

자동 세포 추적을 위한 클러스터 세포 분리 알고리즘

조미경* · 심재솔**†

* 동명대학교 미디어공학과, ** 영남대학교 기계공학부

Cluster Cell Separation Algorithm for Automated Cell Tracking

Mi Gyung Cho* and Jaesool Shim**†

* Dept. of Media Engineering, Tongmyong Univ.,

** School of Mechanical Engineering, Yeungnam Univ.

(Received May 2, 2012 ; Revised December 24, 2012 ; Accepted December 26, 2012)

Key Words: Analysis of Cell Image(세포 이미지 분석), Cell Tracking(세포 추적), Cell Separation(세포 분리), Ellipse Fitting(타원 근사), Bio-Nano(바이오나노)

초록: 광학 현미경을 통해 일정한 시간 간격으로 얻은 세포 이미지로부터 세포 변화를 자동적으로 추적 및 분석하는 것이 세포 트래킹이라고 한다. 세포 변화 과정에서 이웃에 있는 세포들이 겹쳐져 있는 상태를 클러스터라고 하며 세포트래킹에서 클러스터를 다시 세포로 분리하는 작업은 매우 중요하다. 본 논문에서는 타원 근사법을 기반으로 클러스터를 분리하기 위한 알고리즘을 제안한다. 클러스터의 외곽선을 추출한 후 외곽선의 오목정점을 이용하여 클러스터를 라인 세그먼트들로 분리한 다음 휴리스틱을 이용하여 라인 세그먼트들을 결합해 가며 근사 타원을 생성한다. 실험 결과 두 개의 세포가 겹쳐진 클러스터의 경우 평균적으로 91%, 세 개의 세포가 겹쳐진 경우 평균적으로 84% 그리고 겹쳐진 세포의 개수가 네 개 이상인 경우 약 73%의 정확도로 클러스터를 분리해 주었다.

Abstract: An automated cell tracking system is used to automatically analyze and track the changes in cell behavior in time-lapse cell images acquired using a microscope with a cell culture. Clustering is the partial overlapping of neighboring cells in the process of cell change. Separating clusters into individual cells is very important for cell tracking. In this study, we proposed an algorithm for separating clusters by using ellipse fitting based on a direct least square method. We extracted the contours of clusters, divided them into line segments, and then produced their fitted ellipses using a direct least square method for each line segment. All of the fitted ellipses could be used to separate their corresponding clusters. In experiments, our algorithm separated clusters with average precisions of 91% for two overlapping cells, 84% for three overlapping cells, and about 73% for four overlapping cells.

1. 서론

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. 일정한 시간 간격으로 광학현미경을 통해 촬영한 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 때문이다. 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.

배양과정에서 세포는 시간이 흐름에 따라 새롭게 나타나기도, 퇴화(Apoptosis)되기도 하며 한 개 이상의 이웃한 세포가 부분적으로 겹쳐서 생성되는 클러스터(Cluster)를 만들기도 하고 클러스터는 다시 몇 개의 세포로 분리되기도 한다. 세포 트

† Corresponding Author, jshim@ynu.ac.kr

래킹 알고리즘이 신뢰성이 있기 위해서는 잘 분리된 세포뿐만 아니라 몇 개의 세포들이 부분적으로 겹쳐서 형성된 클러스터에 대해서도 필요한 분석을 정확하게 할 수 있어야 한다. 현존하는 세포 트래킹 알고리즘은 클러스터를 한 개 이상의 세포들로 분리하는 단계에서 정확한 세포 식별에 실패하므로 세포 배양의 초기 단계에서 다양한 세포의 운동성을 구별해 낼 수 없는 결과를 초래한다.⁽¹⁾ 따라서 세포 트래킹 시스템의 정확성은 클러스터를 얼마나 정확하게 분리하였는가에 의존적임을 알 수 있다.

세포의 배양 과정에서 연속적으로 얻은 이미지(형광물질이 태그되지 않은 세포 이미지 포함)로부터 싱글 단위로 세포의 수를 정확하게 분리하기 위한 알고리즘은 세포 배양 및 세포 관련 실험에 필요한 정확한 농도 등을 알아 낼 수 있고, 초기 단계의 배양에서는 세포 분열시 분열하는 세포 수 자체에 대한 정보를 얻을 수 있다. 특히 장시간 배양된 경우 세포와 세포 간의 접촉시 싱글 단위의 세포의 운동성 변화를 초래하므로 세포들 간의 접촉이 전혀 없는 세포와 세포들 간의 접촉이 있는 세포 사이의 운동성 차이에 대한 비교 등에 유용하게 사용할 수 있다.

