Analysis of Blood Cell Detection Using Radial Symmetry and Ellipse Fitting Approaches

Hemanta Kumar Bhuyan1,*, Bodapati Sri Sai Divya2,*, Gangavarapu Sony Krishna Sree3,*
1,2,3Department of Information Technology, Vignan's Foundation for Science, Technology & Research (Deemed to be University), Guntur, Andhra Pradesh, India
*hmb.bhuyan@gmail.com, *srisaidivyabodapati@gmail.com, *sonykrishnasree@gmail.com

Abstract. This paper addresses essential hematology that is automatically segmenting blood cells. We consider a unique hybrid ellipse fitting (EF) based blood-cell segmentation technique that combines noniterative-geometric and algebraic approaches to overcome current concerns with inaccurate seed-point identification, noise, and over-segmentation. We propose a hybrid EF approach based on least-squares (LS) for improved segmentation performance. To enhance segmentation while retaining a reasonable EF, we advise utilizing the specified value for the residue offset and computing the minor and major axes using the residue and offset residue variables. Different evaluation matrices such as precision, and F1 score are used where our approach performs better than existing EF algorithms. This approach is computationally effective and might be used in cybernetics and medicine.

Keywords: Blood cell detection, residue offset, ellipse fitting (EF), hematological disorder, segmentation

1 Introduction

Involuntary segmentation is a vital stage in many image processing. Different diseases such as sickle cell anemia (SCA), acute lymphoblastic leukemia (ALL), and acute myeloid leukemia (AML) are under hematological diseases [1-4]. All gain from its application in identifying, categorizing, and diagnosing the illness. Because of SCA, red blood cells (RBCs) [1,5] can have a wide range of shapes. However, white blood cells (WBCs) are damaged by AML and ALL. Therefore, the diagnosis of hematological diseases relies heavily on cell morphology study [6].

For the segmentation of overlapping objects cells, elliptical fitting (EF) has emerged as a popular method in recent years [1,5,6,7]. To successfully segment overlapping RBCs, González-Hidalgo et al. [5] focused on identifying concave points and then employing an ellipse correction technique. Image feature data can be detected by using a few feature selection approaches [8,9]. The classification approaches can be considered from [10,11,12]. This approach demonstrates that the EF approach does not work well [13,14,15]. An excellent LS-based EF method was presented by Prasad [16] et al. According to the reported results, provides superior performance than [17]. An effective EF method (EFM) for separating overlapping objects in shadow photos was presented by Zafari [18] et al.in 2015. Bounded erosion (BE) and FRS have both proven successful in retrieving seed points. The
evidence for contours has then been gleaned through the integration of edge and seed point data \cite{19,20,21,22} to estimate contour and segment overlapping objects. The optimization model is used in \cite{23,24} for maximum segmentation. The machine learning approach is used for disease as per \cite{25}. The IOT-based approaches \cite{26,27,28} are used to transfer data through various nodes. The three main parts of the system are the EF, the extraction of contour evidence, and the detection of seed points.

1.1 Contribution

The following are the main contribution of the proposed approach.

❖ Using a modified high-boosting operation based on the Laplacian of Gaussian (LoG) for improving image quality (deblurring, sharpening, and noise reduction).
❖ For effective seed-point detection, quick radial symmetry after a restricted opening. (BO-FRS).
❖ An innovative geometric EF method in which residue and residue-offset values provide a precise approximation of major and minor axes.

The following is the article's structure. In Section 2, the suggested approach for segmenting RBCs is covered. We can see how difficult it is to compute the suggested EFM in Part 3. The performance evaluation is included in Section 4. The final section of this paper summarizes the whole paper.

2. Proposed Method

2.1 Schematic model

A schematic representation of the proposed technique for accurate blood cell segmentation is presented in Fig. 1, and the following steps are involved to process our work. Blur and undesired noise reduce the quality of the collected image, which could lead to an incorrect diagnosis \cite{1}. In this study, a modified high-boosting procedure based on the LoG filter \cite{28} is used to eliminate noise from a blurred image \cite{29}. Different methods for this process are represented as follows. When the input color image is depicted by \( C_I(x, y) \) and the preprocessed image is depicted by \( F_I(x, y) \). A weighting factor \( k \) (when \( k > 1 \)) is used in equ. (1).

\[
F_I(x, y) = C_I(x, y) + kL_G(x, y)
\]  

To efficiently improve image quality, the weight factor \( k \) should be carefully chosen. For this work, let's set \( k = 1.5 \). LoG on the complex integer \( C_I(x, y) \) is written as \( L_G(x, y) \). The standard deviation is represented by \( \sigma \), choosing appropriately to maximize image quality. In this case, we use a 7 x 7 LoG filter with a value of \( \sigma = 1 \).

