Cell Splitting with High Degree of Overlapping in Peripheral Blood Smear

Ngoc-Tung Nguyen, Anh-Duc Duong, and Hai-Quan Vu

Abstract—Determining the number of cell in blood sample, especially with high degree of overlapping, are remaining an obstacle need to be overcome, however. This study presents a new method for splitting (extracting) clumped cells into single (individual) cells supporting useful information for classification and detection infected cells. The proposed method is mainly focused on rapidly detecting central point using the distance transform value. Additionally, a boundary-covering degree of each center point was applied to select the best potential center points. Single cell extraction is employed in order to estimate the average size of cell. Finally, an effective algorithm was designed split correctly and speedily. As the result, the F-measure generally reaches 93.5% and more than 82% clumped cells can be tolerated in the condition of non-distorted shape and well-focused images. 1

Index Terms—Clumped cells splitting, central point detection, cell size estimation, blood smears.

I. INTRODUCTION

Malaria, a mosquito-borne infectious disease caused by a eukaryotic protest of the genus Plasmodium, is still considered a serious infectious disease in global human health [1]. According to the World Malaria Report 2008 of World Health Organization, there were estimated 250 million cases among 3.3 billion people at risk, causing nearly a million deaths. Prompt and accurate parasitological confirmation of malaria diagnosis is very important aspect in curing malaria disease. Because of limited microscopy accuracy, inaccurate dilution, loss of antigen during preparation and storage, and the natural variation in the ratio of parasite density to antigen concentration [2] also emotional problems and fatigue degrade the expert’s performance, it usually results in inaccuracy in diagnose malaria case. Therefore, an automated image analysis system would support to prevent these limitations associated with manual determination.

Blood is naturally precipitated in the ambient temperature and as blood sample is not treated well, it is very easy for red blood cells to form clumped clusters that usually result in inaccuracy in malaria diagnosis. Recently, there are several methods researching on clumped clusters splitting. These approaches and methods is all based on the Gonzalez’s fundamental knowledge [19]. These methods are conducted efficiently with simple cells such as white blood cells, especially Dorini and Liao [20] - [23]. However, as using these methods for other cells like red blood cells, new attributes and parameters appear, making the cell-separating problem become more complex and cannot be solved easily.

In the literature, clump-splitting methods are available such as: morphology operation with Di Ruberto first user [6]-[7], and then developed by Fabio Scotti [24] and Rao Mohana [25]; watershed techniques [3]-[8] and [26]-[27]; concavity analysis with Yeo [9], Wang [10], Kumar [11][12] and [28][29] using boundary information; and model-based approaches Kyoung-Mi [13], Glodia Diaz [14][15] and Sinha [30][31].

These approaches still do remained some drawbacks which are 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. Also, the concavity analysis methods is too sensitive in using threshold to recognize regions or concavity pixels along border since at lower threshold, the object will be over-split and under-split at higher threshold. Likewise, model-based approach requires not only initialization of model parameters but also considerable computational expense in matching among the variable objects with templates.

In this paper, our main objective is to propose a method to split variation of cell overlap or the clumped clusters in blood images by employing a method to detect central point rapidly. This approaching method has some advantages: (1) it could reduce the disadvantage of concavity analysis methods and model-based approaches by using neither threshold nor templates; (2) with assumption of red blood cell is disk-shaped structure, this method is able to detect other distorted structures and independent on cell size; (3) the individual cell can be fully extracted for further steps such as classification and recognition infected cell.

The paper is organized as follows: section 2 gives the detail of process of proposed method. Section 3 describes the data set and experiments; and then presents the evaluation results. Section 4 presents the conclusion of this study.

II. PROPOSED METHOD

Splitting process is followed by four stages presented in Figure 1:

1) Red blood images are decomposed and transformed into binary images where the distance information is...
extracted.

2) Cell central points are detected by using the value from the Euclidean distance transform and then boundary-covering degree is applied to these points for obtaining high potential central points.

3) Cell size estimation in which cell size are estimated through the single cells in order to get average cell size.

4) In the step of clumped cell splitting, the result from average cell size determination along with the set of potential central points, all overlapping cells are divided into individual cells.

Figure 1. The flowchart of the proposed method

A. Image decomposition

This process is to decompose original images with complex features into simpler form of binary images representing the mask of foreground. The images shown in Figure 2 are used to illustrate the process. This stage includes five steps as follows,

Because the chromatic characteristics of blood smear may change depending on capturing-time and camera setting or samples preparation, colour factors were reduced from 3-channels (Red-Green-Blue) to 1-channels (Grey-level) by the equation bellow

\[
I^{gry} = \text{average}(I^R, I^G, I^B)
\]  

(1)

In Figure 2, histogram shows a bimodal distribution in which the lower peak range represents background (darker) pixels and the higher peak range represents foreground (brighter) pixels and optimum value between two peaks is needed to distinguish between foreground and background.