영상 처리 분야에서 서로 붙어있는 물체를 자동적으로 분리하기 위한 다양한 알고리즘이 제안되었는데 그 중 몇 개는 세포 이미지에서 클러스터를 분리하기 위한 것이다. 제안된 알고리즘들은 모양정보를 기반으로 하는 알고리즘, Watershed 기반 알고리즘, 모폴로지 기반 알고리즘, 타원 근사를 이용한 알고리즘 등이 있다.⁽²⁻¹²⁾ 모양정보 기반 알고리즘은 세포 모양에 대한 다양한 사전 지식과 수학적 틀을 기반으로 하며 주로 클러스터 경계선(Contour)에 나타나는 오목한 부분을 분석함으로써 클러스터를 몇 개의 세포로 분리한다.⁽²⁻⁴⁾ 이 방법은 클러스터 경계선에 포함된 사소한 잡음에도 민감하게 반응함으로써 알고리즘의 성능이 매우 저하될 수 있는 단점이 있다.

Watershed 기반 알고리즘은 모양 정보 기반 알고리즘에 비해 좋은 결과를 주지만 클러스터의 영역이 넓게 분포되어 있거나 세포들이 겹쳐진 부분에서 존재하는 강도(Intensity) 기울기(Gradient)가 낮을 경우 일반적으로 실제보다 많이 분할(Over-segmentation)하거나 적게 분할(Under-segmentation)하는 문제점을 가지고 있다.⁽⁵⁾ 모폴로지(Morphology)

연산을 기반으로 한 알고리즘들도 제안되었는데 이러한 알고리즘들은 뭉쳐진 세포들에서 강도 기울기를 찾기 위해 수학적 모폴로지 연산들을 순차적으로 적용함으로써 클러스터를 분리한다.⁽⁶⁾ 이 방법도 Watershed 기반 알고리즘과 같이 뭉쳐진 세포 경계 영역의 기울기가 적으면 좋은 결과를 주지 못한다.

타원 근사를 이용한 클러스터 분리 알고리즘도 제안되었는데 앞선 다른 방법들보다 성능이 우수한 것으로 알려져 있다.⁽⁷⁻¹²⁾ 하지만 본 연구에서 실험한 결과 이 방법들은 Fig. 1과 같이 광학현미경으로부터 촬영한 세포가 타원과 유사한 모양이 될 때까지 충분히 확대한 이미지들에 대해서만 좋은 성능을 보여 주었다. 반면 촬영 시 충분히 확대하지 않으므로 Fig. 2와 같이 타원 형태가 제대로 나타나지 않는 세포 이미지에 대해서는 근사 타원을 올바르게 구해 주지 못했다. 이것은 근사 타원을 구하기 위해 고유값(eigenvalue)과 고유벡터(eigenvector)을 계산하는 과정에서 다양한 형태의 오류가 발생하기 때문이다.

본 연구에서는 Fig. 2에서 보는 것처럼 이미지 상에 있는 세포와 클러스터의 크기가 매우 작아서 제대로 된 타원 모양이 드러나지 않은 경우에도 클러스터 분리 작업이 가능한 타원 근사 기반 클러스터 분리 알고리즘을 제안하였다. 세포와 클러스터의 크기가 매우 작게 나타나는 데이터는



Fig. 1 Ellipse fitting result of an ellipse-like cluster

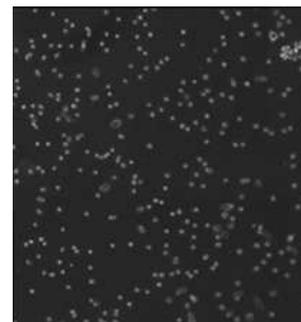


Fig. 2 Cell cluster size used in this paper

수백 개 혹은 수천 개의 세포를 하나의 이미지에 포함하여 관찰해야 되는 경우에 나타난다. 제시한 알고리즘은 데이터의 특징으로 나타나는 타원 근사 과정에서의 오류가 발생할 경우 라인 세그먼트들을 결합해 가는 방식을 사용하여 오류를 해결하였다.