\[
L_G(x, y) = \frac{-1}{\pi \sigma^4} \left(1 - \frac{x^2 + y^2}{2\sigma^2}\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)
\]

2.2 Removal of Undesired Blood Cells

The suggested approach uses morphological processes and k-means clustering-based color segmentation \cite{30,31} to enhance segmentation reliability and accuracy with a focus on removing junk cells. Methods of machine learning, such as k-means clustering, require no human intervention. Thus, we consider algorithm 1 for detecting WBC and Platelets.
Fig. 1. Processing of segmented image

Algorithm 1 (Identification of WBC)

Input: RGB Image
Output: RGB Image with WBC as well as platelets

Begin
1: Translate color images into L*x*y color space.
2: Start the number of clusters as n=6.
3: Use k-means clustering for classification based on “x*y” color space.
4: Design pixel level as per cluster index.
5: Evaluate the resultant original RGB from color-image (W).

End

Algorithm 1 is also used to isolate white blood cells and platelets from an image. Furthermore, the optimal number of clusters for the detection of WBCs is n = 6. In the final RGB color image, the isolated RBCs alone constitute a unique color channel.

2.3 Overlapping-Cell Segmentation

When a small number of cells overlap, as shown in Pseudo Code 2, this article proposes a novel method for quickly segmenting the cells involved. There are three main phases to it: (a) locating the seed point, (b) extracting evidence for the contours and (c) estimating the contours.

(a) Locating seed point: We considered the critical processes for separating overlapping cells. Combining these convex cells $C_j$ yields a grayscale image $G_l$. The mathematical formula for $G_l$ is

$$G_l = \bigcup_{j=1}^{n} C_j$$

The basic goal of adopting bounded opening (BO) is to increase each convex cell $C_j$ independence. Any $C_j$ in $G_l(k-1)$ that is connected to BO is changed into a cell $R_j$ in the $k^{th}$ iteration of BO in each $k-1$ iteration of $G_l$. $R_j$ is expressed as follows:

$$R_j = C_j^k D(0,1)$$
D(0, 1) in this case denotes a closed disc-shaped structuring element with a radius of one. A union of all earlier iterations makes up the final image in BO's kth iteration \( R_j \).

\[
G_i^{k+1} = \bigcup_j R_j \quad (5)
\]

Fast radial symmetry (FRS) is a computationally light and resource-friendly feature extraction technique. It primarily utilizes local radial symmetry (LRS) to identify potential seed points for segmentation [18]. For each pixel in gradient image I, let \( d \) range from \([R_{MIN}, R_{MAX}]\). Estimating the magnitude (Md) and orientation (Od) projection images is a primary focus of FRS. To do so, it first makes estimates of the percentages of pixels that are \( A^{+ve} \) and pixels that are \( A^{-ve} \) that were affected.

\[
A^{+ve}(x, y) = (x, y) + \text{round} \left( \frac{I(x,y)}{|I(x,y)|} \times d \right) \quad (6)
\]

\[
A^{-ve}(x, y) = (x, y) - \text{round} \left( \frac{I(x,y)}{|I(x,y)|} \times d \right) \quad (7)
\]

\[
O_d(A^{+ve}(x, y)) \rightarrow O_d(A^{+ve}(x, y)) + 1 \quad (8)
\]

\[
O_d(A^{-ve}(x, y)) \rightarrow O_d(A^{-ve}(x, y)) - 1 \quad (9)
\]

\[
M_d(A^{+ve}(x, y)) = M_d(A^{+ve}(x, y)) + |I(x, y)| \quad (10)
\]

\[
M_d(A^{-ve}(x, y)) = M_d(A^{-ve}(x, y)) - |I(x, y)| \quad (11)
\]

The calculated value of \( S_d \) is radial symmetry contribution as

\[
S_d = \frac{M_d(x,y)}{K_d} \left( \frac{O_d(x,y)}{K_d} \right) \ast H_d \quad (12)
\]

Where \( \ast \) stands for a convolution operation and \( H_d \) for a two-dimensional Gaussian. Here, \( K_d \) is radial strictness and \( d \) is a scaling factor that normalizes \( M_d \) and \( O_d \). \( \bar{O}_d \) is represented by.

\[
\bar{O}_d(x, y) = \begin{cases} 
O_d(x, y), & \text{if } O_d(x,y) < K_d \\
K_d, & \text{Otherwise}
\end{cases} \quad (13)
\]

By averaging the contributions of symmetry across all radii \( d \) in the range, the FRS-transform is calculated. \([R_{MIN}, R_{MAX}]\).

\[
S = \frac{1}{|N|} \sum_{d\in[R_{MIN},R_{MAX}]} S_d \quad (14)
\]

The projected seed points are then evaluated using centroid and symmetry zones or their average placement.