With the bimodal distribution, it is assumed that two ranges (back- and foreground) are normal (Gaussian) distribution.

The Otsu’s algorithm [16] is widely used for automatic getting threshold \(\tau\). Although there are some other automatic thresholding methods, from our experiments, it showed the better result and robust to different illumination.

Figure 2. (left) the gray-scale image, (middle) its histogram, (right) its binary image

The resulting binary image denoted \(I^\text{mask}\) considering the mask of foreground objects is obtained by following equation.

\[
I^\text{mask} = \begin{cases} 
1 & I^{gry} > \tau \\
0 & \text{otherwise} 
\end{cases}
\]  

(2)

At the end of this step, morphology image processing is applied to remove some noise pixels (small and isolated pixels) and fill in the internal hole of some object since their color are similar to background in particular cases and then the region of each object is annotated by the labeling process. Note that the image may include many objects and they may contain one or more cells, which is called clumped cells. For each regions \(i\) in the image, the area \(A(i)\) are found to use in the later steps.

B. Cell center detection

Recent studies, the information from shape, size and boundary line of a circle object is used to split a cell in red blood image. However, in this paper we proposed another approach to split the red cells from blood image based on the natural feature of cell which is in examination process, detecting central point of a cell is the best way to know what the cell is. Hence, for overcoming our problem, all the information of cell above combined with central point, its particular features of the cell would be extracted easily from the blood image.

Central point is defined as the higher distance value in transformed image from the background (Figure 3). In order to extract the distance value between each pixel inside object with its boundary, Euclidean distance transform is applied to each pixel of foreground image with the following equation.

\[
t(P) = \min_{P} \{d(p, q_k) : P(q_k) = 0 \land 1 \leq k \leq S\}
\]  

(3)

Which \(P\) is the foreground image; each pixel belonging to foreground (value 1) has distance to pixel belonging to back-ground (value 0); \(d(p, q_k)\) is Euclidean distance between two points; and \(t(P) = 0\) for all pixels in background. As the values of pixel in transformed image increase continuously from the boundary of cell to center, Equation 4 is applied to find out the central point.

\[
I^c_\text{xy} = \begin{cases} 
0 \leftrightarrow 3(i, j) \in N : D_y > D_{xy} \\
1 & \text{otherwise}
\end{cases}
\]  

(4)

Define \(N\) is 8-connected-points around \((x, y)\), \(D\) is the transformed image, and \(I^c\) is the set of central point.

Figure 3. Comparison of an original image with its Euclidean distance transformation. The dark red color represents for the highest value in transforming image.

As using equation 4, many central points are obtained, some of which are the false and are not really true central points, and hence, boundary-covering degree is applied to filter and select the central point of potential interest. This degree is exactly the number of pixels in boundary of object that is matching with the internal circle whose center is the current central point. The boundary-covering degree \(W_{xy}\) is computed as following equation.
\[
W_{xy} = \sum_{(i,j):D_i} (P_{w(i)} - D_i) \text{, } D_{ij} > S, \text{ } P_{\text{max}} = D_{\text{max}}
\]  

(5)

Define \( R \) is the pixels belonging to the internal circle where the center is the point \((x, y)\) and the radius is the magnitude of this point in transforming image \( D \).

C. Cell size estimation

Average cell size is crucially necessary information for splitting clumped erythrocytes split into individual cells. Instead of determine through the user pre-definition or granulometry [6] – a statistical method using morphology operation to estimate the average value of object area, single cell in blood image is detected automatically.

It is defined by experts that each single cell need to meet following conditions: (1) it is approximately disk-shaped and (2) meets the equation \( A_R < 1.5A_F \), where \( A_R \) is a real area of object obtained by connecting neighbor pixels and \( A_F \) is the theoretical area obtained by formula \( A_F = \text{peak}^2\pi \) (\text{peak} is the maximum value of radius among the internal circle covering the object).

As this concept is applied, the proportion of true single cell detection is approximate 90% and in some case the inaccuracy happened because detected cell area is much smaller than the real one, though, this theory is preferred to estimating average cell size but combined with additional assumption that the image has an adequate number of individual cells and noise factor must be reduced.

D. Cell splitting procedures

Figure 4. The splitting process. Red color stands for the cover-ring of found single cells. From left-up to right-down images show iteration of the procedure.

At the beginning of this step, the central points of interest are sorted by descendant order of their boundary-covering degree value. According to expert definition, a single cell region is formed as a potential central point meets the following requirements:
1) Otherw Its boundary-covering degree is supposed to exceed a quarter of the boundary-covering degree of an average single cell, and,
2) Its covering area, obtained from connecting points inside the cover-ring of central point that excluding the previous detected area, needs to be greater than one third area of an average single cell.

