CHAPTER 3 – RESEARCH METHODS

3.1 Materials and Method

Thin blood smears images for this research were prepared by Eijkman Institute of Molecular Biology Indonesia. All sample images come from a lot of observation at many places in Indonesia (Kalimantan, Sabang, Papua, Sumba, and the other places). The *Plasmodium* species used in this research is *Plasmodium falciparum*.

The blood samples were gathered during several steps. First, the workers hold the patient's left hand with the palm upwards and then clean the third finger from the thumb or middle finger with alcohol. Then the worker pricked the lateral side of the finger since the parasite concentration is fairly constant in internal and peripheral blood. The worker gently squeezed the finger to drop the blood. The first drop was not used and next 2 or 3 blood drops were collected like the figure 3.1.

Figure 3.1 Taking Blood Sample (CDC et al., n.d.)

Thin blood smears were gathered by spreading the blood to run along the edge of slides by a spreader. Good quality of smears is a round smear of about 1 cm in diameter and to form that kind smears the worker slid the spreader at angle of 45^0 like in figure 3.2. Then allow the sample to dry in room temperature by 2 hours. Then dipped the slides into Giemsa staining solution with distilled water for 30-40 minutes and dry it at room temperature for about 4 days.

Figure 3.2 Spreading Blood (CDC et al., n.d.)

3.2 Proposed System

The proposed system for this research can be used to distinguish erythrocyte from the other component like leucocytes, thrombocytes and artefacts. The result of this system will be used as input image so it can be used by a system that designed for malaria identification. The scheme of algorithm used in this system can be seen in Figure 3.3.

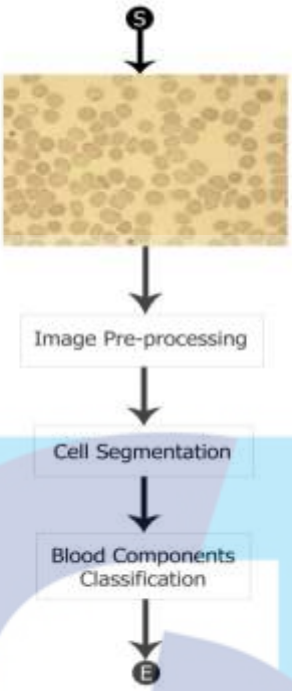

Figure 3.3 Proposed System

The input images for this system are RGB colour thin blood smears images. Image pre-processing will prepare the input image so it can be used for next step (cell segmentation). In image segmentation, all the cells are separated to be a single cell in one image. The last, system will define the single cell is healthy erythrocyte, infected erythrocyte, leucocyte, or the other components.

3.2.1 Image Pre-Processing

The purpose of image pre-processing step is to prepare the input image so it can be used for the next step. First of all, choosing the best quality image for further processing is important to get clear result from the proposed system. Because of grayscale image is more flexible to manipulate than RGB image, the RGB image which is the input image will be converted to grayscale image with windows software called IrfanView. In Figure 3.4, image pre-processing stage was shown.

