Tracing and Recognition of Cell Phases

Donggang Yu
James Cook University
Information Technology
Townsville, QLD 4811
Australia

Tuan Pham
James Cook University
Information Technology
Townsville, QLD 4811
Australia

Xiaobo Zhou
Harvard Medical School
HCNR Centre for Bioinformatics
MA 02215
USA

Abstract: We present in this paper some new and efficient algorithms for the tracing and recognition of cell phases for high-content screening. The conceptual frameworks are based on the morphological structures of cells where a series of morphological structural points are established. Furthermore, we address the issue of touching cells and then propose morphological techniques for cell separation and reconstruction.

Key–Words: High-content screening, morphological structures, image contours, image tracing, image reconstruction.

1 Introduction

The tracing and recognition of cell phases using fluorescence microscopy images play an important role for any automated high-content screening that helps scientists to better understand the complex process of cell division or mitosis [1]-[4]. High content screening concerns with the tracking of cell cycle progression (interphase, prophase, metaphase, and telophase), which can be identified by measuring nuclear changes. The most difficult task of such analysis is finding different stages during cell mitosis [5]-[7]. A typical nuclear migration during cell division is shown in Fig. 1. We discuss in this paper some

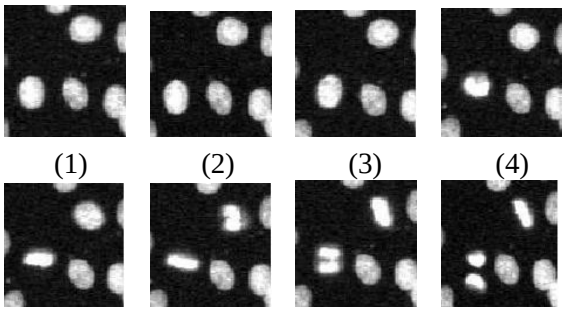


Figure 1: Part sample of one cell-cycle screening (frame time:15 minutes).

novel algorithms for the tracing and recognition of cell phases uisng several mophological models. We also address effective methods for separting and reconstructing touching cells. The rest of this paper is organized as follows. Section 2 describes the preprocessing of cell images. Section 3 presents morphological models for cell tracing and recognition. Section 4 discusses the separation and reconstruction of touching cells. Finally, Section 4 concludes the findings of our work.

2 Preprocessing of Cell Images

2.1 Smooth following and linearization

The Ostu's method [8] is used to separate binary cell images in cell-cycle screening and shown in Fig. 2. The description of image contour is important for the shape analysis and recognition of image. Many methods and algorithms are developed for the description of contours in the past [9] [10]. We introduce some efficient preprocessing algorithms. The chain code set

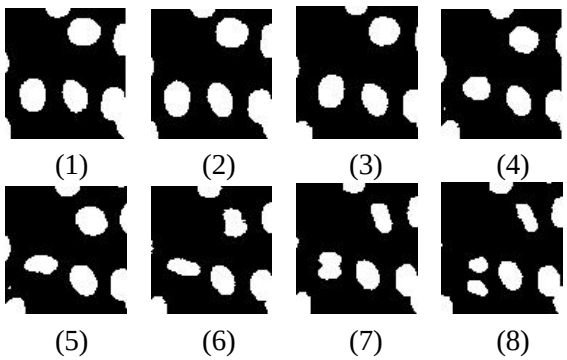


Figure 2: Binarization of images in Fig. 1.

of contour k is represented as:

$$C_k = \{c_0, c_1 ... c_i, ... c_{n-1}, c_n\} \quad (1)$$

where i is the index of the contour pixels. The difference code, d_i, is defined as:

$$d_i = c_{i+1} - c_i. \quad (2)$$

In smooth followed contours, $|d_i|$ equals 0 or 1 [11]. The smoothed contour can be converted to a set of lines which consist of ordered pixels. Suppose that the direction chain code set of the smoothed contour is

$$\{c_l^{ln}[i] \quad (i = 0, ...(n_l^{ln} - 1))\}, \tag{3}$$

where ln is the ln-th line of a smoothed contour and n_l^{ln} is the number of points of the ln-th line. A linearized line has the following property: [11]
if

$$d_{ij} = c_l^{ln}[i] - c_l^{ln}[j] \quad (i = 0, ...k-1), (j = 0, ...k-1), \tag{4}$$

then

$$\mid d_{ij} \mid \leq 1 \quad (i = 0, ...k-1), (j = 0, ...k-1). \tag{5}$$

Therefore, a linearized line contains only two elements whose chain codes meet Equation (5). Two element codes of the linearized line are represented by $cdir1$ and $cdir2$ respectively [11]. The contours in Fig. 1 are smoothly followed, linearized, and shown in Fig. 3 where the spurious points in contours are removed and character "Y" is the first point of each linearized line.

