Regular paper

Performance Comparison of Two Ellipse Fitting-Based Cell Separation Algorithms

Migyung Cho[*], *Member*, *KIICE*

Department of Game Engineering, Tongmyong University, Busan 48520, Korea

Abstract

Cells in a culture process transform with time and produce many overlapping cells in their vicinity. We are interested in a separation algorithm for images of overlapping cells taken using a fluorescence optical microscope system during a cell culture process. In this study, all cells are assumed to have an ellipse-like shape. For an ellipse fitting-based method, an improved least squares method is used by decomposing the design matrix into quadratic and linear parts for the separation of overlapping cells. Through various experiments, the improved least squares method (numerically stable direct least squares fitting [NSDLSF]) is compared with the conventional least squares method (direct least squares fitting [DLSF]). The results reveal that NSDLSF has a successful separation ratio with an average accuracy of 95% for two overlapping cells, an average accuracy of 91% for three overlapping cells, and about 82% accuracy for four overlapping cells.

Index Terms: Cell separation, Ellipse fitting, Splitting overlapping cells

I. INTRODUCTION

Cells in a culture process transform with time. The analysis of the behavior of cells such as stem cells from the images taken during the early stage of a cell culture process is an important research topic in the fields of biology and medicine [1, 2]. Recently, automatic cell tracking has been studied to observe cell characteristics such as apoptosis, aggregation, migration, and stretching in the cell differentiation process. In the cell culture process, two more cells may touch or overlap each other and be separated from one cell [1-4].

A reliable cell tracking system should be capable of tracking well not only in individual cells but also in overlapping cells. The performance degradation of most existing cell tracking algorithms takes place when two and more different cells are overlapped or overlapped cells are divided. These errors are frequently compounded with the misclassification of the split and merge of the overlapping cell division or fusion events, resulting in erroneous cell lineages [1-4].

However, cell objects in the cell images used in the published studies were very clear and large and had an ellipse-like shape when viewed under a high-magnification optical microscope.

studies, researchers have used images similar to Fig. 1(a) whose background is clear and cell size is large. Fig. 1(b) contains many cells and the image quality is degraded.

In this paper, we propose a separation algorithm for overlapping cells, which is based on two different ellipse fitting methods, namely the conventional direct least squares method and the improved direct least squares method, to split even multiple small overlapping cells. The proposed algorithm can be briefly described as follows:
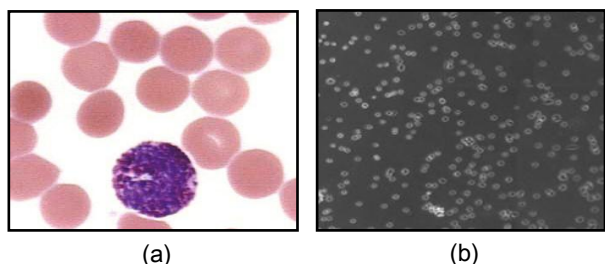
Fig. 1. Two different cell images: (a) one of cell image types which usually has been used in other researches [5] and (b) one of the images used in the present research

We used a mixed threshold method to extract a cell from two or more overlapping cells even in an unclear image with background noise and poor illumination. First, an input cell image, which is a color or a grayscale image, was converted into a binary image by using the modified mixed threshold method. The background color of the image was white, and objects such as individual cells and overlapping cells were black in color. Second, we compared the conventional ellipse fitting with the improved ellipse fitting method based on an enhanced direct least squares method for tens of thousands of multiple small overlapping cells, as shown in Fig. 1(b).

II. MIXED THRESHOLD METHOD

To separate cell objects from the background, we need the binarization process for the given cell image. Upon binarization, pixels that have a higher intensity than the threshold $t(x, y)$ are represented as objects and the other pixels are denoted as the background. The threshold function $g(x, y)$ is defined as follows:

$$g(x,y) = \begin{Bmatrix} 0 & if \ g(x,y) > t(x,y) \\ 255 & otherwise \end{Bmatrix}. \quad (1)$$

Automatic thresholding automatically selects an optimal threshold value for distinguishing objects from the background on the basis of their intensity distribution. There are two categories of automatic thresholding: global thresholding and local thresholding. Global thresholding selects a single threshold value from the histogram of the entire image. This method is simple, but it produces marginal noise for unevenly lighted images. In local thresholding, the threshold is determined for each pixel on the basis of its own gray value and the gray values of its neighborhood. This approach is called the adaptive thresholding method because it is adaptive to the local image characteristic. The adaptive threshold method has a high computational cost but gives good results even for a severely degraded image.