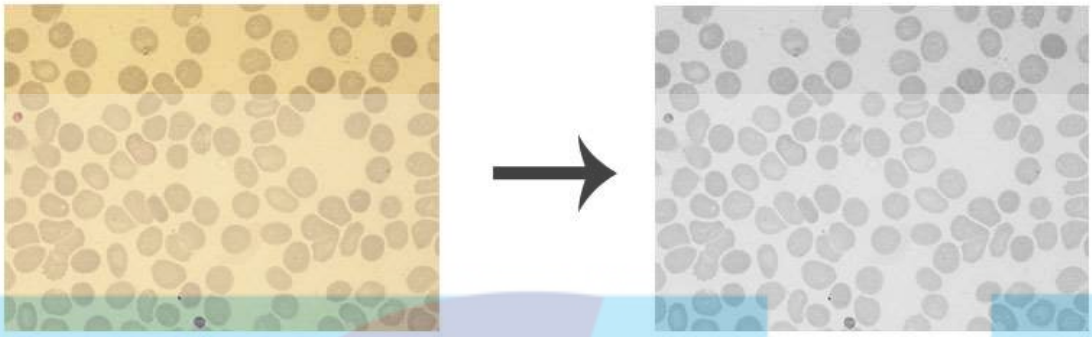

Figure 3.4 RGB Image to Grayscale Image

3.2.2 Blood Cell Component Segmentation

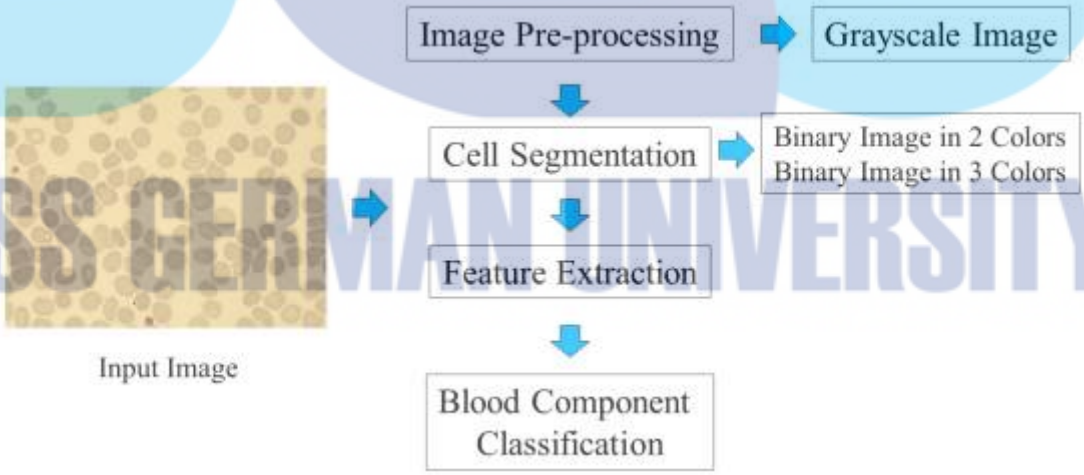

Figure 3.5 Cell Segmentation Diagram

Figure 3.5 is the big image for this system until cell segmentation step and Figure 3.6 is the flow chart of the cell segmentation step.

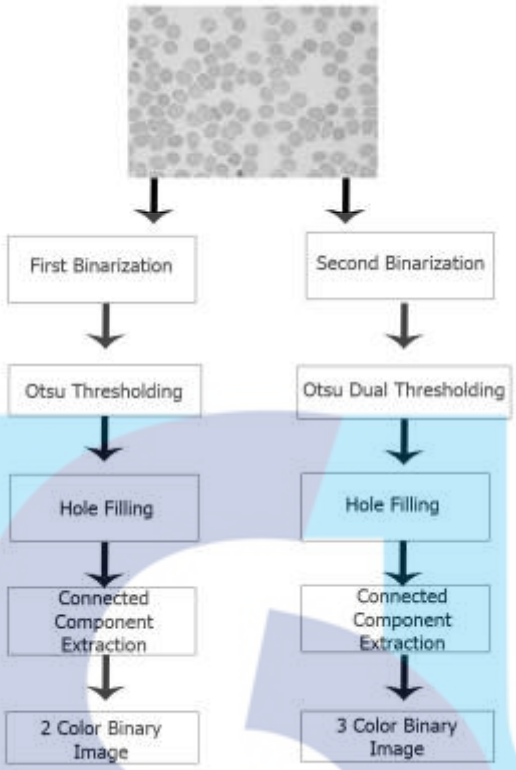

Figure 3.6 Cell Segmentation Flow Chart

The next algorithm for this system was pursued with image segmentation that identifies and segments erythrocytes, leucocytes and the other components from the background. All foreground objects were separated in order to extract the erythrocyte. Many researchers used distance transform and watershed algorithm for blood cell segmentation, which are commonly used in general image segmentation. However, in this study, the algorithm was deigned to be heavily relied on thresholding that convert the grayscale image into a binary image. Distance transform and watershed algorithm were used to handle the overlapping cell.

Figure 3.7 is the histogram of grayscale input image. It shows a representative pixel intensity distribution of the input image. The x-axis represents the pixel intensity level while y-axis represents number of pixels assigned to each pixel intensity level. It can be simply observed object and

background pixels had different regions of intensity level. *R1,* which contains all objects pixels and *R2,* which contains all background pixels. To accurately differentiate between background and object, finding the right thresholds value is essential.