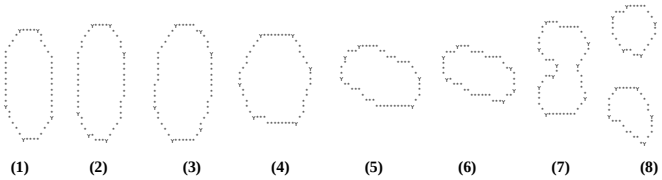


Figure 3: Smooth following and linearization of contours (bottom-left cell in the cell-cycle screening) (refer to Fig. 2).

2.2 Structural points of smoothed contours

The structural points are defined and detected based on the structure patterns of element codes of two lines [11]. Assume that $line[ln]$ is the current line and that $line[ln-1]$ is the previous line.

Definition 1. The convex point in the direction of code 4 (represented with the character "∧")
If the element codes 3, 4 and 5 occur successively as a group of neighborhood linearized lines, then one convex point can be found as follows:
if $cdir1$ of $line[ln]$ is code 4, $cdir2$ is code 5 and the direction chain code of the last pixel of $line[ln-1]$ is code 3, then the first pixel of the current line $line[ln]$ is a convex point which is represented with "∧".

Definition 2. The concave point in the direction of code 4 (represented with the character "m")
If the element codes 5, 4 and 3 occur successively as a group of neighborhood linearized lines, then one concave point can be found as follows:
if $cdir1$ of $line[ln]$ is code 4, $cdir2$ is code 3 and the direction chain code of the last pixel of $line[ln-1]$ is code 5, then the first pixel of the current line, $line[ln]$, is a concave point which is represented with "m". Similar to Definitions 1-2, other structural points, con-

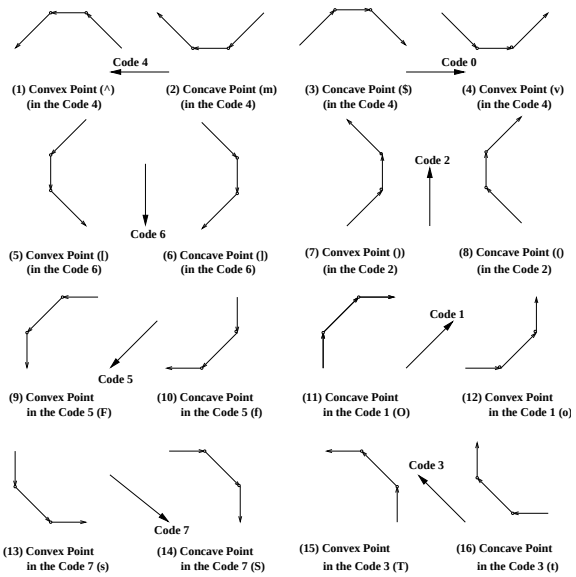


Figure 4: Structural patterns of structural points.

vex points "v", "[', ")", "F", "o", "T", "s" and concave points "$", "]", "(", "f", "O", "t" and "S" can be defined, found and shown in Fig. 4 respectively. These structural points describe the convex or concave change in different chain code directions along the contour, and they can therefore be used to represent the morphological structure of contour regions.
The series of structural points of cell images (see Fig. 3) can be found and shown in Fig. 5 based on the above algorithm For the outer contour in Fig. 5(1), the series of structural points is:
"∧" → "F" → "["→ "s" → "v"→ "o"→ ")" → "T".
It is clear that the contour shape is convex polygon, where it can be approximately defined as a ellipse.
For the outer contour in Fig. 5(7), the series of structural points is:
"∧" → "F" → "[" → "s"(convex) → "S" → "]" → "f" (concave) → "F"→ "["→ "s" → "v"→ "o"→ ")" (convex) → "(" (concave) → ")"→ "T" (convex).
It is clear, that the above contour shape consist of of two pairs of convex and concave change.