(b) Select Contour evidence: Consider the identified seed points to be represented by the set \( P = \{p_1, p_2, ..., p_n\} \). To identify how an edge point \( e_i \) in the set \( E = \{e_1, e_2, ..., e_r\} \) is related to a specific pair of seed points \( p_k \), we can use the definition of the relevance factor \( \text{rel}(e_j, p_k) \):
The variable is a parameter that varies between 0 and 1 and is used to control the influence of the distance term and the divergence term based on the association technique. The divergence term is given by $\text{div}(e, p_k)$, which measures the angle between the tangent to the contour at the edge pixel $e_j$ and the vector from $e_j$ to $p_k$. The pixel with the highest relevance value is selected as the next seed point.

$$g(x) = \begin{cases} |e_j - p_k|, & \text{if } l(e_j, p_k) \subset l_f \\ \infty, & \text{otherwise.} \end{cases}$$ (16)

Div is used to represent the difference between the direction at $e$ and the direction of $l(e_j, p_k)$ (e,p). It is written as

$$\text{div}(e_j, p_k) = \frac{l(e_j, p_k)g(e_j)}{||l(e_j, p_k)||} \|g(e_j)||$$ (17)

The edge-to-seed-point relationship plays an important role in the segmentation process as it helps to connect edge points to their corresponding seed points.

(c) Contour Measurement: The final contour is obtained by filtering out any extraneous points that are not part of the cell boundary.

$$\frac{(x-x_0)\cos \theta_0-(y-y_0)\sin \theta_0)^2}{m^2} + \frac{(x-x_0)\sin \theta_0+(y-y_0)\cos \theta_0)^2}{n^2} = 1$$ (18)

Here, $m$ is a semi-major axis in length, $n$ is a semi-minor axis in length, $\theta_0$ is the orientation angle (i.e., the angle between the x-axis and major axis), $C(x_0, y_0)$ is the center equation (18) here it represents an ellipse, provided the following conditions are satisfied:

- $C1: m, n \in \mathbb{R}^+$
- $C2: m > n$
- $C3: \theta_0 \in [0, \pi) \cup (\pi, 2\pi]$  
- $C4: x_0, y_0 \in \mathbb{R}$. (19)

The equation can be written as

$$a(x-x_0)^2 + \beta(y-y_0)^2 + \gamma(x-x_0)(y-y_0) = (m^*n)^2$$ (20)

Where,

$$a = 0.5((m^2 + n^2) - (m^2 + n^2) \cos(2\theta_0))$$
$$\beta = 0.5((m^2 + n^2) + (m^2 + n^2) \cos(2\theta_0))$$
$$\gamma = (m^2 - n^2)\sin(2\theta_0).$$ (21)

The following methods illustrate the EF mathematical model. Assume that the five-dimensional vector is $E$ with the pixel sequence $P_k(x_k', y_k') \ k = 1 : N$ and the parameters $m, n, \theta_0, x_0,$ and $y_0$ in ellipse. The vector $Y$ holds the y-coordinates of the pixels.

$$\bar{E} = [m, n, \theta_0, x_0, y_0]$$ (22)
$$Y = [\begin{array}{cccc}
-y_1^2 & -y_2^2 & \cdots & -y_N^2
\end{array}], \ Y \in \mathbb{Z}^N$$ (23)
$$H = [h_1, h_2, h_3, h_4, h_5]^T, \ H \in \mathbb{R}^5$$ (24)
$$G: E \rightarrow H; X : H \rightarrow Y.$$ (25)
The new real-valued variables are contained in the five-dimensional vector $h_1$ through $h_5$. It's used to create two new maps, $G$ and $X$, from the given nonlinear mapping $F: E \rightarrow Y$. A linear mapping between "bar $H$" and "bar $Y$," denoted by $Y = XH$, is obtained by calculating the elements $h_1$ through $h_5$.

$$Y = XH \quad (26)$$

$G$: Despite the nonlinearity of the $\bar{E} \rightarrow \bar{H}$ mapping, the nonlinearity of the resulting $Y \rightarrow H$ mapping is preserved. Nonetheless, because of the characteristics of $\text{Bar}h$ and $\text{Bar}E$, $G$ is a one-to-one mapping. Hence, $m$, $n$, $\theta_0$, $x_0$, and $y_0$ can be assessed in a variety of ways across $h_1$ to $h_5$. To evaluate the parameters $h_1$ to $h_5$ using residual distance as

$$\sum_{k=1}^{N} = ||\bar{Y} - X\bar{H}||^2 \quad (27)$$

where $||.||$ denotes the vector's Euclidean norm. To estimate the optimal solution for $\text{bar}H$, the residual distance needs to be minimized as per the following optimization problem: minimize $||EH - F||$, where $EH$ is the ellipse matrix. The solution for $\text{bar}H$ can be obtained as $\text{bar}H = \arg\min ||EH - F||$. This model aims to minimize the distance between the estimated ellipse and the actual contour points.