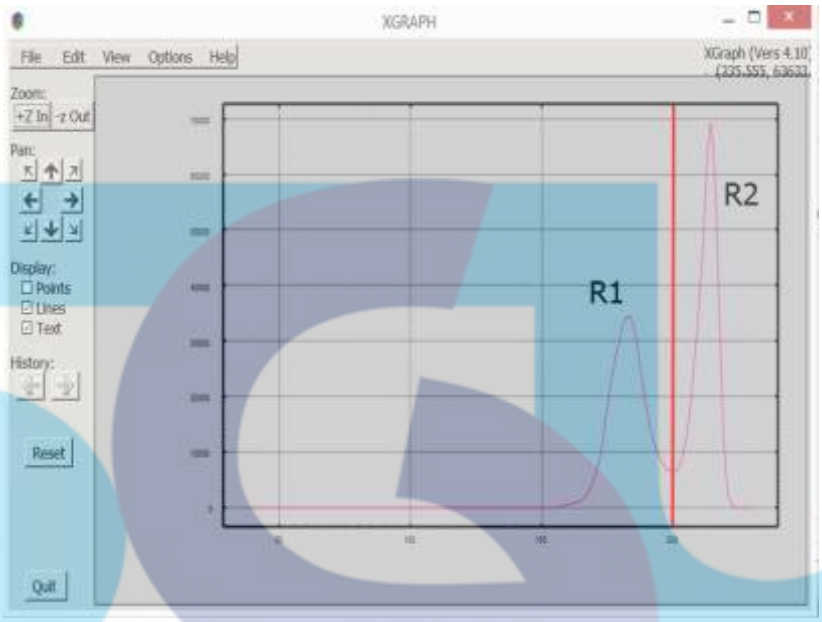

Figure 3.7 Histogram

Based on the intensity distribution, background and object can be divided by the threshold value. In this study, there were two types of thresholding used to get the threshold value. Otsu thresholding that produce single thereshold value and Dual otsu that produce two threshold values.

Basically Otsu's method is an image processing technique that can be used to convert a grayscale image into a purely binary image by calculating a threshold to split pixels into two classes. More generally, Otsu's method can be used to split a histogram into two classes which minimizes the intra-class variance of the data contained within the class.

Figure 3.8 is the result after Otsu thresholding applied to the grayscale image. The image was used for segmentation. The other algorithm, Dual Otsu

thresholding will generate an image with three dimensions of colour. The result of Dual Otsu thresholding will be used for classification stage.

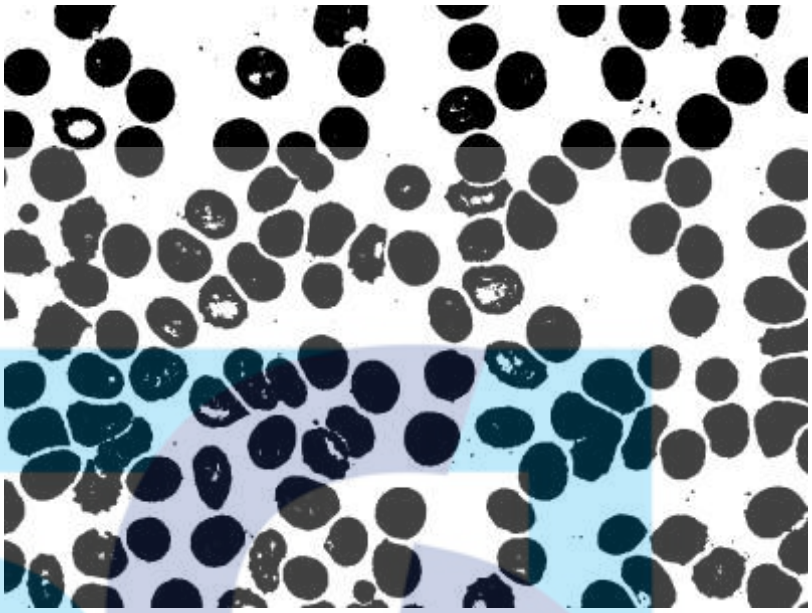

Figure 3.8 Binary Image

The result of Otsu thresholding is an image with two kinds of pixel intensity. "Connected component extraction" algorithm was applied to the binary image. "Connected component extraction" work by giving a label for each object in the image and generate new images from the label. From Figure 3.6, black was the object and white is the background. The algorithm will start by searching the first object and applied dilation at it.

First prepare two groups of arrays, the first one is for saving the temporary image and the other is the new pixels of image. The temporary is the array where the dilation result is saved. When the algorithm start, the temporary array was saved to the new array and then the dilation begin. At the end of dilation, the temporary was compared with the new one. If two of these arrays are same, then the loop is finished. Each cell image will be generated based on the label.

There must be overlapping cells that found in the result of Connected Component Extraction. To handle overlapping cell distance transform and

watershed was applied. The distance transform that suitable for pre-process the image is Euclidean distance. The result of Euclidean distance formula is the new image with new pixel intensity. The center of cell will be shown with high intensity of image (brighter) and the outer is low intensity of image (darker).

After the result from distance transform generated, otsu thresholding was applied. In this case the threshold for otsu was increased by 20%. The threshold was increased to make the brighter area (center of the cell) become smaller and larger the other area.

For example, Figure 3.9 showed the different between image with normal threshold and the image with larger threshold value. The center area of each cell is smaller and that made the search area of watershed become larger.

Figure 3.9 Image with Normal Threshold and Larger Threshold

Image with larger threshold was used as input image for watershed algorithm. In this case, watershed searched for boundary of the image and the meeting point between each cell. Figure 3.10 is the illustration of this step.