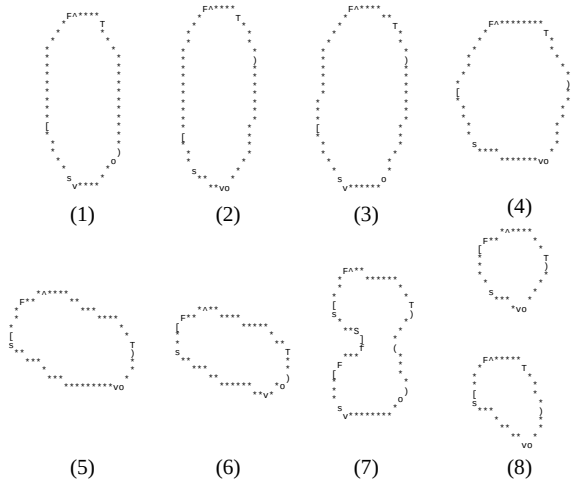


Figure 5: Cell morphological structures of different phases (refer to Fig. 3).

3 Tracing and recognition of different cell phases

3.1 Morphological model of cells

In order to trace progresses of cells, it is necessary to recognize the cell shape of different cell phases. There mainly are two classes of cell shapes, ellipse and barbell, for one cell in our cell-cycle screening. The ellipse shapes can be three types, skew, horizontal and vertical. Some recognition models of cell shapes can be described as following based on their morphological structures:

If a cell shape is an ellipse, there are no concave structural points on the outer contour of the cell contour. Furthermore, two models of ellipse can be described as following.

Morphological model 1: Ellipse shapes $e_{(5,1,2,6)}$ and $e_{(6,2,1,5)}$.

For these shapes, the number of group of codes, (codes 5, 1, 2 and 6), is largest.

Let $c_{5,6,1,2}$ be the total number of codes 5, 6, 1 and 2, c_t be the total number of all codes, $c_{5,1}$ be the total number of codes 5 and 1, and $c_{6,2}$ be the total number of codes 6 and 2, on the outer contour of the cell image respectively. If (1) the above Condition is met; (2) its outer contour mainly consists of chain codes 5, 6, 1 and 2 ($c_{5,6,1,2} \geq \frac{1}{2}c_t$); (3) the number of chain codes 5 and 1 is more than that of chain codes 6 and 2 ($c_{5,1} \geq c_{6,2}$), then the cell image shape is recognized as the shape $e_{(5,1,2,6)}$, otherwise ($c_{5,1} < c_{6,2}$) the cell image is recognized as the shape $e_{(6,2,1,5)}$.

In this model, the cell image shape, $e_{(5,1,2,6)}$, is a skew ellipse in the direction of code 5 or 1, and the cell image shape, $e_{(6,2,1,5)}$, is a vertical ellipse mainly.

Morphological model 2: Ellipse shapes $e_{(7,3,0,4)}$ and $e_{(0,4,3,7)}$.

For these shapes, the number of group of codes, (codes 7, 3, 0 and 4), is largest.

Let $c_{7,0,3,4}$ be the total number of codes 7, 0, 3 and 4, c_t be the total number of all codes, $c_{0,4}$ be the total number of codes 0 and 4, and $c_{7,3}$ be the total number of codes 7 and 3, on the outer contour of the cell image respectively. If (1) the above Condition is met; (2) its outer contour mainly consists of chain codes 7, 0, 3 and 4 ($c_{7,0,3,4} \geq \frac{1}{2}c_t$); (3) the number of chain codes 7 and 3 is more than that of chain codes 0 and 4 ($c_{7,3} \geq c_{0,4}$), then the cell image is recognized as the ellipse shape $e_{(7,3,0,4)}$, otherwise ($c_{7,3} < c_{0,4}$) the cell image is recognized as the shape $e_{(0,4,3,7)}$.

In this model, the cell image shape, $e_{(7,3,0,4)}$, is considered as a skew ellipse in the direction of code 7 or 3, and the cell image shape, $e_{(0,4,7,3)}$, a horizontal ellipse.