제시한 알고리즘은 기존 명암의 일관성 (Coherence)에 의한 미세 파티클에 대한 추적에 흔히 사용되는 PIV (Particle image velocimetry)에 비해서는 아주 빠른 분석을 기대할 수가 없지만, 보다 정확성이 요구되고, 다양하며, 미세한 세포 이미지 형상에서나, 두 개 이상의 서로 다른 세포의 기작을 연구하는 공생배양 (Co-culture)에서 형광물질을 이용한 세포 이미지에서도 유용하게 사용될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 광학현미경으로부터 얻은 이미지에 필요한 전처리 작업을 간략하게 설명하고 3장에서는 타원 근사를 위한 최소 자승법을 소개한다. 그리고 4장에서는 클러스터 분할을 위해 본 논문에서 제안한 타원 근사 기반 알고리즘에 대해 설명하고 마지막으로 5장에서는 실험 결과를 분석하고 6장에서는 결론을 맺는다.

2. 클러스터 모양 추출을 위한 전처리

광학 현미경으로부터 얻은 이미지는 얼룩과 분비물(Glands), 세포질(Cytoplasm) 등을 포함하므로 전처리 작업이 필수적이다. 본 연구에서는 다섯 단계의 전처리 작업을 수행하여 세포 이미지에서 클러스터 모양 정보를 추출하였다. 첫째, 이미지의 잡음을 제거하고 컬러이미지를 이진 이미지로 변환하는 이진화과정을 거친다. 이진화작업은 촬영할 때 발생할 수 있는 불균일한 조명으로 인해 단일 임계 값을 사용할 경우 원치 않는 결과가 나올 수 있다. 따라서 본 연구에서는 영상 히스토그램을 이용하여 세포 이미지들에 대해 적응적 (Adaptive) 임계 값을 자동 계산하여 이진화작업을 수행하였다.

두 번째 전처리 작업은 이미지에 생성된 구멍 (Holes)을 메우는 작업이다. 세포 배양과정에서 시간이 지남에 따라 밀착(Attached)된 세포의 크기가 커짐과 동시에 세포 가운데 부분에는 형광물질이 나타나지 않는 현상이 생기게 된다. 이러

한 부분은 이진화작업 과정에서 구멍을 생성하는데 구멍을 매우지 않으면 클러스터 모양 정보를 올바르게 생성하지 못함으로 잘못된 분리 작업을 하게 된다. 구멍의 크기가 작기 때문에 본 논문에서는 flood fill 알고리즘을 적용하여 구멍을 메웠다.

세 번째 과정은 아포토시스(Apoptosis)된 세포들을 제거하는 것이다. 세포배양과정에서 시간이 지남에 따라 퇴화하는 세포가 발생하는데 이러한 세포들은 닫힌 형태의 모양이 아니므로 두 번째 전처리 과정 후에 짧은 선분 형태나 열린 다각형 모양으로 나타난다. 퇴화된 세포는 처리 대상이 아니므로 이러한 것들은 제거해 주어야 한다. 이를 위해 모폴로지(Morphology) 열림 연산을 수행하였다. 열림 연산의 경우 가는 선들을 제거하여 퇴화된 세포들을 제거하는 작업도 수행해 줄 뿐 아니라 세포 가장자리의 매끄럽지 못한 모양들을 다듬어 주는 역할도 수행한다. 네 번째 전처리 과정은 클러스터 분리를 위해 필요한 외곽선 모양(Shape)을 추출하는 작업이다. 외곽선 검출은 여덟 방향을 검사하는 터틀(Turtle) 알고리즘을 사용하였다. 검출된 외곽선은 세포나 클러스터를 나타내는 닫힌 다각형의 모양으로 나타난다.

마지막 전처리 작업으로 외곽선에 대한 다각형 대략화(Polygon approximation) 작업을 수행한다. 이 작업은 세포나 클러스터의 외곽선 모양을 매끄럽게 만들어 준다. 대략화 작업을 간단하게 설명하기 위해 클러스터의 외곽선을 $C = \{p_1, p_2, \dots, p_n\}$, $p_i = (x_i, y_i)$ 로 표현하자. 다각형 대략화는 시작점 $i = 1$ 부터 선분 $\overline{p_i p_j}$, ($j = i + 2$)의 길이가 임계값(본 논문에서는 3의 값을 사용)보다 크면 i 값을 1증가시켜 다음 단계로 가고 그렇지 않으면 p_{i+1} 을 경계선

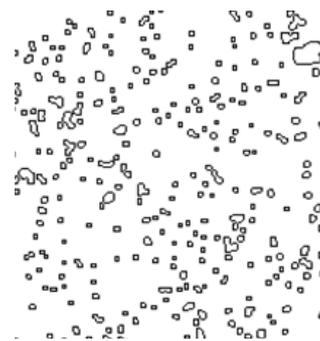


Fig. 3 Contours produced by preprocessing steps

에서 제거하고 j 을 1 증가시켜 다음 단계로 간다. 이러한 과정을 경계선의 마지막 점까지 수행한다. Fig. 3은 Fig. 2의 세포 이미지에 대해 전처리 작업을 수행한 결과 생성된 세포와 클러스터 모양 정보를 보여준다. 전처리 작업에 대한 더 세부적인 내용은 참고문헌^(11~13)을 참고하기 바란다.