Otherwise, this central point and other adjacent points were removed, and the next central point is continuously evaluated. In a few cases, as the area unifying by formed regions smaller than the average cell area, they will be merged together. For preventing from reassessment of evaluated regions, invalid points near examining central point are removed before proceeding to the next central point. This iterative detective process is sequentially repeated for each valid central point and the steps of splitting process are illustrated in Figure 4.

The algorithm for splitting clumped cells:

1: \( P(x,y) = \text{Sort} - \rightarrow W(x,y) \); 
2: \( \text{Set average}W = \text{the boundary-covering degree of an average single cell} \); 
3: \( \text{Set averageArea} - \text{area of an average single cell} \); 
4: for each \((x,y) \in P\)-sorted 
5: if \((x,y) = \text{invalid} - \rightarrow \text{continue} \); 
6: if \( W(x,y) < 1/4 \text{average}W \rightarrow \text{continue} \); 
7: \( \text{Area}(x,y) = \{p \in \text{Circle}(x,y), \text{radius} = D(x,y)) \} / \{ \text{recorded regions} \} \); 
8: if \( \text{Area}(x,y) < 1/3 \text{averageArea} \rightarrow \text{continue} \); 
9: \( \text{Neighbour}(x,y) = \{ p\in P \& \in \text{Circle}(x,y), \text{radius} = D(x,y)/6 \} \); 
10: \( \text{MaskOfCell} = \{ p\in \text{Circle}(\text{Neighbour}(x,y), \text{radius} = D(x,y)) \} \); 
11: Save MaskOfCell as a mask of single cell 
12: \( \text{InvalidPoint} = \{ p\in P \&\in \text{Circle}(x,y), \text{radius} = D(x,y)/3*2 \} \); 
13: \( \text{Remove} \text{InvalidPoint} \); 
14: endfor

III. EXPERIMENS AND RESULTS

A. Data set and experiment setup

The evaluation is based on difference sources of image databases including: images are supplied by Ross [17], images database from Singapore project, images are downloaded from international medical image library CDC [18], and images taken from Hospital for Tropical Diseases – Ho Chi Minh City. These are RGB color images of from 640x480 to 1600x1200 pixels resolution. All experiments were performed using the Matlab image processing. The ground-truth for the images was determined by an examiner working at least five years in malaria diagnosis lab in the hospital. Twenty images were randomly selected from the set of more than 800 images for evaluation.

B. Splitting efficiency assessment

Define \( A \) and \( Z \) are the sets of automatically and manually detected erythrocytes, respectively, then precision \( P \) and recall \( R \) is calculated by equations \( P = |A \cap Z|/|A| \) and \( R = |A \cap Z|/|Z| \). The operator \(|\cdot|\) describes the number elements of set. For assessing the balance degree of \( P \) and \( R \), \( F \)-measure is applied

\[
F = 2PR / (P + R)
\]

(6)

According to Table 1, by conducting experiment with random red blood images, based on the result of \( F \)-measure, it showed a very good balance in \( P \) and \( R \) with \( F \)-measure value ranging from around 90% at high degree of overlapping blood image to nearly 99% at other well treatment ones. Likewise, \( P \) and \( R \) value are experienced at high degree from approximate 90% to 100%, in some circumstances, \( P \) and \( R \) degree obtains relatively different value as some location of images could not be detected and poor image quality.
TABLE 1. RESULTS OF SPLITTING PERFORMANCE

| File-name | $|Z_i| \cap |Z_j|$ | $|Z_i|$ | $|Z_j|$ | $P$ (precision) | $R$ (recall) | $F$-measure |
|-----------|-----------------|-------|-------|-----------|-------------|-------------|
| g14       | 278             | 290   | 306   | 95.86%    | 90.85%      | 93.29%      |
| g15       | 272             | 280   | 288   | 97.14%    | 94.44%      | 95.77%      |
| g16       | 282             | 295   | 292   | 95.59%    | 96.58%      | 96.08%      |
| g17       | 273             | 287   | 284   | 95.12%    | 96.13%      | 95.62%      |
| g18       | 287             | 300   | 301   | 95.67%    | 95.35%      | 95.51%      |
| Fal_1     | 33              | 35    | 35    | 94.29%    | 94.29%      | 94.29%      |
| Fal_2     | 31              | 33    | 34    | 93.94%    | 91.18%      | 92.54%      |
| Fal_3     | 35              | 36    | 35    | 97.22%    | 100.00%     | 98.59%      |
| Fal_4     | 30              | 33    | 33    | 90.91%    | 88.24%      | 89.55%      |
| Fal_5     | 31              | 31    | 33    | 100.00%   | 93.94%      | 96.88%      |
| DSC01     | 95              | 104   | 103   | 91.35%    | 92.23%      | 91.79%      |
| DSC02     | 80              | 82    | 83    | 97.56%    | 96.39%      | 96.97%      |
| DSC03     | 151             | 162   | 152   | 93.21%    | 99.34%      | 96.18%      |
| DSC04     | 163             | 175   | 181   | 93.14%    | 90.06%      | 91.57%      |
| DSC05     | 135             | 147   | 145   | 91.84%    | 93.10%      | 92.47%      |
| lores1    | 69              | 74    | 71    | 93.24%    | 97.18%      | 95.17%      |
| lores1    | 39              | 41    | 39    | 95.12%    | 100.00%     | 97.50%      |
| lores1    | 77              | 81    | 83    | 95.06%    | 92.77%      | 93.90%      |
| lores1    | 76              | 79    | 80    | 96.20%    | 95.00%      | 95.60%      |
| lores1    | 71              | 73    | 77    | 97.26%    | 92.21%      | 94.67%      |

