# AN EXTENSIBLE COMPUTER VISION APPLICATION FOR BLOOD CELL RECOGNITION AND ANALYSIS

G.P.M Priyankara[praboda@cse.mrt.ac.lk], O.W Seneviratne [oshani@cse.mrt.ac.lk],
R.K.O.H Silva[omega@cse.mrt.ac.lk], W.V.D Soysa[vidakna@gmail.com],
C.R. De Silva [crdsilva@cse.mrt.ac.lk].
Department of Computer Science and Engineering,
University of Moratuwa,
Sri Lanka.

## Abstract

*Fast and cost-effective production of blood cell count reports is of paramount importance in the healthcare industry. The traditional method of manual count under the microscope yields inaccurate results and put an intolerable amount of stress on the Medical Laboratory Technicians. Although there are hardware solutions such as the Automated Hematology Counter, developing countries like Sri Lanka are not capable of deploying such prohibitively expensive machines in every hospital laboratory in the country. As a solution to this problem, this research project aims to provide a software-based cost effective and an efficient alternative in recognizing and analyzing blood cells.*

**Keywords**: Microscopic Image Analysis, Standard Blood Cell Count, Differential White Blood Cell Count, Image Segmentation, Hough Transformation, Image Moments, Neural Networks

## 1 Introduction

### 1.1 Introduction to blood cell recognition & counting

In almost all of the medical laboratories, blood cell count reports are taken on physicians' recommendation, in order to assist the diagnosis of the particular ailments of the patients. In fact this test is one of the most frequent tests carried out in a medical laboratory.

Producing blood cell reports can be broadly categorized into 2 areas:

1. Standard Count for Red Blood Cells (RBC), White Blood Cells (WBC) and Platelets (PLT)
2. Differential Count for WBC

### 1.2 Standard Count for all cell types

For this type of a count, the MLT is produced with a Hemacytometer shown in the figure, which is also known as the Counting Chamber (CC). The count is obtained for all the different blood cell types (i.e. RBC, WBC and PLT).

Since blood cells are counted per unit volume (per liter), it is vital that the volume of blood, in which the cells are counted, corresponds to a known quantity. Usually, the CC has a special objective slide containing counting grids of the size 3mm x 3mm. In the figure below, there are 2 such counting grids in the middle.
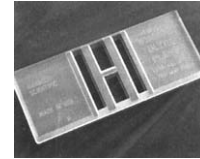


**Figure 1-1 Counting Chamber**

The schematic representation of the counting grid is represented in the figure below. The counting grid is composed of 9 big squares, measuring 1 x 1 mm. From these squares, the central square contains 25 medium sized squares each measuring 0.2 x 0.2 mm. These are further divided into 16 small squares each measuring 0.05 x 0.05 mm. The large central square is also called the "erythrocyte" grid. The squares highlighted in red correspond to 80 small squares, and are used to establish the RBC and PLT counts. The large squares marked in blue are used to establish the leukocyte or the WBC count.
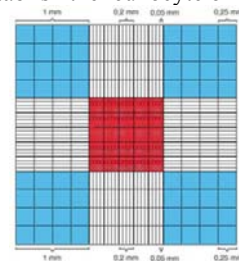


**Figure 1-2 Schematic representation of the counting grid**

In the RBC and Platelet count, the erythrocyte grid in the middle is examined much closely.

## 1.3 Differential Count for WBC

White blood cells in human blood can be put in to five main sub categories as follows. [1]

1. Neutrophils
2. Eosinophils
3. Basophils
4. Monocytes
5. Lymphocytes

In differential counting, the white blood cells are counted and classified in to the above categories and output the number of each of the above types as a percentage.

The above five types are broadly divided in to two categories; the "phils" category and the "cytes" category. This division is done based on how the names of the above types end. Therefore, Neutrophils, Eosinophils and Basophils fall in to the "phils" category while Monocytes and Lymphocytes fall in to the "cytes" category.

The nucleus of each of the above types has a unique shape, and this is the most important feature used in cell classification. In addition to the shape of the nucleus, the "phils" category has granules with in the blood cell where as "cytes" category does not have granules.[1]

### 1.3.1 Types of White blood cells

**Neutrophils**

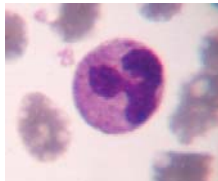Neutrophils are identified based on the blobbed nucleus small granules. The nucleus has 3 to 5 blobs.



**Figure 1.3-3 A Neutrophil**

**Eosinophils**

The Eosinophil is distinguished by its red granules and blobbed nucleus. The granules are larger than that of a Neutrophil.
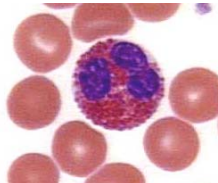


**Figure 1.3-4 An Eosinophil**

**Basophils**

A Basophil is characterized by a lobed nucleus and it is filled by large blue-black granules that sometimes cover the nucleus.
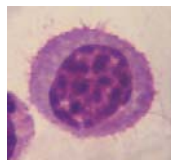


**Figure 1.3-5 A Basophil**

**Lymphocytes**

These have round shaped nucleus. They have large nucleus to cytoplasm ratio. No granules are present.
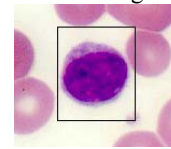


**Figure 1-6 A Lymphocyte**

**Monocytes**

These have horse-shoe shaped nucleus. No granules. The nucleus to cytoplasm ratio is high but less than that of a Lymphocyte.
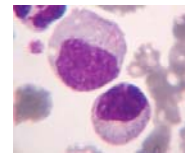


**Figure 1-7 A Monocyte**

## 2 Current Methods for blood cell counting

### 2.1 Manual Method

In general, the MLT would prepare the slide and examine it under the microscope. As explained in HemoSurf [2] and from the requirements of the MLTs, different parameters are taken for each of the cell type counts. These inputs are used in different equations and the count is obtained.