The summary of the EF method uses an LS-based approach to estimate the ellipse contour of each cell. The EF is further fine-tuned by adjusting the parameters $Rn$ and $Rn\prime$, which alter the length of axes as per the significant difference between $(X\bar{H} - \bar{Y})$. By combining the strengths of both methods, this approach aims to improve the accuracy and robustness of the contour estimation process.

3. Experiments

We show the results of experiments comparing the proposed method to other known cell segmentation techniques. The used datasets and performance metrics are then discussed. Finally, the method's performance is analyzed and discussed.

3.1. Data sets

The proposed method is validated using three publicly available datasets: the ALL dataset of the SCA dataset (erythrocytes DB) [32], (ALLIDB1[33]), and the AML dataset [34].

3.2. Performance Measures

For quantitative performance analysis, we place a strong emphasis on several performance metrics, including precision, recall, F1 score, Jaccard coefficient ($J_c$), and Dice similarity coefficient (DSC) [35].

$$
\text{Precision} = \frac{TP}{TP + FP} \quad (28) \\
\text{Recall} = \frac{TP}{TP + FN} \quad (29) \\
\text{F1 Score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} \quad (30) \\
J_c = \frac{|S \cap G|}{|S \cup G|} \quad (31) \\
DSC = \frac{2|S \cap G|}{|S| + |G|} \quad (32)
$$
Where \( I_g \) symbolizes the ground truth picture and \( I_s \) symbolizes the segmented image. The quantity of correctly separated pixels is expressed as true positive (TP). False negative serves as an example of the number of unnoticed pixels. The number of wrongly detected pixels (FP), as indicated, that are false positives. DSC, \( J_c \), and \( A_S \) should all be set to 1.

### 3.3 Performance analysis

The effectiveness of the suggested EF methodology with other existing methodologies is also evaluated. To ensure a more accurate and pertinent comparison, we additionally use the precise pipeline topology that is depicted in the suggested method.

(a) Disposal of White Blood Cells and Platelets: Here we show how effectively undesirable cells can be removed qualitatively. The input image shown in Fig. 2(a), is a mix of platelets and both types of blood cells that are red blood cells and white blood cells.

![Fig. 2. Segmentation of undesirable cells in the sickle-cell dataset](image)

As can be seen in Fig. 2, the technique can successfully separate these undesirable cells are separated by using color segmentation of k-means clustering 2 (b). The final, idealized image is shown in Fig. 2(c), and it consists entirely of RBCs. The precise positioning of the cells is the main objective of seed-point detection.

![Fig. 3: Performance comparison (a) Sickle cell, (b) ALL, (c) AML](image)

(b) Seed point detection: It shows how helpful seed-point detection may be in terms of quality. When compared to BE-fast radial symmetry, BO-FRS provides more accurate identification of seed sites. These illustrations show the poor performance of BE-FRS due to erroneous seed point identification, especially in high aspect ratio cells. Precision, recall, and F1 score are all better for BOFRS than BEFRS, as seen in Fig. 3.

(c) Overlapping Cells: The qualitative results of cell segmentation in the SCA dataset [31]. The quantitative effectiveness of cell segmentation in the datasets for SCA [31], ALL [32], and AML [33] is shown in Fig 4-6. The figure demonstrates that the suggested approach produces the maximum possible DSC, Jaccard score, precision, and F1 score. The hybrid EF strategy is the undisputed champion across all investigations when considering DSC, Jaccard score, precision, and F1.
In addition, it reaches its peak performance in its residue-offset-based assessment of axes. Due to the small size of the usual datasets, Deep Lab has been proven to perform poorly on datasets for hematological disorders. As a result, the suggested hybrid EF technique is a promising substitute for accurately segmenting overlapping cells in datasets for hematological disorders. It is important to note that deep learning algorithms usually require a lot of training data to function well. In contrast, the proposed hybrid EF approach does not require any training images and performs well even with a small dataset like the AML dataset. This is a significant advantage in cases where obtaining a large amount of labeled data is difficult or expensive.

4. Conclusions

The residue-offset factor is introduced to enhance the EF algorithm and improve segmentation outputs. Compared to deep-learning-based semantic segmentation methods, the proposed approach eliminates the need for training images and performs well on small datasets such as the AML dataset. This method combines noniterative geometric and
algebraic approaches, making it computationally efficient. Moreover, the proposed segmentation method is more appropriate in circumstances when there is little to no data available for training because it does not require training data. The noise issue is addressed by the modified high-boosting method based on LoG, and the suggested solution is less sensitive to ambient noise. Additionally, the technique avoids over-segmentation, and the hybrid EF-based contour estimation enables effective segmentation of overlapping cells even in visual conditions with low contrast and visual inhomogeneity.

References