

겹친 세포 분리를 위한 타원 근사 기반 알고리즘

조미경* · 심재술**

*동명대학교 · **영남대학교

An Ellipse Fitting based Algorithm for Separating Overlapping Cells

Migyung Cho* · Jaesool Shim**

*Tongmyong University · **Yeungnam University

E-mail : mgcho@tu.ac.kr* · jshim@yu.ac.kr**

요 약

광학 현미경을 통해 일정한 시간 간격으로 얻은 세포 이미지들로부터 세포의 변화 과정을 관찰하여 어떻게 변화되어 가는지 자동적으로 추적하고 분석하는 것을 자동화된 세포 트래킹이라고 한다. 본 연구에서는 수 천 개 혹은 수 만개의 세포를 하나의 이미지에 포함함으로써 크기가 매우 작아진 세포 클러스터를 분리하기 위한 타원 근사 기반의 알고리즘을 제안하고 개발하였다. 제안된 방법은 클러스터의 경계선을 추출하여 라인 세그먼트들로 분리한 다음 휴리스틱을 이용하여 라인 세그먼트들을 결합해 가며 근사 타원을 생성한다. 실험 결과 제안된 알고리즘은 두 개의 세포가 겹쳐진 클러스터의 경우 평균적으로 91%의 정확도로, 세 개의 세포가 겹쳐진 클러스터의 경우 평균적으로 84%의 정확도를 가지도 클러스터를 분리해 주었다.

ABSTRACT

An automated cell tracking system is automatically to analyze and track changes of cell behaviors in time-lapse cell images acquired from microscope in the cell culture. In this paper, we proposed and developed an ellipse fitting based algorithm for separating very small size overlapping cells in a cell image consisted of thousands or ten thousands cells. We were extracted contours of clusters and divided them into line segments and then produced their fitted ellipses for each line segment. By experimentations, our algorithm was separated clusters with average 91% precision for two overlapping cells and average 84% precision for three overlapping cells respectively.

키워드

세포 이미지 분석, 세포 트래킹, 겹쳐진 세포 분리, 타원 근사, 나노바이오

1. 서 론

세포(cell)는 배양과정에서 시간이 경과함에 따라 변화 혹은 변형되는데 광학현미경을 통해 얻은 세포 이미지에서 이러한 세포의 행동들을 분석하는 것은 생물학이나 의학에서 중요한 연구 분야 중 하나이다[1]. 일정한 시간 간격으로 광학 현미경을 통해 얻은 세포 이미지들로부터 세포의 변화 과정을 관찰하여 어떻게 변화되어 가는지 자동적으로 추적하고 세포의 행동들을 분석하는 것을 세포 트래킹이라고 한다[1]. 최근 들어 자동화된 세포 트래킹에 대한 연구가 활발하게 진행되고 있는데 이는 활용 분야가 다양하기 때문이

다. 예를 들어, 줄기 세포의 경우 환경과 자극에 따라 특정한 기능을 지닌 세포로 분화하게 되는데 줄기 세포의 성장 인자의 작용 메커니즘이나 유도 과정을 자동화된 세포 트래킹 시스템으로 관찰할 수 있다.

세포 트래킹 시스템의 정확성은 클러스터를 얼마나 정확하게 분리하였는가에 매우 의존적이다. 영상처리 분야에서 서로 붙어있는 물체를 자동적으로 분리하기 위한 다양한 알고리즘이 제안되었는데 그 중 몇 개는 세포 이미지에서 클러스터들을 몇 개의 세포로 분리하기 위한 것이다.

본 연구에서는 타원 근사법(ellipse fitting)을 이용한 세포 분리 알고리즘을 제안하고 구현하였다.

현재까지 제안되어 있는 기존의 세포분리 방법과 제안한 방법의 차이점은 처리하는 세포 이미지의 크기에 있다. 기존의 방법은 광학현미경으로부터 이미지를 촬영할 때 충분히 확대하여 촬영함으로써 세포의 모양이 타원에 가까운 모양을 가진 데이터만을 사용하였다[2,3]. 하지만 본 연구에서 제안한 알고리즘은 수 천 개 혹은 수 만개의 세포를 하나의 이미지에 포함함으로써 크기가 매우 작아진 세포 이미지에 대해서도 알고리즘을 적용할 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 광학현미경으로부터 얻은 이미지에 필요한 전처리 작업 과정을 간략하게 설명하고 3장에서는 대표적인 타원 근사 알고리즘인 direct least square 방법을 소개한다. 그리고 4장에서는 클러스터 분할을 위해 본 논문에서 제안한 타원 근사기반 알고리즘에 대해 설명할 것이다. 마지막으로 5장에서는 실험 결과를 분석하고 결론을 맺는다.

