatic malaria detection.
- Testing and evaluating the performance of the system using sample data.

1.5 Scope and Limitations

The main focus of this thesis is to detect malaria infected RBC using image analysis technique and to estimate the number of RBC and Plasmodium parasites. Malaria detection will be considered only for two species namely Plasmodium vivax and Plasmodium

falciparum as availability of data and degree of prevalence being high relative to other species.

1.6 Methods

The following methods have been used for the accomplishment of this thesis.

- **Review of Related Literature**

Before starting the actual work, a deep study was made in the literature written on this area to have a clear picture about the work. Papers written on automatic malaria detection were reviewed to get an understanding of the various techniques and methods of automatic malaria detection.

- **Data Collection**

Sample data was collected using oil immersion views (10 × 1000) of Giemsa stained blood films from AAU, Faculty of Medicine, Department of Pathology. This was captured using a Leica DM750 HD camera mounted microscope. Captured images were assumed to have 1024 pixels × 603 pixels in joint photograph expert group (*JPEG)* format. In addition to this, various images were collected from standard image database that provides medical images tailored for research purpose. These are namely, center for disease control (CDC) [12] and MaMic [13]. These image databases are the result of a real patient's blood specimen that are prepared through the so called Giemsa staining process using microscopic examination.

- **Tools**

MATLAB version R2014a was used to implement the prototype of the system. Visio 2013 was used for designing the system architecture and various diagram.

- **Experimental Evaluations**

For the purpose of evaluating the system in a fair and logical manner, sample data which are collected from different sources were fed into the developed prototype. Accordingly, the system was evaluated by comparing its output against the manually observed phenomena using sensitivity and specificity measures for evaluating diagnostic accuracy and also by conducting a comparison with other algorithm that are frequently used by previous works.

1.7 Application of Results

The outcome of the research brings a good opportunity in simplifying the challenge behind malaria diagnosis process as described below.

- It saves both time and effort of patients and physicians by fastening diagnoses and determining the right cures which surely stops the suffering of the affected person.
- It brings better measurement of parasitemia inside the body of patient using proper quantification and classification of malaria infected red blood cells.
- It enables physicians or blood analysts from false positive due to overload and high concentration of long time microscope observation; it might also save their vision rate due to the refraction behavior of microscope.

1.8 Organization of the Thesis

The remaining of this thesis report is organized as follows. Chapter 2 presents literature review on automatic malaria detection. Chapter 3 introduces related works. This chapter discusses researches that have been conducted on automatic malaria detection. The design of automatic malaria detection is presented in Chapter 4. The evaluation, test results and discussion are described in Chapter 5. Lastly, in Chapter 6 conclusions and future works are pointed out.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter presents the review of literatures on the concepts that are basis for the research. The chapter begins with a brief introduction to malaria diagnosis with this the distinction between manual based and automated version is discussed. The morphological structure of Plasmodium parasite, red blood cell arrangement and overview about approaches to automatic malaria detection are also presented.

This chapter is organized in five sections: Malaria Diagnosis, Automatic Malaria Detection System, Malaria Parasite Life Stage Morphology, digital image processing (DIP) Approaches to Malaria Detection, and Summary.

2.2 Malaria Diagnosis

Malaria diagnosis is in general the process of testing the existence of malaria parasites inside patient's blood. As described in Chapter one there are two ways of diagnosing malaria, these are automated and manual. The manual approach includes sign and symptoms, RDT [14], and microscopic examination. Microscopic examination is the gold standard method of diagnosing malaria that means by examining under the microscope a drop of the patient's blood, spread out as a "blood smear" on a microscope slide. Prior to examination, the specimen is stained (most often with the Giemsa stain [3]) to give the parasites a distinctive appearance. A specimen is a sample of something, like a specimen of blood or body tissue that is taken for medical testing.

The automated version of diagnosing malaria on the other hand is a computerized approach that enables us to detect malaria parasite and to compute parasitemia. In contrast to manual based malaria diagnosis, automated based diagnosis is fast and accurate.

2.2.1 Manual Based Malaria Detection

Detecting the presence of malaria parasites in the examined blood slide is one of the most important tasks in malaria diagnosis [4]. The procedure in manual based malaria detection is performed manually by expert blood analyst by searching for parasites in blood slide using a light microscope [3]. During this process, blood analyst use color as a clue to identify the presence of parasites in stained slide. Here, visual detection and recognition of parasites is possible and efficient due to the use of chemical named Giemsa stain [3]. The staining process slightly colorizes the red blood cells (RBC), but highlights the

parasites, white blood cells (WBC), platelets and artifacts. In Giemsa staining process there are two types of films namely thin and thick blood films [3].

As it can be seen in Figure 2.1 below, a staining object is a blood specimen collected from patients infected by malaria and pass through the so called Gemisa staining. Under microscopic examination stained object resulting non-infectious comprises RBC, WBC, Platelets, and Artifacts. On the other hand, the stained object that results in malaria infection in any of the four species at various life stage. Malaria infection affects only RBC as a result the morphological structure of infected RBC become changed and the parasite bring it to and use it as a host to do part of its life stage. Therefore, our interest is RBC than other components as it hosts the parasite.

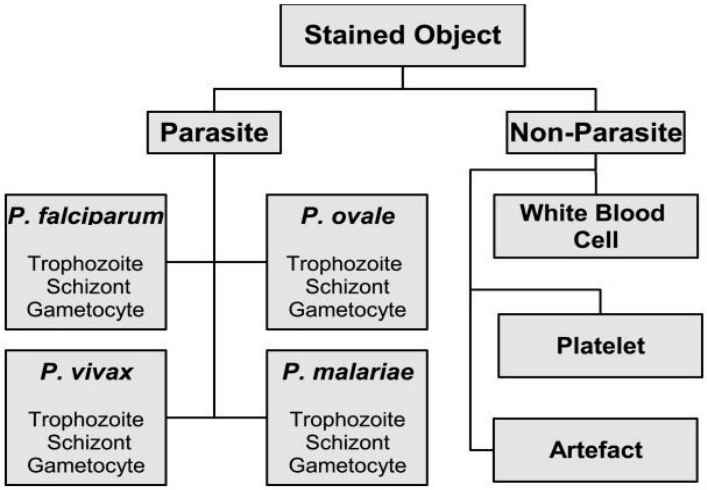


Figure 2.1: Components of stained blood components from malaria infected RBCs under microscope examination

Thus, the whole process requires an ability to differentiate between the malaria parasites and these non-parasitic stained components (RBC, WBC, platelets and artifacts) using visual information. However, the manual recognition method is time consuming and effortful. Therefore, fast and efficient methods are required for detection of malaria parasites in order to prevent the false diagnosis of malaria.

According to WHO basic malaria microscopy [3], to detect parasite starts with the effect the parasite has on the infected red blood cell. The existence of stippling, maurur dots and clefts are indication of malaria parasite in the blood.

2.2.2 Automatic Malaria Detection System

Malaria detection system is a system that enables us to detect malaria parasite. Automatic malaria detection (AMD) system is therefore, an information system designed for the purpose of detecting malaria parasite. The realization of AMD is through employing digital image analyses technique to: detect the occurrence of malaria parasite inside RBC and determine the density of parasite inside the body of patient which is called percentage of parasitemia or simply parasitemia. However, detection of malaria parasite using image processing is challenged by false malaria parasite detection and lack of better measurement in estimating parasitemia. The fact that the color of the parasites, RBC and background regions vary depending on the pH of the buffer used for smear preparation [3], incorrect microscopic settings, and poor staining process results for poor recognition or even for total loose of information. Therefore, a good staining procedure is the key step that needs to be taken in to account before acquiring microscopic image.

2.3 Morphological Structure of Malaria Parasites in Thin Blood Films

Morphology is the study of structural behavior of an object. Therefore, morphology of Plasmodium parasites entails the study of structural behavior of Plasmodium parasites and erythrocytes. Thin blood films enable us to study the different kinds of malaria parasite under microscopic observation. Malaria parasite passes through different steps from early stage to mature state. The morphological structure of the parasite under different stages will also differ.

- **Malaria Species**

As mentioned in Chapter One, there are four species of malaria parasite that naturally infect humans [3]: Plasmodium falciparum is the commonest species in tropical parts of the world. It is responsible for most cases of severe malaria and death. Plasmodium vivax is the commonest species in the cooler parts of the tropics. It is the largest of the human malaria parasites and the cause of much illness and absenteeism. Plasmodium malaria is less common but is found in much of the tropical world. Plasmodium ovale is considered a rare species [3]. In Ethiopia, plasmodium falciparum and plasmodium vivax are most prevalent. According to WHO basic malaria microscopy [3] and standard operating procedures in managing malaria slide [15], to distinguishing between the four species starts with the effect the parasite has on the infected red blood cell.

- ✓ Is the cell enlarged?
- ✓ Does the infected red blood cell have some kind of stippling?

With Plasmodium vivax, good staining shows schuffner stippling or james dots. With Plasmodium falciparum, maurer dots or clefts are less commonly seen.

- **Malaria Life Stage**

There are three levels of parasite stages in different species [3]. These are Tropazoit Stage, Schizont Stage, and Gametocyte Stage. The morphological structure of RBC in those stages are: ring shape, rosette shape and banana shape respectively. However, from the human visual aspect, shape only will not be enough to recognize the type of life stage. Hence, color and size also have a significant role in determining stages. The various life stages in the two malaria specious are explained visually in Figure 2.2 - 2.4 [3].

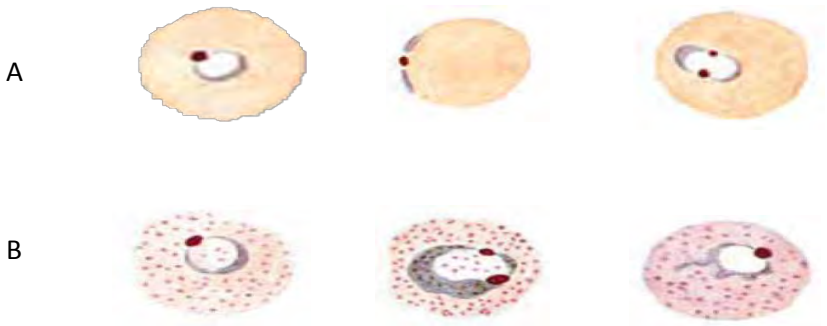


Figure 2.2: Tropazoit stage (A) Plasmodium falciparum (B) Plasmodium vivax

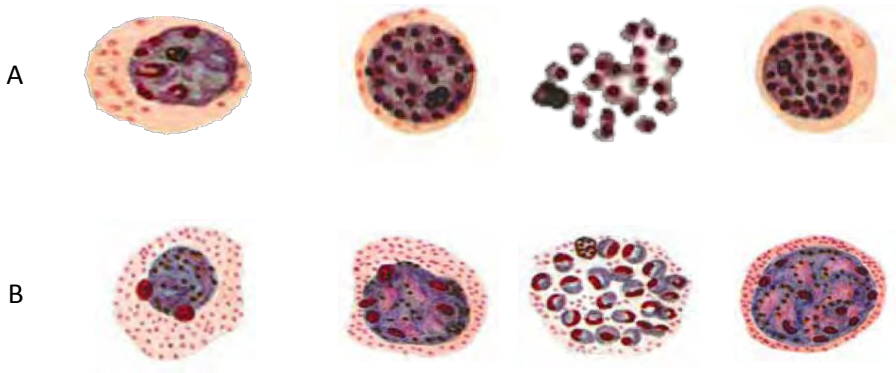


Figure 2.3: Schizont stage (A) Plasmodium falciparum (B) Plasmodium vivax

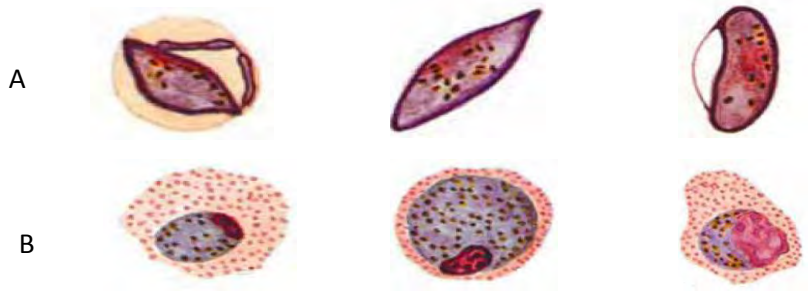


Figure 2.4: Gametocyte stage (A) Plasmodium falciparum (B) Plasmodium vivax

- **RBC Arrangement**

The morphological structure of RBCs under thin blood films microscopic examination vary accordingly. Most of the time RBCs exist in clumped form rather than stand alone.

The fact that these cells become arranged in single, clustered, hidden (broken) is non deterministic. The reason would be firstly, blood films are naturally non manageable in a sense that blood analyst prepare staining procedure without the knowledge of how the expected results look like. Secondly, erythrocytes are dynamic in nature in a sense that they provide oxygen to different tissues of the body hence, it may squeeze itself to deliver oxygen and remove carbon dioxide [6]. A good staining procedure can produces a good result in terms of detecting parasite and visibility but not ensuring the degree of separation among cells and the way RBCs is arranged. The cell might be arranged in such a way that cells might become connected/touched, overlapped due to dynamic nature of RBC, microscopic setting and chromatic nature of microscope, staining quality, etc. As can be seen in Figure 2.5 [12, 13] images are extracted from Giemsa staining thin blood films showing degree of overlap and complex arrangement of RBCs.

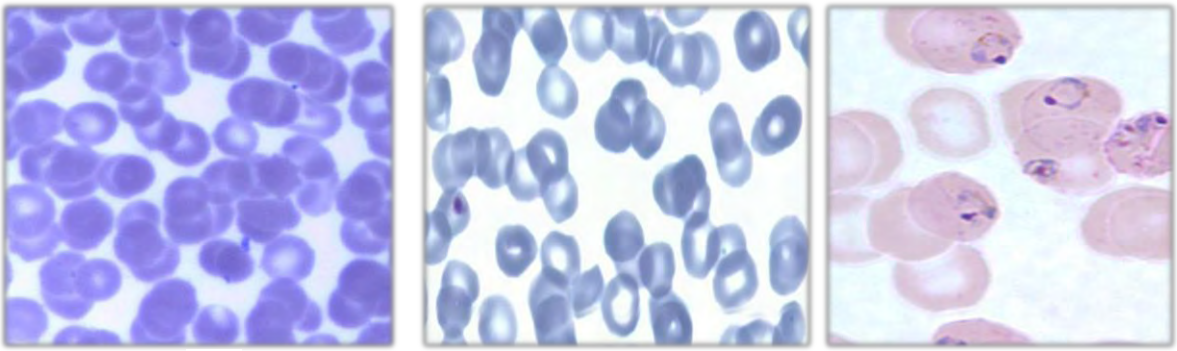


Figure 2.5: Image extracted from thin blood films showing overlap, touch and single instance RBCs

Hence, the arrangement of cells as can be seen from Figure 2.5 would have the following geometrical structure.

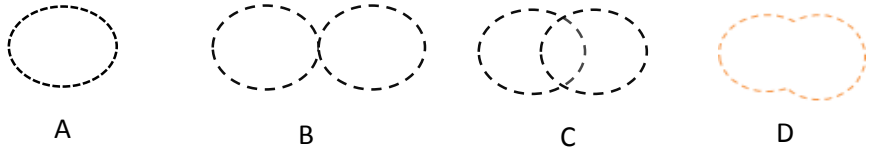


Figure 2.6: Possible overlap cases of RBCs

In A, cells are isolated from the others. Such arrangement is easy to count the number of cells and RBCs with similar attribute is called single instance RBCs. However, in B, C, and D counting becomes so sophisticated that it needs a further segmentation process. RBCs with similar behavior are under the category of complex arrangement and is called overlapped RBCs. Hence, RBC segmentation is challenging because cells might arrange in such a way that it becomes connected (B) which is low degree of overlap, overlapped with partial information like B or overlapped with little information when high degree of overlap is occurred such as in D. The outcome of complex arrangement like overlap cases will be inconsistent in recognition and classification for other DIP processes.

In literature, counting of erythrocytes are determined through several methods. Among it, if we consider the simplest approach, pixel count method [16] is a straightforward technique for counting similar objects in a segmented image. This can be realized through calculating the total number of pixels that are classified as objects. If the objects to be counted are similar in terms of shape and size, then the average area of each object is known. By dividing the number of pixels classified as objects by the known average area of one object, the total number of objects is achieved.

This simple method provides an accurate estimation of the total number of objects in the image, only if the following criteria are met:

- The objects have the same area
- No objects overlap

However, this method is not considered as a reliable method to count the blood cells for two reasons. First, the red blood cells are not necessarily the same size. Overlapping is the second reason to avoid using this method. These two factors reduce the reliability of this method. Clump-splitting methods are also considered by several researchers. Watershed techniques [17, 18], concavity analysis [19], cell center detection [20] are the major ones.

These approaches still have some drawbacks and which need to be improved. For example, watershed technique is relatively time-consuming because of the partitions being over-split in cluster locations and contour parts being merged together. The concavity analysis methods is also too sensitive in using threshold to recognize regions or concavity pixels along border since at lower and higher thresholds, the objects will be over-split and under-split, respectively. Cell center detection also best suit for the case of non-distorted shape and well-focused images. In this research, model based and boundary edge analysis that are responsible for extracting curves are integrated to come up with better estimation of parasitemia.

2.4 Approaches to Automatic Malaria Detection

Digital image processing has been serving as a tool for solving real problems associated with images. Likewise, different researchers employ the DIP approach in their works, and hence, this study also considers its use. An image may be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y, and the amplitude values of f are all finite, discrete quantities, we call the image a digital image. Hence, the field of digital image processing refers to processing digital images by means of a digital computer [21, 22]. Generally, any automatic malaria detection systems include five major processes: image acquisition, pre-processing, segmentation, feature extraction and classification.

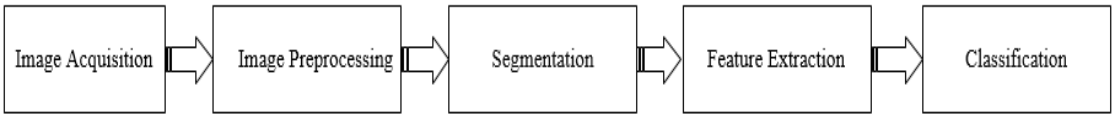


Figure 2.7: DIP process for automatic malaria detection system

2.4.1 Input Data and Image Acquisition

In automatic malaria detection (AMD) system, image acquisition is an activity of acquiring images from microscopic thin blood films through Giemsa staining process using microscopic camera or camera-mounted microscope designed for this purpose.

One of the challenges in microscopy image is the introduction of photometric variant [23] such as shadow/shading, specularities and highlights. Consequently, the occurrence of inconvenient color temperature (illumination) under different environment results in less

quality image. Microscopic image acquisition has been done using microscope under different control environment. It is worth mentioning that adjusting resolution through different objective lens level is a key step before acquiring microscopic image. Further details and sample images regarding the different level of visibility under different objective lenses could be shown in Annex A.

2.4.2 Image Preprocessing

Preprocessing is a general name for operations done on images at the initial level where both input and output are intensity images. The aim of preprocessing is an improvement of the malaria infected RBC image by suppressing unnecessary distortions or enhancing major image features important for further recognition processes. Among the preprocessing activities in automatic malaria detection systems, a set of sequential processes (enhancing image, creating binary image, removing small objects, removing cells at boundary) has been applied to the original image after image acquisition.

2.4.3 Image Segmentation

Segmentation means labeling pixels in an image based on their values and positions. Segmentation subdivides an image to its constituent regions or objects. The level to which the subdivision is carried depends on the problem being solved. That is, segmentation should stop when the object of interest in an application has been isolated [21, 22].

In AMD system, segmentation deals with extraction of RBCs and Plasmodium parasites from the background of a thin or thick smeared microscopic image. Furthermore, overlapping cell isolation is also a further segmentation process that needs a special treatment of cells arranged in complex manner. Hence, before processing detection or classification of Plasmodium parasites, accurate segmentation process needs to be performed.

Numerous methods of segmentation have been proposed from the early years of computer vision. However, it is still one of the most difficult tasks in image processing. There are two main factors that contribute to image segmentation: similarity of the pixels of an object and discontinuity at the boundary of the object [16, 24, 24]. Two different approaches for image segmentation are in common use: edge based and region based methods. The edge based approach attempts to extract the boundaries of the objects or regions, since the edge is recognizable by application of a differential operator. The Sobel, Robert, Prewitt, Laplacian, and Canny operators are mostly used for edge extraction in image processing

applications [21, 22]. If the interest is to extract the whole region of an object, a region-based method is applied. A region-based approach can use any of a variety of strategies for pixel classification. Threshold based is also employed in several image segmentation process it is based on pixel distribution (intensity based).

Furthermore, threshold method can be viewed in two different approaches known as local thresholding where image needs to be subdivided in to a certain block and different threshold is taken for each block in an image. The advantage in local thresholding is that it is suitable in segmenting poorly illuminated image like microscopic image. The second threshold method is adaptive threshold where threshold value changes dynamically over the image. The threshold value is calculated for each pixel and if this calculated value is lower than threshold value then the given pixel is part of the background object otherwise it is part of the foreground object. Like local thresholding, adaptive thresholding has also a significant role to segment images which are affected by uneven illumination. However, both local and adaptive threshold pose a new object specially when the space in the foreground pixels is dominated and affected by poor illumination. Global thresholding chose a threshold for the whole image. In general threshold method is simple but powerful approach for segmenting light object on dark background. This can be done, for example, by choosing proper local threshold value T in a certain block of an image. If the pixel value is greater than T, then that pixel belongs to object region otherwise the pixel belongs to background region. This can divide object pixel into several regions and isolate the object from background [25].

- **Segmentation of Microscopic Image**

The use of a light microscope for the quantitative analysis of specimens requires an understanding of light sources, the interaction of light with the desired specimen, the characteristics of modern microscope optics, the characteristics of modern electro-optical sensors (in particular, CCD cameras), and the proper use of algorithms for the restoration, segmentation, and analysis of digital images. All of these components are necessary if one is to achieve the accurate measurement of analog quantities given a digital representation of an image [18, 26].

Several works [6, 8, 10] used watershed segmentation to isolate overlapped objects. Any grey image can be considered as a topographic surface. If we flood this surface from its minima and, if we prevent the merging of the waters coming from different sources, the

image is partitioned into two different sets: the catchment basins and the watershed lines [22, 27]. If this transformation is applied to the image gradient, the catchment basins should theoretically correspond to the homogeneous grey level regions of this image. However, in practice, this transformation produces an important over segmentation due to noise or local irregularities in the gradient image. Segmenting an image by the watershed transformation is therefore a two-step process: Finding the markers and the segmentation criterion (the criterion or function which will be used to split the regions - it is most often the contrast or gradient, but not necessarily or distance transformation in the image). If the catchment basin is circular in shape then watershed with distance transformation generates a good result [18, 22, 27].

2.4.4 Feature Extraction

After an image has been segmented the detected region needs to be described (represented) in a form more suitable for further processing. Representation of an image region can be carried out in two ways which are by characterizing features of its boundary, i.e. edge of a region or by characterizing features of its interior, i.e. set of pixels constituting the region. Feature extraction is, therefore, a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. Before the feature extraction task is done, the features of the object need to be first detected. And the feature extraction is for feature matching. Therefore feature extraction is a task between feature detection and feature matching [21, 22, 25].

In AMD system, color, texture and geometric features are employed for classification of malaria species, stage and isolation of overlapped cells [4, 8, 28]. Color features are derived from the red, green, blue, hue and saturation components. Texture features are generated from the greyscale image matrices of the red, green and blue components, as well as the intensity component from the hue-saturation-intensity image space [21, 22].

