

세포 주기 자동화 분석을 위한 최적의 분류 알고리즘 검출

나예지*, 이상준**, 왕창원*, 정화영*, 호종갑*, 민세동*
 순천향대*, 선문대**

Detection of optimal recognition algorithm for automated analysis of cell cycle

Ye-Ji Na*, Sang-Joon Lee**, Chang-Won Wang*, Hwa-Young Jeong*, Jong-Gab Ho*, Se-Dong Min*
 Soonchunhyang University*, SunMoon University**

Abstract - 최근 세포 이미지 분석을 자동화 하기 위한 연구가 많이 이루어지고 있다. 이러한 연구는 머신러닝과 함께 수행되어 연구자의 개입 없이도 컴퓨터가 자동으로 분석할 수 있게 되었다. 본 논문은 현미경으로부터 얻은 세포 이미지의 특징을 추출하여 다양한 머신러닝 알고리즘을 적용해보고, 세포 주기 분류의 정확도가 가장 높은 분류 알고리즘을 검출하고자 한다.

부인 M기의 단계로 진행한다고 볼 수 있다. 형광 현미경으로 세포를 촬영하기 위해서는 형광 시약을 사용하여 세포를 염색하는 과정이 필요하며, 본 연구에서는 모든 세포를 관찰하기 위한 형광 시약인 DAPI 용액으로 MKN-28세포를 염색시켰다. 이후 그림 1에서와 같이 형광 현미경으로 촬영된 이미지를 획득하였다.

1. 서 론

생물정보학 분야의 연구자들은 조직세포의 세포학적 현상을 발견하기 위해 현미경을 주로 사용한다. 최근 형광 현미경 이미지 분석에 대한 요구가 증가하고 있는데 이는 생체 세포 내 일어나는 현상들을 영상을 통해 직접 확인하고 분석할 수 있도록 해 주기 때문에 질병의 조기 진단 및 치료에 필수적으로 사용될 수 있기 때문이다. [1,2]

형광 이미지를 통한 분석은 현재도 꾸준히 연구되고 있는 분야로, 살아있는 세포를 실시간으로 트래킹하기 위한 연구[3,4] 및 줄기세포 계수 [5], 세포 분열 패턴의 인식[6] 및 세포 주기 분석[7] 등 각종 세포생리학적 현상 발견을 위해 현미경 장비가 사용되어지고 있다.

그러나 형광 현미경은 가격이 고가이기 때문에 대부분의 연구실에서는 하나의 장비만을 배치하여 사용하고 있는 실정이다. 연구자들은 형광 현미경으로 촬영한 수백여 개의 세포 경향 파악을 위해 직접 육안으로 분석을 수행하며, 이에 많은 시간과 노동력을 투자한다. 하지만 장기간의 분석으로 현미경 장비가 독점될 경우 다른 연구자들의 실험 진행에 막대한 피해가 갈 수 있다. 따라서 현미경 장비로부터 촬영한 세포 이미지를 개인용 PC 및 모바일에서도 분석할 수 있는 프로그램 개발이 요구된다.[8]

현재 상용 세포 이미지 분석 소프트웨어는 특정 현미경 장비와 함께 구매하는 경우가 많으며, 고가의 비용으로 인해 규모가 큰 일부 연구소에서만 사용되고 있다. 또한 상용 소프트웨어를 사용하는 데 있어 연구자가 수동으로 설정해주어야 하는 변수 값들이 존재하는데, 숙련된 연구원들은 각자 자신의 주관적인 기준을 적용하여 설정하기 때문에 결과 데이터가 개인마다 상이하다.

이에 따라, 세포 이미지의 자동 분석을 위해 최근 컴퓨터에 기계학습(Machine Learning)을 적용한 연구가 다수 진행되었다.[9-11] 세포 데이터는 방대한 양이기 때문에 공통성을 갖는 세포끼리 군집분석을 실시한다면 집단 별로 가지는 특성 및 관계를 분석 할 수 있게 되어 시간과 비용을 줄일 수 있다는 장점이 있다.