3. 라인 세그먼트 분리와 근사 타원 생성

전처리 작업이 끝난 후 클러스터 분리를 위해 먼저 클러스터를 라인 세그먼트(Line segment)들로 분리한다. 클러스터를 라인 세그먼트로 분리하기 위해 기존의 연구들에서는 오목 점(Concave point)을 사용하였다.^(7~14) Fig. 1에서 보는 것처럼 두 개 이상의 세포가 겹쳐지면 겹쳐진 영역에서의 점들은 오목 점이 된다. 이러한 성질을 이용하여 클러스터를 구성하는 점들 중 오목 정점이 발견되는 위치에서 클러스터를 라인 세그먼트로 분리하는 것이다. 따라서 분리된 라인 세그먼트의 수는 클러스터에서 발견된 오목 정점보다 하나 더 많게 된다.

분리된 라인 세그먼트들에 대해 타원 근사법을 적용하여 클러스터를 분리한다. 타원 근사를 위한 알고리즘들 중 가장 효율적인 알고리즘은 최소자승법 방법으로 알려져 있다.⁽¹⁵⁾ 클러스터 분리를 위해 발표된 기존의 연구들에서도 타원 근사를 위해 최소자승법을 사용하였다.^(9~12,14) 본 연구에서도 타원 근사를 위해 최소자승법을 기본적으로 이용할 것이다. 최소자승법을 설명하기 위해 먼저 용어들을 정의해 보자. 클러스터 C 은 한 개 이상의 라인 세그먼트 l_i 로 구성되며 l_i 는 여러 개의 점 p_{ij} 로 구성된다. C 는 m 개의 라인 세그먼트로 구성되고 l_i 는 s 개의 점들로 구성된다고 가정한다.

$$C = \{l_1, l_2, \dots, l_m\} \quad (1)$$

$$l_i = \{p_{i1}, p_{i2}, \dots, p_{is}\}$$

이차곡선을 나타내는 식 (2)이 $b^2 - 4ac < 0$ 조건을 만족하면 타원이 된다. 최소자승법의 핵심은 타원과 데이터 점들 사이의 제곱 거리를 최소화하는 매개변수 a, b, c, d, e, f 의 값을 반복법을

사용하지 않고 구하는 것이다.

$$F(A, X) = A \cdot X = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (2)$$

식 (2)에서 $A = [abcde f]^T$ 이고 $X = [x^2 \ xy \ y^2 \ x \ y \ 1]^T$ 이다. 최소자승법은 부등식 $b^2 - 4ac < 0$ 조건을 $4ac - b^2 = 1$ 으로 바꾼다. 그리고 라인 세그먼트를 구성하는 점의 개수가 n 인 경우 근사타원과 점들 사이의 제곱 거리 $E = \sum_{i=1}^n F(X_i)^2$ 을 벡터 형태인

$E = \|DA\|^2$ 로 두고 제곱 거리를 최소로 하는 매개변수 A 를 구한다. 행렬 D 는 라인세그먼트를 구성하는 점들의 x 좌표 값과 y 좌표 값에 대한 행렬로 식 (3)과 같다.

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

$4ac - b^2 = 1$ 조건은 식 (4)로 다시 쓸 수 있다.

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

식 (4)에서 행렬 C 는 식 (5)과 같이 지정한다.

$$C = \begin{bmatrix} 0 & 0 & 2000 \\ 0 & -10000 \\ 2 & 0 & 0000 \\ 0 & 0 & 0000 \\ 0 & 0 & 0000 \\ 0 & 0 & 0000 \end{bmatrix} \quad (5)$$

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

$$E = \|DA\|^2 = A^T D^T D A = A^T \lambda C A = \lambda A^T C A = \lambda \quad (7)$$

식 (7)에 의해, 가장 작은 양의 고유값 λ 에

대응되는 DTD의 고유벡터가 라인 세그먼트를 구성하는 점들을 표현하는 가장 적합한 근사 타원의 매개변수 A가 된다.

