CHAPTER 4 – RESULTS AND DISCUSSIONS

4.1 Dataset of Blood Smears Images

The dataset for this study consist of two different folders. One folder contains all the input images, the result of Connected Component Extraction (CCE), the result of pre-processing images. The other folder consists of all algorithms that used for processing the data.

The input images are RGB image 1280 x 960 pixels in size. The images that used were thin blood smears images from *Plasmodium falciparum* and 104 images were used for testing images. The images were provided by Eijkman Institute of Molecular Biology.

The utilized approach has several advantages including:

- Simplicity Sample images can be easily retrieved by anyone involved in the project. The images can be viewed without any specialized software.

- No information loss the image samples were not rescaled, re-compressed or modified in any other way from the original version. All information contained in the original image is preserved.

- Efficiency The individual image can only be interpreted in relation to the one that has been analyzed by expert microscopist. Therefore, it is important that both original image and the copy of the image in which the blood cell components have been marked are available as part of the dataset.

- Expendability With this simple arrangement of dataset, any addition to the data set does not mean that the previously prepared set needs to be modified.

This study was done with the following programming environments :

- Operating System : Linux Ubuntu 13.10

- Programming Language : C Programming language

- Compiler : GCC Compiler – GCC Version 44.5

Richard Kartiyanta

4.2   Image Pre-Processing

This stage purposed to convert the RGB image into grayscale image. IrfanView (Software) was used to convert the binary image to grayscale image. The image that converted using IrfanView cannot be used directly for further research. When converting the image, IrfanView add a line inside the header and to prepare the image so it can be used at next step programming, deleting the extra line is need.

4.3   Blood Cell Segmentation

The purpose of this stage was to separate all the cells as a single cell in one image. This stage began with converting grayscale image to be binary image using Otsu thresholding algorithm to obtain the binary mask of the image.

When the binary image acquired, smoothing using 3 x 3 median filtering was applied to reduce the noise of image. After the noise in the image removed, hole filling will be applied to get full area of the cell when segmentation is applied.

To prepare the image for further stage, Dual Otsu also applied to the grayscale image to get three dimensions of colour images. The colour is black for medium intensity pixel, grey for low intensity pixel and white for high intensity pixel. After the image generated, median filtering and hole filling also applied.

Connected Component Extraction (CCE) was applied. Input image for CCE is the binary image after Otsu thresholding was applied. The result can be seen in Figure 4.1. The border at each single cell image used to make the cell clear.
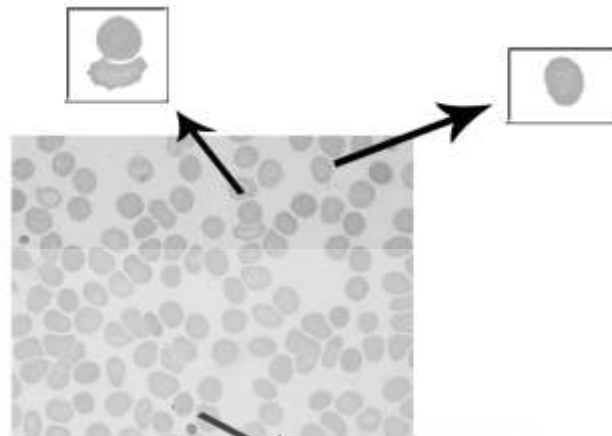
Figure 4.1 Segmentation Result

4.4   Feature Extraction

Identification of erythrocyte in the blood needs a lot of observation to get a satisfied result. The diameter of real erythrocyte is about 6-8 μm and the diameter of image is 70-100 pixels. Further result of the observation for this study is written in a log like Table 4.2. The area of erythrocyte is around 1000 to 7300 pixels and the area beneath that is overlapping cells or leucocyte. Based on the log file, infected erythrocyte and leucocyte has large number of "Range of the pixel value" like output 12 and the overlapping cell has large number of "Area of cell" like output 28.