Morphological model 3: Ellipse shapes $e_{(6,2,7,3)}$ and $e_{(7,3,2,6)}$.

For these shapes, the number of group of codes, (codes 6, 2, 7 and 3), is largest.

Let $c_{6,7,2,3}$ be the total number of codes 6, 7, 2 and 3, c_t be the total number of all codes, $c_{6,2}$ be the total number of codes 6 and 2, and $c_{7,3}$ be the total number of codes 7 and 3, on the outer contour of the cell image respectively. If (1) the above condition is met; (2) its outer contour mainly consists of chain codes 6, 7, 2 and 3 ($c_{6,7,2,3} \geq \frac{1}{2}c_t$); (3) the number of chain codes 6 and 2 is more than that of chain codes 7 and 3 ($c_{6,2} \geq c_{7,3}$), then the adjunctive segment is recognized as the ellipse shape $e_{(6,2,7,3)}$, otherwise ($c_{6,2} < c_{7,3}$) the cell image is recognized as the shape $e_{(7,3,2,6)}$.

Based on the above recognition models, the cell image in Fig. 5(1) is recognized as shape $e_{(6,2,7,3)}$, the cell images in Figs. 5(2-3) are recognized as shape $e_{(6,2,1,5)}$ and the cell images in Figs. 5(5-6) is recognized as ellipse shape $e_{(0,4,3,7)}$ respectively.

Morphological model 4: Barbell shapes.

If (1) there are two concave structural changes; (2) there is one pair of corresponding concave structural points, "∧" and "$" (horizontal), "]" and "(" (vertical), "f" and "O" (skew), or "t" and "S" (skew), then cell image contour can be recognized as the barbell shape. The cell image contour in Fig. 5(7) can be recognized as a barbell shape, and its pair of corresponding concave structural points are structural points "]" and "(". Also, other morphological structures of cell images can be described.

3.2 Cell phases

The different phases of cell are determined based on the shape and size of the cell image in the current

frame and those in its neighboring frames of the cell-screening. As the shape ellipse, its shape change is mainly based on the rate between major and minor axes, and area size of cell image.

Basically, if there are concave changes in the cell image and its size is big, then the cell image is touching cell image. Firstly, if the cell image's size is not too large, then different cell phases can be traced and recognized as follows:

(1) if (a) the cell shape is an ellipse; (b) the area size of cell image is large enough (threshold can be found from statistical analysis of cell-screening), then the cell is a normal changing one (metaphase). In this case, there are two phases:**(1.1)** if its changing trend is that the rate between major and minor axes of the cell image is decreased with time (compare these features between two neighboring frames) , then the phase is changing from prophase to metaphase; **(1.2)** otherwise from metaphase to telophase.

(2) if the rate between major and minor axes of the cell image is large enough (threshold can be found from statistical analysis of cell-screening), there are two phases: **(2.1)** if there are two new small cells at tracing location, then the cell has split; **(2.2)** otherwise the cell will be split soon (telophase).

(3) if the cell shapes are a barbell shape and the area size of cell image is not big, then then the cell will be split (telophase).

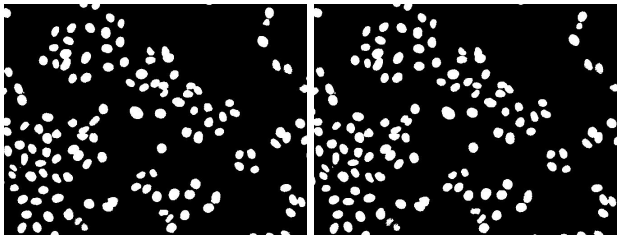
(4) if (a) the cell shape is an ellipse; (b) the area size of cell image is small enough, then the cell is a newly changing one, and its changing trend is that the area size of cell image is increased with time (compare these features between two neighboring frames) (prophase).

(5) if the cell shapes and area are not changed for long time, it should be dead cell.

Based on the above analysis, the cells in Figs. 5(1-5) are at metaphase, the cells in Figs. 5(6-7) are at telophase soon, and the cells in Fig. 5(8) are at prophase. In this way, cell-screening can be traced and recognized dynamically and automatically.