매개변수 $A=[a\ b\ c\ d\ e\ f]^T$ 가 결정되면 식 (2)에 의해 타원의 중심은 식 (8), 타원의 장축과 단축은 식 (9)로 타원의 회전율은 x축에 대한 반시계 방향의 각도로 식 (10)과 같이 구할 수 있다.

$$x_0 = \frac{2cd-be}{b^2-4ac}, \quad y_0 = \frac{2ae-bd}{b^2-4ac} \quad (8)$$

$$axis = \sqrt{\frac{(ae^2+cd^2+fb^2-bde-4acf)}{(b^2-4ac)} \pm \left[\frac{\sqrt{(a-c)^2+b^2}-(a+c)}{2} \right]} \quad (9)$$

$$\phi = \begin{cases} 0 & \text{for } b=0 \text{ and } a < c \\ \pi/2 & \text{for } b=0 \text{ and } a > c \\ 0.5\cot^{-1}((a-c)/2b) & \text{for } b \neq 0 \text{ and } a < c \\ \pi/2+0.5\cot^{-1}((a-c)/2b) & \text{for } b \neq 0 \text{ and } a > c \end{cases} \quad (10)$$

기존의 발표되어 있는 타원 근사를 이용한 클러스터 분리 알고리즘⁽⁷⁻¹⁰⁾을 Fig. 2와 같이 클러스터의 크기가 매우 작은 데이터에 적용한 결과 두 가지 문제점이 발생하였다. 첫 번째 문제점은 어떤 라인 세그먼트들에 대해서는 근사 타원을 구해주지 못하는 경우가 발생한 것이고 두 번째는 구해진 근사 타원이 클러스터 분리에 적절하지 않는 경우이다. 근사 타원을 구해 주지 못한 경우는 식 (6), (7)을 계산하는 과정에서 오류가 발생하는 경우로 세 가지 종류의 오류가 발생하였다. 첫 번째 오류는 식 (7)의 계산에서 양의 정수 값을 가지는 고유값이 존재하지 않음으로 발생하였고, 두 번째 오류는 고유값이 복소수인 경우이며 세 번째 오류는 역행렬 계산과정에서 Inf 값이 계산되어 고유값을 아예 구해 주지 못하는 경우에 발생하였다. 이러한 종류의 오류들은 모두 라인 세그먼트를 구성하는 점의 개수가 다섯 개 이하인 경우에 발생하였다.

두 번째 문제점은 구해진 근사 타원이 적절하지 않는 경우이다. Fig. 4에서 굵은 선은 라인 세그먼트를 의미하고 가는 선은 라인 세그먼트에 대해 구해진 근사 타원을 보여준다. 이러한 근사 타원은 클러스터로부터 분리할 세포의 영역을 제대로 표현해 주지 못한다. 적절하지 못한 근사 타원은 라인 세그먼트를 구성하는 대부분의 점들



Fig. 4 Not appropriate fitted ellipse

의 x좌표나 y 좌표가 동일하여 수직선이나 수평선에 가까운 경우에 생성되었다. 본 연구에서는 근사 타원의 장축의 길이를 LongAxis, 단축의 길이를 ShortAxis라고 정의하고 식 (11)의 AxisRatio가 0.2보다 작으면, 부적절한 근사 타원으로 판단하였다.

$$AxisRatio = \frac{ShortAxis}{LongAxis} \quad (11)$$

기존의 클러스터 분리 알고리즘에서 발생하는 문제점들은 모두 클러스터의 크기가 매우 작은 경우에 발생하는 현상들이다. 본 연구에서는 클러스터의 크기가 매우 작은 데이터에 대해, 최소자승법에 의한 타원 근사를 이용하되 앞에서 언급한 오류들이 최소한으로 발생하도록 하는 클러스터 분리 알고리즘을 제안한다.

4. 타원 근사를 기반으로 한 클러스터 분리 알고리즘

타원 근사를 기반으로 하는 기존의 클러스터 분리 알고리즘들은 클러스터의 외곽선을 구한 다음 오목 정점을 기준으로 클러스터를 몇 개의 라인 세그먼트로 분리하였다. 기존의 연구들에서 사용한 데이터는 Fig. 1과 같이 충분히 확대하여 촬영한 세포 이미지를 사용함으로 오목 정점으로 분리한 라인 세그먼트의 길이와 모양이 매우 짧거나 수직이나 수평 모양에 가까운 것들이 거의 없다. 하지만 수백 혹은 수천 개의 세포가 들어가도록 촬영함으로 클러스터의 크기가 작은 데이터의 경우 이러한 형태의 라인 세그먼트를 매우 많이 포함한다. Xiangzhi⁽⁹⁾이 제안한 알고리즘에 본 논문의 데이터를 적용한 결과 전체 라인 세그먼트의 약 65%정도가 3장에서 설명한 오류들로 인해 근사 타원을 구해 주지 못했다.