2.4.5 Classification

Classification includes a broad range of decision theoretic approaches to the identification of images [21]. All classification algorithms are based on the assumption that the image in question depicts one or more features and that each of these features belongs to one of several distinct and exclusive classes. The classes may be specified priori by an analyst (as in supervised classification) or automatically clustered (i.e. as in unsupervised classification) into sets of prototype classes, where the analyst merely specifies the number

of desired categories. Classification algorithms consist of two phases of processing: training and testing. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, i.e. training class, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features.

In AMD system, classification of malarial infected cell becomes the challenging task. Several classifiers have been reported for the computerized recognition of malaria parasite species and life stage. Bayes classifier and different types of Artificial Neural Networks (ANNs), local linear map, SVM, k-mean and fuzzy system has been extensively used as classifier for blood cell recognition [9, 11, 16, 29].

2.4.6 Fundamentals of Color Image

Color is the most important feature in color image processing. There are two major advantages of using color vision, the first one is that, color provides extra information which allows the distinction between various physical causes for color variations in the world, such as changes due to shadows, light source reflections, and object reflectance variations [30, 31].

Next to this, color is an important discriminative property of objects, allowing to distinguish between different properties [30]. Straightforward application of existing luminance based operators on the separate color channels, and subsequent combination of the results, fails due to undesired artifacts [30]. For example, smoothing a color image with a Gaussian filter blurs the edges, which is also common for luminance based smoothing. In color images linear smoothing introduces new chromaticity. Therefore, an efficient model to handle the cancelation effect is discussed in [30].

Color space or color model refers to a coordinate system where each color stands for a point. The often used color models consist of the RGB (red-green-blue) model, CMY (cyan- magenta-yellow) model, CMYK (cyan- magenta-yellow- black) model and HIS (hue-saturation- intensity) model. RGB model consists of three components. These three components are combined together to produce composite colorful images. Each image pixel is formed by a number of bits. The number of these bits is called pixel depth. A full color image is normally 24 bits, and therefore the total number of the colors in a 24-bit RGB image is 16,777,216. RGB (or true color) images are 3-D arrays that we may consider conceptually as three distinct 2-D planes, one corresponding to each of the three red (R),

green (G) and blue (B) color channels. RGB is the most common color space used for digital image representation as it conveniently corresponds to the three primary colors which are mixed for display on a monitor or similar device [21, 22, 31].

- **Color Structure Tensor**

Tensor analysis is a generalization of the notion from vector analysis [32, 33]. The need for such a theory is motivated by the fact that there are physical quantities that are characterized by complex nature that cannot naturally be described or represented by scalar or vectors. For example, the stress of a point of a solid body due to internal force, the deformation of an arbitrary element of volume of an elastic body, etc. These quantity can be described and represented by mathematical entities called tensors [33]. The concept of vector, like that of vector, is an invariance concept. It means that it is an entity that does not depend on any frame of reference or coordinate system [33]. Tensors describe the local orientation rather than the direction. More precise, the tensor of a vector and the tensor of the same vector rotated over π radian are equal [31].

Color structure tensor [30, 31, 32] models the local structure in color image and computes the linear structure in the images. It also models linearly symmetric structures that are frequently found in images and at the same time it is a good estimator of local orientation [23]. To represent certain geometric properties of images, it associates 2 × 2 symmetric matrices that are tensors to them. A more complete introduction about tensor based feature detection in color images can be found in [23, 30, 31, 32]. Yaregal Assabie and Bigun [34] had also employed the local orientation features in luminance while they built offline hand written Amharic word recognition system. They showed that eigenvectors of an inertia matrix are related to the average of a complex number$\left(D_{xf} + D_{yf}\right)^2$, where $D_{xf}$and D_{yf} are the spatial derivatives in the direction of x and y, respectively and j is imaginary number.

The notion behind structure tensor (ST) in luminance and color is different only in the level of dimension. By this , we mean that in both cases there is a challenge that opposing vectors occur around edges where for one channel the signal decreases while for another the signal increases, and a simple addition of the opposing derivative signals of the different channels reduces the total derivative strength [32, 34]. As a result, opposing vectors cancel each other which in turn yields weak signal or dim highlights that edges become unidentifiable. However, the challenge behind cancelation effect in both

luminance and color had been solved by proposing tensor based gradient for which opposing vectors reinforce one another [32, 34].

Structure tensor in luminance can be derived from the eigenvalues analysis of the known hessian matrix using PCA or from the second order moment I_{20} (the eigenvector from the largest eigenvalues) which is derived from the ST. However, in color images, the three channels in RGB color space [28], namely Red-channel, Green-channel, and Blue-channel would be threated separately and combined in a single channel. Given an image *f,* structure tensor is given by the following formula [32].

$$S = \begin{pmatrix} \overline{f_x^2} & \overline{f_x f_y} \\ \overline{f_x f_y} & \overline{f_y^2} \end{pmatrix} \quad (1)$$

where the subscripts indicates spatial derivatives in the direction of x and y, and the bar _: indicates the convolution with a Gaussian filter.

Besides, tensors can be added for different channels. For a multichannel image $f = (f1, f^2 \ldots f^n)^T$. Hence, the structure tensor in Equation (1) will be rewritten as:

$$S = \begin{pmatrix} \overline{f_x^T f_x} & \overline{f_x^T f_y} \\ \overline{f_y^T f_y} & \overline{f_y^T f_y} \end{pmatrix} \quad (2)$$

where superscript T indicates the transpose operation. For color images, f = (R; G; B) ^{T}. This results in the color structure tensor.

$$S = \begin{pmatrix} \overline{R_x^2 + G_x^2 + B_x^2} & \overline{R_x R_y + G_x G_y + B_x B_y} \\ \overline{R_x R_y + G_x G_y + B_x B_y} & \overline{R_y^2 + G_y^2 + B_y^2} \end{pmatrix} \quad (3)$$

The color structure tensor describes the 2D first order differential structure at a certain point in the image. Eigenvalue analysis of the tensor leads to two eigenvalues which are defined by

$$\lambda_1 = \frac{1}{2}\left(\overline{f_x^T f_x} + \overline{f_y^T f_y} + \sqrt{\left(\overline{f_x^T f_x} - \overline{f_y^T f_y}\right)^2 + \left(\overline{2 f_x^T f_y}\right)^2}\right) \quad (4)$$

$$\lambda_2 = \frac{1}{2}\left(\overline{f_x^T f_x} + \overline{f_y^T f_y} - \sqrt{\left(\overline{f_x^T f_x} - \overline{f_y^T f_y}\right)^2 + \left(\overline{2 f_x^T f_y}\right)^2}\right) \quad (5)$$

Where λ_1 is the first eigenvalue and λ_2 represents the second eigenvalue .

The direction of $\boldsymbol{\lambda_1}$ indicates the prominent local orientation, which is equal to the orientation in the image with maximum color change.

$$\theta = \frac{1}{2} arctan\left(\frac{\overline{2 f_x^T f_y}}{\overline{f_x^T f_x} - \overline{f_y^T f_y}}\right) \quad (6)$$

The two Eigen values λ_1, and λ_2 are values in the local orientation which are the most and least prominent orientation respectively. λ_1 - λ_2 describes the derivative energy in the prominent orientation is corrected for by energy contributed by noise, λ_2. As Bigun pointed out in his book Direction in 2D [26], an ideal linear symmetry is present in the image, when value of the two Eigen values, $\lambda_2 = 0$ and $\lambda_1 > 0$. Besides, the λ's can be combined to give local descriptors. The sum of λ_1 and λ_2 describes the total local derivative energy [31, 32]. I_{20} which is the vector in the direction of largest eigenvalue, can be also computed from the direction field tensor based on the second moment inertia.

$$I_{20} = \left(D_{xf} + D_{yf}\right)^2 \quad (7)$$

where D_{xf} and D_{yf} represent spatial derivative in the direction of *x* and *y* respectively. Equivalently, as Bigun [33] pointed out, the vector I_{20} can be computed directly from the complex tensor as:

$$I_{20} = S(1,1) - S(2,2) + j * 2 * S(1,2) \quad (8)$$

where S is the complex structure tensor defined in equation 2. However, in color tensor the spatial derivative in the direction of x and y should be applied for all color channels as follow: $D_{xf} = (R_x + G_x + B_x)$

$$D_{yf} = \left(R_y + G_y + B_y\right)$$

where R, G, B define the corresponding Red, Green, and Blue channels respectively. Therefore, I_{20} in color structure tensor can be computed directly by applying Equation 3 to equation 7 as follows:

$S(1,1) = \overline{R_x^2 + G_x^2 + B_x^2}$

$S(2,2) = \overline{R_y^2 + G_y^2 + B_y^2}$

$S(1,2) = \overline{R_x R_y + G_x G_y + B_x B_y}$

Hence, we can drive the vector I_{20} for three channels from Equation (8) as:

$$I_{20} = \left(I_{xx} - I_{yy}\right) + j * 2 * I_{xy} \quad (9)$$

where $\boldsymbol{I_{xx}}$ **:** represents the convolution of the sum of the square of the spatial derivative in the direction of x.

$\boldsymbol{I_{yy}}$ **:** represents the convolution of the sum of the square of the spatial derivative in the direction of x.

$\boldsymbol{I_{xy}}$ **:** represents the convolution of the sum of the product of the partial derivative in the direction of both x and y.

Hence, Equation (9) can be used as a measure of symmetry structure in color images.

Color structure tensor is applicable in several image processing activities where preprocessing and feature extraction are the major ones [31, 32]. As a preprocessing, color tensor improves the occurrence of noise due to high frequency around edges of an image [32]. Tensor based feature detection is also possible through local orientation estimation using the complex I_{20} derived from orientation in color tensor and eigenvalue analysis.

- **Photometric Invariant Corner Detection**

Corners in images represent critical information in describing object features that are essential for pattern recognition and identification. In [23] a curvature-based corner detector that detects both fine and coarse features is presented. First, it extracts contours from a canny edge map. Second, it computes the absolute value of curvature of each point on a contour at a low scale and regards local maxima of absolute curvature as initial corner candidates. Third, it uses an adaptive curvature threshold to remove round corners from the initial list. Finally, false corners due to quantization noise and trivial details are eliminated by evaluating the angles of corner candidates in a dynamic region of support.

Harris corner detector is also another method used by several image analyses for detecting corners. The relevance of color structure tensor along with dichromatic reflection model in corner detection is ensuring a corner which is invariant to shadow, shading and reflection [31, 32]. The color Harris operator Hf can be written as a function of the eigenvalues of the structure tensor.

$$\begin{aligned} Hf &= \left(\overline{f_x^T f_x}\,\overline{f_y^T f_y}\right) - \left(\overline{2 f_x^T f_y}\right)^2 \text{-k}\left(\overline{f_x^T f_x} + \overline{f_y^T f_y}\right)^2 \quad (10) \\ &= \lambda_1\lambda_1 \text{ - k } (\lambda_1 + \lambda_2)^2 \end{aligned}$$

where the over line indicates a Gaussian averaging window and λ_1, λ_2 represents eigenvalues and k is a constant .

2.5 Summary

In this chapter, general background regarding malaria and its diagnosis process have been discussed. The distinction between manual and automatic malaria detection is also presented. In addition to this, the morphology of Plasmodium parasites and the structural arrangement of RBCs are introduced. Theoretical backgrounds about digital image processing that are applied to automatic malaria detection such as preprocessing, segmentation, feature extraction, and classification are discussed. Color structure tensor, a newly emerging tool which has a power in better preprocessing and feature extraction capability, has been also discussed in logical manner. On top of that, I_{20} for color structure tensor is able to be formulated on the bases of structure tensor in luminance and from the three channels in true color image. Tools associated with AMD are also discovered. This thesis applies digital image processing to come up with detection of Plasmodium parasites and computation of parasitemia.

CHAPTER THREE: RELATED WORK

3.1 Introduction

The issue of malaria is significant around the world and a number of researchers, using digital image processing, contributed detection of malaria from images of microscopic thin blood films that passed through Giemsa staining procedure. In this chapter, previous works conducted in automatic malaria detection using digital image processing are critically reviewed. Here, only those works whose contributions made great progress to the field are presented. The research work related to automatic malaria detection using digital image processing could be broadly categorized into two main groups which is based on the diagnostic result obtained automation. The first category which is called complete automatic method where detection, classification and parasitemia computation are integrated; whereas the second category is semi-automatic method that excludes parasitemia computation or classification of malaria species and life stage. Gaps that exist in previous works for automatic malaria detection are also been identified.

3.2 Semi-Automated Malaria Detection

Diaz *et al*. [11] evaluates a color segmentation technique for separation of pixels into three different classes: parasite, red blood cell and background, based on standard supervised classification algorithms. Four different supervised classification techniques (KNN, Naive Bayes, SVM and Neural network) are evaluated on different color spaces: RGB, normalized RGB, HSV and YCbCr. But the complex mix of colors present in the parasites makes it difficult to discriminate individual pixels using only color information. The study reported a sensitivity of 94% for detection of infected erythrocytes and 79% for stages identification.

Mausumi *et al.* [5] undertook a research on RBCs detection and counting using Hough transform. The primary objective of the research was to develop automatic segmentation and counting of red blood cells in microscopic blood cells images using Hough transform. Their research work isolated RBC from other blood cells in a given image by using Hough transform. The method they used was by applying various pre-processing techniques like edge detection, spatial smoothing filtering and adaptive histogram equalization to detect and extract the red blood cells from the images. Feature extraction has been done through

the Hough transform method which has been used to find out the red blood cells based on their sizes and their shapes. This isolated the red blood cells from the rest of the image of the blood sample so that further processes like counting can be applied exclusively on them. Their proposed method was cost effective and can be easily implemented. However, the author did not address the treatment of overlapped RBC, although they used anticoagulant liquid to separate the cells to decrease overlapping effect which is not a perfect solution.

Neetu *et al.* [7] carried out a research on malaria detection using morphological and novel threshold selection techniques to identify erythrocytes (red blood cells) and possible parasites present on microscopic slides. However, the authors did not consider occurrence of overlapped cells while segmenting RBCs and the cases where RBCs and parasites are affected by poor illumination.

Abbas and Dzulkifli [28] studied on the enhancement of contrast in the YCbCr color for algorithms like KNN and k-means clustering since in the YCbCr color space, the enhancement results in different colors for malarial parasites. Their proposed solution is rather an initial step for counting blood cells, species identification, life stage identification and leukemia recognition. They did focus only on the enhancement process. They stated that their proposed method was checked on a set of 90 sample images.

3.3 Complete Automated Malaria Detection

Suradkar [4] conducted a research on automated diagnosis or screening of malaria infection in microscopic images of thin blood film smears. The primary objective of the work was to develop a fully automated image classification system, to positively identify malaria parasites present in thin blood smears, and to differentiate species of malaria. The method the author used is that images taken from thin blood smears as a source of input are further processed using image analyses techniques such as histogram equalization for preprocessing and edge detection to count the number of red blood cells. However, to detect infected red blood cells the author used color features from RGB channel directly in which case it has drawbacks in considering poorly illuminated microscopic image. Besides, the author did not consider counting of RBC when the cells become arranged in a pool /overlapped/ or hidden around the edge of the scene, which may bring a deviation from the exact number to be detected.

Renuka *et al.* [6] studied on computerized shape analyses of erythrocytes. The changes in the shape of erythrocyte and its cytoplasm have been determined by shape descriptors and gray scale variation of the cytoplasm by microscopic imaging and image processing tools. They did the shape analyses through the help of shape descriptors based on projected area, perimeter and form factor. Edge enhancements were derived by employing the Sobel operators in the software. Contour extractions of the images are done by successive deletion of the outer most layer of the images. The area and perimeter were determined based on the number of pixels. They used watershed technique to separate overlapped RBCs. The proposed shape analyses have a good implication for identifying RBC morphology and they provide a good technique for detecting edges and highlighting edges. They also recommend a better technique for measuring the severity level of malaria. However, the system has a pitfall in that they used watershed techniques where it has limitation on over segmentation of highly overlapped RBCs.

Panchbhai *et al*. [8] conducted a research on RBCs and parasites segmentation from thin smear blood cell images. They developed automated system for the identification of malaria parasites by using RGB color space model to segment RBC from other blood components. They first separated out the RGB image into three different layers of Red, Green, and Blue. By performing the processing on Green layer, they segmented the infected RBCs out of all through automatic thersholding technique. The selection of thresholding required to analyze the histogram of the given images which gives the best selection of thresholding. But for the automatic selection of thersholding they adopted Otsu algorithm and offset value which is based on global thresholding. This way, they discovered that still effort is needed to come up with a better result. Based on this result, they collected 8 images from [12] and tested its validity by comparing the manual counting and proposed automatic counting. The sample images were not affected by poor illumination and at the same time the degree of overlapping among erythrocytes in the image was less. Finally, the percentage of parasitemia is computed for each case, and the automated counting provides a good result. However, the authors neglected the existence of RBC in an overlapped manner which usually brings a big deviation in achieving the exact number to be detected. Besides, as Otsu method is based on global threshold, the method is recommended if the nature of input image is manageable in a sense that it lacks to select threshold in microscopic image as it is highly affected by illumination.

Ruberto *et al.* [10] proposed a technique of automatically detecting and quantifying malaria parasites infection in blood images of patients. The method employed a modified watershed algorithm to segment erythrocytes. Two alternatives were proposed for classifying parasite stages. One was the use of morphological thinning, where skeletons of parasites images were used to categorize parasites into their respective stages of infection. The second option was the use of color histograms similarity.The efficiency of the segmentation algorithm proposed reduces with the degree of clustering of erythrocytes. Similarly, the accuracy of color histogram similarity for classification of parasites would depend on the imaging parameters and illumination conditions under which the image being probed is taken. The detection accuracy of parasitemia was relatively low at 50%.

Nasir *et al*. [9] proposed a segmentation method based on color image to separate malaria infected RBC. They used C-Y Color model and k-means clustering algorithm. They used 50 images to test their result. The result indicated that k-means clustering using saturation component of C-Y color model had produced the best segmentation performance with segmentation accuracy of 99.19%. However, the k-means clustering is sensitive to color change as it tries to cluster each pixels based on the given centers. Hence, a small change in color will affect the overall result computed like what we have the case in poorly illuminated microscopic images.

Ross *et al.* [29] proposed a technique for automating malaria diagnosis using light microscopy. Camera mounted microscope was used to capture image of Giemsa stained blood slides. Common DIP techniques and neural network classifiers were used. Plasmodium parasites were positively identified with a sensitivity of 81% while the accuracy for species determination was 73%. The common approach such as morphological processing to segment Plasmodium parasite is not enough for the case of segmenting clumped cells. The sizes of erythrocytes were assumed to be circular. This is not always the case. Sometimes erythrocytes shapes are oval or sickled like cell if, for example, there is deficient in hemoglobin the so called disease known as anemia which results in morphological variation on RBCs .

3.4 Summary

This chapter reviewed automatic malaria detection systems that are related to the study at hand. In general, the whole attempts lie down either in fully automatic or semi-automatic malaria detection. In both cases, segmentation of microscopic images affected by poor illumination and treating RBC lying around the border of the image were not considered satisfactorily. Besides, counting of RBCs was not determined for the cases where complex state of arrangements exist. By this, it is meant that further research is needed to complement the gap specified above.

Thus, in this thesis, an attempt is made to consider the integration of different level of segmentation using adaptive technique and color structure tensor threshold to overcome the effect of illumination. Furthermore, model based estimation of overlapping RBCs will be applied for better determination of parasitemia.

CHAPTER FOUR: THE PROPOSED SOLUTION

4.1 Introduction

As discussed in Chapter Three, most of the previous attempts lack accurate measurement in diagnosing malaria. By this, it is meant that that accuracy of parasite detection under poorly illuminated microscopic images and computing parasitemia with similar condition and also with the presence of clustered enterocytes need a special treatment. In this thesis, therefore, an automatic malaria detection system with the design goal that enables to attain accurate measurement in computing parasitemia and to build a better performance in detecting Plasmodium parasites is proposed.

In the following sections, the details about the techniques and the model developed for the proposed solution are described. In Section 4.2, the proposed system architecture for automatic malaria detection is presented. Section 4.3 presents how input image is preprocessed for further activities. Section 4.4 presents the requirement and assumptions for segmentation of microscopic images extracted from thin blood films that passed through Gemisa staining process while Section 4.5 presents feature extraction. Section 4.6 presents malaria parasite detection and Section 4.7 summarizes generalized model for identification of erythrocytes' arrangement. Section 4.8 summarizes the requirement and assumptions made on counting of RBCs and malaria parasites, and finally, Section 4.9 summarizes this chapter.

4.2 Proposed System Architecture

In this section, the proposed system architecture for automatic malaria detection system is illustrated in Figure 4.1. The architecture of the proposed system encompasses all digital image processing that most of DIP applications need to follow. However, different level of segmentation based on adaptive thresholding and color thresholding are employed to render better result. On top of that, integrating edge symmetry usually found from the boundary of RBCs are also applied for modeling RBCs's arrangement. Counting clustered RBCs are also employed through rule based model fitting algorithm that is used to match overlapped RBCs and the template RBC obtained from arrangement model. Hence, those activities made the proposed system architecture different from the previous approach.

The system architecture consists of six major components namely, preprocessing, segmentation, feature extraction, malaria parasite detection, identification of RBCs'

arrangement, and RBCs and parasites counting. The preprocessing component does the job of removing noise and false regions. The output of the preprocessing component is fed into segmentation component. The segmentation component is responsible for isolating foreground object from background. The third component, feature extraction, computes discriminant properties of image constitutes from the result of segmentation. This component consists of three sub components to extract features namely, geometric feature extraction which is responsible for extracting geometric features, color structure tensor (CST) feature extraction in which case 6 features are identified for the purpose of parasite detection, modeling RBCs arrangement, and isolation of overlapped RBCs. Malaria parasite detection, isolates Plasmodium parasites from RBCs and other blood components. This component uses color structure tensor thresholding and color thresholding from RGB of the input image. Identification of RBCs arrangement analyzes how RBCs are arranged and it addresses isolation of RBCs into three categories that are used as an output to RBCs and parasites counting component. The three categories are single instance RBCs, overlapped RBCs and broken RBCs. This component constitutes three sub components: primitive extraction, pattern generation to generate oriented patterns found from RBCs' arrangement and model RBC selection. The result obtained from identification of RBCs' arrangement is used as input to the next component, counting RBCs and parasites, responsible for estimating the number of RBC and Plasmodium parasite. This component is based on the result obtained from identification of RBCs' arrangement and Malaria parasite detection.