Most of the cell images obtained using optical microscopy are images with noise. In particular, we are interested

in a degraded cell image because of the uneven microscope light. For our image, the adaptive threshold algorithm is good, but its computational cost is very high. Therefore, we used a mixed threshold method that modified the multistage adaptive threshold method proposed by Yan et al. [6]. The mixed threshold method chooses between a global threshold and an adaptive local threshold depending on the pixel intensity.

We use a global threshold T such that pixels whose gray values are less than T are classified as the background. If a pixel has a higher intensity than T, we compute the locally adaptive threshold $t(x, y)$ for the pixel. Then, by using the adaptive threshold value $t(x, y)$, we determine whether a pixel belongs to a cell object or the background. A mixed threshold function $t(x, y)$ is defined as follows:

$$t(x,y) = \begin{Bmatrix} T & if \ g(x,y) \leq T \\ \mu_R(x,y) + k \cdot \delta_R(x,y) & otherwise \end{Bmatrix}, \quad (2)$$

where

$$\mu_R = \frac{1}{N}\sum_{(x,y)\in R} g(x,y) \quad (3)$$

and

$$\delta_R(x,y) = \sqrt{\frac{1}{N-1}\sum_{(x,y)\in R} (g(x,y) - \mu_R(x,y))^2}. \quad (4)$$

denote the mean and the standard deviation, respectively. R denotes a $b \times b$ neighborhood region centered at coordinates (x, y), b represents the mask size, and k indicates the constant coefficient. Because we consider very small overlapping cells, we used the constant 3 as the mask size. Further, k was set to a value of less than 1.

Sometimes, cells in a cell image can have holes, as shown in Fig. 1(b). This is observed when fluorescence does not stick close to the floor at the center of the cell. The next preprocessing step is to apply the flood fill algorithm to fill these holes. Finally, we extract the boundaries of the single cells and the overlapping cells. Before applying an ellipse fitting method, we split the boundaries into line segments. We use concave points on the boundaries to do this. This process involves splitting the touching cells into different ellipses, which represents separated cells. A fitted ellipse is used for representing a cell.

III. SEPARATION OF OVERLAPPING CELLS BASED ON ELLIPSE FITTING

A. Direct Least Squares Fitting of Ellipses

We used two types of ellipse fitting algorithms: direct least squares fitting of an ellipse by Fitzgibbon et al. [7]

(hereinafter Fitzgibbon's method) and numerically stable direct least squares fitting of an ellipse by Halir and Flusser [8] (hereinafter Halir's method). Both these methods are non-iterative algorithms for fitting an ellipse to a set of data points. Let us briefly overview Fitzgibbon's method.

An ellipse is a special case of a general conic, which can be described by an implicit second-order polynomial as follows:

$$F(a, x) = a \cdot x = ax^2 + bxy + cy^2 + dx + ey + f = 0, \quad (5)$$

with an ellipse-specific constraint

$$b^2 - 4ac < 0, \quad (6)$$

where $a = [a, b, c, d, e, f]^T$ denotes the coefficients of the ellipse and $x = [x^2, xy, y^2, x, y, 1]^T$ represents the coordinates of the points of the line segment. Under appropriate scaling, the inequality constraint, shown in (6), can be changed into an equality constraint as follows:

$$4ac - b^2 = 1. \quad (7)$$

The fitting of a general conic may be approached by minimizing the sum of the squared algebraic distance of points. Therefore, the ellipse-specific fitting problem can be described as follows:

$$\min_a \|Da\|^2 \text{ subject to } a^T Ca = 1. \quad (8)$$

Here, D denotes a matrix of the size $N \times 6$ that presents the coordinates of all points and C represents the constraint matrix of the size 6×6.

$$D = \begin{pmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N^2 & x_N y_N & y_N^2 & x_N & y_N & 1 \end{pmatrix}, \quad (9)$$

$$C = \begin{pmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (10)$$

By using the Lagrange multiplier λ and differentiating, we obtain the following system of simultaneous equations:

$$2D^T Da - 2\lambda Ca = 0, a^T Ca = 1. \quad (11)$$

This may be rewritten as follows:

$$Sa = \lambda Ca, a^T Ca = 1, \quad (12)$$

where S denotes the scatter matrix DTD. With the generalized eigenvectors of (12), we can get six real solutions (λ_J, a_j) for S. Further, the eigenvector of S corresponding to the minimal positive eigenvalue λ represents the best-fit ellipse for a given line segment.