4 Separation and reconstruction of touching cells

A key problem for identifying the size and shape of the cell nuclei is that they are touching each other . For example, the images of two frames are shown in Fig. 6. We can see that the size and shape of some cells cannot be found because these cell images are touched. Therefore, we have to find which cell images are touched, how many cell images are touched, where these separation points, how the touched cell image are separated and reconstructed.



(1) (2)

Figure 6: Binary images of two frames in one cell-cycle screening (frame time:15 minutes).

4.1 Morphological structures of touching cell images

Based on the prior knowledge, the cell shape of cell-cycle screening images can approximate as an ellipse before it is divided. Therefore, if two or more cells are touched, there are is one concave structural point at least on its outer contour. Also, its size is larger than that of one cell image as touching cell image consists of two or more cells. There are nine groups of touching cell images in Fig. 6(1). Three groups of touching cell images and one telophase cell image in Fig. 6 are shown in Fig. 7. They can be smoothly followed, linearized and extracted structural points, and shown in Fig. 8 based on preprocessing algorithms. Based on the definition of structural points, one concave point means a concave change in the direction of one chain code on the contour. Let a series of concave structural points on the outer contour of touching cell images is

(1) (2) (3) (4)

Figure 7: Binarization images of three touching cell images and one telophase cell image.

$$S_{cc} = \{s_{cc}(0), s_{cc}(1)...s_{cc}(i), ...s_{cc}(n-1), s_{cc}(n)\} \quad (6)$$

where $s_{cc}(i)$ is the structural point number of the i-th concave structural point on the contour, and there are n concave structural points on the contour. It is clear that $s_{cc}(i) < s_{cc}(i+1)$. In fact, one concave change on the contour may consists of several closest concave structural points. For example, if there exists $s_{cc}(i+1) - s_{cc}(i) = 1$ and $s_{cc}(i+2) - s_{cc}(i+1) = 1$, then that means one concave change consists of three concave structural points, $s_{cc}(i)$, $s_{cc}(i+1)$ and $s_{cc}(i+2)$. In this case, these three concave structural points should be merged into one group of concave structural points. After the above merging processing for S_{cc}, a

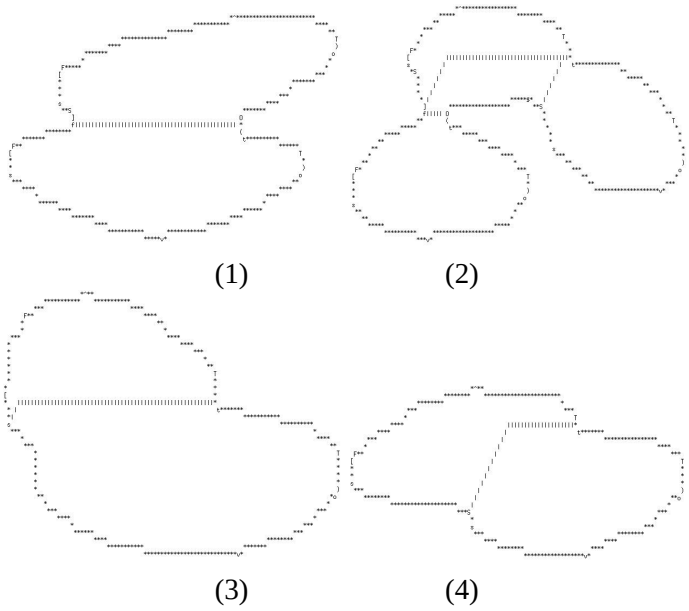


Figure 8: The extracting structural points of the images in Figs. 8(1-4).

series of groups of concave structural points (S_{cg}),

$$S_{cg} = \{s_{cg}(0), s_{cg}(1)...s_{cg}(i), ...s_{cg}(k-1), s_{cg}(k)\} \quad (7)$$

can be found, where k is the number of groups and $k \leq n$. For example, nine concave structural points in Fig. 8(2) are merged into three groups of concave structural points. The morphological patterns of touching cell images can be determined based on the number of groups of concave structural points. If $k = 1$ or $k = 2$, two cells are touched. If $k = 3$, three cells are touched. If $k = 4$, four cells are touched.

4.2 Separation points of touching cell images

The method of searching separation points can be described as follows.