II. 전처리 작업과 타원 근사 방법

다섯 단계의 전처리 작업이 수행되었다. 첫째, 잡음을 제거하고 컬러이미지를 이진 이미지로 변환하는 과정을 거친다. 두 번째 전처리 작업은 이미지에 생성된 구멍(holes)을 메우는 작업이다. 세 번째 과정은 아포토시스(apoptosis)된 세포들을 제거하는 것이다. 네 번째 과정은 클러스터 분리를 위해 필요한 모양을 추출하는 작업이다. 마지막으로 외곽선에 대한 다각형 대략화 작업을 수행한다. 이 작업은 경계선의 모양을 매끄럽게 만들어 준다.

그림 1 (a)은 원 이미지의 일부를 (b)은 전처리 작업 후 생성된 이미지를 보여준다. 전처리 작업에 대한 더 세부적인 내용은 참고문헌 [4]을 참고하기 바란다.

타원 근사를 위한 알고리즘들 중 가장 효율적인 알고리즘은 direct least square 방법으로 알려져 있다[5]. 본 연구에서는 타원 근사를 위해 direct least square 방법을 사용하였다. 타원 근사 방법을 설명하기 위해 먼저 용어들을 정의해 보자. 세포나 클러스터 C는 한 개 이상의 라인 세그먼트 l_i 로 구성되며 l_i 는 여러 개의 점들 p_{ij} 로 구성된다. C는 m개의 라인 세그먼트로 구성되고 l_i 는 s개의 점들로 구성된다고 가정한다.

$$L = \{l_1, l_2, \dots, l_m\}$$

$$l_i = \{p_{i1}, p_{i2}, \dots, p_{is}\} \quad (1)$$

이차곡선을 나타내는 식 (2)이 $b^2 - 4ac < 0$ 조건을 만족하면 타원이 된다. Direct least square 방법의 핵심은 타원과 데이터 점들 사이의 제곱 거리를 최소화하는 매개변수 a, b, c, d, e, f의 값을 구하는 것이다.

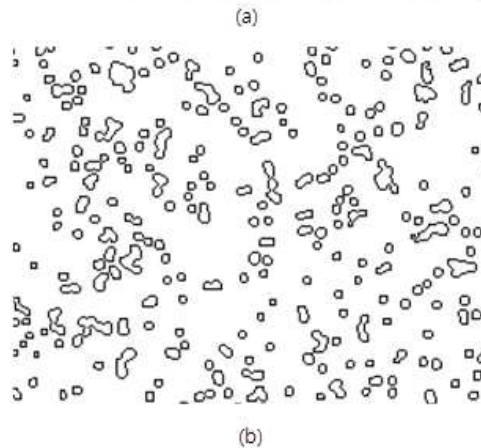
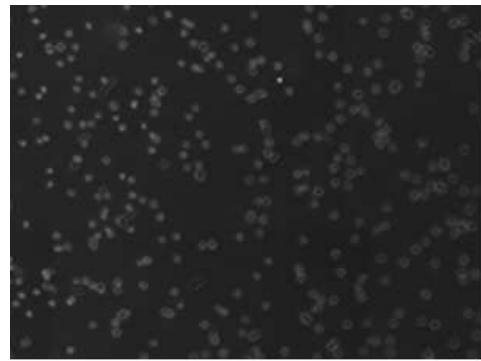


그림 1. (a) 원 세포 이미지 (b) 전처리 작업 후 세포 이미지

$$F(A, X) = A \cdot X$$

$$= ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (2)$$

여기서 $A = [a \ b \ c \ d \ e \ f]^T$ 이고 $X = [x^2 \ xy \ y^2 \ x \ y \ 1]^T$ 이다. 부등식 조건을 간단하게 풀기 위해 direct least square 방법은 부등식 조건을 $4ac - b^2 = 1$ 으로 바꾸고 점의 개수가 s인 경우 제곱 거리 $E = \sum_{i=1}^s F(X_i)^2$ 을 벡터 형태인 $E = \|DA\|^2$ 로 두고 제곱 거리를 최소로 하는 매개변수 A를 구한다. 행렬 D는 식 (3)의 s×6 행렬이며 s는 각 라인 세그먼트를 구성하는 점의 개수이다.

$$D = \begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_s^2 & x_s y_s & y_s^2 & x_s & y_s & 1 \end{bmatrix} \quad (3)$$

$4ac - b^2 = 1$ 조건은 식 (4)로 다시 쓸 수 있다.

$$A^T C A = 1 \quad (4)$$

행렬 C 는 식 (5)과 같다.

$$C = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (5)$$

Bookstein[13]는 $E = \|DA\|^2$ 을 최소로 하는 A 의 값은 식 (6)의 일반화된 고유값(eigenvalue) 수식으로 풀 수 있음을 증명했다.

$$D^T D A = \lambda C A \quad (6)$$

식 6에서 λ 는 $D^T D$ 의 고유값이다. 식 (6)을 이용하여 타원과 라인 세그먼트를 구성하는 점들 사이의 제곱 거리의 합은 식 (7)과 같이 유도할 수 있다.

$$\begin{aligned} E &= \|DA\|^2 = A^T D^T D A \\ &= A^T \lambda C A = \lambda A^T C A = \lambda \end{aligned} \quad (7)$$

식 (7)에 의해 가장 작은 양의 고유값 λ 에 대응되는 $D^T D$ 의 고유 벡터(eigenvector)가 라인 세그먼트를 구성하는 점들의 집합을 표현할 수 있는 가장 적합한 근사 타원의 매개변수 A 가 된다.