All previously reported ellipse fitting-based methods that split overlapping cells are based on the direct least squares fitting of an ellipse [9, 10].

B. Numerically Stable Direct Least Squares Fitting of Ellipses

According to Halir and Flusser [8], the constraint matrix C is singular and the matrix S is also nearly singular. Therefore, the computation of the eigenvalues is numerically unstable and can produce incorrect results. In fact, when we used the generalized eigenvectors of (12), we could not obtain a minimal positive eigenvalue for a number of line segments. This implied that many of the overlapping cells could not be split into their separated cells.

To overcome these drawbacks of the original least squares method, we used Halir's numerically stable direct least squares fitting of an ellipse. By using Halir's method, we could rewrite (12) as the following set of equations:

$$\begin{aligned} Ma_1 &= \lambda a_1 \\ a_1^T C_1 a_1 &= 1 \\ a_2 &= -S_3^{-1} S_2^T a_1, \\ a &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \end{aligned} \quad (13)$$

where M denotes the reduced scatter matrix of the size 3×3.

$$M = C_1^{-1}(S_1 - S_2 S_2^{-1} S_2^T a_1), \quad (14)$$

where

$$D = (D_1 | D_2), \quad (15)$$

$$S = \begin{pmatrix} S_1 & S_2 \\ S_2^T & S_3 \end{pmatrix} \text{ where } \begin{aligned} S_1 &= D_1^T D_1 \\ S_2 &= D_1^T D_2 \\ S_3 &= D_2^T D_2 \end{aligned}, \quad (16)$$

$$C = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} \text{ where } C_1 = \begin{pmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 0 \end{pmatrix}. \quad (17)$$

Finally, we can find the eigenvector a_1 of matrix M, which yields a minimal non-negative value. The values of a_1 represent the coefficients of the best-fit ellipse of the given line segment.

C. Separation Algorithm of Overlapping Cells

The proposed algorithm is based on the ellipse fitting method to split the overlapping cells. The procedure is given

below:

Algorithm (OverlappingCellPolygon CP)

$C = \emptyset, S = \emptyset$;
find all the concave points of CP;
divide CP into different segments L_i by using the concave
point and $C = \{L_1, L_2, .. L_n\}$;
for each $L_i \in C$
    if the number of points on $L_i \leq 3$ then
     delete L_i from set C and add L_i to set S;
end for
for each $L_k \in S$
  for each $L_i \in C$
   find the fitted ellipse E_{ki} for $L_{ki} = L_k \cup L_i$;
   (call two methods: ellipse fitting method by
    Fitzgibbon and that by Halir)
   end for
  choose L_i with the best-fitted ellipse for L_k;
  delete L_k from set S;
  replace L_i with L_{ki} in set C;
end for
for each $L_i \in C$
  find the fitted ellipse E_i for L_i;
  (call two methods: ellipse fitting method by
   Fitzgibbon and that by Halir)
end.

The algorithm was implemented in Visual C++ and MATLAB.

IV. EXPERIMENTAL RESULTS

To verify the performance of the proposed algorithm, we used optical microscope images taken at intervals of 30 minutes during the cell culture process. The quality of data was degraded with severe noise and uneven microscope light. Further, there are tens of thousands of cells and overlapping cells in an image.

Fig. 2 shows the results of separating the overlapping cells. The second column shows the preprocessing results and the boundary information, and the third and fourth columns present the splitting results obtained by applying Fitzgibbon's and Halir's ellipse fitting method in the proposed algorithm, respectively. Both the methods gave good results for two overlapping cells. However, for three or more overlapping cells, the proposed algorithm using Fitzgibbon's fitting method could not find the fitted ellipses in many overlapping cells. The third and fourth columns show the erroneous results obtained by an unstable computation of the eigenvalues such as infinite or complex numbers.
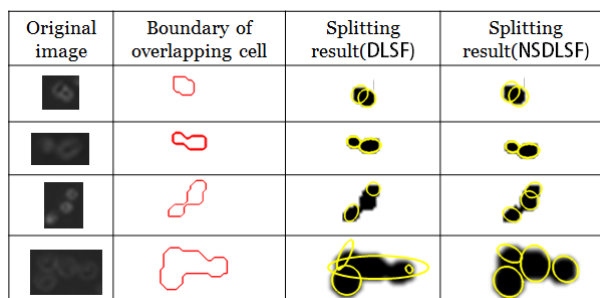