Case 1 ($k = 1$):

If $k = 1$), there is one group of concave points, $s_{cg}(0)$. Suppose $s_{cg}(0)$ contains p concave points, $s_{cg0}(0), ...s_{cg0}(p-1) p < 4$. For each concave point, find its match convex structural points which are defined as its corresponding convex structural points in the approximate reverse direction of chain code. For example, if $s_{cg0}(0)$ is concave structural point "∧", then its match convex structural points are "s", "v" and "o". Let the number of the corresponding match convex structural points for all $s_{cg0}(0), ...s_{cg0}(p)$ be q, and they are represented as $s_{cv}(0), ...s_{cv}(q-1)$. We can determine separation points which make minimum distance between one pair of one concave structural points in $\{s_{cg0}(0), ...s_{cg0}(p)\}$ and one convex structural point in $\{s_{cv}(0), ...s_{cv}(q-1)\}$. That is

$$\{s_{cg0}(m), s_{cc}(n)\} = mini\{|s_{cg0}(i), s_{cc}(j)| i < p, j < q\}, \quad (8)$$

where $s_{cg0}(m)$ and $s_{cc}(n)$ are selected separation points.

Case 2 ($k = 2$):

If $k = 2$), there are two groups of concave points, $s_{cg}(0)$ and $s_{cg}(1)$. Suppose the number of concave structural points in $s_{cg}(0)$ is p_0, and in $s_{cg}(1)$ is p_1 respectively. In this case, we can determine separation points which make minimum distance between one pair of one concave structural point in $\{s_{cg0}(0), ...s_{cg0}(p_0)\}$ and one in $\{s_{cg1}(0), ...s_{cg1}(p_1)\}$. That is

$$\{s_{cg0}(m), s_{cg1}(n)\} = mini\{|s_{cg0}(i), s_{cg1}(j)| i < p_0, j < p_1\}, \quad (9)$$

where $s_{cg0}(m)$ and $s_{cg1}(n)$ are selected separation points.

Case 3 ($k > 2$):

If $k > 2$), there are more than two groups of concave points, $s_{cg}(0)$ ... $s_{cg}(l) l > 2$. In this case, we can determine each pair of separation points which make minimum distance between each pair of one concave structural point in $\{s_{cgx}(0), ...s_{cgx}(p_x)\}$ and one in $\{s_{cgy}(0), ...s_{cg1}(p_y)\}$, where $\{s_{cgx}(0), ...s_{cgx}(p_x)\}$ and $\{s_{cgy}(0), ...s_{cg1}(p_y)\}$ are neighboring groups of concave structural points. That is

$$\{s_{cgx}(m), s_{cgy}(n)\} = mini\{|s_{cgx}(i), s_{cgy}(j)| i < p_x, j < p_y\}, \quad (10)$$

where $s_{cgx}(m)$ and $s_{cgy}(n)$ are selected separation points. For example, if $k = 3$), there are three pairs of groups of concave structural points, $s_{cg}(0)$ and $s_{cg}(1)$, $s_{cg}(1)$ and $s_{cg}(2)$, and $s_{cg}(2)$ and $s_{cg}(0)$ respectively. Based on the above algorithm, we can find all separation points of images in Figs. 8(1-4). We can find related separation lines (see Figs. 8(1-4)) and the coordinate data of related arcs which are shown in Figs. 9(2,3), 10(2,3) and 11(2,3,4) based on these separation points and series of points of the contour. These contours of touching cell images are shown in Figs. 9(1), 10(1) and 11(1). The cell image in Fig. 8(4) is not touching based on the detection rule of the phase telophase because it is a barbell shape and its area is not big.

4.3 Reconstruction of touching cell images

We have found the coordinate data of all related arcs which are separated based on the above algorithm. As all cell shapes are approximately as an ellipse, touching cell images can be reconstructed based on these separated arcs. The reconstruction method is direct

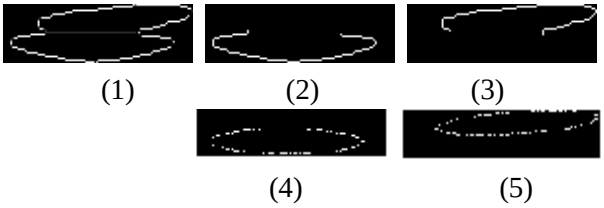


Figure 9: The contour, separated arcs and reconstructed ellipses of sample touching cell image 1.