III. 클러스터 분리 알고리즘

본 연구에서 사용한 그림 1의 세포 이미지에 Xiangzhi 등이 제안한 알고리즘[2]을 적용한 결과 타원 근사 과정에서 오류가 발생하여 근사 타원을 생성하지 못하는 라인 세그먼트들이 많았다. 본 연구에서는 세포의 크기가 작은 경우에도 오류 없이 타원 근사 작업을 수행할 수 있도록 라인 세그먼트를 결합하여 근사 타원을 구해주는 클러스터 분리 알고리즘을 제안한다.

알고리즘: 타원근사를 이용한 클러스터 분리

- [단계 1] 라인 세그먼트의 집합 $L = \emptyset, S = \emptyset$ 으로 초기화한다.
- [단계 2] 클러스터를 오목 정점을 만날 때마다 라인 세그먼트 l_i 로 분리한다. 하나의 클러스터는 $L = \{l_1, l_2, \dots, l_m\}$ 로 분리된다. 모든 라인 세그먼트 l_i 에 대해, l_i 을 구성하는 점의 개수가 네 개 이하이면 l_i 을 L 에서 제거하고 집합 S 에 추가한다.
- [단계 3] 집합 S 가 공집합이 아니면 집합 S 에 속한 각 $l_i \in S$ 에 대해 단계 4-단계5를 수행하고 공집합이면 단계 6을 수행한다.

[단계 4] $l_i \in S$ 에 대해, 집합 L 에 속한 모든 라인 세그먼트 $l_j \in L$ 와 l_i 을 결합하여 $l_j \cup l_i$ 에 대한 근사 타원 E_j 을 구한다.

[단계 5] 근사 타원 E_j 와 라인 세그먼트 $l_j \cup l_i$ 의 거리 $d_j = \text{dist}(E_j, l_j \cup l_i)$ 을 구한다. 가장 작은 거리 값 d_k 을 구하여 라인 세그먼트 $l_i \in S$ 을 $l_k \in L$ 과 결합하여 새로운 라인 세그먼트 $l_k \in L$ 을 구한다. l_i 는 라인 세그먼트 집합 S 에서 제거한다. 집합 S 가 공집합이 아니면 단계 4를 수행한다.

[단계 6] 집합 L 에 있는 모든 라인 세그먼트에 대해 근사 타원을 구한다.

[단계 7] 생성된 근사 타원에 대해 *AxisRatio* 값을 계산하여 0.2 이하인 근사 타원 E_k 이 있는지 검사한다. 존재하면 E_k 에 해당되는 라인 세그먼트 L_k 을 L 에서 제거하고 집합 S 에 추가한다.

[단계 8] 집합 S 가 공집합이 아니면 단계 3을 수행하고 공집합이면 프로그램의 종료한다.

단계 5에서 $d_j = \text{dist}(E_j, l_j \cup l_i)$ 는 $l_j \cup l_i$ 을 구성하는 모든 점 p_u 과 근사 타원 E_j 사이의 거리값으로 식 (13)과 같이 계산하였다.

$$\text{dist}(E_j, L_k) = \sum_{u=0}^n F(A, p_u \in L_k) \quad (13)$$

식 (13)에서 A 는 근사 타원 E_j 의 식 2에 대한 매개변수 $[a b c d e f]^T$ 이다.

IV. 실험 결과 및 결론

제안된 알고리즘의 성능을 실험하기 위해 쥐의 3T3 세포 배양 과정에서 30분 간격으로 촬영한 광학현미경 이미지들을 사용하였다. 세포 이미지들에 대해 전처리 과정을 거친 후 두 개 이상의 세포가 겹쳐진 클러스터만을 분류하여 본 논문에서 제안한 클러스터 분리 알고리즘을 적용하였다. 사용된 세포 이미지는 그림 1과 같이 크기가 매우 작으므로 클러스터 분리 결과를 정확히 확인하기 위하여 결과를 확대하였다.