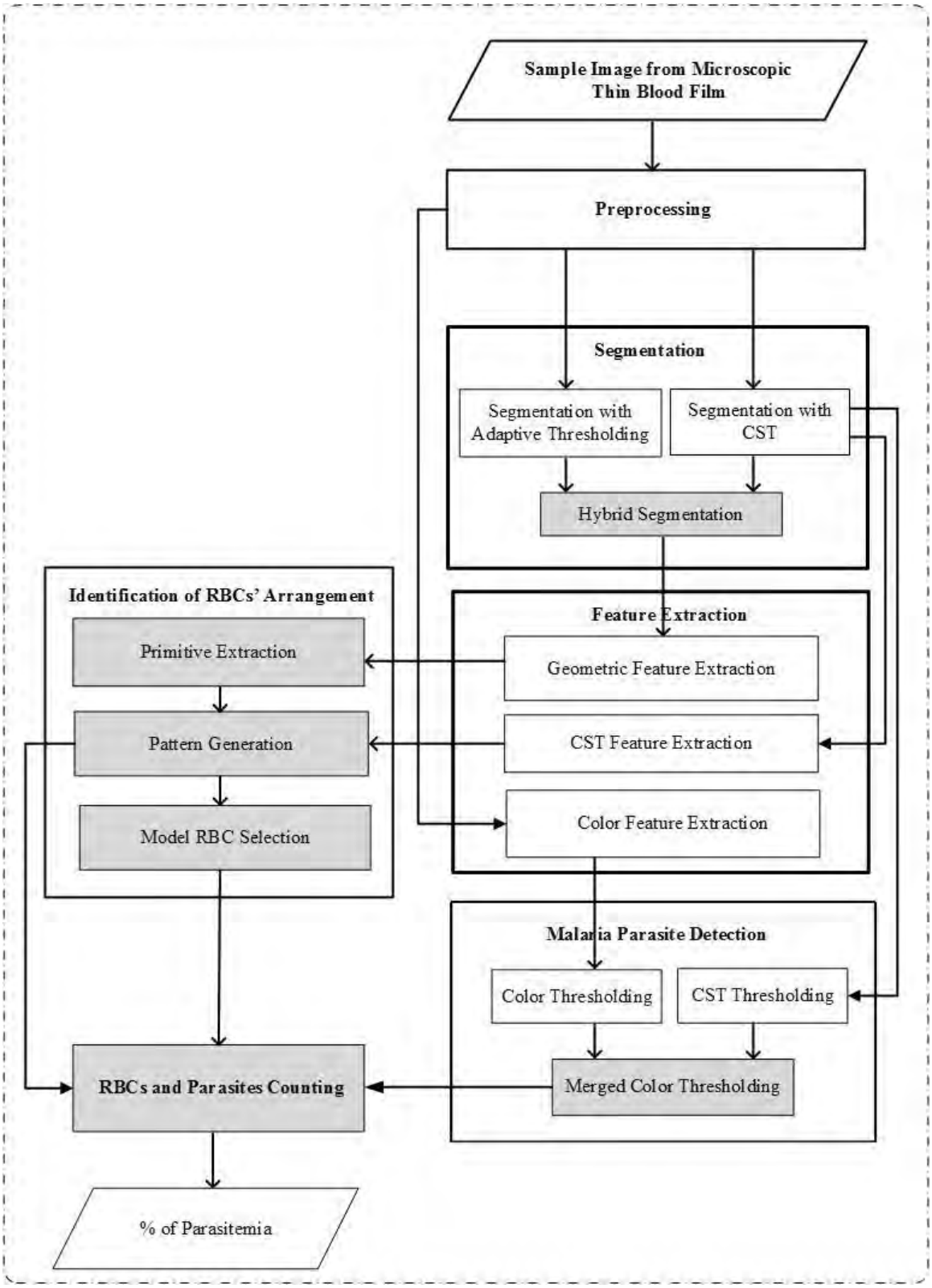
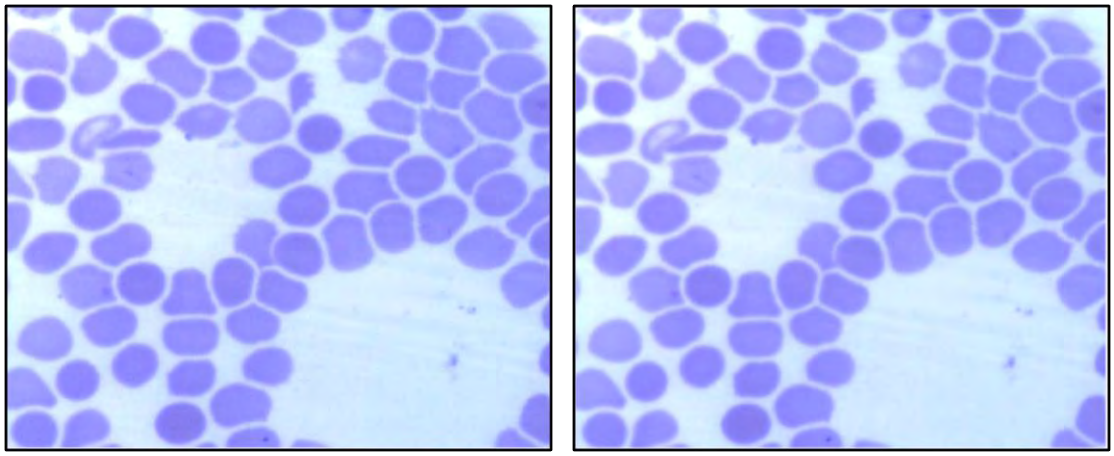


Figure 4.1: The general architecture of proposed system

4.3 Preprocessing

After image is acquired using camera mounted microscope, it is essential to perform preprocessing. The aim of pre-processing in the proposed architecture is to improve the quality of images of microscopic thin blood films by suppressing unnecessary parts or by enhancing major image features that are important for further processes. Though noise is not a usual challenge, the most frequent noise introduced in microscopic image includes introduction of highlights and specular reflection sometimes introducing paper and salt noise. Median filter is a nonlinear filter used to remove such noisy images. Therefore, median filter in color image has been applied on separate channel and recombined to a single channel.



A B

Figure 4.2: Preprocessing (A) original image (B) preprocessed image

Input: A sample malaria infected microscopic image I
Output: An image free of noise
Get original image I
Separate I in to 3 channel I_R, I_G, I_B
Apply median filter on each channel I_R, I_G, I_B
Combine separate channel I_R, I_G, I_B in to I
Return I

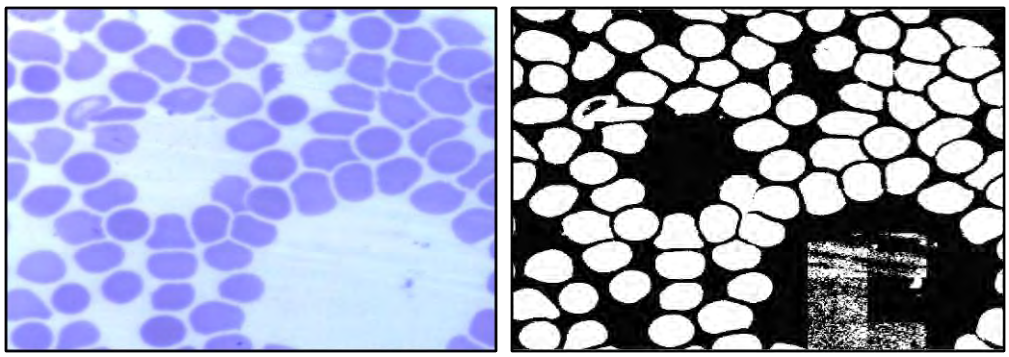
Algorithm 4.*1*: Image preprocessing

4.4 Segmentation

This component is responsible for isolating background from foreground objects and is done after the image is preprocessed. One of the previous challenges while segmenting microscopic image extracted from thin blood films is occurrence of uneven illumination due to existence of light sources emitted from microscope and external sources. As a result, the rate of false detection or wrong analysis will increase due to existence of incomplete cells such as wrong detection of RBC as sickled one. In the proposed architecture, the segmentation is employed by integrating adaptive threshold and color structure tensor threshold.

4.4.1 Segmentation with Adaptive Thresholding

Adaptive thresholding realizes the implementation of adaptive segmentation and is capable to overcome the effect of illumination. There are several methods to compute adaptive thresholding. With acknowledgment to the Wellner's adaptive thresholding, in which a threshold that is varied across the image relative to the local mean, or median, at that point in the image is selected and used to determine whether a pixel should be part of the foreground pixel or not, the study has applied it. However, adaptive segmentation alone sometimes poses introduction of new objects as shown in Figure 4.3. As described in Algorithm 4.2, adaptive threshold is selected for each pixel with a specified window size. Since RBCs are closely localized and sometimes sparsely populated, a 32 pixel window size could be chosen to isolate foreground from background. Hence, the window size and the threshold being selected, the mean of pixel’s intensity value have been applied to render better result.



A B

Figure 4.3: The result of segmentation using adaptive thresholding

```
Input: Preprocessed microscopic binary image I
Output: Foreground pixels

    For each pixel location J  in I
         Get  adaptive threshold T1
            If    I(J) < T1
                  I(J)=0;
              Else
                   I(J)=1;
            End
    End
    Return I
```

Algorithm 4.2: Segmentation using adaptive thresholding

4.4.2 Segmentation with CST

Color structure tensor is carried out to amplify linear structure in color images. The significant role of color structure tensor is to measure linear symmetry and to model the local structure in a color image. It amplifies linear structure in color images. Since edges are characterized by high frequency, linear structure is commonly found there and could be realized through color structure tensor. CST can be implemented by filtering of the orientation tensor.

Figure 4.4 shows the result obtained from color structure tensor and comparison made between the two eigenvalues where (A) is the original image, (B) is the first eigenvalue, (C) is the second eigenvalue and (D) is the difference of the two eigenvalue. In (D), a clean and visible information can be seen in relation to (B) and (C).

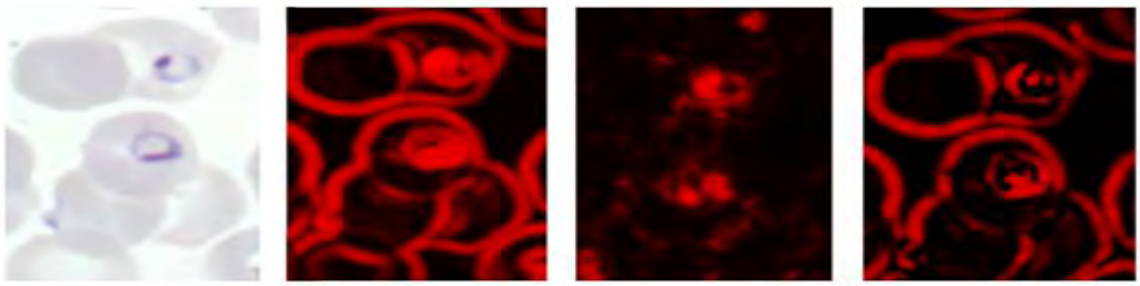
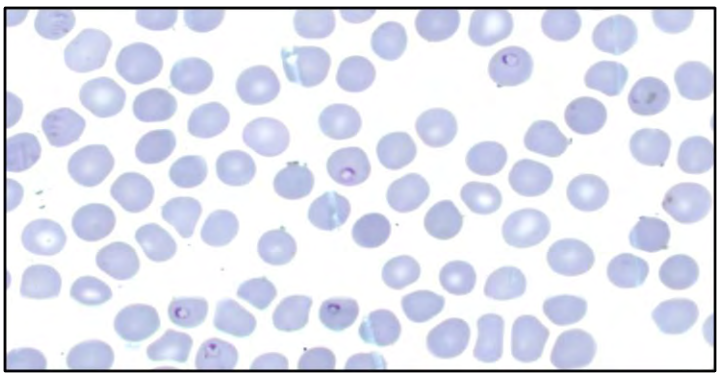


Figure 4.4: Image of sample blood and eigenvalues in RGB color space

An ideal linear symmetry occurs in an image when the two eigenvalues are characterized by being attaining minimum and maximum threshold. Hence, in Figure 4.4 (D) ideal linear symmetry is attained in the case where there is a single and well-identified direction of isocurves. Equivalently, noise introduced in I_{20} can be suppressed by applying pixel-wise operation on D as follows:

$$I_{20\_modified} = (\lambda 2 - \lambda 1).* I_{20} \quad (11)$$

where I_{20} has been computed from Equation (9) and implemented using Algorithm 4.5, λ_1 and λ_2 represents the eigenvalues computed from Equation (4) and Equation (5), respectively and implemented using Algorithm 4.5. Consider Figure 4.5 which is affected by specular highlights and occurrence of tiny microscopic objects. The result obtained from color structure tensor namely, I_{20} and its magnitude, is shown in Figure 4.5.



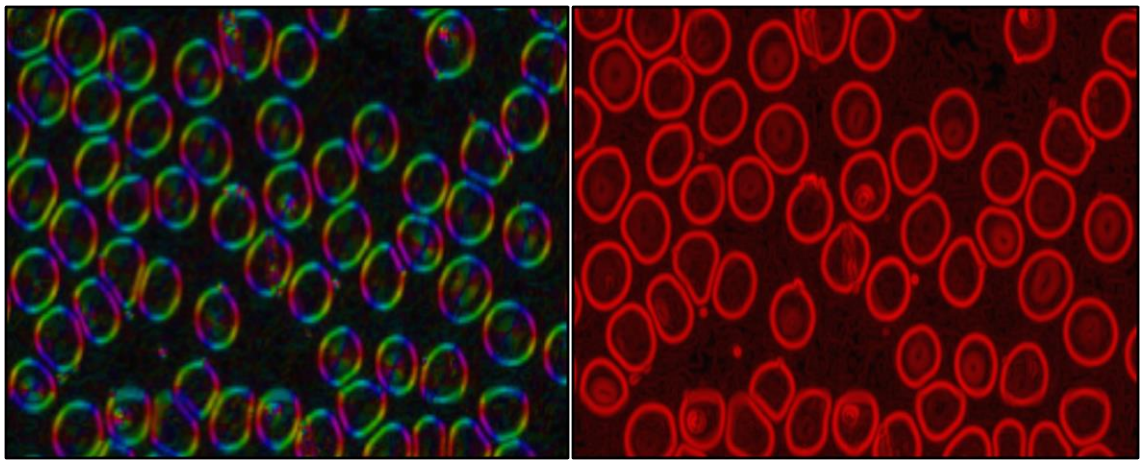
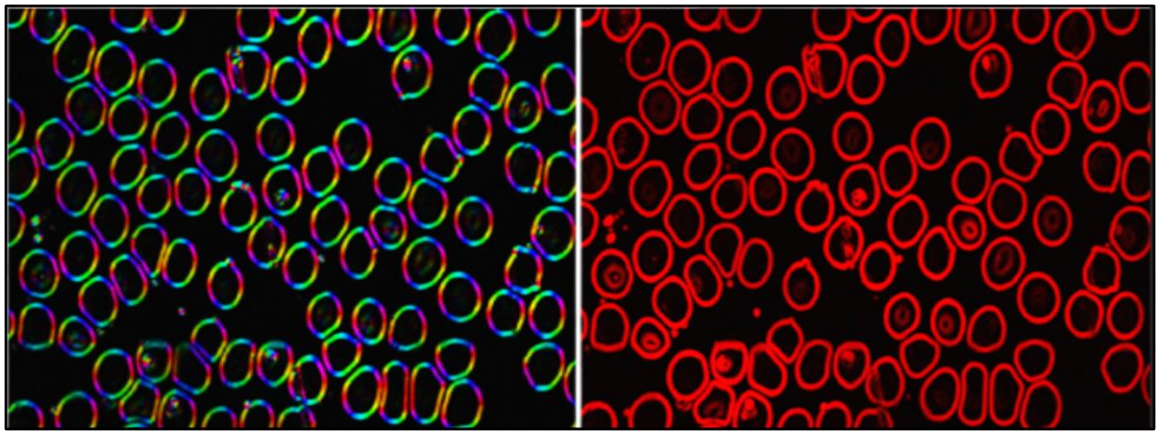


Figure 4.5: Plasmodium falciparum positive slide

Figure 4.6: (A) The result of complex I_{20} in RGB (B) Magnitude of I_{20} in RGB

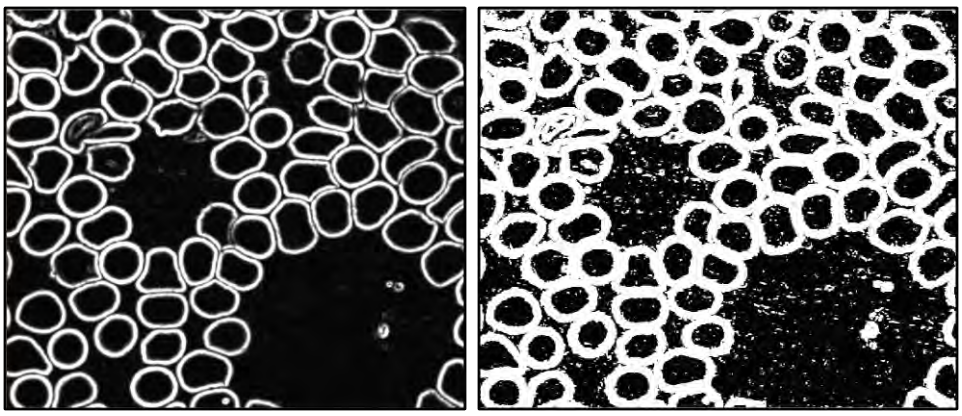
Pseudo code generation to suppress blur introduced in the above scenario is presented as follows. In Figure 4.6, the derivative energy in the prominent orientation is not corrected for by energy contributed by noise λ_2. On the other hand, Figure 4.7 is obtained after applying ideal linear symmetry explained in Equation (11).



A B

Figure 4.7: (A) The result of complex I_{20} after suppressing noise introduced by lowest eigenvalue and (B) Magnitude information

Since linear structure in background region is poor as compared to the foreground object, color structure tensor's threshold should also be selected to overcome the pitfalls from adaptive threshold. The result of the normalized magnitude of I_{20} as shown in Figure 4.8 (A) is used to select a threshold. As a result, background of the image and pixels around the hole region are less amplified as compared to pixels near edges. This property enable us to select a threshold to isolate edges with background and pixels around hole region. In Figure 4.8 (B) a threshold T = 0.01 is used to isolate edge pixels from background and pixels around hole region.

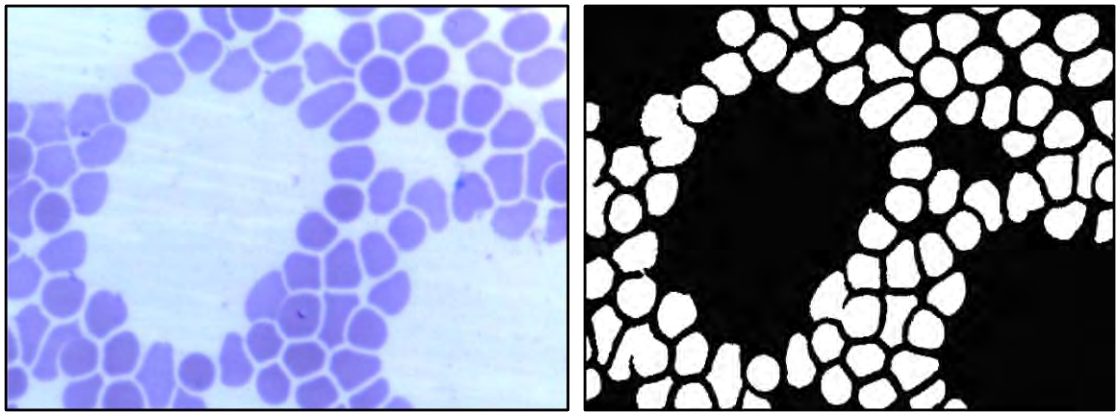


A B

Figure 4.8: The result of segmentation using CST thresholding

4.4.3 Hybrid Segmentation

The relevance of hybrid segmentation is to remove unpredictable objects created after adaptive segmentation is done, and therefore, a threshold is selected using normalized magnitude of I_{20}. Besides, further morphological filling followed by removal of tiny objects using morphological erosion and dilation operations is crucial steps to produce better result. Figure 4.9 (A) shows original image, and (B) shows the integration of adaptive and color structure tensor segmentation. Algorithm 4.3 shows hybrid segmentation obtained after applying CST threshold on the result from adaptive segmentation.



A B

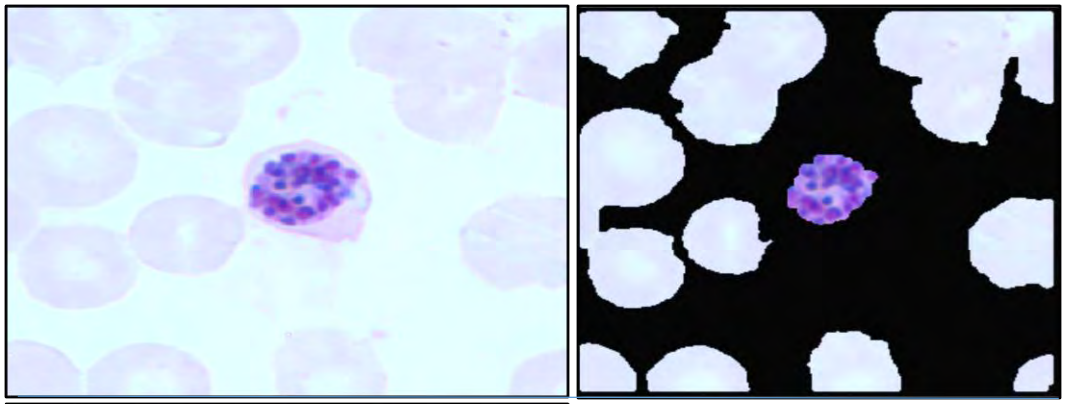
Figure 4.9: Result of adaptive segmentation after applying hybrid thresholds

```
Input:  The result of adaptive segmentation I and
normalized magnitude of I20 Mag
Output: Hybrid segmentation

Get a threshold T2 that is used by CST segmentation
     For each pixel location J  in I
              If  I(J) > 0 and Mag(J) < T2
                  I(J)=0;
              End
     End
     Apply morphological filling on I
     Remove small objects
     Return I
```

Algorithm 4.3: Hybrid segmentation

A



B

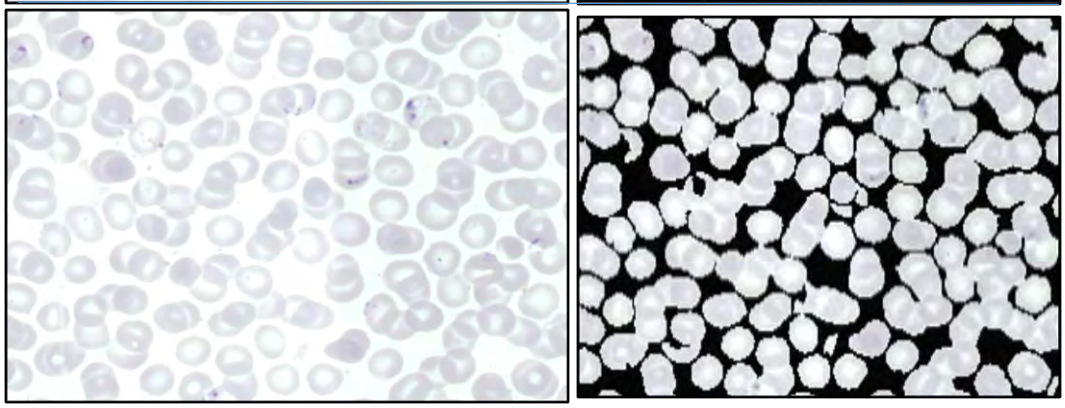
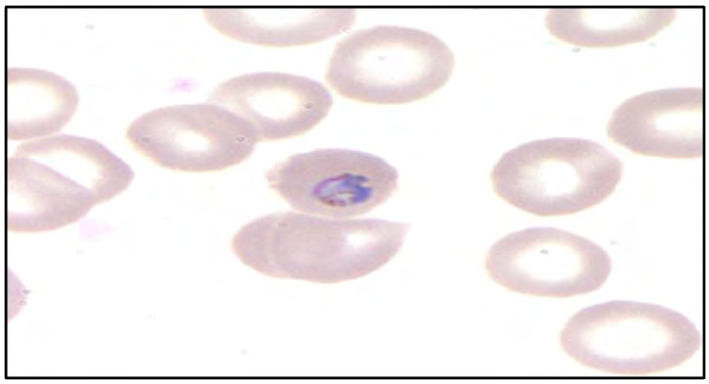
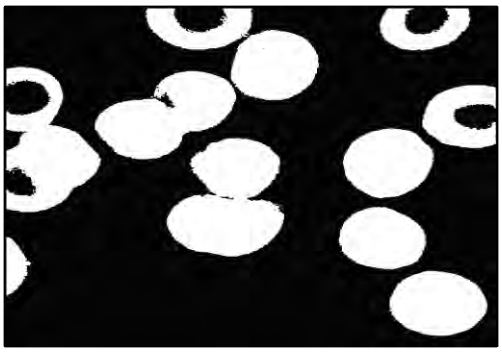


Figure 4.10: Examples of isolation of foreground object and background from poorly illuminated image using adaptive and color structure tensor thresholding

Besides, treating RBCs lying around the border of the scene was also challenging. Algorithm 4.4 is the solution for the challenge. Here, the objective is to retain RBCs lying around the border of the image from appearing as sickled. These RBCs are not really sickled, but they appear to be sickled because the color of their hole region is similar to that of the background making threshold selection difficult. Therefore, this algorithm could solve such a problem by proper segmentation without making the RBCs appear sickled or broken. Algorithm 4.4 shows proper segmentation of RBCs lied around the border of the image. Figure 4.11 (A) shows original image (B) shows the result of segmentation before Algorithm 4.3 is applied and (C) shows the result of segmentation after Algorithm 4.4 is applied.