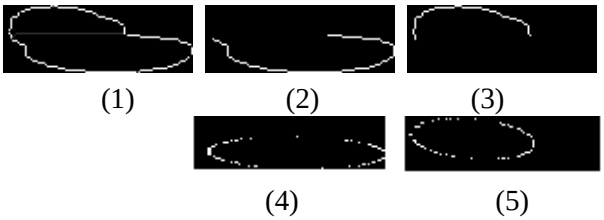


Figure 10: The contour, separated arcs and reconstructed ellipses of sample touching cell image 2.

least square fitting of ellipses [12]. The the reconstructed cell images are shown in Figs. 9(4,5), 10(4,5) and 11(5,6,7) based on the coordinate data of separated arcs respectively. These reconstructed cells can be used for tracing and recognition of different cell phases.

5 Conclusion

An efficient and new method has been developed for tracing and recognition of cell-cycle screening based on morphological structures. The morphological structures of cells are described based structural points. The different cell phases are determined in terms of cell shape recognition, cell size between neighboring frames of cell-screening. The touching cell images are recognized, segmentation points are detected, and separated cell images are reconstructed. Our algorithms have has been used to process the cell database which consists of 480 frames. Average division time is 1830 minuets (tracing the cell from prophase to telophase). All phases of cells of the database can be traced and recognized. Our method is efficient and new because morphological structure models of cell images are constructed, and these models simulate artificial intelligence.

References:

[1] S. Fox, Accommodating cells in HTS, *Drug Discovery World*, **5**, 2003, pp. 21–30.

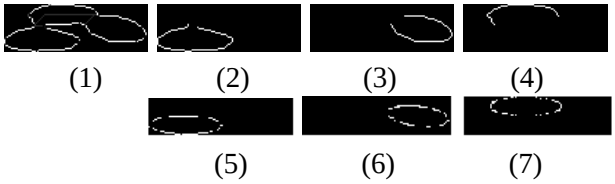


Figure 11: The contour, separated arcs and reconstructed ellipses of sample touching cell image 3.

[2] Y. Feng, Practicing cell morphology based screen, *European Pharmaceutical Review,* **7**, 2002, pp. 7–11.

[3] R. Dunkle, Role of image informatics in accelerating drug discovery and development, *Drug Discovery World,* **5**, 2003, pp. 75–82.

[4] J.C. Yarrow et al., Phenotypic screening of small molecule libraries by high throughput cell imaging. *Comb Chem High Throughput Screen,* **6**, 2003, pp. 279–286.

[5] X. Chen, X. Zhou and S.T.C. Wong, Automated segmentation, classification, and tracking cancer cell nuclei in time-lapse microscopy, *IEEE Trans. on Biomedical Engineering,* in press.

[6] T.D. Pham, D. Tran, X. Zhou, S.T.C. Wong, An automated procedure for cellphase imaging identification, *Proc. AI-2005 Workshop on Learning Algorithms for Pattern Recognition,* 2005, pp. 52–29.

[7] T.D. Pham, D.T. Tran, X. Zhou, and S.T.C. Wong, Classification of cell phases in time-lapse images by vector quantization and Markov models, *Neural Stem Cel l Research, ed. Erik V. Greer, Nova Science, New York,* 2006.

[8] N. Ostu, A thresholding selection method from graylevel histogram, *IEEE Trans. Systems Man Cybernet*, **SMC8**, 1978, pp. 62–66.

[9] F. Moktarian, A. K. Mackworth, A Theory of Multiscale Curvature-Based Shape Representation for Planer Curvature Angles, *IEEE Trans. Pattern Analysis Mach. Intell.* **14** (8), 1992, pp. 789–805.

[10] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis and Machine Vision, *Chapman & Hall Computing, Cambridge,* 1993.

[11] D. Yu, H. Yan, An efficient algorithm for smoothing, linearization and detection of structure feature points of binary image contours, *Patt. Recog.*, **30**, **(1)**, 1997, pp. 57–69.

[12] A. Fitzgibbon, A. M. Pilu, and R. B. Fisher, Direct Least Square Fitting of Ellipse, *Term Analysis and Machine Intelligence,* **21**, 1999, pp. 476–480.