#### 2.1.1 Manual Red Blood Cell (RBC) Counting

In a manual RBC count, 10 µl of blood is diluted in 1990 µl of dilution solution. This results in a dilution of 1:200. This suspension is usually well-mixed and be immediately placed into the counting chamber. After approximately 3 minutes, the RBCs will have settled, and the MLT begins counting the RBCs in 80 small squares.

The calculation is achieved by following the formula below using these factors:

i. The number of RBCs counted in the small squares
ii. The dilution of the cell solution
iii. The number of counted small squares
iv. The volume above one small square

$$\frac{RBCs}{\mu l} = \frac{number\ of\ counted\ RBCs(i)\ x\ dilution(ii)}{number\ of\ counted\ squares(iii)\ x\ volume\ above\ one\ small\ square(iv)}$$

**Equation 2-1 Calculation of RBC per micro-liter**

In 80 small squares, around 400 RBCs are counted for normal values. This yields a coefficient of variation (variability) of ±5%. This constitutes the highest acceptable amount of random error (accuracy).

#### 2.1.2 Manual White Blood Cell (WBC) Counting

In the manual WBC count, 50 µl of blood is mixed together with 950 µl dilution solution. This constitutes a dilution of 1:20. The RBCs will be lysed (i.e. cells are

destroyed by bursting), and the WBC nucleus is stained. The counting chamber is immediately filled after mixing. After 2 minutes, the MLT begins counting the WBCs in the 4 large squares.

Calculation of the WBC count is achieved by following the formula below using these factors:

i. The number of WBCs counted in the big squares
ii. The dilution of the cell solution
iii. The number of counted big squares
iv. The volume above a big square

$$\frac{WBCs}{\mu l} = \frac{number\ of\ counted\ WBCs(i)\ x\ dilution(ii)}{number\ of\ counted\ squares(iii)\ x\ volume\ above\ one\ big\ square(iv)}$$

**Equation 2-2 Calculation of WBC per micro-liter**

Since the WBC count shows greater physiologic variations, a coefficient of variation of ±10% is accepted.

In addition since normoblasts (i.e. nucleated precursors of RBCS which would get erroneously counted as WBCs) are not recognized in this procedure, a large number of normoblasts can invalidate the result. This is also true for mechanical WBC counts. The number of normoblasts can only be accurately identified in a blood film and is expressed as the number of normoblasts per 100 WBCs. If there are more than 5 normoblasts per 100 WBCs, a correction must be used.

The normoblast correction is achieved by using the formula below:

$$Corrected\ WBC\ count = \frac{Automatically\ counted\ WBCs}{100 + Microscopically\ counted\ normoblasts} \times 100$$

**Equation 2-3 Normoblasts correction for WBC count**

### 2.1.3 Manual Platelet (PLT) Counting

The PLT count is very similar to the RBC count. However unlike in the RBC count, the solution has a dilution of 1:20 and the RBCs are completely lysed before analysis. The suspension is then mixed and put into the counting chamber. The chamber is left in a moist environment for 20-30 minutes so the platelets can settle without the chamber drying. Like in the RBC count, 80 small squares are counted.

Calculation of the PLT count is achieved by using the formula below using these factors:

i. The number of PLTs counted in the small squares
ii. The dilution of the cell solution
iii. The number of counted squares
iv. The volume above a square

$$\frac{PLTs}{\mu l} = \frac{number\ of\ counted\ platelets(i)\ x\ dilution(ii)}{number\ of\ counted\ squares(iii)\ x\ volume\ above\ one\ small\ square(iv)}$$

**Equation 2-4 Calculation of PLT per micro-liter**

A coefficient of variation of ±10% is acceptable for PLTs and at least 100 platelets must be counted.

### 2.1.4 Drawbacks of the manual method

- Visual inspection of microscopic images is time consuming and exhaustive. If the counting process is interrupted, the MLT has to start over again from the scratch.
- Cell analysis is realized by an experienced MLT by comparing what she sees with images of cell types she is familiar with. An amateur MLT would have to check with medical literature to confirm on the cell types to determine the count of a given sample. Thus these manual methods are susceptible to human fatigue that can easily result in errors.
- After the blood cell slides have been analysed, they are kept away. There is no quick and easy way of retrieving analyzing lot of images for future reference as with a computerized system.