A



B

C

Figure 4.11: The result of segmentation before and after applying Treating objects lying on the border of the image

```
Input: RBCs on the border of the image in binary
Output: Properly segmented binary image L

Get a threshold T2 that is used by CST segmentation
Assumption: The Labeled Object is not empty
Extract edge from input using canny edge detector: EDG
Chose pad points on EDG 2-pixel inward direction:
Pad_Pts
    For each object Object in EDG
    Chose a starting point SP on Pad_Pts by testing
    zero  value across vertical and horizontal line on
    Object
    Initialize trace_point=0;
    For each point pt on the pad clockwise direction
        If   Object(pt) != 0 and  trace_point==0
```

```
            Trace_point=1;
        Elseif   Object( pt ) != 0 and  trace_point==1

              Trace_point=0;
         End
         If trace_point==1
            Object(SP)=1
         End
        Apply a  morphological fill operation on Object

        Apply a morphological erosion followed by
        dilation  on  Object: L
        Modify object in EDG by  Object
     End
    End
    get I20 from Algorithm 4.5 and modify I20 (Pad_Pts)
    Uniform hue and magnitude
    Return  EDG
```

Algorithm 4.*4*: Treating objects lying on the border of the image

4.5 Feature Extraction

This component is responsible to extracting features that are used to describe normal RBCs and malaria infected RBCs. Besides, discriminant features to identify the state of RBC arrangement is also identified. The most important features used to detect RBCs and malaria infected RBCs are color, shape and others gained from color structure tensor. In this research, contour based and region based feature extractions are used to model RBC arrangement and parasite detection respectively.

4.5.1 Geometric Features Extraction

This sub component is responsible for extracting geometric features of RBCs. Geometric feature is one of the most important features to represent the spatial distribution of pixels. Among the most important geometric features, extreme points and area are used to analyze the morphology of RBCs. Extreme points are used to determine connector or joints when primitives are extracted whereas area feature is used to compare and contrast RBCs at various structural arrangement. These features can be determined as follows.

Area of RBC: determined after RBC image is segmented and can be calculated by counting the number of pixels that are found in a specific labeled red blood cell.

Extreme points: specifies the extreme points in the region and can be determined by computing extreme points in the top-right, top-left, left-top, left-right, bottom-left, bottom-right, right-top, and right-bottom of segmented object. Those points are determined based on Algorithm 4.9.

4.5.2 CST Feature Extraction

This component is responsible for extracting six features from color structure tensor namely, magnitude of complex tensor, orientation, sign magnitude, discontinuity, eigenvalues λ_1 and λ_2 and invariant Harris corner. The linear strength or magnitude of regions occupied by Plasmodium parasites is stronger than others (RBCs and background) do. Hence, the detection of parasites is based on color of parasites and magnitude of $\boldsymbol{I_{20}}$. Besides, orientation, sign magnitude, discontinuity, eigenvalues λ_1 and λ_2 and invariant Harris corners are used to model red blood cell arrangement so as to improve segmentation of overlapped RBCs. Algorithm 4.5 [23, 24, 26] computes $\boldsymbol{I_{20}}$, orientation, magnitude and eigenvalues based on Equation 4 – Equation 9.

```
Input: Microscopic color image
Output: Orientation, I20, λ1, λ2, magnitude
    Extract the three channel from input color image
    :R, G ,B
    Extract spatial derivative with respect to x and y:
    Rx, Gx, Bx, Ry, Gy, By applying 2 1-D integrative
    filters
    Compute Cxy = 2*(Rx.*Ry + Gx.*Gy + Bx.*By);
    Compute Cxx = Rx.*Rx + Gx.*Gx + Bx.*Bx;
    Compute Cyy=Ry.*Ry + Gy.*Gy + By.*By;
    Apply averaging on cxx, Cyy, and Cxy with large
    Gaussian to generate I20
    I20= Cxx - Cyy + j.*Cxy; j=sqrt(-1)
    Orientation = .5*atan2(Cxy, Cxx - Cyy);
```

```
D=sqrt((Cxx - Cyy).^2 + Cxy.^2 + eps);
λ1=Cxx + Cyy + D;
λ2=Cxx + Cyy - D;
Return Orientation, I20, λ1, λ2, magnitude
```

Algorithm 4.5: Extraction of features from color structure tensor

Orientation: In color context orientation determines the direction to the dominant channel which is obtained by averaging it and would be employed by color structure tensor to avoid cancelation effect imposed by gradient direction. As it is shown in Figure 4.15, RBC is represented using oriented patterns where edges are represented using RGB. The RGB value is obtained by computing orientation at each pixel location based on Equation 6 followed by modulating it using π radian to give the direction, and the value component is obtained by computing the magnitude of I_{20} at that particular pixel location leaving saturation maximum. Hence, each cell has four unique opposite edges. Algorithm 4.6 implements orientation using color structure tensor.

```
Input: A labeled binary RBCs
Output: Oriented edges (curves)
    Get Orientation from color structure tensor: Orient
    Get labeled microscopic Binary image: I
    For each object in I(j),j=1:N, N=Total labeled
        number
        Binary_edge = edge(I(j));
        Oriented_Edge=Orient.* Binary_edge;
    End
    Return  Oriented_Edge;
```

Algorithm 4.6: Extraction of oriented edges

Sign Magnitude: This feature considers the degree of positivity and negativity in a given oriented curves. This characteristic is an important discriminant feature because RBCs that come alone without in aggregate form constituting such character unless and otherwise it is affected by distortion. Hence, positive sign magnitude implies the ratio of the number of positively oriented pixels from the whole oriented pixels in a given curve.

$$SM^{+} = NS^{+} / (NS^{+} + NS^{-})$$
$$SM^{-} = NS^{-} / (NS^{+} + NS^{-}) \quad (11)$$

where NS represents number of signs, the + sign indicates positively oriented pixel and the – sign indicates negatively oriented pixel.

The following algorithm is used to compute sign magnitude in a given oriented edge (curve)

```
Input: Oriented Edge
Output: Sign Magnitude
    Get oriented edge using Algorithm 4.5: OE
    Initialize  SM+= SM- = 0, the number of positive and
      negative sign is assigned a value of zero
      initially
    For each pixel value in  OE (j),j=1:N,N=Total
        number of oriented pixels
         If OE >=0
               SM+ =  SM+ + 1;
            Else
               SM- =  SM- + 1;
         End
    End
    SM+ = SM+/( SM+ +  SM-);
    SM- = SM-/( SM- +  SM+);
    Return  SM-, SM+
```

Algorithm 4.7: Computation of sign magnitude

Discontinuity: Discontinuity **i**s a measure of how much oriented patterns are consistent to its values in the curve. It can be directly computed from the bounding box operation performed on each extreme points defined on the curve. Occurrence of discontinuity is the sign of irregular curves; the case in which overlapping is identified. Detail explanation on how to compute discontinuity is given in pattern generation component.

Magnitude of the complex tensor: The magnitude information enables to detect existence of parasite inside RBC. Magnitude value is always positive and it is a measure of how much strong is the value of intensity. Malaria parasite relative to RBC is

characterized by strong magnitude. Hence, magnitude is a good discriminant features to detect parasite. Algorithm 4.5 shows how magnitude of complex tensor is computed.

Invariant Harris Corner Detector: This feature is one of the most important features to identify corners in a given color image. Corner detector algorithm is applied to identify turning points while isolating overlapped cells. Eigenvalues are employed to detect corner points as described in Equation 10.

The feature space therefore is composed of the following attributes: Orientation, Sign magnitude, magnitude of complex tensor, discontinuity, eigenvalues, area, extreme points, harris corners, hue, saturation, and value.

4.5.3 Color Feature Extraction

Color is one of a descriptive measure of Plasmodium parasite detection. As described in Chapter Two, the HSV color model is the other most commonly used color model. In this model, color is described by three components: hue, saturation and value (intensity). The colors of the RGB space are usually not easy for humans to interpret. However, the hue, saturation and value space, HSV color space is, by contrast, intuitive. Hue is an attribute associated with the dominant pure color such as pure blue, pure red, etc. Saturation is the amount of white light that is mixed with a hue while intensity (value) is defined as a measure of the brightness of light. By this, it is meant that saturation measures the quality of pure color. The color features for Plasmodium parasites are obtained from HSV.

4.6 Malaria Parasite Detection

This component is responsible for separating parasite from RBCs. Malaria parasite detection is achieved by using color features extracted from preprocessed image and a threshold selected from color structure tensor. Besides, the result of segmentation is used as an input to this component.

4.6.1 Color Thresholding

This sub component is used to select color thresholds after color is extracted from the preprocessed image. Plasmodium parasites have a range of color in HSV. Hue and saturation components are selected to determine occurrence of parasites. For plasmodium falciparum and vivax a range of hue between 0.72 and 0.79, a range of saturation between

0.18 and 0.66 are used as thresholds for color. These thresholds are obtained after several experimentation on the two Plasmodium species at different life stages. However, these thresholds are not always enough to detect plasmodium parasites due to the existence of RBCs and other blood components that have similar threshold value. Hence, the integration of color and CST thresholds are used to offer better results.

4.6.2 CST Thresholding

The linear structure in regions where parasites are localized is strong. Magnitude of $\boldsymbol{I_{20}}$ is stronger for parasites than other objects as shown in Figure 4.12, and Figure 4.13 shows the direction of pixels in the dominant orientation. The arrow length tells us about the magnitude and the direction of the vectors represents the change in hue. In Figure 4.13 (A), the magnitude of $\boldsymbol{I_{20}}$ is too weak which is an indication of background pixels.

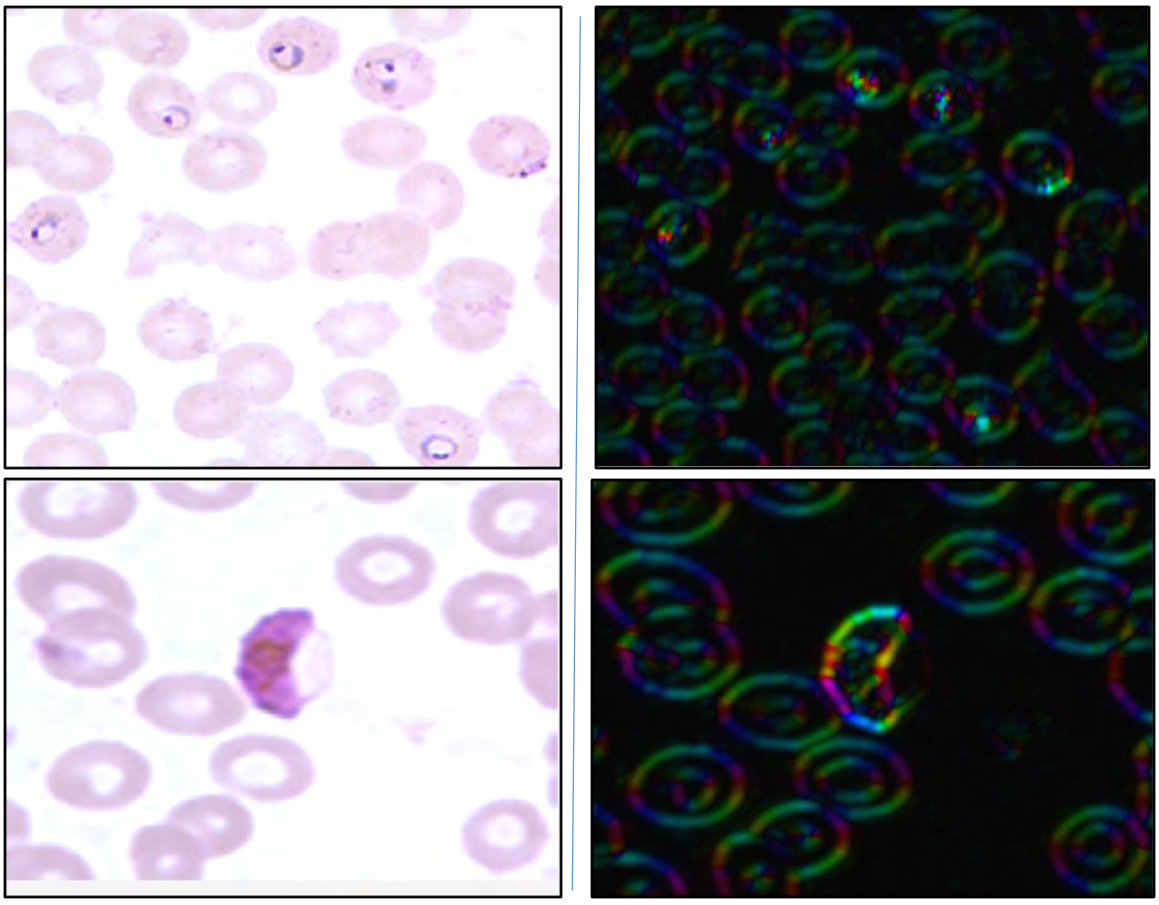


Figure 4.12: Visual detection of parasites in color structure tensor (A) original image and (B) $\boldsymbol{I_{20}}$ in RGB

In (B), the magnitude of I_{20} is better than (A) but still weak as compared to the magnitude of I_{20} in C which is stronger and an indication of Plasmodium parasites. The smallest eigenvalue is also capable to visualize parasites with strong magnitude than other objects. However, if the color of the parasite is affected by poor illumination, the magnitude will become weaker.

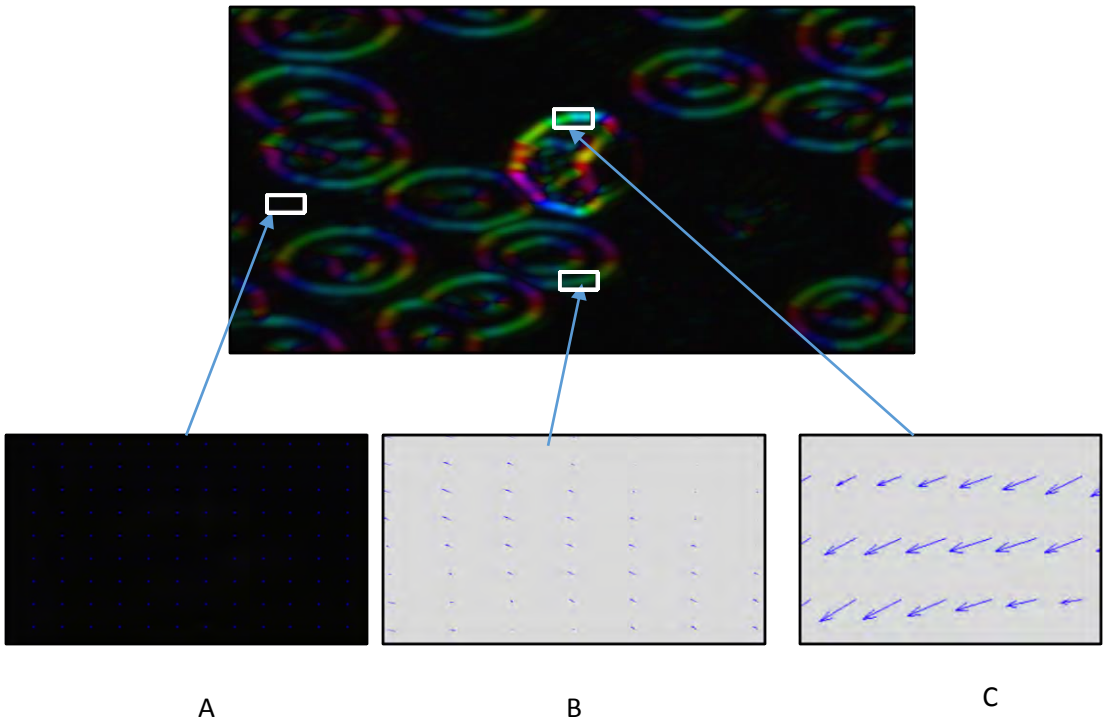


Figure 4.13: The degree of magnitude strength of I_{20} at different pixel location

4.6.3 Merged Color Thresholding

This component is responsible for merging the threshold obtained from color thresholding and CST thresholding as shown in Algorithm 4.8.

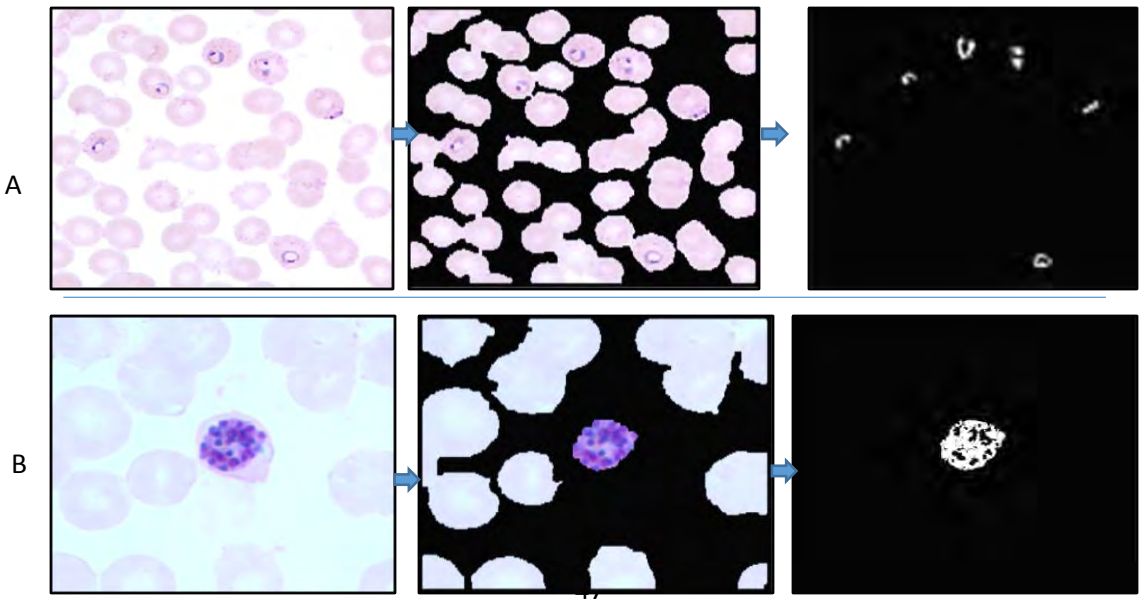
```
Input:  Foreground pixels (FG)
Output: Parasite(PAR)

Get normalized magnitude threshold : MAGmin, Mamax
Get range of color threshold from Hue and Saturation of
Plasmodium parasite microscopic image : T41,T42,T43,T44
Where T41= hue min,T42= hue max,T43=saturation min and
T44=saturation max
```

```
Get empty instance of RBC from FG  :PAR
    For each pixel J  in FG
       If((Hue(J) > T41 && Hue(J) <  MAGmin  &&
          T42))&&((magnitude(J) >  MAGmin &&  magnitude(J)
           Saturation(J) > T43 &&  Saturation(J) > T44))
                      PAR( J)=1
       End
    End
    Label each parasite
    Apply morphological dilation followed by erosion
    Compute area A from PAR
    Remove objects A <= 10 pixels
    Return PAR
```

Algorithm 4.8: Isolation of parasite from foreground pixels

Figure 4.14 is the result of segmentation from relatively unevenly illuminated positive sample images while detecting Plasmodium falciparum ring stage, schizont stage, and gametocyte stage. The detection process as discussed in the proposed architecture passed through three process: preprocessing, isolation of foreground and background and isolation of RBC and parasites. Similarly the result of segmentation from positive sample test images for Plasmodium vivax species is presented in Annex B.



C

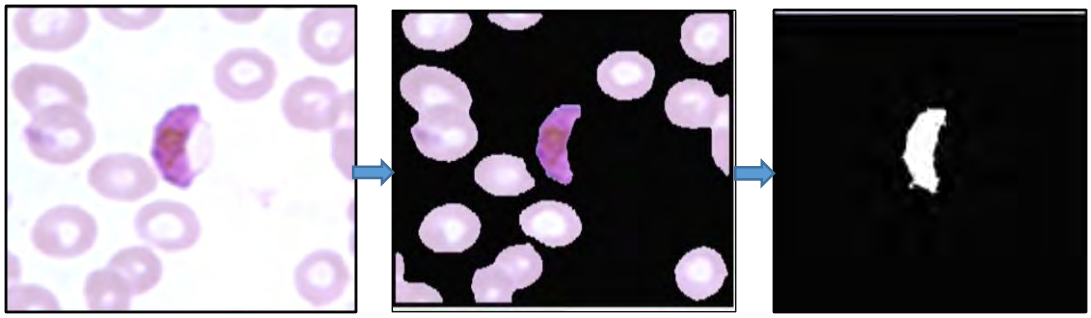


Figure 4.14: Detection of plasmodium falciparum (A) Ring stage (B) Schzont stage (C) Gametocyte stage

4.7 Identification of RBCs' Arrangement

This component is responsible for modeling arrangement of red blood cells by using the result obtained from feature extraction component. As it is shown in Figure 4.15, the oriented patterns are obtained from the complex $\boldsymbol{I_{20}}$. In Figure 4.15 (C) the fact that the color perceived in RGB for opposite edges look symmetric is because the displaying function modulate the hue of the screen pixels by angle obtained from $\boldsymbol{I_{20}}$ linearly so that the range [0,PI] corresponds, to hues defined by the CIE color standard, with zero representing the red color. Saturation of the screen elements is put to 1. Then, the value of screen elements is modulated by magnitude of $\boldsymbol{I_{20}}$.

Oriented patterns are usually found in symmetric image. The symmetric structure of the edges in an image brings a good opportunity for determining the way the objects are arranged. Hence, we exploit it for modeling arrangement. Edges from any arrangement is formed from the so called primitives known as curves and/or lines. A curve in mathematics is, generally speaking, an object similar to a line but that need not be straight. Thus, a line is a special case of curve (null curvature).

Therefore, in this research, RBCs or its state of arrangement is modelled by integrating the variation of hue observed on the edge, estimation of area or size of RBCs, and the behavior of primitives within a particular arrangement.