Figure 3.10 Grayscale Image, Euclidean Distance Image, Image After Thresholding, Image After Watershed

3.2.3 Feature Extraction

The purpose of feature extraction is to generate characteristic to distinguish an object with the other object. In this study, feature extraction was used to get the characteristic of healthy erythrocytes, infected erythrocytes, thrombocytes, leucocytes, artefacts and overlapping cells. Later, the result of this stage was used as the condition in determining which component is the cell.

In this stage, observation was done in two sets of features. The first feature is the area of cell, the result of this feature can be used to determine which cell is erythrocyte, leucocyte, thrombocyte, artefact, or overlapping cell. The other feature is pixel intensity levels, the result of this feature can be used to determine which cell is healthy erythrocyte or infected erythrocyte.

3.2.4 Classification

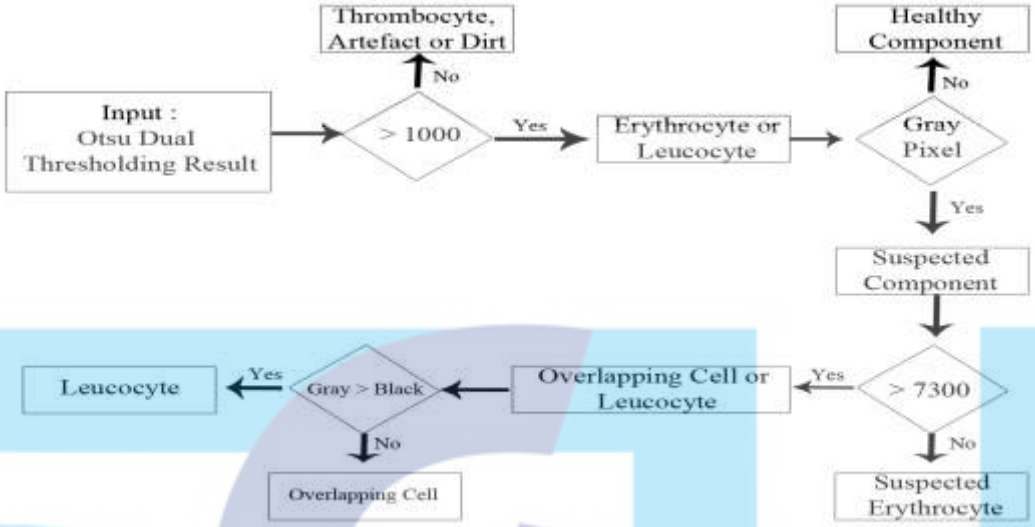

Figure 3.11 Classification Flow Chart

Figure 3.11 describe the flow of classification stage. In this stage, the system will tell which component is the cell. Is it thrombocyte, erythrocyte, leucocyte, or artefact. The algorithm that used for classification was focused at pixel intensity level.

The Dual Otsu will generate image with three kinds of colour image (grey, black and white). This image will be used to determine the components.

- 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 gray 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 gray 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 gray pixels founded than black pixels and the pixel area is beneath some threshold value.

There are three conditions in recognizing the components. First, recognize the component as erythrocyte, recognize the component as suspected erythrocyte and recognize the component as infected erythrocyte. Figure 3.11 is the example.

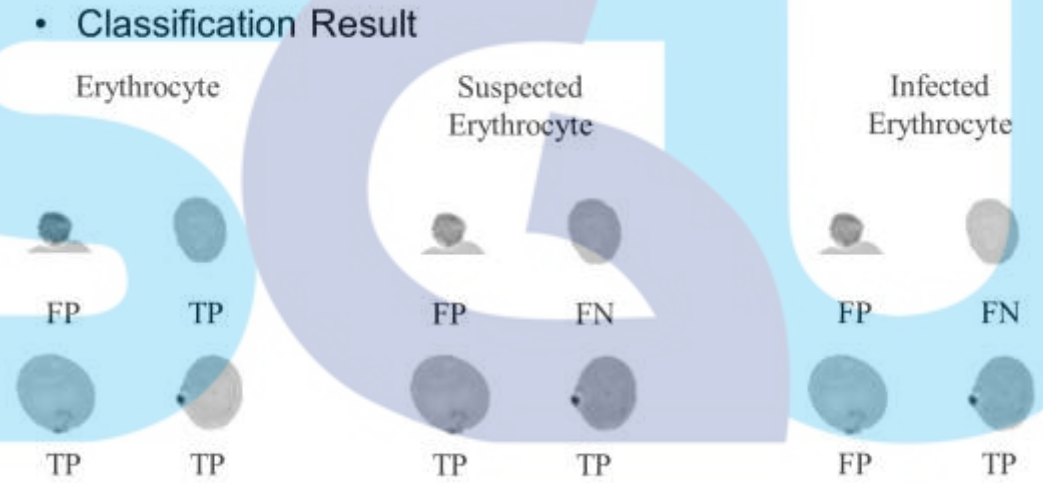

Figure 3.12 Classification Example