### 2.2 *CellaVision*

This is a commercial product developed by CellaVisionAB. The CellaVision DM Analyzer (3) consists of a fully-automated system of counting blood cells. To analyze a sample of blood the following steps are undertaken:

1. The vials containing the blood are bar-coded for identification.
2. All the vials are fed in to the Automated Hematology Counter for the Standard Count.
3. Samples are taken for morphological review, where a piece of hardware known as the "slide maker" or "stainer" is used to get blood on to a thin film on a slide.
4. These slides are kept in a special container and placed in a hardware which is used to automate the manual differential count.
5. The slides are then moved under a microscope using a robot arm and images of the WBCs are taken.
6. The images are analysed and classified accordingly.

### 2.2.1 Drawbacks of CellaVision

- This product is not widely available.
- The cost of a CellaVision unit is unbearable for medical laboratories in developing countries.

## 3 Aims of the Lohitha

Lohitha is a software that can be used for recognizing and analyzing blood cells and produce blood count reports. Lohitha is capable of performing standard counts which comprises of RBC counts, WBC counts, PLT counts and differential counts.

The operation of Lohitha is purely based on image processing and computer vision technologies. The input is an image of the already prepared slide containing a film of blood, taken from a special camera attached to an ordinary microscope. The software would not consider about the preparation of the slides to be viewed through a microscope. It assumes that the preparation of the slides is

done by trained professionals and the image taken through a microscope is identical to what the MLT sees through a microscope. Therefore, unlike in CellaVision [3], Lohitha does not replace the MLTs. They are required to prepare the slides as they normally do. The main objective of Lohitha is to provide a software solution which is cost effective as well as efficient for countries like Sri Lanka to be widely utilized in the healthcare industry.

The software is designed to be extensible, tiered and a highly efficient with great consideration on ease of use for the end user.

## 4 Lohitha Functionality

The methodology followed with the overall processing of a microscopic blood smear image is shown in the following figure.
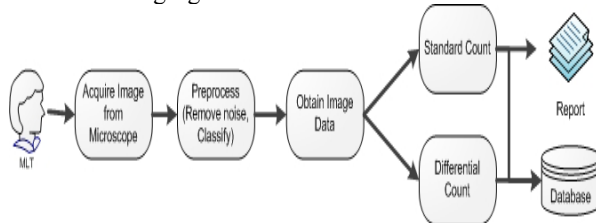


**Figure 4-1 Overall flow of processing**

Based on the type, an image going through this flow of processing would end up in the 'Standard Count' or in the 'Differential Count'. Furthermore, standard count can be performed in the automatic mode, where the application would count the cells in the entire image or in the interactive mode, where it would require the MLT to select a region of interest and the locations of few blood cells in the image.
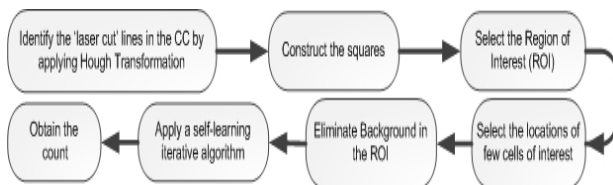


**Figure 4-2 Standard Count Process**

In the differential count, WBCs are differentiated based on the shape of their nucleus. The percentage area of the nucleus and the color of the granules are used to identify between the different types of WBCs.
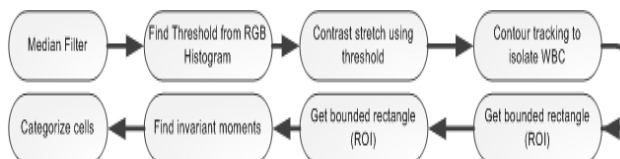


**Figure 4-3 Differential Count Process**

### 4.1 Functionality of the System

Basic functionality of Lohitha consists of following:
- Standard count

Standard count is performed for RBC, WBC and platelets. Given the image, the user will have the option of either going in to the automatic mode or the interactive mode. In the automatic mode, the application will determine the standard squares for the count operation, whereas in the interactive mode the user will have to select the regions in the image (i.e. squares preferred for counting) and perform the count.

- Differential count

Differential count is performed for WBC only. It provides a breakdown of the individual counts for lymphocytes, monocytes, neutrophils, eosonophils and basophils along with the percentage accuracy. The user has the option of selecting the image of his/her preference from the thumbnail view and loading in to the main image container for analysis.

- Report generation

After getting the blood count, user can request to prepare blood count reports. These reports can be either standard reports which consist of RBC and Platelet count or differential count reports. All these report formats will conform to HL7 or HIPAA, standards for health care specific data exchange between computer applications.

## 5 Extensible architecture

The system is designed on the conventional three tiered architecture composed of the Presentation Logic tier which contains all the logic related to the user interfaces and the display options, the Application Logic tier which contains all the logic related to blood cell counting, recognition, reporting, login and logging, and the Data Access tier which handles all the interactions the application has with the database.

The logic related to counting and recognition of blood cells dominates the system. The presentation logic plays the next important role and the system has a very thin data access layer. Therefore, there was no need to go for more than three tiers since that would create a number of very thin layers and it would be difficult to manage the system.