따라서 본 연구에서는 형광 현미경으로 촬영한 세포 이미지를 기반으로 각 세포의 주기를 자동으로 군집화 및 분류해주는 프로그램 개발을 목표로 하여, 1차적으로 이미지의 특징 정보를 추출하여 다양한 기계학습 분류기에 데이터를 적용해보고 그 성능을 비교하는 연구를 진행하였다.

2. 본 론

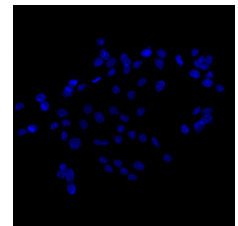
2.1 세포 이미지 획득

실험에 사용된 세포는 MKN-28 세포로, 몽골의 70세 여자 노인의 위암세포이다. 세포 이미지는 Laser Scanning Confocal Microscope (Carl Zeiss LSM 700, Jena, Germany) 장비로 촬영되었다. 위 장비의 해상도는 sms 72 DPI이고, 촬영 배율은 200배로 촬영하였으며 촬영된 이미지는 512x512 크기의 16bit TIFF 파일이다.

본 실험에서의 목적은 암 세포에 특정 약물을 처리 시, 세포의 분열 과정을 확인하여 약효가 있는지 판단하는 데에 있다. 따라서 MKN-28 세포에 항암제로 많이 쓰이는 Doxorubicin 약물을 처리하여 세포 주기를 관찰하고자 한다. 세포 주기는 분열기(M phase)와 간기(Interphase)로 이루어져 있으며, 간기는 다시 DNA가 복제되는 S기와 휴지기인 G1, G2기로 구성된다. 세포의 주기는 형광 강도가 상대적으로 클수록 후반



(a) 컨포칼 현미경
 ‘출처: ZEISS사의 LSM700’

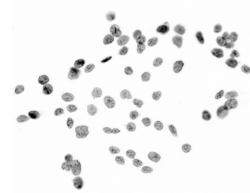


(b) MKN-28 세포 이미지

<그림 1> Confocal 현미경으로부터 얻은 MKN-28 세포

2.2 이미지 특징 추출

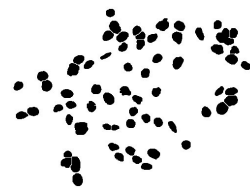
현미경으로부터 얻은 이미지에서 그림 2의 (a)~(d)에 나타난 과정을 거쳐 특징 값을 추출하였다. 주기를 분류하기 위한 특징 값은 각 세포의 크기 정보(Area), 형광 강도(Intensity), 평균 형광 강도의 값(Avg_Intensity), 최대 형광 강도(Max_Intensity), 그리고 최소 형광 강도 값(Min_Intensity)으로 설정하였다. 이미지에서 위의 5가지 특징을 추출하기 위한 영상 처리 과정을 아래 그림에 나타내었다.



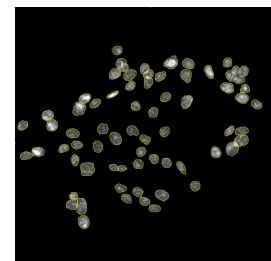
(a) 그레이 스케일 및 역변환



(b) 이진화



(c) 워터셰드 분할



(d) 라벨링 및 경계선 추출

<그림 2> 영상 전처리 과정

먼저 입력된 칼라 영상을 그레이 스케일 처리하였으며, 잡음 제거를 위해 이미지에 Mean 필터를 적용한 후 그림 2의 (a)와 같이 역상 처리하였다. 그림 2의 (b)에서는 컴퓨터가 세포를 객체로써 인지할 수 있도록 적용적 임계값 설정 알고리즘인 오츠(Otsu) 이진화를 수행하였다. 이

진화 된 이미지에서 여러 세포들이 하나로 겹쳐있는 영역은 영상 분할 알고리즘인 워터셰드(Watershed) 방법을 적용하여 그림 2의 (c)와 같이 단일세포로 분할하였다. 마지막으로는 이미지 내의 세포 개수 및 intensity 값, 크기 정보를 얻기 위해 라벨링(Labeling)을 수행하였으며, 라벨이 붙여진 각 세포의 경계선을 추적하였다. 그림 2의 (d)는 원래의 칼라 이미지를 그레이 스케일로 변환한 영상에 추출된 경계선을 함께 표시한 결과이다. 위와 같은 과정을 통해 영상에서 5가지 특정 정보를 추출할 수 있었다.