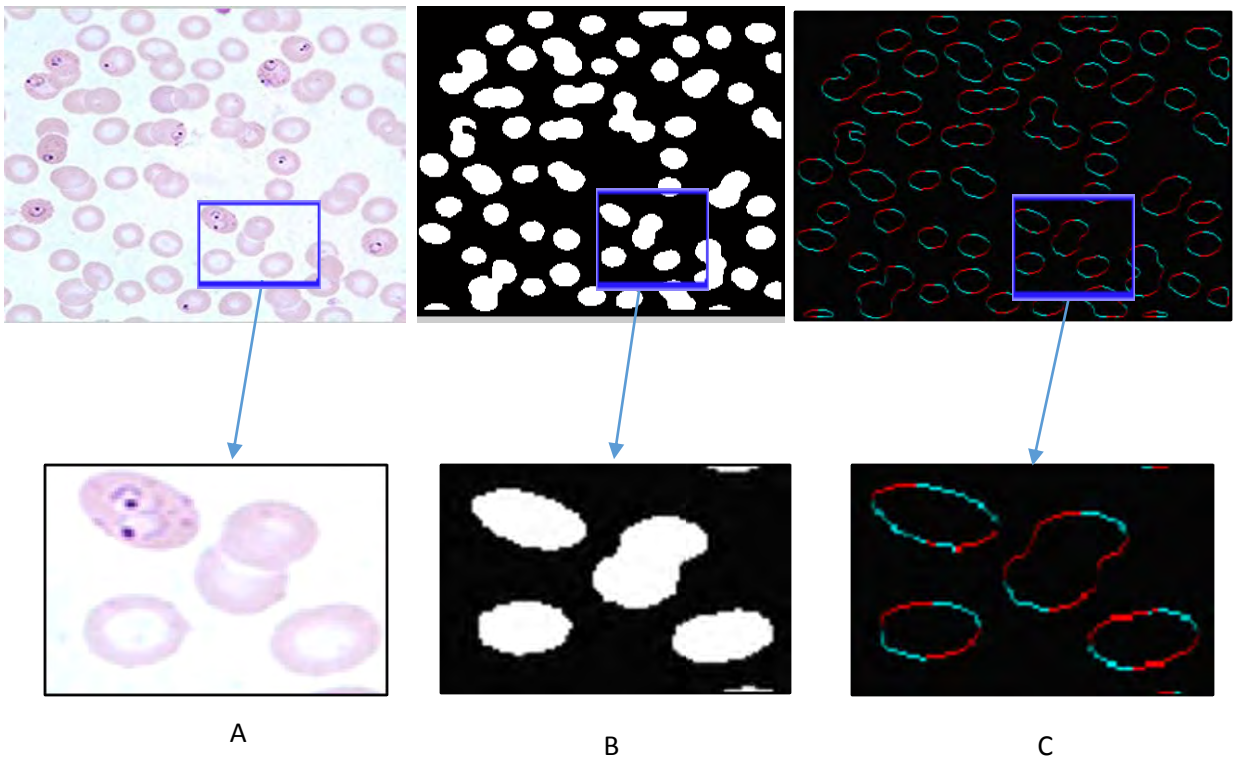


Figure 4.15: Structural representation of RBC using oriented curve (A) Original image (B) binary representation (C) Curves using RGB

4.7.1 Primitive Extraction

In this component structural representation of red blood cells that exist in aggregate or single instance form is modeled. Considering structural shape representation, RBC has bi-concave disk shape that are made up of four simple edges or oriented curves relative to the four extreme points. Extreme point in this context defines the geometric point that any object contains with relative to the eight extreme locations. However, in this problem domain we used the mean points of each respective extreme points in the left, top, right and bottom points consequently the number reduced to four extreme points. Any edges of RBCs in a given arrangement may have therefore, the following elements as a primitive.

Figure 4.16: Possible primitives in RBCs' arrangement

The four sides of RBCs therefore, can have either combination of (A) and connectors (C), or combination of (B) and connectors (C) or the combination of the fourth primitive (D) and connector (C) or the combination of (A, B) and connector (C) or combination of (A, D) and connector (C) or combination of (D, B) and connector (C) or combination of all items. The non-connector primitives have the following major features to discriminate its arrangement type: length, orientation, and eccentricity.

The morphological structure of RBC will be affected by malaria parasite infection. Consequently, the size may increase, the edge may become distorted, and part of cell may become nucleated in a sense that its hole is filled by cytoplasm and chromatin. Furthermore, its surface may be filled by small dots known as stippling.

Broken cell occurs while a snapshot is extracted from thin blood films and some parts of the cells reside on the border region of the image. As a result, the cell becomes partially broken and therefore analysis become difficult unless and otherwise proper segmentation is made. The manual approach to consider broken RBCs under the process of counting is based on size proximity. With regard to automatic malaria detection, previous attempts did not consider the issue of considering broken cell for analysis. The challenge that being incomplete and being affected by distortion and noise results in biased estimation in the process of RBC counting.

In this section, three types of arrangement of cell are modeled. A cell in this notion shall be known in terms of collection of only RBC; whereas WBC, platelets and artifacts are not part of the model. Hence, WBC, platelets and artifacts are discarded as foreign materials. RBCs' arrangement deals with the occurrence of RBCs as single, aggregate (overlapped) and/ or broken whether they are affected by parasite or not. Before modeling RBCs' arrangement, it is important to distinguish among the different ways of arrangement as depicted in Figure 4.17. RBCs' arrangement is featured by the shape from outer boundary of the cell. The shape representation is determined by primitives having four edges and corresponding connectors. Each and every edge is made up of a simple curve, complex curve, and/or a sharp line. Therefore, detection of overlapped, broken, or single instance cell is based on features associated with curves.

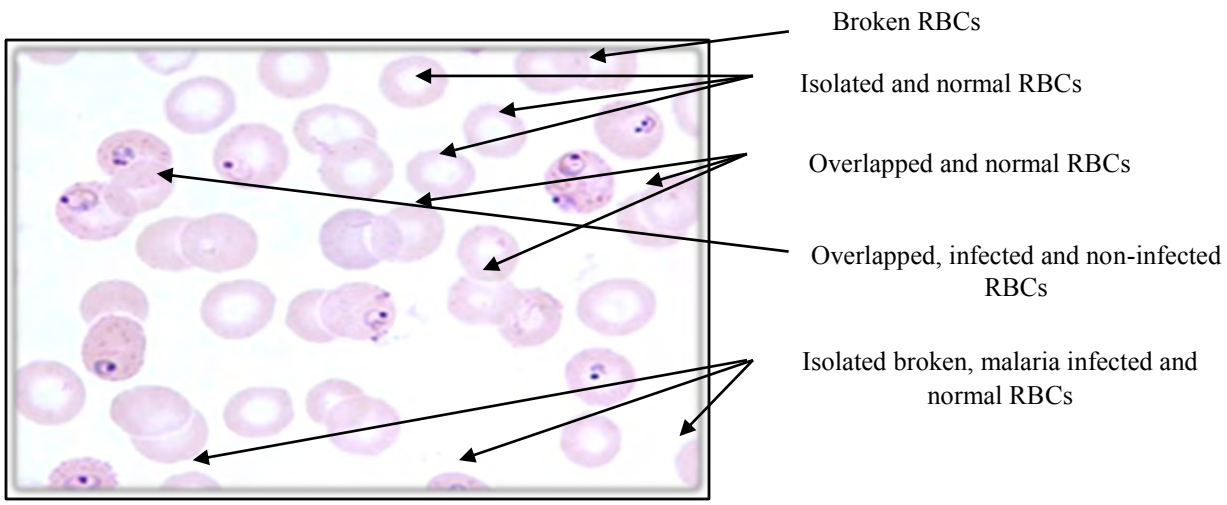
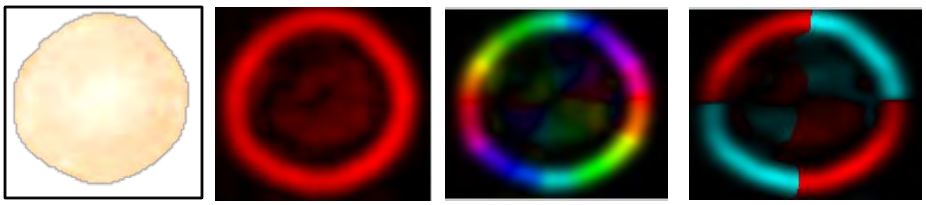


Figure 4.17: Different arrangement of RBCs seen in microscopic image

Single RBC Instance Arrangement

As it is depicted in Figure 4.18, RBCs under single instance arrangement comprises the following key attributes.



A B C D

Figure 4.18: Single instance RBC and features from color structure tensor

In Figure 4.18, (A) shows single instance morphological structure of RBC/Uninfected, (B) shows smoothed version of RBC which is magnitude of $\boldsymbol{I_{20}}$ when modulated by HSV, (C) shows $\boldsymbol{I_{20}}$ which is complex symmetric information computed from color structure tensor when hue is modulated on (0 - 2π) and (D) shows local orientation modulated on (-π/2 - π/2).

In (D), clear hue is observed around the edge of RBC which is the result of orientation from the most prominent direction. Any edge of a given arrangement has four simple or complex curves (depending up on the nature of the cell) and four nodes. RBCs' arrangement determination is, therefore, based on the additive or combined feature set obtained from color structure tensor. A single instance RBC has four simple curves, a

simple curve in this notion deals with a smooth curve with no more than one turning point and it can be characterized by the property that it holds uniform oriented pattern, or the hue in the non-adjacent edges is almost uniform unless and otherwise it is affected by distortion. Besides, when determining the element using bounding rectangle, there is no introduction of common or extra edges or occurrence of discontinuity in the non-adjacent edges. In Figure 4.19 (B) and (C) the four dots with strong intensity value indicate joints whereas the opposite hue in (B) indicate the four primitive edges. Opposing edges are symmetric with π to each other in a sense if one of the edge is positively oriented so does the other opposite edge. This property is a key determinant in modeling arrangement of RBCs. (D), in turn, shows the direction field in the dominant orientation.

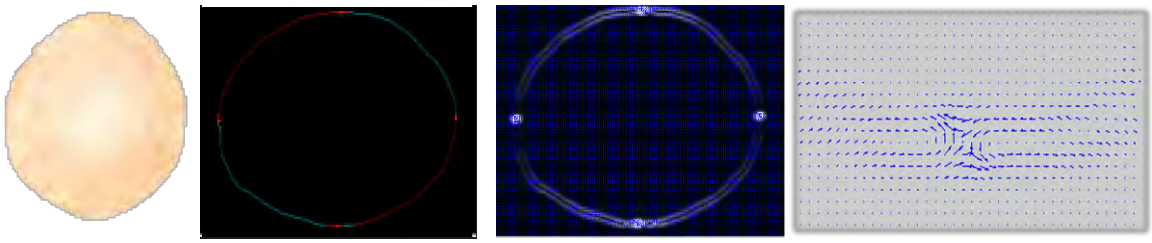
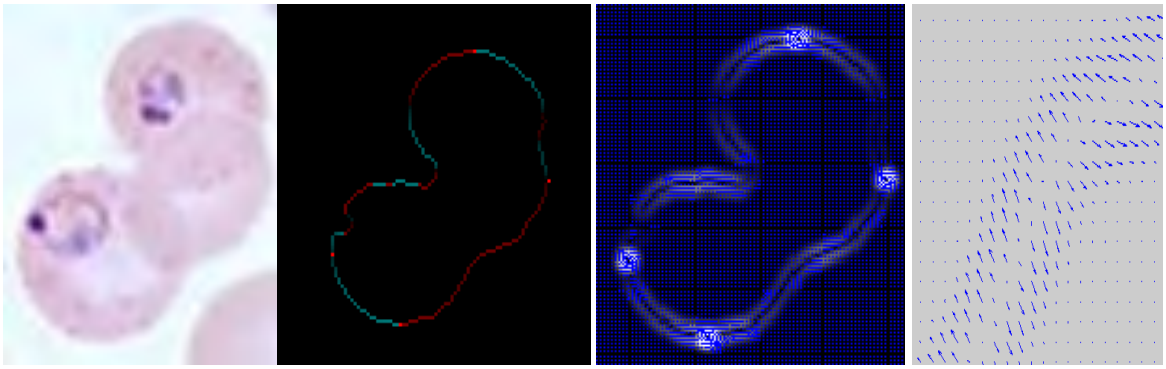


Figure 4.19: Detection of curves and joints in single RBC

Overlapped RBCs' Arrangement

Overlapped RBCs form irregular edges, where each primitive may constitute several turning points. Mixed orientation occurs when edges orient positively and negatively for specific points on the edge. Upon investigating the spatial distribution of each point inside bounding rectangle, discontinuity occurs due to the fact that some points might be outside the bounded rectangle or a new piece of edges which are not part of the edge might get introduced. In Figure 4.20, examples of three overlapped RBCs, hue of the four edges, and the hue vectors in the most prominent direction is shown.



A B C D

Figure 4.20: Detection of curves and joints in single RBC

Broken RBCs' Arrangement

Broken RBCs can be also determined through edge analysis. One of this study's contribution is the treatment of broken RBCs that lie around the border of the image where it has also significant role in determining parasitemia. Therefore, it is considered that while computing percentage of parasitemia. As it is mentioned in Algorithm 4.1, a unique hue (using complex representation) is assigned at the boundary of the whole image so that part of the cells which are broken would be possible for further detection and analysis. Note that arrangement might exist in different form. In this work, the complex type of arrangement such as nested RBCs where RBCs are hidden by other RBCs would be taken as one of the limitations that future works will consider. Following this, we have developed an algorithm to model RBCs arrangement. The area of single RBC is totally determined by the adjustment (setting) made on the microscope. In the proposed system RBCs observation below 64x lenses objective is found to be unreliable for detecting parasite.

```
Input:  RBC Arrangement
Output: Eight primitives(four connectors and four curves)

        Get the binary of RBC arrangement, RBC_Binary
        Extract edges from  RBC_Binary, Edge
        Get oriented patterns of  Edge according to
             Algorithm 4.4, RBC_Oriented
```

```
    Extract the eight extremes points from  Edge
      (Top_left,    Top_right,    Left_top,    Left_bottom,
      Bottom_left, Bottom_right, Right_bottom, Right_top)
    Compute Connectors by computing mid-points of the
     eigh corresponding extremes: Jo,J1,J3,J4
    Extract Boundary points from Edge
    Lo= Spatial Points and intensity value between Jo,J1;
    L1= Spatial Points and  intensity value between J1,J2;
    L2= Spatial Points and intensity value between J2:J3;
    L3= Spatial Points and  intensity value between J3:J0;
  Return Jo,J1, J2, J3, Lo,L1, L2, L3
```

Algorithm 4.9: Generate the four primitives and connectors

4.7.2 Pattern Generation

This component is responsible for categorizing the type of arrangement into three: single instance RBC, overlapped RBCs and RBCs lying on the border. Pattern generation combines the result obtained from primitive extraction and CST feature extraction. Each primitive edge is converted into a curve using orientation features .Then, the curve is analyzed using sign magnitude algorithm, discontinuity algorithm and the algorithm that checks whether part of the curve is broken.

As mentioned in Section 4.6, sign magnitude measures the degree of positivity and negativity in a given oriented curve. Positive sign magnitude implies the proportion of the number of positively oriented pixels out of the whole oriented pixels in a given curve. Sign magnitude for a given oriented edge is determined based on Algorithm 4.7.

Discontinuity is a measure of continuity when boundary points are categorized into primitive edges. The continuity without losing a certain point among a curve is, for example, an indication of part of a non-overlap edge. This is because the four primitive edges are extracted using a bounding box in which case distorted edges become outside of the bounding box. It can be computed by employing finite differences using first order calculus on the spatial information. D= Diff (points) computes the difference among points and put into array D. Hence, occurrence of zero value indicates discontinuity among points. A pattern is generated based on the respective result found from the analyzed curves

and used as a source of information to classify a given arrangement into single instance RBC, overlapped or clustered RBCs and broken RBCs lying on the border of the image. Figure 4.21 shows the process of generating a pattern using curve analysis. The primitives, four connectors and the four edges, would be generated based on Algorithm 4.9. The arrangement patterns are generated based on analyzing curves using sign magnitude, discontinuity and border touch point measure.

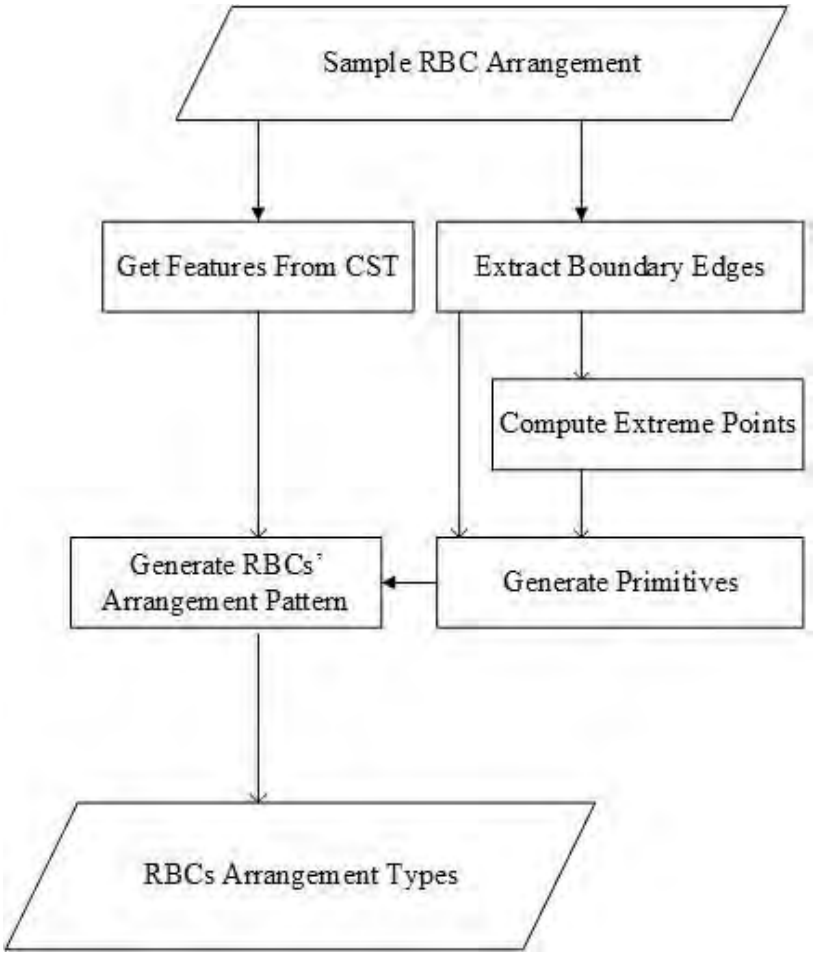


Figure 4.21: A flow chart showing extraction of types of RBCs' arrangement

4.7.3 Model RBC Selection

This sub component does extraction of model RBC from a given arrangement. As discussed in Section 4.8, the selection of representative RBC as a template is based on the result obtained from pattern generation. The structural arrangement of normal single RBC comprises consistent oriented patterns with respect to the four connectors. Hence, the study used this assumption to identify the best representative RBC from the whole arrangement.

However, since there might be cases where a given arrangement constitutes all candidates with complex arrangement, a default normal RBC is used to maintain such cases. The selection of template RBC is done using Algorithm 4.10.

```
Input:  RBC Arrangement
Output: Model RBC

     Get the binary RBC arrangement, RBC_Binary
     Get initial model RBC, init_model_RBC
     If all RBC_Binary is overlapped
        Model RBC = init_model_RBC
     Else
      Previous_SM=0;
      For each labeled arrangement in RBC_Binary
        Initialize sign magnitude: Sum_SM=0;
        Extract primitive curves from  RBC_Binary  using
          Algorithm 4.9:Prim
        For each primitive prim
            Calculate current sign magnitude according to
              Algorithm 4.7:CSM
            Sum_SM = Sum_SM +  CSM;
        End For
        IF  previous_SM <  Sum_SM
            Model_RBC = LArr;
            Previous_SM = Sum_SM;
        End IF
       End FOR
     End IF
     Return  Model_RBC;
```

Algorithm 4.10: Computing selection of model RBC

4.8 Counting RBCs and Parasites

This component performs estimation of number of RBCs and Plasmodium parasites. Counting of RBCs is focused on segmenting overlapped RBCs since counting isolated RBCs is straight forward. Hence, the interest in this component is counting the overlapped RBCs which in turn need further process of segmentation. This component inputs result obtained from pattern generation, model RBC selection and malaria parasite detection. Hence, counting RBCs and parasites is realized through analyzing curves obtained from a template RBC and overlapped RBCs identified from pattern generation. In Figure 4.23, oriented patterns in (D) is used to represent overlapped RBCs in (A).

When arrangement is extracted, the oriented edge in single instance is different from overlapped case. From the human vision point of view, two RBCs are overlapped because two overlapped RBCs contain features associated to size proximity, curvature of edges or occurrence of one or two turning points around the boundary region, occurrence of non-homogeneous region or change of color due to overlap around the center of the region. Complex arrangement also makes oriented edges with various structural arrangement. Figure 4.22 shows overlapped RBCs' arrangement. (A) shows the original fragment in RGB, (B) shows the binary equivalent, (C) shows the edge after canny edge detector is applied and (D) shows the result of orientation while it is displayed in RGB.

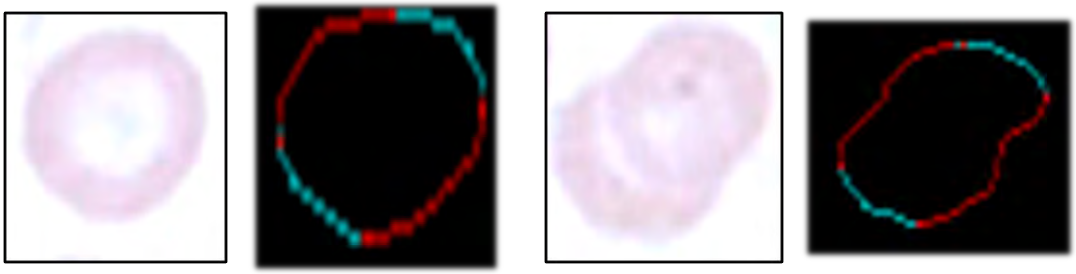
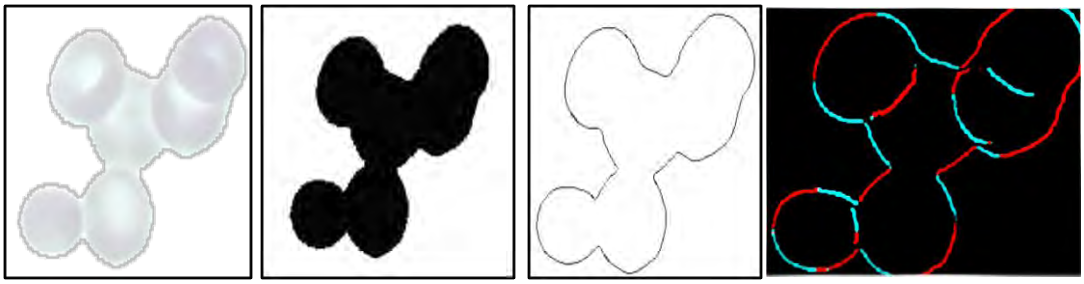


Figure 4.22: Comparison of curves in (A) single instance RBC and (B) two overlapping RBCs using hue of orientation



A B C D

Figure 4.23: Oriented patterns in more than two overlapping RBCs using hue of orientation

Each oriented edge (curve) in the edges of the RBCs' arrangement has features associated with how overlap is determined. Hence, it is possible to estimate the number of overlapping cells through curve analysis. The most important features associated with curve for determining counting has been identified. A curve may contain several corners that are the point at which two overlapped RBCs are identified. Moreover, the symmetric property of the curve that is obtained from orientation can be used to categorize oriented edges in particular RBCs' arrangement.

Every curve in the edges of a given arrangement is profiled in such a way that the geometric centroid, the distance of centroid from non-adjacent or symmetric edges, the length of the curve, are considered.