The business logic which composes of identifying and counting different types of cells is plugged in the form of DLLs and COMs into the presentation layer. The data layer is implemented separately.
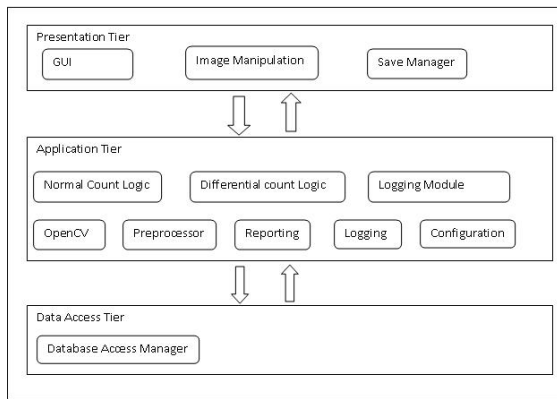
**Figure 5-1 Abstract System View**

# 6 Standard Count

The Standard Count is used to count the amount of RBCs, WBCs and PLTs in a given microscopic image. The Intel Open Computer Vision library is used extensively to obtain the functionality required to identify the lines. Based on the lines identified all the repetitive lines are removed and the squares are constructed. The MLT then has the freedom to select the square of his or her choice and obtain a count. The type of cell being counted is determined by the input from the user, i.e. whether RBC, WBC or PLT.

## 6.1 Handling Squares

The square handling algorithm is mainly concerned about constructing the squares from the lines detected. The overview flow of the process is as follows.
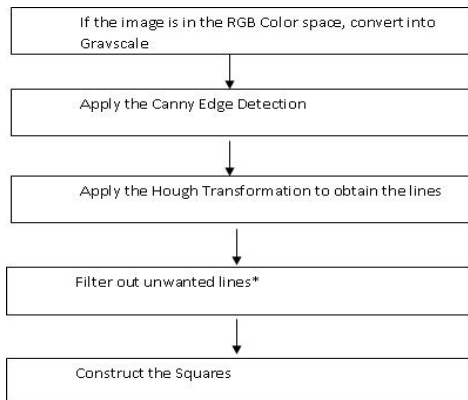


**Figure 6-1 Top level flow chart for detecting the squares**

Filtering out unwanted lines is a special requirement, since care has to be taken for the following items:

- The 3 lines close to each other in the outline erythrocyte grid that demarcates a particular square should be detected.

- Lines formed due to dust particles or any other impurities should be neglected.

- Very rarely, some cells may align in such as way that would form a line. These too have to be neglected.

- Compensate for rotation of the grid.

Correction for the above factors is achieved by primarily considering the shape and dimension of the counting grid. The algorithm for detecting unwanted lines in the image is as follows:
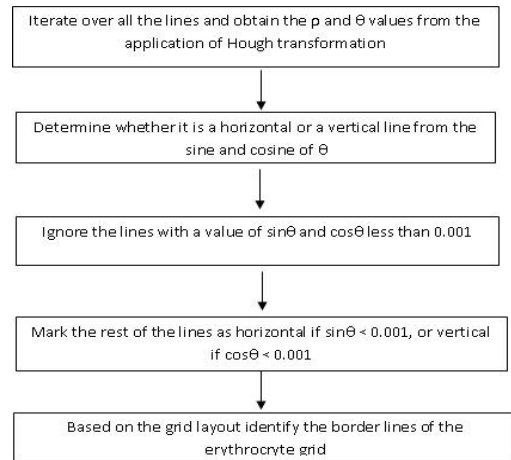


**Figure 6-2 Flow diagram for filtering lines**

Once these algorithms have been applied to the original counting chamber image, the lines will be identified as in figure. The image appears in black because the canny edge detection is applied. The lines drawn in red are all the lines detected, and the squares are identified in green.
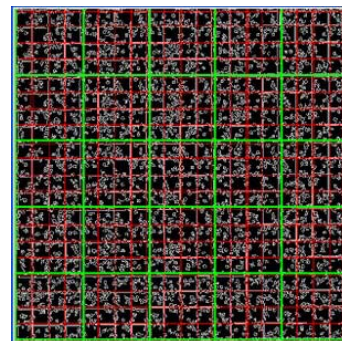


**Figure 6-3 Standard Count applied to the erythrocyte grid**

## 6.2 Automatic Count

This method is used to obtain a cell count with minimum user intervention. Once the image is input the image is pre-processed extensively to remove the noise which would occasionally consist of dust particles and air bubbles. Then the contours in the image are tracked, and a statistical method is used to filter out the unwanted items. This is done by calculating the areas of the contours and counting only the contours which falls within one standard deviation of the mean of the contours. With a considerable

number of sample images, it is found out that this method yields an acceptable accuracy.

When the count is finished the user is presented with the output image which draws only the contours which have been counted. This gives an indication as to how accurate the count has been.