2.3 세포 군집화

기계학습 분류기를 적용하기 전에 먼저, 이미지에서 추출된 각 세포의 특징들을 유사성에 의해 묶기 위해 군집화를 수행한다. 군집화는 계층적(hierarchical) 방법과 비계층적(non-hierarchical) 군집방법이 있으며, 본 연구에서는 비계층적 군집 방법 중의 하나인 K-평균법으로 군집화 하였다. K-평균법은 군집의 개수를 계층적 군집방법으로 먼저 정하고 K-평균법을 이용하여 개체를 재배치 할 수 있어 세포 데이터와 같이 여러 개체들에 대한 군집방법으로 유용하며 오차를 최소화하는 알고리즘이다.

다음 표 1에 군집의 수를 4개로 정하고 K-평균법을 이용하여 군집화 된 세포의 개수 및 비율을 나타냈다.

<표 1> K-means 방법으로 군집화한 결과

이미지	세포 개수	그룹a (G1)	그룹b (S)	그룹c (G2)	그룹d (M)
control	67개	26개(39%)	19개(28%)	9개(13%)	13개(19%)
treated	188개	35개(19%)	48개(26%)	63개(34%)	42개(22%)

표 1에서의 이미지는 MKN-28세포에 아무 물질도 처리하지 않은 control 이미지와, Doxorubicin 약물을 처리한 treated 이미지 두 가지 종류에 DAPI염료를 염색시켜 컨포칼 현미경으로 촬영한 여러 장의 샘플 이미지 중 하나의 대표 이미지이다. 두 종류의 이미지에서 추출한 각 세포는 서로 유사한 특징 값을 기반으로 그룹 a~d의 군집으로 군집화 되었으며, 이는 차례대로 G1기, S기, G2기, M기를 나타낸다. control 이미지의 군집화 결과는 세포 주기 단계 별로 G1기에 해당하는 세포가 39%, S기에 해당하는 세포가 28%, G2기 13%, M기 19%로 나타났으며, treated 이미지의 군집화 결과는 G1기 19%, S기 26%, G2기 34%, M기 22%로 나타났다. 이를 통해 control 이미지에 비해 treated 이미지에서의 세포 주기 분포가 후반부인 G2기와 M기에 많이 군집되어 있다는 것을 확인할 수 있다. 이는 MKN-28세포의 사멸을 유도하는 Doxorubicin 약물 처리로 인해 treated 이미지의 DAPI의 염색 강도가 상대적으로 증가했다고 볼 수 있다. 두 이미지에 k-means 알고리즘을 적용하여 군집화 한 결과 오차제곱합(sum of squared errors)은 5.538로 나타났다.

2.4 기계학습 분류기 성능 분석

군집화로 인해 각 군집으로 묶인 세포들은 기계학습의 분류 방법으로 세포 주기를 분류했으며, 분류 알고리즘의 결과와 정확도를 기준으로 성능을 비교하였다. 데이터 분류(Classification) 기법은 이미 분류된 집단을 학습 데이터로 사용하여 새로운 객체가 입력되었을 때, 어떤 집단으로 결정될지 테스트하는 작업이다. 현재 제안된 여러 가지 머신러닝 분류 모델 중 일반적으로 사용되는 5가지 분류기인 RF(Random Forest), NB(Naive Bayes), KNN(K-Nearest Neighbor), SVM(Support Vector Machine), MLP(Multilayer Perceptron) 를 물질 처리 되지 않은 이미지에 적용한 결과를 아래 표에 나타내었다.