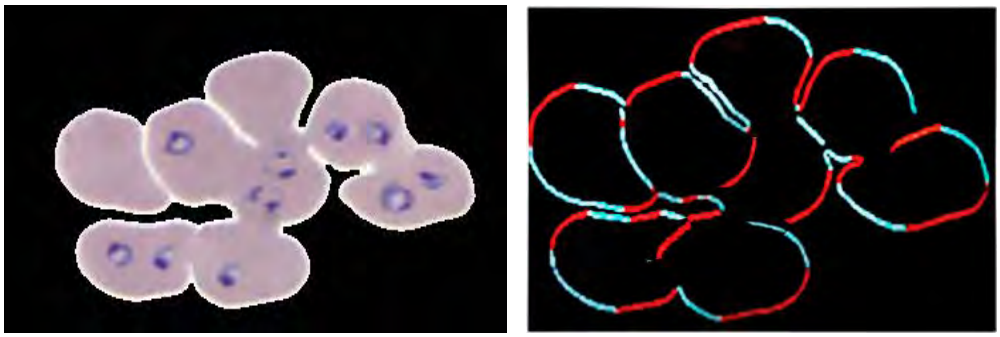
In this component, therefore, model fitting algorithm which is responsible for matching the template obtained from the sub component RBC model selection and pattern generation is developed. Hence, counting RBCs and malaria infected RBCs are based on curve weight estimation that are designed after determining the type of arrangement. Based on the identified arrangement, a single instance RBC has four simple curves namely Lo, L1, L2 and L3. As it has been discussed in Section 4.8, the intensity value in each of the four edges is in either of the two states known as positively or negatively oriented. The four simple curves constitute two types of hue or oriented patterns namely positively or negatively oriented edges. Curves can be analyzed by the hue in the oriented patterns. Based on the arrangement model, determination of number of RBCs in a given overlapped RBCs is possible through analyzing curves in the corresponding edges.

In Appendix C, the algorithm for fitting a model is performed using features extracted from the separated curve and the equivalent feature selected from template RBC. Hence,

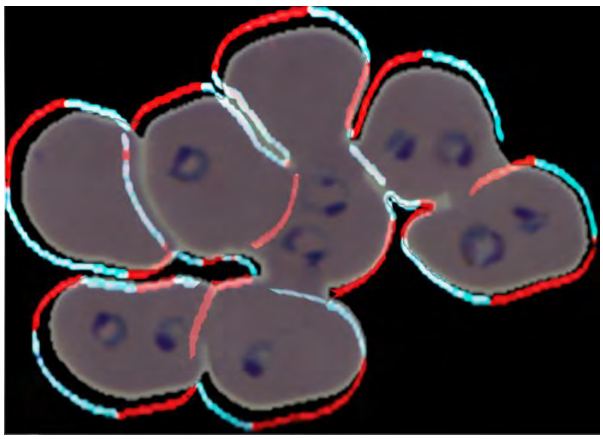
for the comparison to be made between template RBC and curves in the arrangement, the curve length is found to be a good measure to determine the weight. A dynamic threshold that is used to compare the respective curves is chosen based on the curve length determined from model RBC. This has resulted in developing a system which is adaptable to various objective lenses.

Edge extraction is possible through edge detection algorithm. Canny edge detector is applied to extract edges even from the internal region of a given arrangement. Besides, harries corner detectors are simple ways to define corner points or turning points. A minor correction to handle false corners are also determined through invariant harries corner detector. Invariant harries corner detector is one application of color structure tensor as discussed in Chapter Two. It enables the detection of corners that are invariant to specular reflection. The algorithm to compute overlap isolation and determination of counting in a given arrangement is shown in Annex C.

Parasites are also counted on the bases of arrangement pattern. Counting RBCs does not address the status of RBC being affected by Plasmodium parasite or not. A single RBC might be filled by pieces of Plasmodium parasites. This might happen due to the challenge behind accuracy in segmentation or the possibility that more than one parasites invade a single RBC. However, during counting the each of these pieces of parasites should not be counted separately. Hence, parasites are counted as per infected RBC. The reason behind this is that parasitemia is computed on the bases of infected RBC and non-infected one. Our novel algorithm to count Plasmodium parasite addresses counting them in the case of overlapped RBCs and also considers the case when parasites are far apart from each other. The algorithm with some sample data is figured out in Annex C. For clarification, the algorithm could point out counting parasite under the three types of arrangement. It first counts parasites in single instance RBC in such a way that any single instance RBC could be treated as a single infected RBC despite the number of Plasmodium parasites. Similarly, in broken RBCs lying around the border of the image Plasmodium parasites are counted regardless of RBC sizes. However, counting Plasmodium parasites under clustered RBC needs a further step which is taken to be one of notable contribution in this work.



A B



C

Figure 4.24: Counting parasites in overlapped RBCs

Plasmodium parasite are easily detected in the clustered RBCs because the developed algorithm could indicate them through the existence of edges or demarcation lines between the two parasites. This is shown in Figure 4.24. In (A) original image is shown, in (B) the oriented edges used to isolate the overlap is shown using RGB and in (C) the demarcation lines isolating the parasites is shown. We acknowledge the works done by mat works as optimized version of Bresenham line algorithm enabled us to collect spatial points between two arbitrary coordinate points. Hence, Bresnham points between the two parasite centroid is mapped to the oriented curves resulting from counting algorithm. Then, tracing those spatial points gained from Bresnham line algorithm on the oriented curves enabled us to determine whether the two Plasmodium parasites containing the two centroids are part of a single RBC or not. This way, accuracy of parasitemia is optimized. A sample result regarding counting process is shown in Annex D. Parasitemia is the outcome, and it is determined by dividing the number of infected erythrocytes by the total number of erythrocytes in the input image.

4.9 Summary

In this chapter, the design of automatic malaria detection system is discussed in detail. The components that constitute the system, the relationship among these components, and the responsibility of each component are presented. The design is composed of 6 components; preprocessing, segmentation, feature extraction, malaria parasite detection, RBCs arrangement identification and counting RBCs and parasites. In the next chapter the experimentation and its procedure will be presented.

CHAPTER FIVE: EXPERIMENT

5.1 Introduction

Experimental evaluation ensures the realization of the proposed architecture. It is an integral part of the development of any system. This chapter presents evaluation of performance of parasite detection and estimation of the number of RBCs and parasites. To conduct the experiment as a procedure data set elements were assessed and prototype was implemented for evaluation purpose. In the subsequent pages the procedures and the results are discussed.

5.2 Dataset

Sample image data is taken from microscopic thin blood films through Giemsa staining process. This image is acquired through the help of a camera mounted microscope along with computer. In this study therefore, original malaria infected microscopic images is captured from the thin blood smears of Plasmodium vivax and Plasmodium falciparum samples.

Figure 5.1: Required laboratory equipment for image acquisition process

The evaluation was conducted using images collected from different sources as depicted in Table 5.1. The MaMic [13] image database provides thin smear blood sample for identification of Plasmodium vivax. CDC [12] is an educational resource designed for health professionals and laboratory scientists. Images are also captured using Leica DM750 HD Camera mounted microscope at Addis Ababa University, School of Medicine, Department of Pathology. The five slides are taken from Ethiopian Health institution

research center. To generate the images, the slides are examined using 100 X oil immersion objective (1000 times magnification of the object) is used to capture the microscopic image of blood slides. The lens in the objective turret of the microscope is at 100x magnification, and the eyepiece or ocular lens is at 10x magnification. So the 100x objective stands for 100×10=1000-time magnification. 100x objective lenses is used because sometimes Plasmodium parasites at ring stages are very tiny due to the effect of uneven illumination hides part of the parasite which makes parasites less visible and unpredictable.

Table 5.1: Data description

Serial. No	**Source**	**Magnification**	**Spatial resolution**	**Image format**	**Quantity**
1	Leica DM750 HD Camera mounted microscope	100 x and 63 x oil immersion	2048 x1536	jpeg	84
2	CDC [12] Standard image database	100 x oil immersion	400x260	jpeg	42
3	MaMic [13] Standard image database	100 x oil immersion	1050 x 603	jpeg	84

5.3 Implementation

Experiments were carried out using the prototype developed with MATLAB 2014a on AMDA6-3420M APU Radeon(tm) HD graphics 1.5GHz, 4 GB of RAM. A sample image captured from camera mounted microscope and how parasitemia is computed is shown in Figure 5.2 which is the graphical user interface of the developed prototype

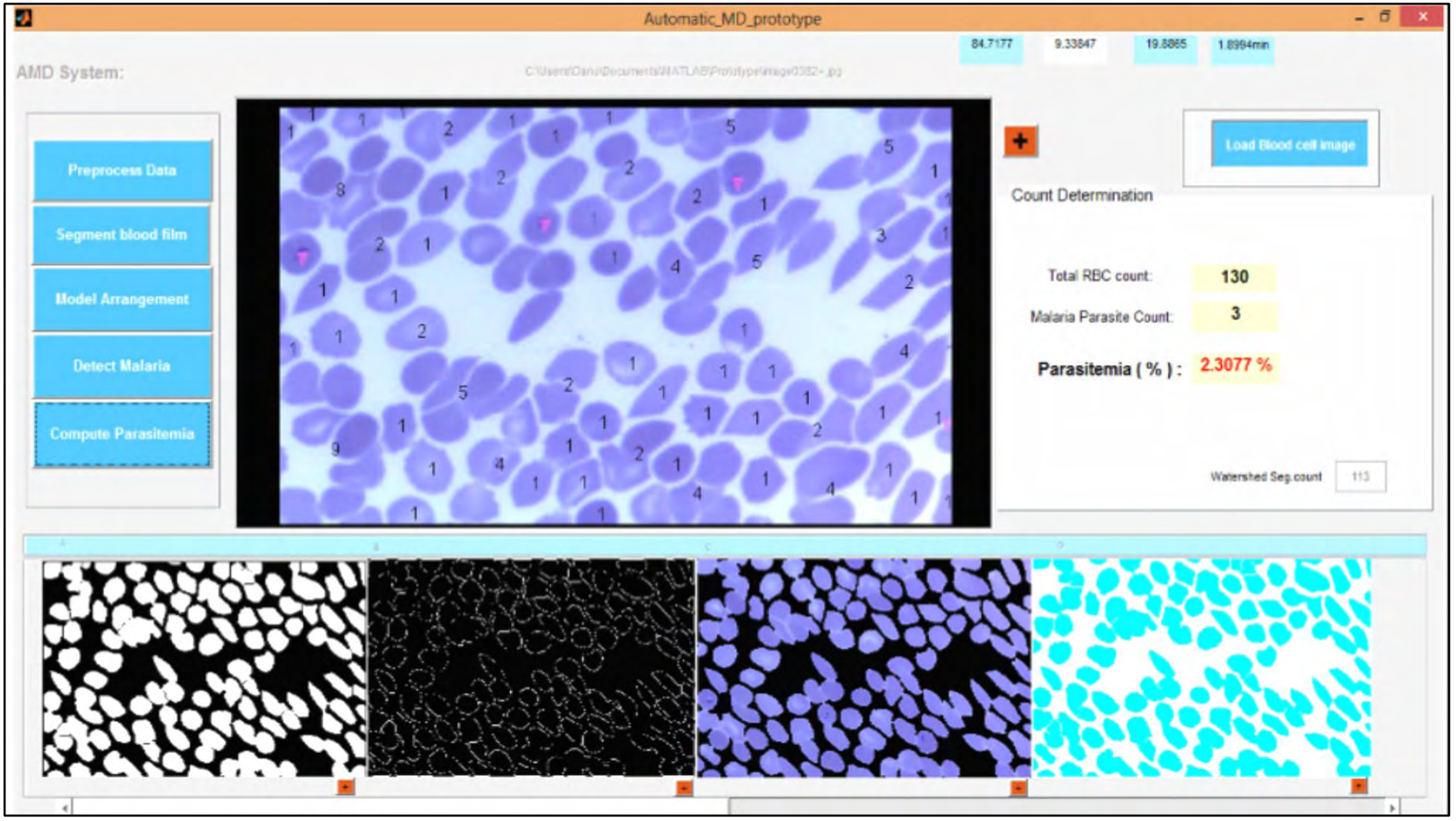


Figure 5.2: Screenshot of the user interface of the developed prototype

5.4 Test Results

Before evaluation of accuracy in Plasmodium parasite detection and parasitemia computation, the whole images are categorized into positive tests and negative tests using expert. As a result, out of the whole population (210 ground truth images), 32 images score negative tests meaning Plasmodium parasites are not observed on the image. The rest 178 images are infected by Plasmodium parasites and score positive tests of which 13 images are found to be gametocyte stage, 8 images are categorized under schizont stage, the rest 157 images are found to be ring stage of Plasmodium falciparum and vivax.

5.4.1 The Result of Accuracy in Plasmodium Parasites Detection

The detection accuracy would be performed by comparing the manual observed data with the automated result since the computer cannot determine automatically which cells are correctly or falsely detected or not detected at all. The truly and falsely detected cells are identified by visual inspection.

Table 5.2: Accuracy of detecting Plasmodium parasites

	Ring Stage	Gametocyte	Schizoints	Total
Number of Infected Images Correctly Detected (TP)	148	12	7	167
Number of Infected Images Wrongly Detected (FN)	9	1	1	11
Number of Non-Infected Images Correctly Detected (TN)	29	0	0	29
Number of Non-Infected Images Wrongly Detected (FP)	3	0	0	3
Sensitivity (%)			94	
Specificity (%)			91	
Accuracy (%)			93	
PPV (%)			98	

The TP (True Positive) and TN (True Negative) represented the number of erythrocytes diagnosed by the diagnosis system correctly either as infected or not infected. The FP (False Positive) represented the number of erythrocytes diagnosed by the system as

infected while as they were actually not infected. Finally, FN (False Negative) represented the number of erythrocytes classified as not infected while they were actually infected.

Sensitivity address the proportion of the samples that are classified as positive among all the positive samples and usually called true detection rate. In contrast to this, specificity measures the proportion of the samples that are classified as negative among all the negative samples.

Positive predictive value (PPV) is the proportion of the positive samples of all that are classified as positive. Negative predictive value is the proportion of the negative samples of all that are classified as negative [35]. Equation (13), (14), (15), (16), and (17) shows how those measures is computed.

$$\text{Sensitivity} = TP/(TP + FN) \tag{13}$$

$$\text{Specificity} = TN/(TN + FP) \tag{14}$$

$$\text{Accuracy} = \text{TP} + \text{TN}/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) \tag{15}$$

$$PPV = TP/(TP + FP) \tag{16}$$

From the results, the overall detection accuracy was 94%. Out of the total 210 erythrocytes images used, only 11 were falsely detected. This therefore, indicates that using the integrated threshold method on color structure tensor and color features extracted from Plasmodium parasite can be an effective tool for detection of Plasmodium parasites thinking that the presence of less quality images such as parasite with junk and aberrant morphology is also under consideration. The algorithm developed in this work has also the advantage of compatibility with poor illumination in images.

5.4.2 Segmentation Result of RBCs and Parasites

The accuracy of segmenting erythrocytes and Plasmodium parasites is done through the help of comparison of manual count with that of automated count in the proposed algorithm. Manual count is done before segmentation in contrast to this, automated count is done after segmentation. Furthermore, the method proposed in this work is compared with watershed segmentation approach using distance transformation. For the experimental evaluation 5 candidate images are used. The selection of samples was based on the criteria that enabled to ensure how the proposed algorithm is accurate estimator of parasitemia. The samples are chosen considering the level of illumination that the image is affected by, degree of overlapping and the variation in parasite presence. Dealing with this, all samples through visual inspection as human visual system adapts change in illumination was chosen.

Though visual inspection is a crucial step in determining the count in clumped cell, false region might also be detected as biasness or concentration are the natural behavior of human being. Hence, our truth value bases on the manual observed data. However, to make it more realistic and more prone to error, the same microscopic image is analyzed by three blood analyst and the average result was considered as a good measure for making comparison between manual and automated parasitemia computation.

Hence, to come up with the main objective which is performance issue, images are selected from the second and third sources (MaMic medical image database and CDC) and from the first source proportionally. The first two images are chosen from the first source data. The second two sample images are taken from MaMic image database, and the last two images are taken from CDC image database.

Sample images along with counting process is shown in the Annex B. As shown in Table 5.3 watershed based segmentation and the proposed algorithm to count RBCs are compared relative to manually computed result.

Table 5.3: Test result on sample images

Serial No	Sample Images	Manual Count Before Segmentation		Automated Count after Segmentation			%Parasitemia		Relative Error for Counting RBCs (%)		Relative Error for Computing Parasitemia (%)
				WSA	PA						
		NRBC	NoP	NRBC	NRBC	NoP	Ground Truth	PA	WSA	PA	
1	GT_img1++	131	3	113	128	3	2.29	2.34	2.23	1.06	2.34
2	GT_img2++	138	5	114	144	5	3.62	3.47	12.44	4.17	4.17
3	CDC_img1++	48	4	34	43	4	8.33	9.30	29.81	8.80	11.63
4	CDC_img2++	51	7	42	49	7	13.73	14.29	17.02	1.75	4.08
5	Mam_img1++	162	12	137	173	12	7.41	6.94	18.37	10.86	6.36
6	Mam_img2++	179	6	153	173	6	3.35	3.47	11.72	4.46	3.47
							Mean (%) Error		**6.11**	**2.07**	**2.14**

Table 5.3 depicts experiment carried out on a collection of 6 images taken from CDC, MaMic and first source data. The double plus sign shows that the test is positive meaning RBCs are infected by Plasmodium parasites. The NoP, NRBC, WSA, PA represents number of Plasmodium parasite, number of RBC, watershed segmentation approach, and proposed approach respectively. Since the computer cannot automatically determine which cells are correctly or falsely detected or not detected at all, the truly and falsely detected cells are identified by visual inspection. The relative percentage error is computed from the mean of the relative percentage error from each fragment in a given input image. A sample

mean percentage error for Mam_img2++ and a sample counting result comparison between watershed segmentation and the proposed counting algorithm is shown in Annex D.

In this new accuracy measure, the counting error is the proportion of the difference between the numbers of algorithm-counted cells, regardless of true or false results, to the true number of cells that are obtained using manual count.

The mean percentage error or relative error (MRE) is computed using Equation (14):

$$MRE = \frac{100\%}{n}\sum_{1}^{n}(x_{ia} - x_{im}) \qquad (14)$$

where n represents the number of samples, x_{ia} represents the i^{th} automatically measured value in the i^{th} input image, and x_{im} represents the i^{th} manually measured value in the i^{th} input image.

As can be seen from the results, the measuring mean percentage error of counting using watershed segmentation is 6.11; whereas the measuring mean absolute error of counting using the proposed algorithm was 2.07. Clearly, the relative error in the proposed counting algorithm is considerably smaller than the other one.

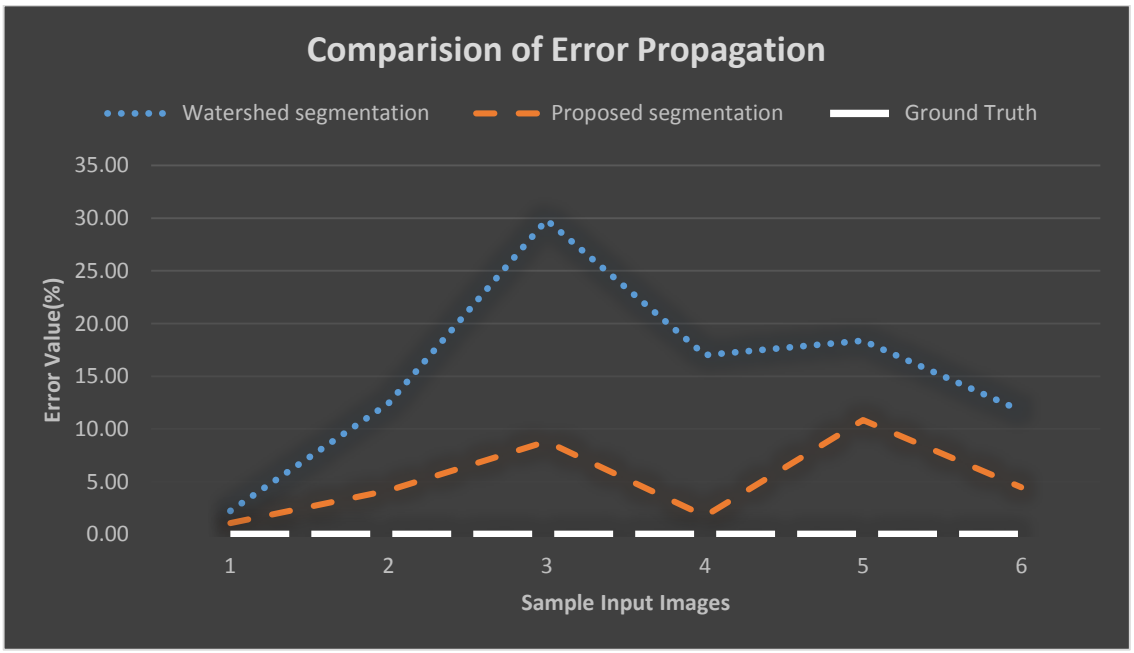


Figure 5.3: Error propagation in counting RBCs

Figure 5.3 shows comparison of error propagation in counting using watershed segmentation and the proposed counting algorithm relative to the ground truth. Dealing with parasitemia computation, Figure 5.4 shows comparison of parasitemia in the proposed approach and the ground truth parasitemia. The result shows that false detection of parasites and over segmentation of RBCs prevail in watershed than the proposed approach.

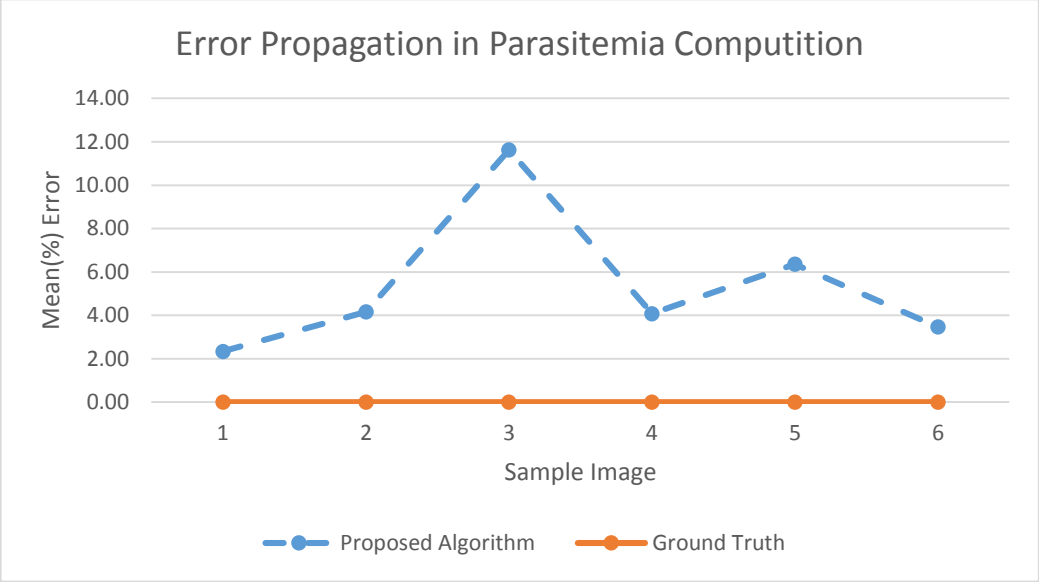


Figure 5.4: Error propagation in parasitemia

5.4.3 Measuring Efficiency of Algorithm to Compute Parasitemia

A test has been conducted to measure the efficiency of the proposed system. This could be realized by testing the time it takes the algorithms to detect Plasmodium parasites and compute parasitemia. The total time the algorithms took is computed as the sum of the time taken by each algorithm to accomplish its task i.e., the time taken to segment RBCs and parasites, the time taken to model RBCs and the time taken to compute parasitemia. 10 sample images having different level of overlap have been used. The experiment indicated that a very short time was needed to deal with images having little or no overlap and the more the degree of the overlap in the image, the longer it takes to accomplish the task. However, on average the time elapsed by the proposed algorithm to accomplish diagnosis is 0.54 minute on the specified machine specification as mentioned in the implementation section. This was achieved even with the presence of poor illumination and complex RBCs' arrangement.