The following flow chart summarises the steps followed in the automatic count.
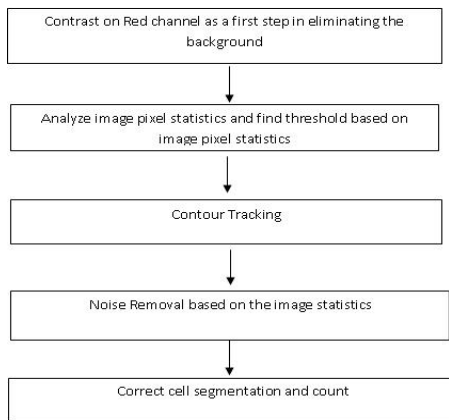


**Figure 6-4 Automatic Cell Counting method**

Once the image has been processed with the 'Automatic Count' the MLT will be able to see the counted blood cells marked in blue as shown in the image below.
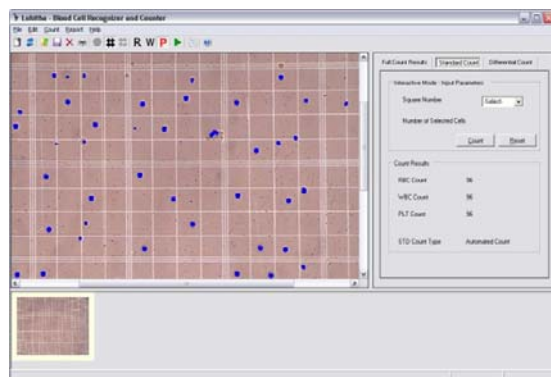


**Figure 6-5 After performing the automatic standard cell counting**

### 6.3    Interactive Count

Interactive count is a method to recognize and count the blood cells in a blood sample.

To recognize a blood cells in the given image, first the system has to know what you call a blood cell. The user provides positions of the center for several cells (5-7 is usually enough) by clicking on them. Optionally he/she gives the typical cell dimension in number of pixels that fits across a cell. Or else the system will use the default value for cell dimension.

By these user provided information, the system gets to know the pixel values of a blood cell.

Usually in a blood sample which is prepared to get the Red Blood Cell Count (RBC) following cells and particles will appear,
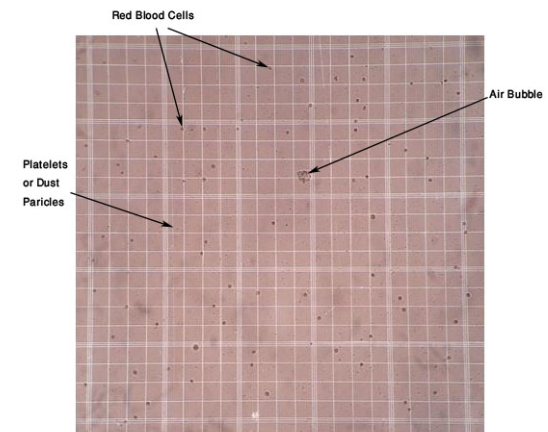
- Blood Cells
- Dust Particles
- Air Bubbles



**Figure 6-6 Image of a Blood Sample prepared for RBC with impurities**

Images as above are first preprocessed by taking into consideration the pixel values of a typical blood cell which is provided by the user. Therefore this preprocessing will be specific to the image in interest.Preprocessing of the image consists of following steps,

Step 1: Threshold the image using the average pixel value of a blood cell as the threshold value

Step 2: Apply the Median filter and convert the image to a binary image

Step 3: Dilate the image by two iterations

Step 4: Erode the image by two iterations

**Last two steps make sure to eliminate all the dust particles. Now the image consists only of blood cells and air bubbles which are very bigger than blood cells.**

To detect red blood cells in the preprocessed image, contours of the objects in the image are extracted and the area of each contour is calculated and compared with the area of blood cell to determine whether it is a blood cell or not. The identified number of blood cells is taken as the RBC of the blood sample in interest.
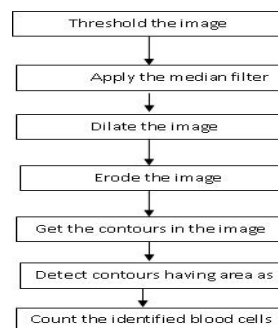


**Figure 6-7 Flow chart of the Interactive count process**

## 6.4    Differential Count

Differential Count is one of the core functionalities in Lohitha. Unlike in standard counting, a normal glass slide is used and there are no specific counting areas drawn in the slide. Approximately 50 – 100 white blood cells are counted and differentiated in to the five categories of white blood cells. The results are given as a percentage of the total counted cells.

Given a set of images of a stained blood, obtained through a camera attached to a biological microscope, Lohitha is able to process the images, recognize the white blood cells and categorize them with an acceptable accuracy.

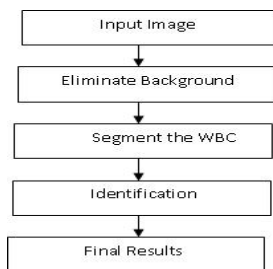The flow of work involved in differential counting is shown below.



**Figure 6-8 White Blood Cell Count Work flow chart**

### 6.4.1    Background Elimination

In order to segment the white blood cells in a given image for further categorization, first the white blood cells have to be segmented. To do that, the white blood cells have to be isolated from the background by eliminating the background.
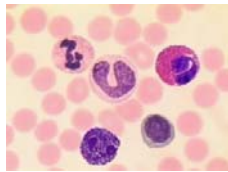


**Figure 6-9 Input *image for Differential Counting containing white blood cells***

An input image for differential counting mainly consists of the following portions, as shown in the figure 4.2.1-1

      1. Red blood cells
      2. White blood cells
      3. Background (contains the blood plasma, glass of the slide, dust particles etc.)

Therefore, in isolating the white blood cells, it is needed to remove not only the Background, but also the red blood cells in the image.