본 논문에서는 근사 타원을 구해 주지 못하거나 Fig. 4와 같이 적절하지 못한 근사 타원을 생성한 라인 세그먼트에 대해, 동일 클러스터 내 다른 라인 세그먼트와 결합하는 방식을 사용하여 근사 타원을 다시 구해 주도록 하는 클러스터 분리 알고리즘을 제안한다.

알고리즘: 타원근사를 이용한 클러스터 분리
 [단계 1] 라인 세그먼트들로 구성된 집합 C , S 를 $C = \emptyset, S = \emptyset$ 으로 초기화한다.

[단계 2] 클러스터의 외곽선을 따라가며 오목 정점이 나타날 때마다 분리하여 분리된 라인 세그먼트를 집합 C 에 추가한다. 분리가 끝나면 클러스터는 $C = \{l_1, l_2, \dots, l_m\}$ 가 된다. 모든 라인 세그먼트 l_i 에 대해, l_i 을 구성하는 점의 개수가 다섯 개 이하이면 l_i 을 C 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 4를 수행하고 공집합이면 단계 6을 수행한다.

[단계 4] $l_i \in S$ 에 대해, 집합 C 에 속한 모든 라인 세그먼트 $l_j \in C$ ($1 \leq j \leq m$)와 l_i 을 결합하여 $l_j \cup l_i$ 에 대한 근사 타원 E_k ($1 \leq k \leq m$)을 구한다.

[단계 5] 근사 타원 E_k ($1 \leq k \leq m$)와 라인 세그먼트 $l_j \cup l_i$ 의 거리 $d_k = dist(E_k, l_j \cup l_i)$ 을 구한다. 가장 작은 거리 값 d_k 을 가지는 두 라인 세그먼트 l_i 와 l_j 를 결합하여 새로운 라인 세그먼트 l_k 을 생성한다. l_k 의 점의 개수가 다섯 개 이상이면 l_k 를 집합 C 에 포함시키고 아니면 집합 S 에 포함시킨다. 결합된 l_i 는 라인 세그먼트 집합 S 에서 제거하고 l_j 는 집합 C 에서 제거한다. 집합 S 가 공집합이 아니면 단계 4로 간다.

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

[단계 7] 생성된 근사 타원에 대해 식 (11)의 *AxisRatio* 값을 계산하여 0.2 이하인 근사 타원 E_m 이 있는지 검사한다. 존재하면 E_m 에 해당되는 라인 세그먼트 l_m 을 C 에서 제거하고 집합 S 에 추가한다.

[단계 8] 집합 S 가 공집합이 아니면 단계 4로 가고 공집합이면 프로그램의 종료한다.

단계 5에서 $d_k = dist(E_k, l_j \cup l_i)$ 는 $l_j \cup l_i$ 을 구성하는 모든 점 p_u 과 근사 타원 E_k 사이의 거리값으로 식 (12)과 같이 계산하였다. 식에서 A 는 근사 타원 E_k 의 식 2에 대한 매개변수 $[abcdef]^T$ 이다.

$$dist(E_k, l_j \cup l_i) = \sum_{u=0}^n F(A, p_u \in l_j \cup l_i) \quad (12)$$

5. 실험 결과 및 분석

제안된 알고리즘의 성능을 실험하기 위해 쥐의 3T3 세포 배양 과정에서 30분 간격으로 촬영한 광학현미경 이미지들을 사용하였다. Fig. 2는 촬영한 이미지의 한 부분이며 하나의 이미지 안에 수백 개의 세포들과 클러스터들이 나타난다. 전처리 과정을 거친 후 두 개 이상의 세포가 겹쳐진 클러스터만을 분류하여 본 논문에서 제안한 클러스터 분리 알고리즘을 적용하였다. 세포 이미지에서 클러스터만을 분류하기 위한 방법은 참고문헌을 보기 바란다.⁽¹¹⁻¹³⁾ 본 논문에서 제안한 전처리 작업과 클러스터 분리 알고리즘은 C++과 매트랩으로 구현하였다.

Fig. 5은 클러스터 분리 결과를 보여준다. 그림에서 (a)는 원래 이미지를, (b)은 전처리후 생성된 이미지를, (c)은 클러스터의 외곽선 추출 결과를 보여주고 (d)는 제시한 알고리즘에 의해 분리된 결과이며 명확하게 보여 주기 위해 확대하였다.