Table 4.1 Log of Feature Extraction Result

Picture Name	Area of cell	Range of pixel value	Max. pixel value	Min. pixel value	Standard Deviation
Output 1	4099	30	169	199	5
Output 2	4411	29	171	200	3
Output 3	1400	23	173	196	4
Output 4	3203	31	165	196	5
Output 5	4861	30	166	196	5
Output 6	3975	32	164	196	6
Output 7	2288	33	163	196	6
Output 8	4722	28	170	198	5
Output 9	4674	39	160	199	5
Output 10	4966	25	171	196	4
Output 11	4885	28	168	196	5
Output 12	5314	86	116	202	8
Output 13	4460	32	164	196	4
Output 14	6032	49	150	199	4
Output 15	5045	43	153	196	9
…	…	…	…	…	…
Output 28	9389	52	154	206	6

Richard Kartiyanta

4.5   Classification

From the log file like in Table 4.1, the system will classify each component
with condition like below:

- Healthy erythrocyte will be determined if there are all the pixels founded in the
  black pixels.
- Infected erythrocyte will be determined if there are part of pixels founded in
  the black pixels, part of pixels founded in the grey pixels and the pixel area is
  below some threshold value.
- Suspected overlapping cell will be determined if there are part of pixels
  founded in the black pixels, part of pixels founded in the grey pixels and the
  pixel area is beneath some threshold value.
- Leucocyte will be determined if there are part of pixels founded in the black
  pixels, part of pixels founded in the gray pixels, more grey pixels founded than
  black pixels and the pixel area is beneath some threshold value.

The result of classification can be seen in Table 4.2. 0 means healthy
erythrocyte, 1 means infected erythrocyte, 2 is overlapping cell, 3 is
suspected overlapping cell and 4 is leucocyte.

Table 4.2 Log of Classification

Picture Name	Area of cell	Status
Output 1	4099	0
Output 2	4411	0
Output 3	1400	0
Output 4	3203	1
Output 5	4861	0
Output 6	3975	0
Output 7	2288	0
Output 8	4722	0
Output 9	4674	0
Output 10	4966	0
Output 11	4885	0
Output 12	5314	1
Output 13	4460	0
Output 14	6032	0
Output 15	5045	0
…	…	…
Output 28	9389	2

To make the observation of infected cell easier, the healthy erythrocyte cell was removed from the log and from the data that already collected this is the example of report log. The input name is 12405662t005. The image input image can be seen in Figure 4.2, the example of suspected infected erythrocyte can be seen in Figure 4.3, and the log of suspected erythrocyte can be seen in Table 4.3 below.
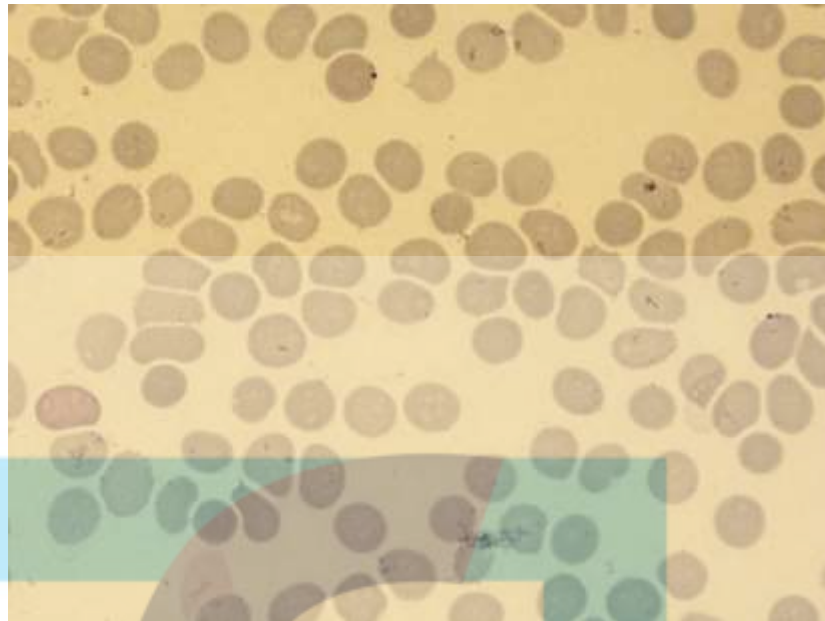
Figure 4.2 Image 12405662t005

Output 26   Output 19   Output 62   Output 55   Output 93

Figure 4.3 Suspected Cell Example (Border is Added to Clear The Region of Each Image)

Table 4.3 Log of Suspected Cell

Picture Name	Area of cell	Status
Output 10	4693	1
Output 13	5083	1
Output 19	4792	1
Output 25	4520	1
Output 26	14279	3
Output 39	5813	1
Output 40	5191	1
Output 51	5149	1
Output 55	4667	1
Output 61	6001	1
Output 62	9578	3
Output 77	5801	1
Output 80	4970	1
Output 93	12629	3
Output 97	5300	1

For image output 26, output 62, output 93, were the suspected overlapping cell that founded in image 12405662t005 and distance transform plus watershed algorithm was applied to them. After the results were generated, rule for classification was applied to each cell. The result of infected cell can be seen in Figure 4.4 and the rest of separated images were healthy erythrocyte, the images are in three colour dimensions. The classification log can be seen in Table 4.4.