When slides are prepared for differential counting, they are stained with some chemicals, so that the nucleolus of the white blood cells takes a unique color as shown in the figure. This specific color comes handy in background elimination.

The RGB histogram for the image is used for background elimination. The algorithm for generating the RGB histogram is shown below:

```
rgb_histogram[256];
foreach pixel in the image
   foreach channel  /* Red, Blue and Green */
      gray_val = channel_intensity_of the pixel;
      rgb_histogram[gray_val]++;
   end
end
```

**Figure 6-10 Algorithm to Calculate the RGB Histogram**

The RGB histogram obtained for the above input image is shown below
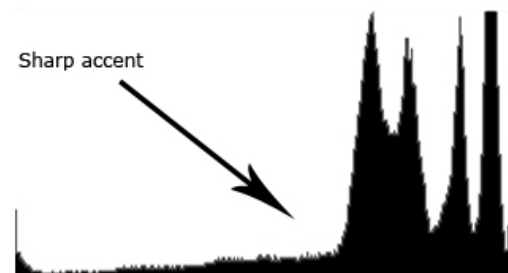


**Figure 6-11 RGB Histogram**

The RGB histogram shows a sharp accent at a certain point when traversing the histogram from left to right. This feature is prominent in the RGB histogram for all the input images for differential counting. When the starting point of the sharp accent is found out and corresponding gray value of the histogram is used as a threshold value and contrast stretching is applied on the image, the following output is obtained.
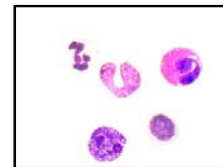


**Figure 6-12 Contrast stretched and background eliminated**

The algorithm to find out the threshold in the RGB histogram depends on the gradient of accent. An optimum gradient is found out by testing with many input images with many gradients. Using that, the threshold is found and then linear contrast stretching is applied on the image. The contrast stretching algorithm is shown below.

```
threshold;  //the threshold found from the RGB histogram
foreach image pixel
   foreach channel
      gray_value = (255 * channel_intensity) // threshold;
      new_channel_value = gray_value > 255 ? 255 : gray_value;
   end
end
```

**Figure 6-14 Contrast Stretching Algorithm**

### 6.4.2    White Blood Cell Segmentation

Even though eliminating the image background isolates the white blood cells in the image, for further processing the location of each and every white blood cell has to be identified. That is the white blood cells that appear on the image have to be segmented. The approach taken for segmenting the cells is shown below.
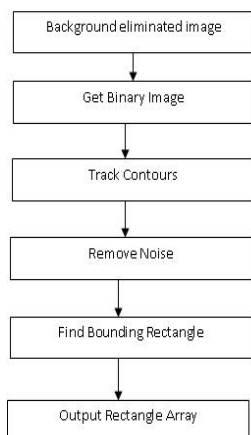


**Figure 6-15: Cell segmentation work flow**

The binary image (i.e. an image only contains gray values 0 and 255) of the background eliminated image is obtained because a binary image can be used to find edges which separate the objects (i.e. the cells in this case) from the background. A sample image is shown below.



**Figure 6-16 Input image after making binary**

After obtaining the binary image, edge detection is applied to detect edges. The Canny edge detection algorithm is used for edge detection. OpenCV provides its implementation of Canny edge detection and Lohitha uses this canny implementation.

Once the edges are detected, the edges have to be fine tuned, and contour tracking is applied. After contour tracking, all the cells can be separately identified. But contour tracking returns contours for areas which are not white blood cells. This happens due to the colored image segments that are left in the image after background elimination. Therefore, this noise has to be removed before further processing.



**Figure 6-17 Contour tracked Image**

To remove the noise, the rectangle that bounds each contour (bounding rectangle) is found and the area of that rectangle is calculated. The noise contours have very less area compared to the areas of the bounding rectangles of the cells. Depending on this difference, all the noise contours are detected and removed.

The contours that are left are the ones belonging to the white blood cells that needed to be analyzed. When an input image is given, each of the blood cells has to be analyzed separately in order to classify them. Therefore, the bounding rectangle is calculated for each cell contour and an array of bounding rectangles are returned, which will be used in the cell identification functions.
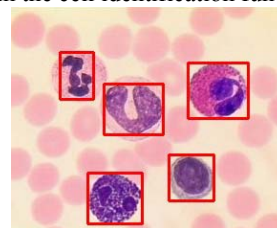


**Figure 6-18 Image with bounding rectangles**

Without processing an entire image,the Region Of Interest (ROI) can be set in that image and only that region can be processed. OpenCV implements the ROI as a rectangular square. Therefore, once the white blood cell locations are returned as a rectangle array, for each blood cell, the ROI of the image can be set, and then process only that portion which reduces the cost of operation in a great deal.

### 6.4.3    Improved Algorithm to eliminate background and segment white blood cells

Since the approach explained above to eliminate the background of an image containing white blood cells is mainly based on the RGB histogram, the method directly depends on the illumination of the image since depending on the illumination, the shape of the histogram changes. At the same time, the efficiency of the algorithm was quite low when a number of images were subjected to processing and when the image resolution is high, since for each image three histograms for channels Red, Green and Blue has to be calculated and using them, the RGB histogram should be created.