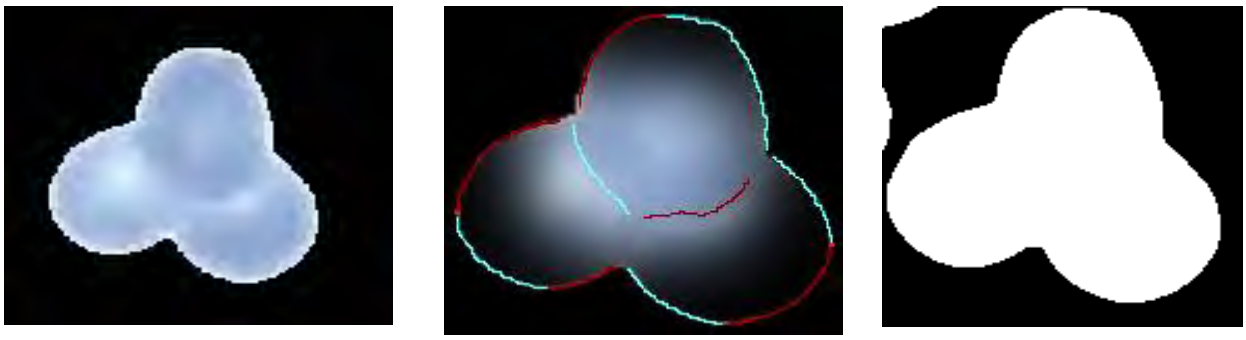


Figure 5.5: Comparison of segmentation

Figure 5.5 shows a sample RBCs fragment where it is difficult to isolate using watershed techniques but possible using the proposed segmentation approach using oriented edges. In (A) the original image is shown and (B) shows segmentation in the proposed algorithm whereas (C) shows the result of segmentation using watershed algorithm. More comparison is shown in Annex D.

5.5 Discussion

In diagnosing malaria, it is amenable that pathologists' skills have an important role in the results accuracy of detecting parasites and computing parasitemia. In this regard, the proposed system could elevate the level of accuracy that depended on pathologists' skills. For the experiment we collected the images from CDC [12] and MaMic [13]. However, some of these images were not sufficient in amount and variety. Hence, it was necessary to collect actual data that is malaria-stained blood smears images infected by the Plasmodium vivax and falciparum from Pathology Department of School of Medicine at Addis Ababa University. Our novel algorithm to detect Plasmodium parasite and compute parasitemia was tested using sample data selected from the ground truth and the two other sources mentioned previously. We applied statistical approach to test the performance of the proposed system. On top of that, a parallel comparison of two automated approaches has been performed with respect to the ground truth. This, however, is for the case of computing parasitemia. In contrast to this, comparing parasite detection algorithm with other works is difficult due to variability in data and environment setting.

The performance of parasite detection has been determined through specificity and sensitivity measures. From the result, the overall detection accuracy was 93 %, while sensitivity was 94% and specificity being 91%. Out of the total 210 erythrocytes images (approximately 29,400 RBCs) used, only 14 erythrocytes images were falsely detected. This result is found to be good enough by taking worst scenario into consideration.

Regarding computing parasitemia, 6 composite images were tested to check counting performance. An error measure was used to check the accuracy of the proposed counting algorithm. The mean percentage error to count red blood cells was 2.07 from the ground truth which is found to be a good result. However, the performance reduces as the degree of overlap increases. The comparison between the proposed counting algorithm and counting based on watershed segmentation technique in relation to the ground truth also showed that the proposed algorithm comes up with better results. The mean percentage error to count red blood cells using watershed segmentation was 6.11 from the ground truth. This showed that the proposed algorithm is less affected by over segmentation and under segmentation relative to watershed teqnique.

Though some of the related works, [9] and [11], claim that they could achieve better sensitivity and specificity results than this work, theirs are constrained as they didn't consider worst cases scenarios such as counting clumped RBCs with high degree of overlapping, treating RBCs lying on the border of the image, detecting Plasmodium parasites which are affected by poor illumination, etc. which this study took into account with higher level of accuracy. Moreover, the mentioned works used limited number of sample images; whereas this work dealt with various sample images from different sources.

However, challenges are still reflected while computing parasitemia in the cases where there existed high degree of cell overlapping. Another problem experienced related to parasitemia calculation included debris or occurrence of rubble on top of the erythrocytes and throphozoite rings with aberrant morphology.

CHAPTER SIX: CONCLUSION AND FUTURE WORKS

6.1 Conclusion

A system for detecting Plasmodium parasites and computing parasitemia using images of thin blood smears stained with Giemsa was developed.

Consequently, the biggest detraction of microscopy, namely its dependence on the skill, experience and motivation of a human technician, is solved. The use of an automated digital microscope, which would allow entire slides to be examined, would allow the system to make diagnoses with a high degree of certainty.

The significance of the developed system has brought change in the process of diagnosis in terms of accuracy and efficiency. The level of accuracy was measured from two perspectives. Parasite detection and counting of RBCs and parasites. It could be learned that the system has a 94% accuracy in parasite detection and a mean percentage error of 2.07 to conduct counting. When this level of accuracy was compared to that of watershed segmentation approach, the newly developed system was found to excel. This implies the newly developed system could alleviate the problems that the field faced before in its dependence on expert's skills and experience and delay in diagnosis.

As a recommendation, the implementation of this work in real scenario will be advantageous as it could be proven to produce a result which is closer to what an expert does.

6.2 Contribution to the Knowledge

As a contribution to knowledge, the proposed system offered a systematic approach in the process of automatic malaria detection. Considering preprocessing activity, color structure tensor is employed due to its power in good smoothing capability and also being capable to magnify detail elements. Next, the proposed novel segmentation algorithm is based on adaptive threshold and threshold selected from color structure tensor. Hence, segmenting microscopic image that is affected by poor illumination was made possible with this algorithm. Color structure tensor is also applied as a tool for extracting features that have high discriminant power in detecting Plasmodium parasites. In this work, therefore the following notable contributions are made.

- A novel segmentation algorithm has been developed to isolate complex red blood cells' arrangement that yielded improved parasitemia.
- A novel counting algorithm has also been produced to estimate Plasmodium parasites in overlapped RBCs.
- A method for successful detection of parasites in poorly illuminated images has been made possible.
- A method for properly segmenting RBCs lying around the border of the image has been found out.
- An algorithm that enabled a simplified parasitemia computation that is compatible with various objective lenses is built.

Besides, the developed model has also enabled counting of Plasmodium parasite under complex RBCs' arrangement, which had been remained a challenge.

6.3 Future Work

In this section we insight two key points that remain a challenge in the field of microscopy image analysis, and of course were challenges in this work. Microscopic objects are sensitive to change in illumination and at the same time are easily affected by change in setting made on microscope. A multi–dimensional adaptation towards microscopic image analysis will bring a significant role in determining objects in different state. The researchers believe future study should consider it to achieve better result in analyzing microscopic image. The other insight from the reference to proposed system is that treatment of complex overlapped cells such as nested or clumped erythrocytes should be further investigated.

REFERENCES

[1] WHO, "*World Malaria Report*", Switzerland, 20 Avenue Appia, 1211 Geneva 27, 2015.

[2] FDREMH, *Malaria Diagnosis and Treatment Guidelines for Health Workers in Ethiopia* (2nd ed.), Ethiopia, 2004.

[3] World Health Organization. *Basic malaria microscopy: Learners Guide (2nd ed.)*. Switzerland, 2009.

[4] T. Suradkar "Detection of Malaria Parasite in Blood Using Image Processing," *International Journal of Engineering and Innovative Technology (IJEIT)*, Vol. 2, No. 10, pp. 124-126, 2013.

[5] M. Mausumi, G. Rahul, and M. Mukherjee "Detection and Counting of Red Blood Cells in Blood Cell Images using Hough Transform," *International Journal of Computer Applications.,* Vol. 53, No. 16, pp. 18-22, 2012.

[6] R. Renuka, V.Rajagopal, M. kumar, and G. Magesh "Computerized Shape Analysis of Erythrocytes and their Formed Aggregates in Patients Infected With P.vivax Malaria," *Advanced Computing: An International Journal (ACIJ).*, Vol. 2, No. 2, pp. 71-77, 2011.

[7] A. Neetu, P. Sapnojit, and A. Bibhudendra "Advanced Image Analyses Based System for Automatic Detection and Classification of Malaria Parasite in Blood Images," *International Journal of Information Technology and Knowledge Management,* Vol. 5, No. 1, pp. 59-64, 2012.

[8] V. Panchbhai, B. Damahe, V. Nagpure, and N. Chopkar, "RBCs and Parasites Segmentation from Thin Smear Blood Cell Images," *Image, Graphics and Signal Processing*, Maharashtra, Vol. 10, pp. 54-60, 2012.

[9] A. Nasir, Y. Mashor, and Z. MohamedColour, "Color Image Segmentation of Malaria Parasites in Thin Blood Smears Using C-Y Colour Model and K-Means Clustering," *The 2nd International Malaysia-Ireland Joint Symposium on Engineering, Science and Business,"* pp. 99-107,2012.

[10] D. Ruberto, R. Dempster, A. Khan, and B. Jarra "Analysis of Infected Blood Cell Images Using Morphological Operators," *Image and Vision Computing ,* Vol. 20**,** pp. 133-146, 2002.

[11] G. Díaz, A. González, E. Romero, "A Semi-Automatic Method for Quantification and Classification of Erythrocytes Infected with Malaria Parasites in Microscopic Images," *Journal of Biomedical Informatics,* Vol. 42, pp. 296-307, 2009.

[12] Center for Disease Control and Prevention, "Malaria sample images," retrieved from http://www.cdc.gov/dpdx/malaria/gallery.html#pfalringformtrophs, Last accessed on June 10, 2014.

[13] The MaMic Image Database, "Thin smear blood samples for identification of Plasmodium Falciparum," retrieved from http://fimm.webmicroscope.net/research/ momic/mamic, Last accessed on June 4, 2014.

[14] K. Edward, "Malaria Diagnosis in the Community: Challenges and Potential Role of Rapid Diagnostic Tests (RDTs) in the African Region*," African Journal of Health Sciences*, Vol. 14, pp. 114-117, 2007.

[15] S. Kamolrat "Standard Operating Procedure: Malaria Slide Management", *Mahidol Oxford University Research Program.* Version 10. 2013

[16] R. Mehdi , " A Fuzzy Approach for Cell Counting in Poorly Illuminated Images to a Cell-Phone Microscope", Unpublished Master's Thesis, Department of Electrical and Electronic Engineering , California State University, 2012.

[17] M. Norberto, O. Carlos, J. Juan , S. Andres , V. Isabel , M. Jose, and P. Francisco , "Applying Watershed Algorithms to the Segmentation of Clustered Nuclei," *Cytometry 28*, pp. 290-297, 1997.

[18] R. Sameer, "Implementation of Watershed Based Image Segmentation Algorithm in FPGA," *Institute of Parallel and Distributed Systems*, Department of Parallel Systems. Thesis Nr. 3256, 2012

[19] S. Kumar, H. Ong, S. Ranganath, C. Ong, T. Chew, "A Rule Based Approach for Robust Clump Splitting," *Pattern Recognition*, Vol. 39, 2006.

[20] N. Tung, A. Duc, and H. Quan "Cell Splitting with High Degree of Overlapping in Peripheral Blood Smear," *International Journal of Computer Theory and Engineering,* Vol. 3, No. 3, pp. 473-478, 2011.

[21] Digital Image Processing, Image Analysis: Part 2. Olivier.bernard@creatis.insa-lyon.fr

[22] R. Gonzalez and R. Woods, *Digital Image Processing*, 2nd ed, Upper Saddle River, NJ: Prentice Hall, 2001.

[23] V. Joost, G. Theo, G. Jan-Mark " Edge and Corner Detection by Photometric Quasi-Invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 27, No. 4, pp. 625-630, 2005.

[24] M. Alasdair, *An Introduction to Digital Image Processing with Matlab*, 2004.

[25] R. Gonzalez, R. Woods, S. Eddins, *Digital Image Processing Using MATLAB*, Pearson Prentice Hall Pearson Education, Inc, New Jersey, USA, 2004.

[26] M. Wolniak, "BSCI 427 Principles of Microscopy Fall 2004 Syllabus: Principle of Light-Microscopy," retrieved from: http://www.life.umd.edu/cbmg/faculty/wolniak/ wolniakmicro.html, Last accessed on May 12, 2014.

[27] M. Norberto, O. Carlos, J. Juan, S. Andres, V. Isabel, M. Jose, and P. Francisco, "Applying Watershed Algorithms to the Segmentation of Clustered Nuclei," *Cytometry 28*, pp. 290-297, 1997.

[28] N. Abbas, and M. Dzulkifli, "Microscopic RGB Color Images Enhancement For Blood Cells Segmentation in YCbCr Color Space for K-Means Clustering," *Journal of Theoretical and Applied Information Technology.* Vol. 15, No. 1, pp. 117-125, 2013.

[29] N. Ross, C. Pritchard, D. Rubin, and A. Duse, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears, " *Medical & Biological Engineering & Computing*, Vol. 44, pp. 427-436, 2006.

[30] J. Van, T. Jeversr,"Tensor Based Feature Detection For Color Images," *The Intelelegent Sensory Information Systems*, University of Amsterdama.

[31] V. Joost, K. geboren, “Color Features and Local Structure in Images,” *The Intelelegent Sensory Information Systems*, University of Amsterdama., 2005.

[32] W. Fredrik, “Feature Extraction Based on a Tensor Image Description,” *Linköping Studies in Science and Technology*, Linköping University, 1991.

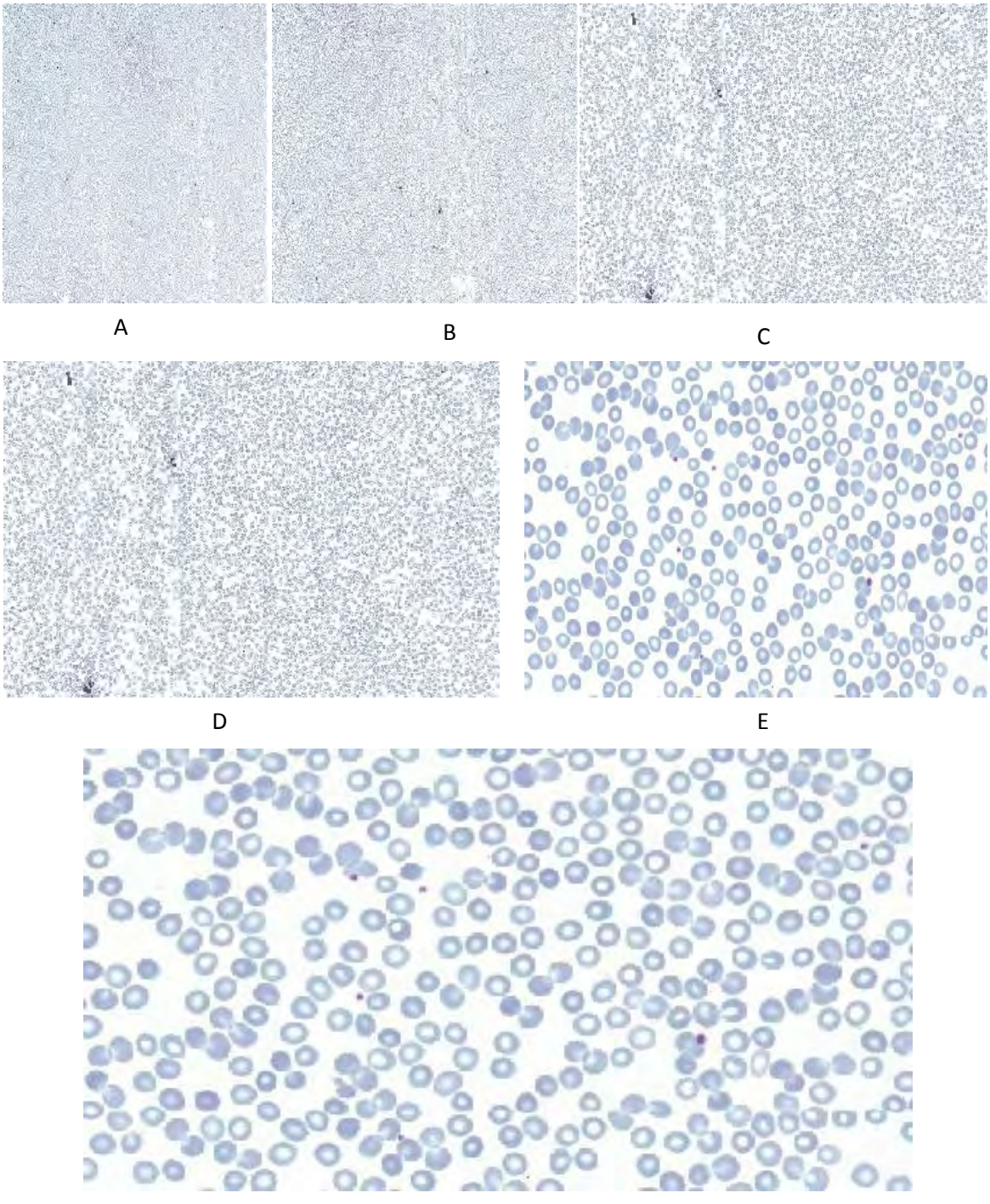
[33] J. Bigun, “A systematic introduction to image processing and computer vision,” *Springer*, 2006.

[34] Yaregal Assabie and J. Bigun, “Offline Handwritten Amharic Word Recognition,” *Pattern Recognition Letters*, 32 (2011), 1089-1099, 2011.

[35] Loong TW: Understanding sensitivity and specificity with the right side of the brain. *BMJ* 2003, 327:716.

ANNEXES

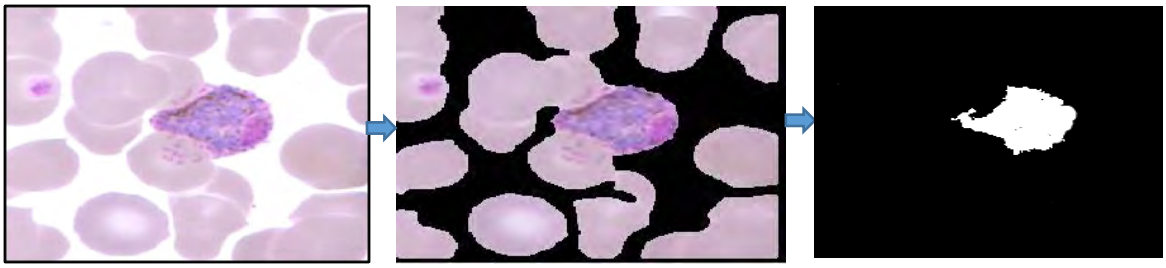
Annex A: Sample Image Extracted from Microscopic Thin Blood Films



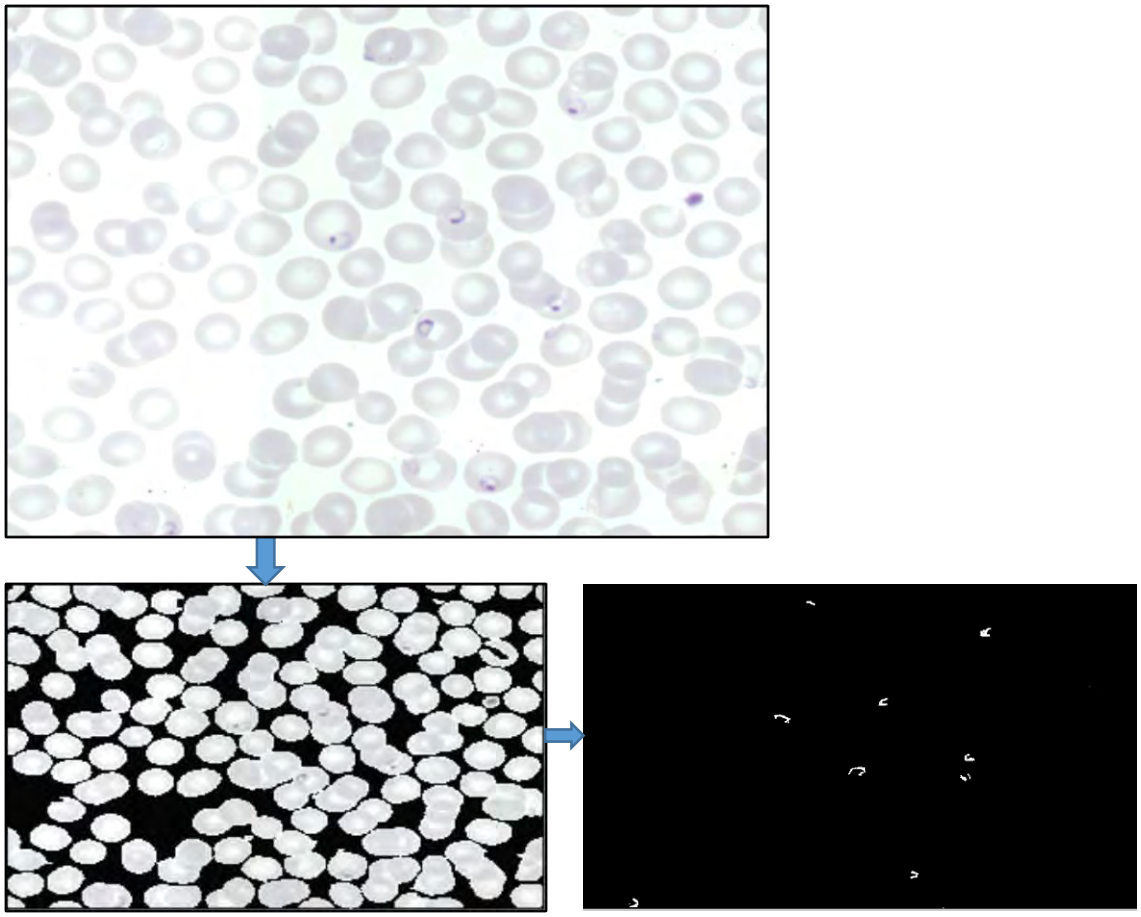
A B C D E F

Result of images taken from microscope through Gemisa staining process at different Objective lens level A: 2X, B: 5X, C: 10X, D: 20X, E: 40X, and F: 100X, source: http://fimm.webmicroscope.net/research/momic/mamic

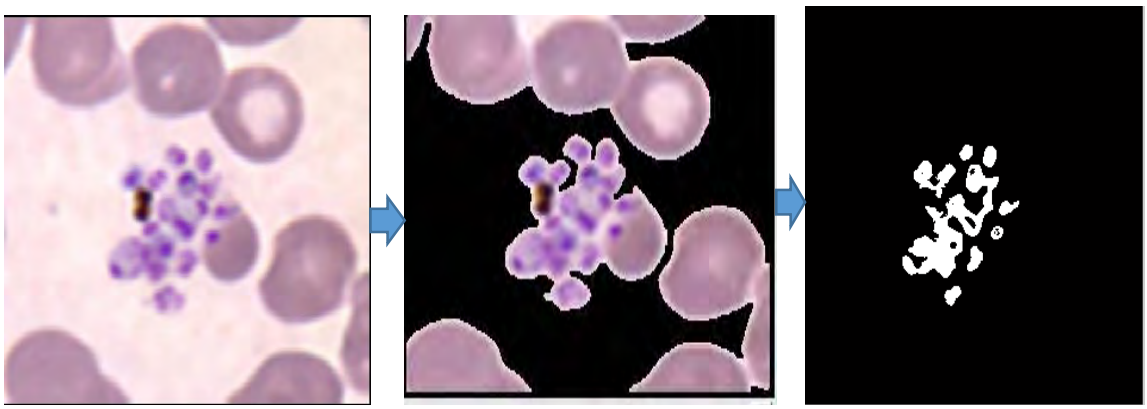
Annex B: Sample Poorly Illuminated Images Taken from Positive Blood Tests and its Segmentation Result