Output 26-3   Output 62-2   Output 93-2   Output 93-3   Output 93-4

Figure 4.4 Result of Suspected Cell

Table 4.4 Log of Classification

Picture Name	Area of cell	Status
Output 26-3	4723	1
Output 62-2	5467	1
Output 93-2	4343	1
Output 93-3	4898	1
Output 93-4	3198	1

From Table 4.3 and Table 4.3, the distance transform and watershed
algorithm runs well for lot of image but there were error appear if the concavity
from the cell was not clear, like Figure 4.5.

Figure 4.5 Unclear Concavity Image

After the distance transform applied, the image would be like Figure 4.6. This
kind of cell was detected as one image because only one cell detected. There
is a dot at right corner of the grayscale image and because of that the status
of this image determined as infected one. In this case, two infected cells was
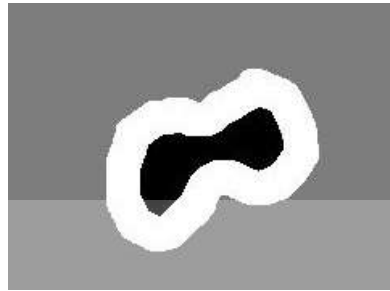counted and stored in the log.

Figure 4.6 Image After Thresholding

104 images of thin blood smears image were used to test this algorithm. The result from 104 images can be seen in the appendix. From the result, the measure of sensitivity (recall), positive predicted value (PPV or precision) and accuracy (F1) can be calculated.

$$Recall = \frac{TP}{TP + FN} \qquad (4.5 - 1)$$

$$Precision = \frac{TP}{TP + FP} \qquad (4.5 - 2)$$

$$F1 = \frac{2}{(\frac{1}{Recall}) + (\frac{1}{Precision})} \qquad (4.5 - 3)$$

From Formula (4.5 – 1, 2, 3), the accuracy for identified erythrocyte is 99.6% and for identified infected erythrocyte is 14.5%. TP or True Positive means the actual cell and the prediction are positive. FN or False Negative means the actual cell detected as positive but the prediction is negative. FP or False Positive means the actual cell detected as negative but the prediction is positive.

To calculate the F1 or the accuracy, recall and precision needed to be calculate first. Only calculating recall or precision cannot determine the accuracy of the system. Only calculating recall means the system does not considered about the false positive from the result. Otherwise, only calculating

precision means the system does not considered about the false negative from the result.

From the data that gathered in recognizing erythrocyte, total TP from 104 images is 10.153 cells. Total FN is 0 and total FP is 73. So the recall, precision and F1 were calculated like Formula (4.5 – 7, 8 ,9).

$$Recall_{Erythrocyte} = \frac{10.153}{10.153 + 0} = 1\ (100\%) \quad (4.5 - 4)$$

$$Presicion_{Erythrocyte} = \frac{10.153}{10.153 + 73} = 0.99286\ (99.3\%) \quad (4.5 - 5)$$

$$F1_{Erythrocyte} = \frac{2}{(\frac{1}{1}) + (\frac{1}{0.99286})} = 0.99641\ (99.6\%) \quad (4.5 - 6)$$

The system also calculates the accuracy in recognizing infected erythrocyte. Based on the data that gathered, total TP is 136 cells. Total FN is 0 and total FP is 1611. So the recall, precision and F1 were calculated like Formula (4.5 – 7, 8 ,9).

$$Recall_{Suspected} = \frac{136}{136 + 0} = 1\ (100\%) \quad (4.5 - 7)$$

$$Presicion_{Suspected} = \frac{136}{136 + 1611} = 0.07784\ (7.8\%) \quad (4.5 - 8)$$

$$F1_{Suspected} = \frac{2}{(\frac{1}{1}) + (\frac{1}{0.99286})} = 0.14445\ (14.5\%) \quad (4.5 - 9)$$