C. Overlapping degree assessment for evaluating robustness

In addition, for evaluating the robustness of the proposed method with respect to the differences of overlapping degree, randomly images were taken for testing and then splitting ability was determined individually for each cell cluster and compared to the overlapping degree as follows

$$\phi_{\text{overlap}} = \frac{B_i}{|Z_i| \mu(B_j)} \quad \forall j: \text{single cells}$$

(7)

where $|Z_i|$ is the numbers of manually detected cells in clusters, $\mu(B_j)$ is the mean of boundary pixels of single cells, and $B_j$ is the boundary pixels of cluster. The meaning of the measurement is to relate the actual boundary pixels covering a cluster to the boundary pixels of the same number of single cells. The lower $\phi_{\text{overlap}}$ is (i.e. the lower boundary pixels in a cluster), the greater overlapping degree is. (Figure 5)

More than 30 images with 269 clumped clusters including approximately 1000 red blood cells were investigated and the result is displayed in Figure 6. The scatter plot shows very good splitting performance for various clumped objects, as $\phi_{\text{overlap}}>0.65$, potency to employ true splitting is considerably high. Among the number of above clusters, there are 48 false splitting clusters and the proportion of successful splitting is up to 82%. Most splitting errors were due to cells with irregularly shaped and cells with different illumination.

IV. DISCUSSION AND CONCLUSION

Our approach gained a high result in splitting cell images with lower complexity and time-performing as compared with other processing methods reported from recent papers. These results demonstrate that the proposed technique is able to perform well with variable degrees of overlapping, as observed in clumped regions of Figure 7. Furthermore, the method is robust against hidden regions as long as they exist at least fifty percents of boundary information. Even better, distorted cells as well as their different size can be exactly split by this method.
could affect the segmentation accuracy as well. The ‘fill-hole’ method as mentioned in the process is just a heuristic method with acceptable accuracy degree. For improving the accuracy, a size filter was applied. If the value of points inside of holes is twice as compared to the average of single cells, it is determined not a hole of that single cell and is removed. Otherwise, it is filled hole. See the left clusters in the Figure 8. However, in particular cases of 3-cell or 4-cell clusters, the accuracy in hole determination is fell, as a result incorrect in splitting cells and appear some unexpected cells. See the right clusters in the Figure 8.

Figure 8. (a) Result of cell splitting from original image (the brightness of background is inconsistency from top to bottom, lighter at boundary and darker in the center) (b) Result of thresholding from original image (the information is lost at top boundary of image whereas the many noises presence at the center)

Nevertheless, the cell size estimation depends on individual cells in an image, their distribution and density must have great enough to do statistic the cell size. Due to single cells will be counted in blood image for estimating size, if the number of cells are too many or a few and/or amount of noise is considerable significant then the average single cell size will be affected. That will also impact on the next step - cell splitting.

Even though, many methods to detect the parasite in blood samples have been developed and implemented widely by many researchers such as Tek [32]-[35], Glodiaz [36], Halim[37] and [38][39], they still meet obstacles in cell segmentation, especially to separate clumped cells with high overlapping degree. We do believe with the proposed method will improve the accuracy of infected cell detection.

By conducting many experiments on different sources of images, the proposed approach can successfully split the clumped cells even though the samples are high degree of overlapping. With ability of implementing in a low-level programming to speed up the performance, the method can be applied for routine evaluation of red blood samples in laboratory.

ACKNOWLEDGEMENT

This work is part of the national key project no. KC01.16/06-10 supported by the Ministry of Science and Technology, Vietnam.

REFERENCES


[34] Tek, F. Boray and Dempster, Andrew G. and Kale, Izzet, “Malaria parasite detection in peripheral blood images”, Medical Image Understanding and Analysis (MIUA), 2006.