Fig. 2. Separation results of overlapping cells (1st column: original images, 2nd column: its boundary, 3rd column: Fitzgibbon's fitting method, and 4th column: Halir's fitting method). DLSF: direct least squares fitting, NSDLSF: numerically stable direct least squares fitting.

Table 1. Performance of the proposed algorithms (unit: %)

No. of overlapping cells	Proportion	Accuracy (ellipse fitting)	
		Fitzgibbon et al. [6]	Halir and Flusser [7]
Two overlapping cells	64	91	95
Three overlapping cells	20	84	91
Four overlapping cells	13	73	82

Proportion = (current overlapping cells / all overlapping cells) × 100.

Table 1 shows that the successful separation ratio was higher when the ellipse fitting method by Halir and Flusser [8] was used in the proposed algorithm. This is because of the fact that Fitzgibbon's fitting method produced incorrect results for three or more overlapping cells among the many overlapping cells. In the proposed algorithm using Halir's method, the overlapping cells were split with an average separation accuracy of 95%, 91%, and 82% for two overlapping cells, three overlapping cells, and four overlapping cells, respectively.

V. CONCLUSION

In this study, we proposed a separation algorithm for overlapping cells, which is based on two different ellipse fitting methods, namely the conventional direct least squares method and the improved direct least squares method. Through various experiments, the improved least squares method (numerically stable direct least squares fitting [NSDLSF]) is compared with the conventional least squares method (direct least squares fitting [DLSF]) and has proved to give a better result. The results reveal that NSDLSF has a successful separation ratio with an average accuracy of 95% for two overlapping cells, an average accuracy of 91% for three overlapping cells, and about 82% accuracy for four overlapping cells.

REFERENCES

[1] J. Fan, Y. Zhang, R. Wang, and S. Li, "A separating algorithm for overlapping cell images," *Journal of Software Engineering*, vol. 6, no. 4, pp. 179-183, 2013.

[2] R. Bise, K. Li, S. Eom, and T. Kanade, "Reliably tracking partially overlapping neural stem cells in DIC microscopy image sequences," in *Proceedings of MICCAI 2009 Workshop on Optical Tissue Image Analysis in Microscopy, Histopathology and Endoscopy (OPTIMHisE)*, London, pp. 67-77, 2009.

[3] K. Z. Mao, P. Zhao, and P. H. Tan, "Supervised learning-based cell image segmentation for p53 immunohistochemistry," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1153-1163, 2006.

[4] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," *Image and Vision Computing*, vol. 20, no. 2, pp. 133-146, 2002.

[5] The Cell: an image library [Internet], http://www.cellimageli brary.org/.

[6] F. Yan, H. Zhang, and C. R. Kube, "A multistage adaptive thresholding method," *Pattern Recognition Letters*, vol. 26, no. 8, pp. 1183-1191, 2005.

[7] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476-480, 1999.

[8] R. Halir and J. Flusser, "Numerically stable direct least squares fitting of ellipses," in *Proceedings of the 6th International Conference in Central Europe on Computer Graphics and Visualization (WSCG)*, Plzen, Czech Republic, pp. 125-132, 1998.

[9] M. G. Cho and J. Shim, "Ellipse based cell separation algorithm for tens of thousands of overlapping cells," *International Conference on Convergence Technology*, vol. 1, no. 2, pp. 200-204, 2012.

[10] X. Bai, C. Sun, and F. Zhou, "Splitting touching cells based on concave points and ellipse fitting," *Pattern Recognition*, vol. 42, no. 11, pp. 2434-2446, 2009.

Migyung Cho

received her B.S. degree and Ph.D. in the area of design and analysis of computer algorithms from Department of Computer Science, Pusan National University, Pusan, in 1994 and 1998, respectively. She has been an associate professor at Tongmyong University in Pusan since September 2002. Her research interests include design and analysis of computer algorithms, image processing, and computer simulation of nano-bio systems.

219