전처리 작업과 클러스터 분리 알고리즘은 모두 C++로 구현하였다. 그림 1은 클러스터 분리 알고리즘을 적용한 결과들을 보여준다. 그림 2에서 (a)는 원래 이미지를, (b)은 이진화 작업과 전처리 후 생성된 이미지를, (c)은 경계선 추출 결과를

보여주고 (d)는 타원 근사에 의한 클러스터 분리 결과를 확대하여 보여준 것이다. 원래 이미지가 선명하지 못하지만 이진화 작업과 전처리 과정을 통해 세포 정보를 정확하게 추출하였음을 볼 수 있다. 그림 2-3과 2-4의 경우 전처리 작업으로 인해 떨어져 있던 세포들이 겹쳐지는 현상이 발생하였는데 이는 전처리 과정에서 사용한 모폴로지 연산 때문이다.

실험 결과 그림 2-1과 같이 두 개의 세포가 겹쳐진 비교적 단순한 클러스터의 경우 평균적으로 91%의 정확도로 클러스터를 분리해 주었다. 그림 2-2이나 2-3과 같이 세 개인 경우 평균적으로 84%의 정확도를 보여 주었다. 서로 겹쳐진 세포의 개수가 네 개 이상으로 더 복잡해지면 정확도가 현저히 떨어졌다. 또 다른 문제점은 그림 2-5의 경우처럼 클러스터의 모양이 복잡해질수록 생성된 근사 타원의 모양에 있어서도 품질이 떨어지는 현상이 발생하는 것이다.

본 연구에서는 다른 논문에서 사용하지 않은 크기가 매우 작은 세포 이미지를 위한 클러스터 분리 알고리즘을 제안하고 개발하였다. 제안된 알고리즘은 타원 근사법을 기반으로 하며 분리된 세포의 개수에 대한 정확도가 Xiangzhi 등이 제안한 알고리즘보다 평균적으로 높았다. 하지만 다섯 개 이상의 세포가 서로 겹쳐져 있는 복잡한 경우에는 정확도와 분리된 세포의 품질에 대한 개선의 여지를 보여 주었다. 향후 이러한 단점을 개선하기 위한 연구를 계속 수행하고자 한다.

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2012-H0301-12-2002)

참고문헌

[1] Ryoma Bise et al., "Reliable Tracking Partially Overlapping Neural Stem Cells in DIC Microscopy Image Sequences," MICCAI Workshop on Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy Imperial College London, September 24, pp. 67-77, 2009.
 [2] Xiangzhi Bai, Changming Sun, Fugen Zhou, "Splitting touching cells based on concave points and ellipse fitting", Pattern Recognition 42, pp. 2434-2446, 2009.

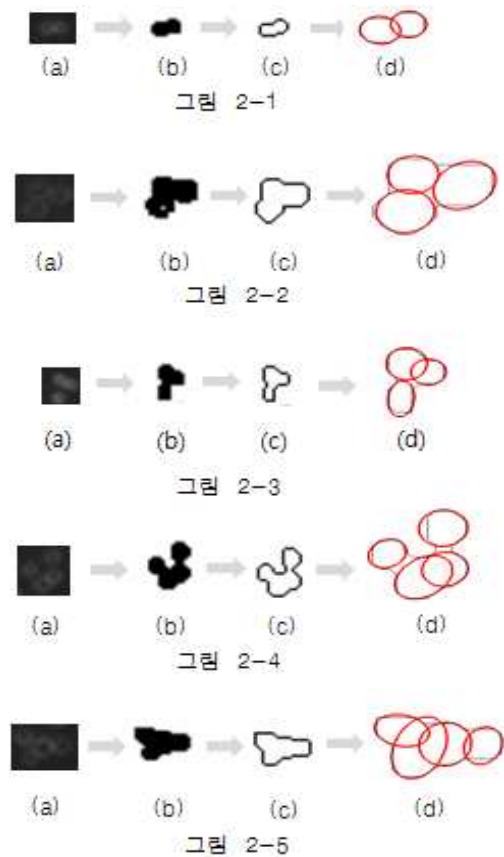


그림 2. 클러스터 분리 결과: (a)원래 이미지 (b)이진화 작업과 전처리 후 이미지 (c)경계선 추출 결과 (d)타원근사에 의한 클러스터 분리

[3]Sonal Kothar et al, "Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques," IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009.[2] Xiangzhi Bai, Changming Sun, Fugen Zhou, "Splitting touching cells based on concave points and ellipse fitting", Pattern Recognition 42, pp. 2434-2446, 2009.
 [4] 조미경, 심재술, 문상준, 김진석, "오목정점을 이용한 셀 및 클러스터 구분과 계수," 한국정보과학회 KCC2011 학술대회 논문집 C권, pp.184-187, 2011.
 [5] Andrew Fitzgibbon, Maurizio Pilu, and Robert B. Fisher, "Direct Least Square Fitting of Ellipses", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 5, pp. 476-480, 1999.