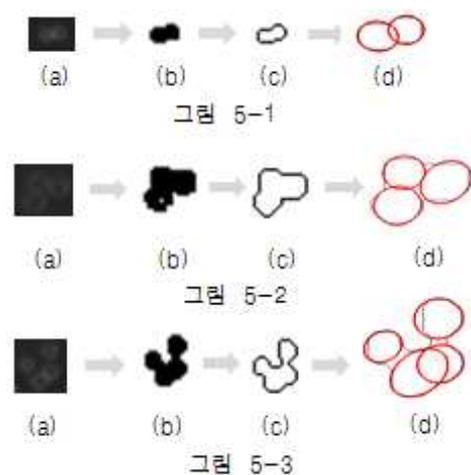


Fig. 5 Cluster Separation result: (a) Original Image (b) Image after preprocessing (c) Contour of cluster (d) Cluster separation by our algorithm

원래 이미지가 희미할 뿐 아니라 세포 가운데 부분의 강도가 매우 약하여 마치 구멍이 생긴 것처럼 보이지만 전처리 과정을 통해 (b)와 (c)의 클러스터 정보를 정확하게 추출함을 볼 수 있다.

Table 1은 실험에 사용한 이미지들에 존재하는 세포와 클러스터의 수를 보여준다. 두 개 이상의 세포가 겹쳐서 형성된 클러스터의 비율이 전체 이미지에서 약 30% 이상이므로 클러스터의 정확한 분리가 세포 트래킹 시스템의 신뢰도에 매우 큰 영향을 미칠 수 있음을 알 수 있다. Table 2는 클러스터의 종류별 비율과 본 논문에서 제시한 알고리즘과 Xiangzhi⁽⁹⁾이 제안한 알고리즘의 성능을 보여준다. 표에서 보는 것처럼 세 개 이하의 세포가 겹쳐서 형성된 클러스터가 전체 클러스터의 83%, 네 개 이상의 세포가 겹쳐서 형성된 클

러스터는 전체의 17%를 차지하였다. 알고리즘의 성능은 클러스터의 종류별로 다르게 나타났지만 본 논문에서 제시한 방법이 Xiangzhi⁽⁹⁾이 제안한 알고리즘보다 정확도가 10% 이상 높게 나타났다. 이것은 Xiangzhi의 알고리즘은 많은 라인 세그먼트에 대해 근사 타원을 구해주지 못하는 오류가 발생하였기 때문이다.

제시한 알고리즘의 성능을 살펴보면 두 개의 세포가 겹쳐진 클러스터의 경우 평균적으로 91%의 정확도로, 세 개의 세포로 겹쳐진 클러스터의 경우 평균적으로 84%의 정확도로, 서로 겹쳐진 세포의 개수가 네 개 이상으로 클러스터의 모양이 복잡해지면 약 73%의 정확도로 클러스터를 분리해 주었다. 네 개 이상의 세포가 겹치는 클러스터의 경우에 정확도가 떨어지는 이유는 제시한 알고리즘에 대해서도 근사 타원을 구해 주지 못하는 경우가 발생했기 때문이다. Fig. 6은 여섯 개의 세포로 겹쳐진 클러스터의 분리 결과를 보여 준다. 분리된 세포의 개수는 정확하지만 생성된 근사 타원의 모양에 있어서 정확도가 떨어지는 것을 볼 수 있다. 이것은 클러스터의 외곽선을 라인 세그먼트로 분리하는 작업에 대한 연구가 더 필요함을 보여 준다.

Table 1 Ratio of Cluster in experimental data

데이터종류	1	2	3	4	5
단일 세포 수	561	689	758	636	694
클러스터 수	303	289	271	291	261
전체 수	864	978	1029	927	955
클러스터 비율(%)	35	30	26	31	27

Table 2 Performance of our algorithm and ratio of kinds of clusters

클러스터 종류	두 개의 세포가 겹침	세 개의 세포가 겹침	네 개 이상의 세포가 겹침
비율(%) (해당 클러스터 수/전체 클러스터 수)*100	62	21	17
분리 정확도(%) - Xiangzhi 방법	79	71	64
분리 정확도(%) - 제안한 방법	91	84	73

6. 결론 및 연구방향

본 연구에서는 다른 논문에서 제시한 방법으로 해결하기 어려운 크기가 매우 작은 세포 이미지를 위한 클러스터 분리 알고리즘을 제안하였다. 제안된 알고리즘은 최소자승법을 적용한 타원 근사법과 휴리스틱을 기반으로 한다. 제시한 알고리즘은 겹쳐진 세포의 개수가 세 개 이하인 경우는 평균적으로 약 89.5%의 정확도로 클러스터를 분리해 주었고 네 개 이상인 경우 정확도가 73%로 떨어졌다. 하지만 Xiangzhi등이 제안한 알고리즘에 비해 10% 이상의 높은 정확도를 보여 주었다.