As a result, the white cell blood segmented image contained areas which were not really white blood cells, and a separate algorithm has to be written to identify and remove this erroneous segmented areas. This introduced a lot of inefficiencies to the approach.

Therefore, an algorithm that doesn't depend on the illumination of the image was in need and the following approach was implemented.
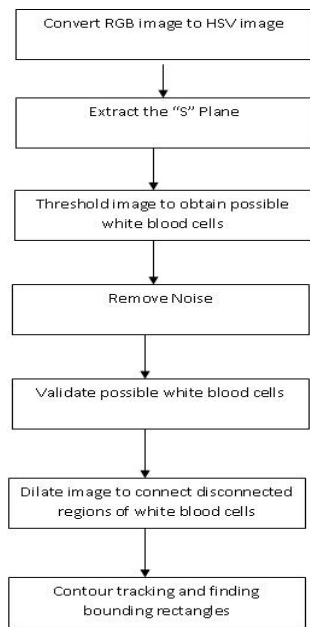
8

**Figure 6-19 Improved approach to background elimination and WBC segmentation work flow**

As explained before, the illumination of the image has to be removed from the image. Therefore, the image is converted form RGB color space to HSI (Hue, Saturation, Illumination) color space and the "Saturation" plane is extracted.
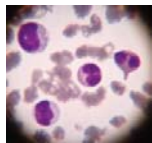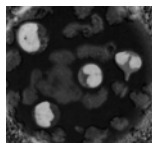


**Figure 6-20 Sample Image**



**Figure 6-21 "S" plane of the image**

As the Figure 4.2-13 shows, the white blood cells have a higher gray level than the rest in the "Saturation" plane of the HSV image. Once this is obtained the image is thresholded to eliminate background and obtain possible white blood cell segments.



**Figure 6-22 Thresholded "S" plane of the image**

Once the "S" plane is thresholded, instead of the white blood cell segments, there are noise segments present. To eliminate the noise, median filter is applied to the image. But this only removes the noisy segments of which has a smaller area compared to the area of a white blood cell. But there are noisy segments that have a relatively larger area. To remove them, every segmented area that is left after applying median filter is subjected to a validation test.

**Validating possible white blood cell segments**

As shown in figure 4.2.3-2, the nucleus of stained white blood cells has a distinctive color as a result of the chemicals applied in staining. Through statistical analysis of the nucleus areas of the images, it is found out that in each and every pixel that belongs to a nucleus of a white blood cell, the following is true;

Gray value of Red Channel > 70
Gray value of Blue Channel > 70
Gray value of Green Channel < 70

Using this statistical data each and every segmented area in the image is evaluated to check whether they actually represent a nucleus of a white blood cell. The algorithm for validation is shown below;

```
//Input is thresholded and noise removed "S" plane of the
image
#define red_threshold          70
#define green_threshold        70
#define blue_threshold                  70

foreach segmented area
   foreach pixel
      if(redChannel > red_threshold && blueChannel >
blue_threshold
         && greenChannel < green_threshold) then
                Make all channel values 0;
         //fills the area with black color
         end
   end
end
```

**Figure 6-23 Algorithm for validating the segments**



**Figure 6-24 Image after validating the segments**

As the figure above shows, due to the thresholding, the nucleus of some white blood cells are segmented in to two. To reconnect the segmented nucleolus of the white blood cells, the image is dilated. After dilation, segmented nuclei (if there's any) reconnects to form a single segment and others remain as they are.
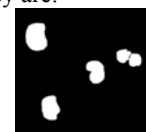


**Figure 6-25 Thresholded image after dilating the validated segments**

9

Now, only the white blood cells are segmented accurately. The resulting image is subjected to contour tracking and then the bounding rectangles are calculated. The final output is shown below.
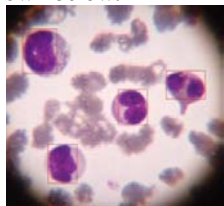


**Figure 6-26 Original image after dilating the validated segments**

As shown in the above figure, the new approach is able to segment precisely the nuclei of the white blood cells. The algorithm doesn't depend on image illumination, the size of the white blood cells or the back ground of the image, since it is based on the "Saturation" plane of the HSI image and since it uses statistical data to remove erroneous segments.

## 7    Reporting

Lohitha keeps all the configuration data, as well as the count specific data in XML format. Once the application has been installed at a particular laboratory, it is first required to configure the laboratory data. These data will be kept in the ''configLab.xml' file and it will be useful in generating the report. In addition, there are parameters such as Dilution Factor, Default Squares, the Volume above the squares in the counting chamber, and the Cell Dimension all of which would be useful in the final calculation of the cell count per micro-liter. These configuration parameters are kept in the "configParams.xml" and are used where necessary.

Before a cell count is initiated Lohitha requires the MLT to enter the patient details through the form shown below.



**Figure 7-1 Entering details necessary for performing the count**

Once the count has finished the application will automatically save the count result in a file called "{SPECIMEN_ID}.xml". This file could later be used to retrieve the information related to the count performed.

Then the MLT also has the option to generate the report. The content for the report is aggregated from several XML files and is output in HTML format with an XSL transformation.