<표 2> 5가지 기계학습 분류기의 세부 정확성

Classifier	TP rate	FP rate	Precision	F-measure
RF	86.6%	4.6%	88.3%	86.7%
NB	86.6%	4.1%	87.5%	86.6%
KNN	80.6%	6.9%	81.1%	80.5%
SVM	95.5%	1.1%	96.1%	95.6%
MLP	92.5%	1.9%	93.8%	92.7%

SVM 분류기에서는 분류 정확도가 96.1%로 가장 높은 성능을 보였으며, 다음으로는 MLP 분류기가 93.8%로 나타났다. Decision Tree 모형을 기반으로 예측오차를 줄이는 방법인 RF 분류기의 정확도는 88.3%로 가장 높은 KNN 분류기에 비해 약 7.8% 낮게 나타났다. 한편 81.1%로 가장 낮은 정확도를 보인 분류기는 KNN 이다.

3. 결 론

본 논문에서는 컨포칼 현미경으로부터 촬영된 세포 영상에서 각 세포의 주기를 판단하기 위한 특징 값을 추출하여 이를 k-means 기법으

로 군집화 한 후, 5가지 기계학습 분류기에 적용하였다. 이를 통해 다양한 분류기 중 SVM 분류기에서 분류 정확성이 96.1%로 다른 4개의 분류기에 비해 가장 높게 나타났다. 본 논문에서 제안한 방법을 이용하여 세포 분석 프로그램을 만든다면 연구자의 개입 없이도 자동으로 이미지 분석을 할 수 있으며, 보다 빠르고 정확성이 높은 분석 결과를 얻을 수 있을 것으로 기대한다. 추후에는 본 논문에서 세포의 주기 분류의 정확도가 가장 높은 결과를 보이는 SVM 알고리즘을 적용한 분석 프로그램을 구현하여 제안한 방법이 실제 연구소나 실험 현장에서 잘 작동되는지에 관한 검증이 요구된다.

Acknowledgements

이 논문은 2014년 교육부와 한국연구재단의 지역혁신 창의 인력양성 사업의 지원을 받아 수행된 연구임(NRF-2014H1C1A1066998).

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 ICT융합고급 인력과정지원사업의 연구결과로 수행되었음(IITP-2016-H8601-16-1009).

[참 고 문 헌]

- [1] 광IT융합산업 워킹그룹, "임상 전 단계 전신 형광 이미지 기술 분석 및 동향", 2011년 광 IT융합산업 보고서, 한국광산업진흥회, 2011.
- [2] Cho, Migyung, and Jaesool Shim. "Development of a real-time Analysis System of Microchip Fluorescence Images based on Server-Client." Journal of the Korea Institute of Information and Communication Engineering 17.5 (2013): 1239-1244.
- [3] Rabut, Gwénaél, and Jan Ellenberg. "Automatic real time three dimensional cell tracking by fluorescence microscopy." Journal of microscopy 216.2 (2004): 131-137.
- [4] Dzyubachyk, Oleh, et al. "Automated analysis of time-lapse fluorescence microscopy images: from live cell images to intracellular foci." Bioinformatics 26.19 (2010): 2424-2430.
- [5] Faustino, Geisa M., et al. "Automatic embryonic stem cells detection and counting method in fluorescence microscopy images." Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on. IEEE, 2009.
- [6] Harder, Nathalie, et al. "Automated recognition of mitotic patterns in fluorescence microscopy images of human cells." Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on. IEEE, 2006.
- [7] Funabiki, Hironori, et al. "Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast." The Journal of cell biology 121.5 (1993): 961-976.
- [8] 조미경 외. "모바일용 형광이미지 분석 통합관리 시스템 개발", 한국정보통신학회 제31회 2012년 춘계종합학술대회, VOL 16 NO.01 PP. 0916 ~ 0919 (2012)
- [9] Sommer, Christoph, and Daniel W. Gerlich. "Machine learning in cell biology - teaching computers to recognize phenotypes." J Cell Sci 126.24 (2013): 5529-5539.
- [10] Wang, Meng, et al. "Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy." Bioinformatics 24.1 (2008): 94-101.
- [11] Sommer, Christoph, et al. "Learning-based mitotic cell detection in histopathological images." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.