제시한 알고리즘이 네 개 이상의 세포로 구성된 클러스터에 대해 정확도가 떨어지는 이유는 계산 과정에서 근사 타원을 구해주지 못하는 라인 세그먼트들이 나타나기 때문이다. 또한 생성된 타원의 형상비율(aspect ratio)이 1:5이하이면 부적절한 타원으로 처리하여 라인 세그먼트를 재결합하도록 하였는데 실험 데이터에 따라 결함에



Fig. 6 Cluster separation for a cluster with more four cells

따라 정확한 근사 타원을 구해주는 경우도 있었지만 그렇지 못한 경우도 발생했기 때문이다. 따라서 라인 세그먼트를 결합하고 형상비율을 데이터의 종류에 따라 조절하는 등 알고리즘의 개선이 필요하다.

앞으로 클러스터에 대한 라인 세그먼트 분리 방법과 결합 방법 및 타원 모양 제한을 위한 형상비율 개선 등으로 알고리즘을 개선하여 클러스터 분리 정확도를 높일 뿐만 아니라 분리된 모양의 정확도도 개선하기 위한 연구를 계속 수행하고자 한다.

후 기

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A1A2009392)

참고문헌

- (1) Bise, R., Li, K., Eom, S. and Kanade, T., 2009, "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, pp. 67~77.
- (2) Wang, W., 1998, "Binary Image Segmentation of Aggregates Based on Polygonal Approximation and Classification of Concavities," *Pattern Recognition*, Vol. 31, pp. 1503~1524.
- (3) Jin, Y.X.C., Jayasooriah, S.H.O. and Sinniah, R., 1994, "Clump Splitting Through Concavity Analysis," *Pattern Recognition Letters*, Vol.15, pp. 1013~1018.
- (4) Kumar, S., Ong, S.H., Ranganath, S., Ong, T.C. and Chew, F.T., 2006, "A Rule-Based Approach for Robust Clump Splitting," *Pattern Recognition*, Vol. 39, pp. 1088~1098.
- (5) Mao, K.Z., Zhao, P. and Tan, P., 2006, "Supervised Learning-Based Cell Image Segmentation for P53 Immunohistochemistry," *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 6, pp. 1153~1163.
- (6) Ruberto, C.D., Dempster, A., Khan, S. and Jarra, B., 2002, "Analysis of Infected Blood Cell Images Using Morphological Operators," *Image and Vision Computing*, Vol. 20, pp. 133~146.
- (7) Talbot, H. and Appleton, B.C., 2002, "Elliptical Distance Transforms and the Object Splitting Problem," *Proceedings of International Symposium on Mathematical Morphology*, Sydney, Australia, pp. 229~240.
- (8) Yu, D., Pham, T.D., Zhou, X. and Wong, S.T.C., 2007, "Segmentation, Recognition and Tracing Analysis for High-Content Cell-Cycle Screening," *Proceedings of International Symposium on Computational Models for Life Sciences*, Queensland, Australia, pp. 66~75.
- (9) Bai, X., Sun, C. and Zhou, F., 2009, "Splitting Touching Cells Based on Concave Points and Ellipse Fitting," *Pattern Recognition*, 42, pp. 2434~2446.
- (10) Kothari, S., Chaudry, Q. and Wang, M. D., 2009, "Automated Cell Counting and Cluster Segmentation Using Concavity Detection and Ellipse Fitting Techniques," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*.
- (11) Cho, M., Shim, J., Kim, J. and Moon, S., 2011, "Classifying and Counting Cells and Clusters Using Concave Points," *Korea Computer Congress 2011* Vol. C, pp. 184~187.
- (12) Cho, M. and Shim, J., 2012, "An Ellipse Fitting based Algorithm for Separating Overlapping Cells," *Proceeding of The Korea Institute of Information and Communication Engineering 2012 Spring*, pp. 912~915.
- (13) Cho, M., Kim, J. and Shim, J., 2011, "Optical Microscope Image Processing for Automated Cells Counting," *Journal of The Korea Institute of Information and Communication Engineering*, Vol. 15, No. 12, pp. 2494~2495.
- (14) Cho, M. and Shim, J., 2012, "Ellipse based Cell Separation Algorithm for Tens of Thousands of Overlapping Cells," *The 2nd International conference on Convergence Technology 2012*, pp. 205~208.
- (15) Fitzgibbon, A., Pilu, M. and Fisher, R. B., 1999, "Direct Least Square Fitting of Ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 5, pp. 476~480.