**Figure 7-2 Blood Count Report**

## 8    RESULTS AND FINAL ANALYSIS

### 8.1    Comparison

#### 8.1.1    Standard count

The following table summarises the best result achieved out of all the images used for testing. Note that for interactive standard count have taken 10 count results on the same image with random clicks on cell centres and used statistics to interpret the result.

| Count Type | Actual Value | Lohitha Value | Percentage Error |
|---|---|---|---|
| Automatic | 74 | 78 | 0.054 % |
| Interactive | 74 | 74 (Best), 74.7 (Mean), 3.05 (SD) | 0.009 % |

**Figure 8-1 Standard Count best results**

The following table summarises the results of the tests that have been performed with 8 sample images. It has to be noted that, based on the location of the click in the Interactive Standard Count, the result would vary.

| Count Type | Mean Percentage Error |
|---|---|
| Automatic | 0.074 % |
| Interactive | 0.267 % |

**Figure 8-2 Standard Count Accuracy – mean percentage error**

#### 8.1.2    Differential Count

For the differential count, more than 100 microscopic images were analysed. In all these images cell boundaries were identified, and the invariant moments

were calculated as a first step in successfully differentiating the WBCs with an acceptable accuracy. Specifically, the results achieved were:

- Number of WBCs in the image
- Neutrophils, Eosinophils, Basophils (the "phils" category) were clearly separated from Lymphocytes and Monocytes (the "cytes" category) based on the significant differences in the nucleus
- Lymphocytes were differentiated from Monocytes
- The different cell types in the "phils" category were not able to be separated from each other properly

The following matrix shows the percentage of accuracy in which Lohitha recognizes each type of white blood cells.

| Output | Input cell identified (as a percentage) | | | | |
|--------|-----------------|-----------|----------|------------|-----------|
| Input | Neutro phils | Eosino -phils | Baso phils | Lympho -cytes | Mono- cytes |
| Neutro- phils | 85% | 14% | 0% | 0% | 1% |
| Eosino- phils | 13% | 80% | 0% | 0% | 7% |
| Baso- phils | 0% | 0% | 12% | 61% | 27% |
| Lympho- cytes | 0% | 0% | 3% | 82% | 15% |
| Mono- cytes | 0% | 2% | 3% | 18% | 77% |

**Figure 8-3 Percentage Accuracy achieved with the Differential Count**

As shown in the above table, except for Basophils, Lohitha can recognize other four types of cells with a higher percentage of accuracy.

Lohitha does the WBC recognition using artificial neural networks and currently the neural networks are trained using a sample of 130 cells. So increasing the number of samples will minimize the error in the neural net and the current accuracy can be improved by a great margin.

Since the percentage of Basophils in human blood is very low compared to other four types of cells, the sample only had 5 images of Basophils which is not enough by any standard to train an artificial neural net to identify Basophils. This is the reason for Lohitha identifying Basophils with a very low accuracy. Therefore, as explained before, increasing the number of samples to train the neural net will increase the accuracy of identification.

## 9 Evolvements

- **Train the neural network used in the Differential Count with more samples**

We worked under the constraints of the images that were available to us. However, with more sample images we believe the recognition of cells would be better.

- **Fine tuning the algorithm for Differential count**

The most advanced functionality of Lohitha is the differential count. We had to do a huge amount of research on it.

Ultimately we decided to use the moment invariants of the nucleus in the white blood cells to differentiate them. At the moment we are carrying out the research to find a more accurate and efficient method for differential count. In the next phases of development we'll be implementing these methods.

WBC differentiation features we identified through our latest research are as follows,

**Feature set 1:**
1. Difference between nucleus and cytoplasm R/B ratios:
2. "Cyan-shifted" R/B ratio of the cytoplasm
3. Cyan-factor of the cytoplasm
4. Nucleus blob area
5. Compactness-like measure of the nucleus blob
6. Steepness of the left side of green intensity histogram of the nucleus

**Feature set 2:**
1. Shifted TUB ratio of the cytoplasm
2. Shifted R/B ratio of the nucleus
3. Red-blue HUE of the cytoplasm
4. Standard deviation of nucleus green intensity
5. Standard deviation of nucleus red intensity
6. Nucleus smoothness as standard deviation of distances between nucleus centre and borders

## 10 Conclusion

Lohitha is a very domain specific application. Biological blood cells have their own distinct characteristics with regard to the shape and size of the nucleus, granularity, etc. Lohitha includes an effective and an efficient method for recognizing and counting blood cells as a practical alternative to the manual blood cell counting. It is much less tedious and very much effective than the mentioned manual method. In addition, it provides a very cost effective alternative to commercial applications.

## References

[1] Carboni, Daniela Tagliasacchi and Giorgio. BLOOD CELLS. [Online] April 1999. http://www.funsci.com/fun3_en/blood/blood.htm. Accessed: Sep 2006

[2] Leukocyte at eMedicine Dictionary. eMedicine Dictionary. [Online] http://www.emedicine.com/asp/dictionary.asp?keyword=leukocyte. Accessed: Sep 2006

[3] HemoSurf - An Interactive Hematology Atlas.
[Online] Division of Instructional Media, Institute for
Medical Education, School of Medicine, University of
Bern.
http://www.aum.iawf.unibe.ch/HemoSurf/Demo_E/La
b/count_manual.htm. Accessed: Nov 2006