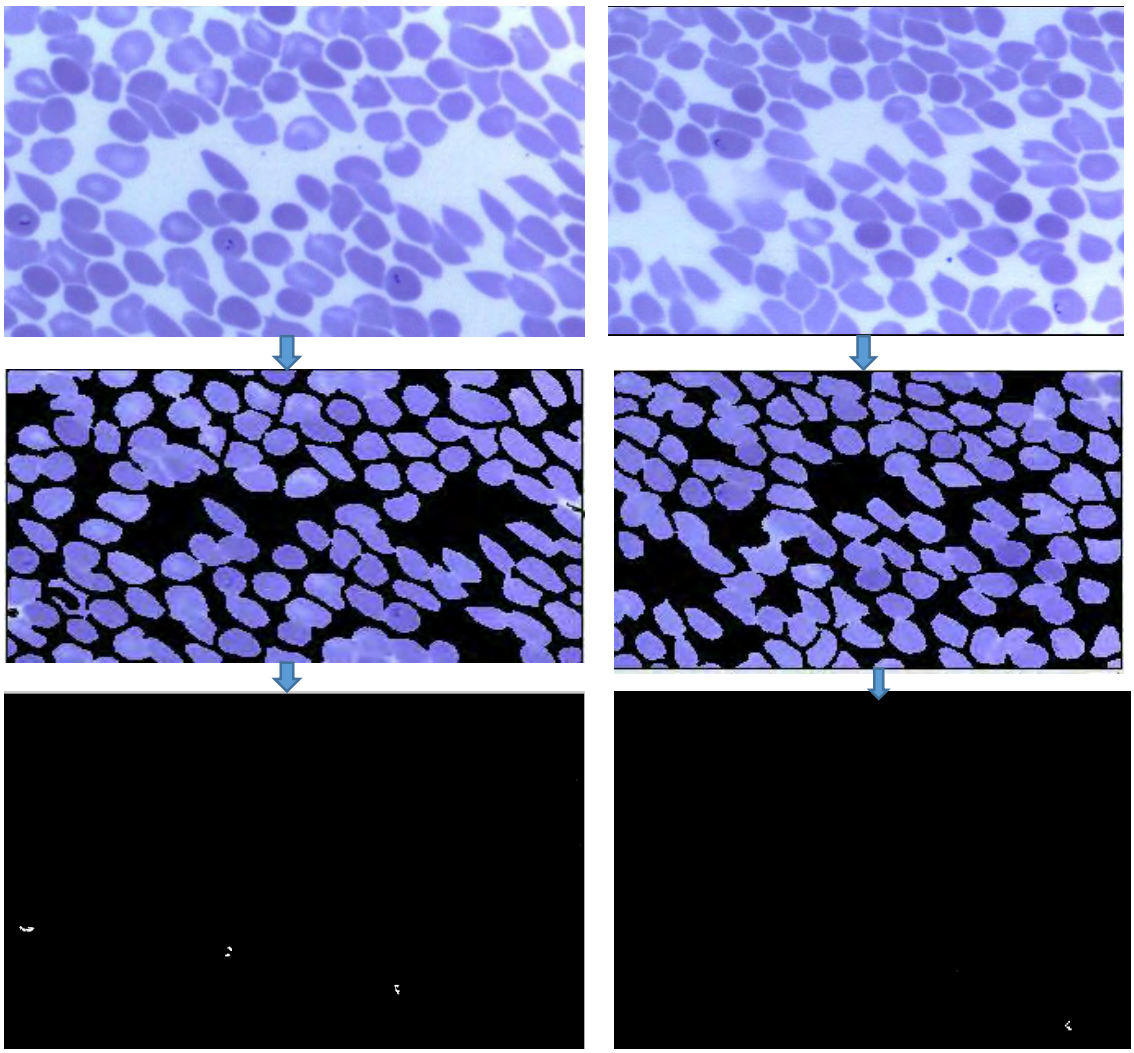
Detection of plasmodium vivax gametocyte stage in poorly illuminated



Detection of plasmodium vivax , tropazoit stage in poorly



Detection of plasmodium vivax schizont, stage in poorly illuminated microscopic image



Detection of plasmodium falciparum tropazoit stage in poorly illuminated microscopic image

Annex C: Algorithm to Isolate Overlapped RBCs and Count Plasmodium Parasites

```
Input:  Non-single instance RBC Arrangement in binary
Output: Estimated Number of RBC
    Get overlapped arrangement R from arrangement model
    Get orientation using Algorithm 4.4
    Get a model oriented pattern T from arrangement model
        based on Algorithm 4.8
    Extract edges Ege_val from arrangement R
    Compute RGB equivalent RGB_val for the binary R
    Extract internal Edges IE using Canny edge detector
    from  RGB_val
    Select candidate internal edges CIE from IE and
    remove others using proper threshold
    Get the combined edges All_Edge from Ege_val and CIE
    Apply Harris corner detector on RGB_val and get
    Extract oriented curves Oriented_Curve from All_Edge
    using Algorithm 4.5
    Separate Oriented_Curve with detected corners corners
    Extract feature from each  Oriented_Curve
    Extract feature from template T
    Extract oriented curves T_Curve from T using Algo.
    4.5
    Separate the whole  Oriented_Curve in to curve+ and
    curve-
    Separate T_Curve in to T+ and T-
    For each separate curve C in  curve+ and  curve-
       For each curve c' in the separated curve  T_Curve
          If c'==  T+
            R_measur = modelfit(c' , curve+);
             Else
            R_measur = modelfit(c' , curve-);
         End
```

```
        End
                If  R_measur < threshold t
                    Assign a wait wt for curve in c;
                End

      Divide the total number of  wt in c in to two
   End
     Count= 0.5*( wt in curve +,  wt in curve-);
   Return  Count;
```

A

```
Input: RBC Arrangement, Oriented Curves from Algorithm 4.1
Output: Estimated Number of Plasmodium Parasite

Get labeled _Single_instance_RBC, Get labeled
_Overlapped_RBC

Get labeled _Broken_RBC, Get Labeled parasite: L.par
Get Centroid of Parasite: P.cent, Get Fore Ground Object :
  Parasite_count=0;
  For each Parasite in L.par
     If F.Obj( P.cent)==0
          Label_Id = L.par (P.cent);
          L.par(L.par == Label_Id)=0;
     End
  End
  Switch  arrangement Type
  Case: Single Instance RBC;A1=Labeled_Single_Instance_RBC
  Case: Overlapped RBC;  A2= Labeled_Overlapped_RBC
  Case: Hidden RBC;  A3= Labeled_Broken_RBC
  End
   For each Lebeled_Arrangement R
      If  R == A1
          For each RBC in A1
```

```
            For each parasite in L.par
               If   A1 (P.cent ) !=  0
                    Label_ID= A1 (P.cent );
                     A1 (A1== Label_ID)=0;
                     Parasite_count= Parasite_count + 1;
                End
             End
End
  Else If R ==A3
    For each RBC in A3
       For each parasite in L.par
         If   A3 (P.cent ) !=  0
              Label_ID= A3 (P.cent );
              A3 (A3  == Label_ID)=0;
              Parasite_count= Parasite_count + 1;
        End
     End
    End
  Else
  For each clustered RBC in A2
     For each parasite in L.par
          Prev. cent = P.cent;
       If A2 (Prev. cent) !=  0
           For each parasite in L. par
       If eculedean_distance(p.cent, Prev. cent) < min
          Nerest.cent = P.cent;
       End
    If  Prev. cent !=  Nerest.cent
        ListOfPoints = bresnham_distance (Prev. cent
        ,Nerest.cent );
    End
     If  Oriented_Curves(ListOfPoints!=0 ) &&
          Parasite_count= Parasite_count + 1;
```

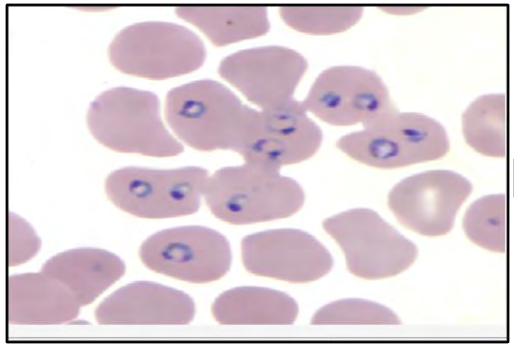
```
            countStatus=1;
        Else If Oriented_Curves(ListOfPoints ==0 ) &&
                countStatus==1

    End
   End
  End
 End
End
```

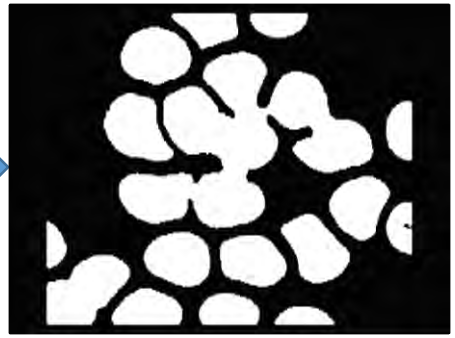
B

A: Overlap separation and Estimating the number of red blood cells in arbitrary complex arrangement

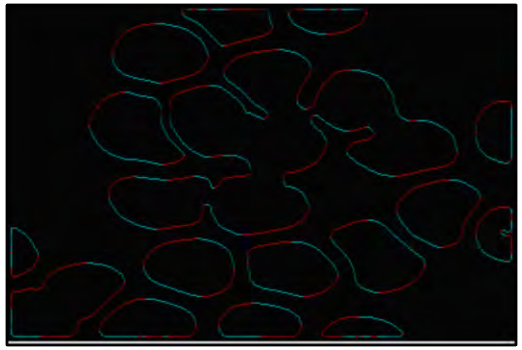
B: Counting of Parasites



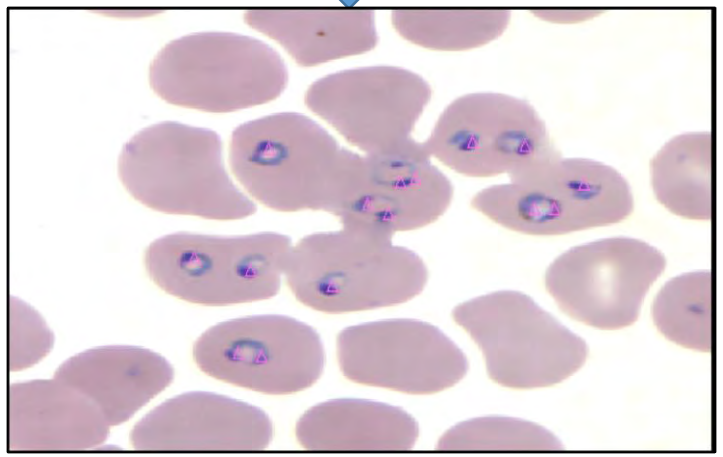
Original Image



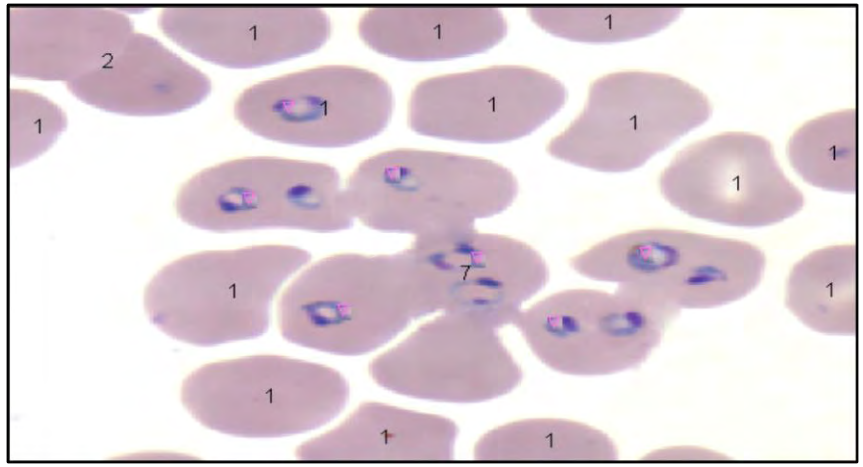
Segmented Result



Modeling using Oriented Edges

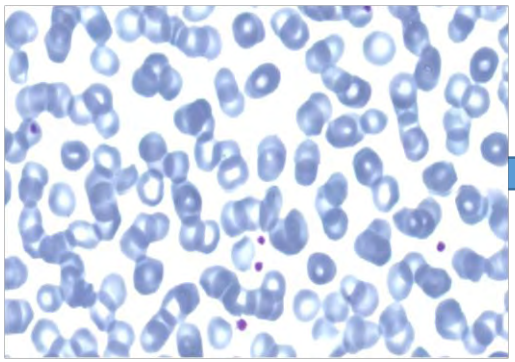


Parasite Detection

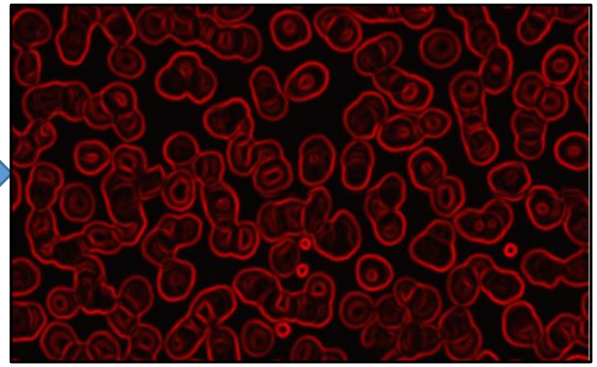


Parasite and RBCs Count Using the proposed algorithm

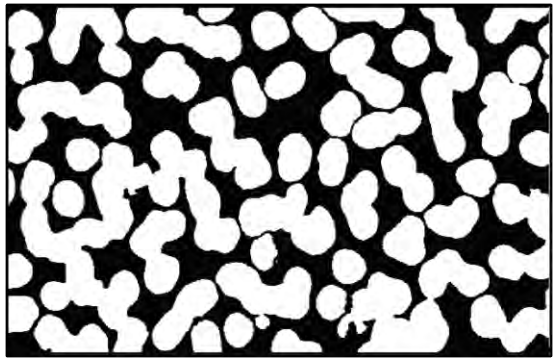
Annex D: Sample RBC Counting Result from Complex RBCs Arrangement and Sample Source Code



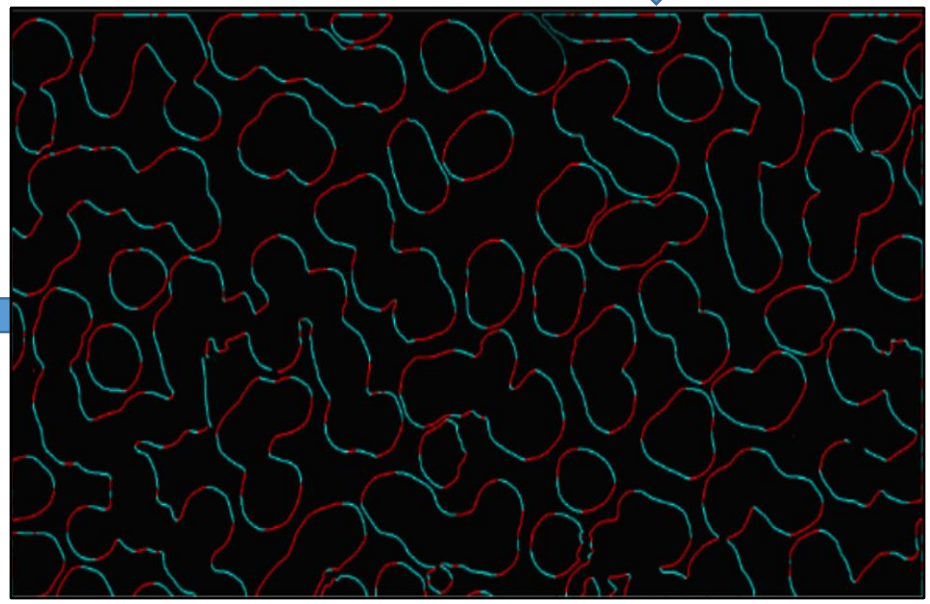
(A). Input Microscopic Image



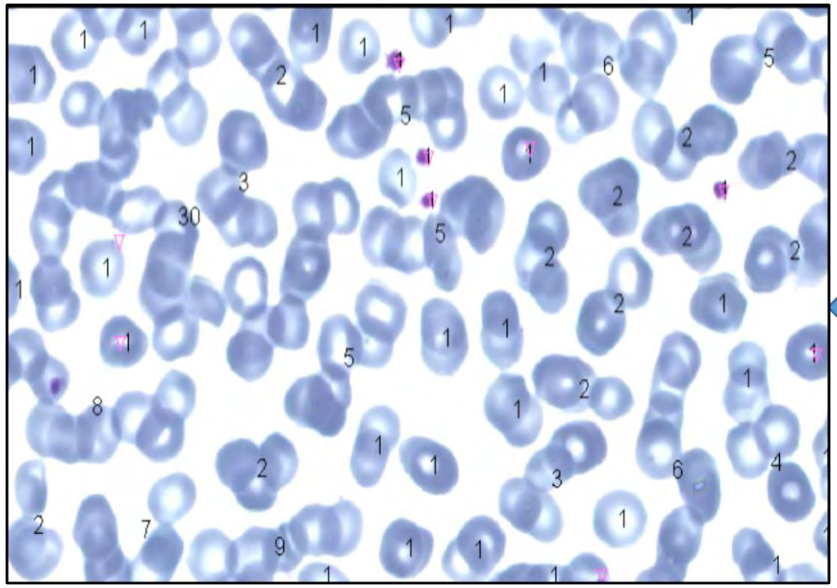
(B). Magnitude of $\overrightarrow{I_{20}}$



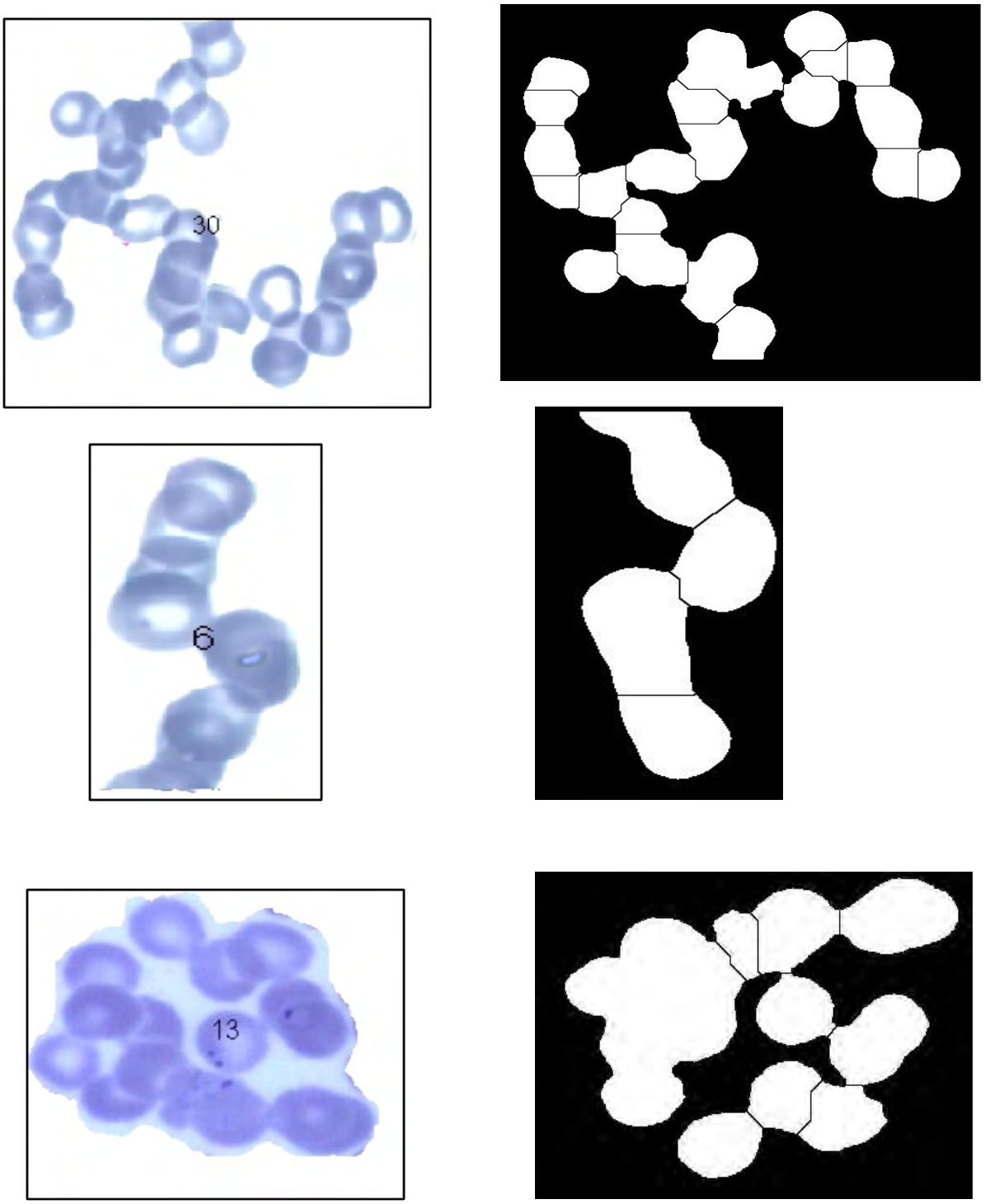
(C). Segmented Image



(D). Modelling RBCs using oriented Edges



(E). The result of estimating RBCs from complex arrangement



A sample complex arrangements and estimated number of RBCs using the proposed segmentation algorithm

`Fragment No	Ground truth Count	Watershed Segmentation Count	Proposed System Count	Proposed Algorithm (RE)	Watershed Segmentation(RE)
FR-1	28	21	30	0.07	0.25
FR-2	7	2	6	0.14	0.71
FR-3	7	4	6	0.14	0.43
FR-4	8	4	8	0.00	0.50
FR-5	7	4	7	0.00	0.43
FR-6	6	9	2	0.67	0.50
FR-7	6	4	5	0.17	0.33
FR-8	3	1	3	0.00	0.67
FR-9	4	2	2	0.50	0.50
FR-10	6	3	5	0.17	0.50
FR-11	3	2	2	0.33	0.33
FR-12	3	2	2	0.33	0.33
FR-13	3	3	2	0.33	0.00
FR-14	3	2	2	0.33	0.33
FR-15	4	3	4	0.00	0.25
FR-16	2	2	2	0.00	0.00
FR-17	3	1	1	0.67	0.67
FR-18	5	2	3	0.40	0.60
FR-19	2	1	1	0.50	0.50
FR-20	2	1	1	0.50	0.50
FR-21	5	3	5	0.00	0.40
FR-22	3	3	3	0.00	0.00
FR-23	2	1	1	0.50	0.50
FR-24	2	2	2	0.00	0.00
FR-25	2	1	2	0.00	0.50
FR-26	2	2	2	0.00	0.00
FR-27	2	2	2	0.00	0.00
FR-Single Instance	26	26	26	0.00	0.00
			Mean % Error	10.86	18.37

A sample test conducted in Mam_img2 ++ and the result of error introduced in individual fragments and its percentage of mean relative error

Signed Declaration Sheet

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and that all source of materials used for the thesis have been duly acknowledged.

Declared by:

Name: _____

Signature: _____

Date: _____

Confirmed by Advisor:

Name: _____

Signature: _____

Date: _____