

푸리에 형태표현자와 SVM 을 이용한 U87 세포의 형태학적 분류기 모델구축

강미선, 김정식, 김명희*
이화여자대학교 컴퓨터학과

e-mail : Mesun@ewhain.net, gitamen@gmail.com, mhkim@ewha.ac.kr

Binary Classifier Construction for U87 Cell Shapes using Fourier Shape Descriptor and SVM

Mi-Sun Kang, Jeong-Sik Kim, Myoung-Hee Kim
Dept. of Computer Science, Ewha-Womans University

요 약

본 논문에서는 위상차 현미경 영상 내 U87 세포의 정확한 형태학적 분류를 위한 이진 분류기 구축 방법을 제안한다. 본 방법은 Fourier descriptor 기반 세포형상 표현을 SVM 이진분류기 구축에 사용함으로써 분류 대상인 원추형과 원형세포에 대해 영상 내 세포의 위치와 회전, 크기의 변화에 대해 강인한 분류성능을 제공한다. 본 실험을 통해 polynomial 커널에서 학습된 SVM 분류기가 linear, RBF, sigmoid 에 비교하여 가장 정확한 분류 성능을 보임을 확인하였다. 본 연구는 논문상 기준인 두 종류의 세포 형태 분류기를 기반 프레임워크로 삼아 좀더 다양한 세포 형태를 분류할 수 있도록 개선된다면 악성뇌종양의 전이억제치료에 효과적인 전이행동분석에 도움을 줄 수 있을 것으로 기대된다.

1. 서론

일반적으로 혈관을 통해 다른 신체부위로 쉽게 전이되는 특성을 갖는 암 세포는 종양 치료를 어렵게 하는 주요 요인중 하나로 인식되고 있다[1]. 최근들어 현미경영상기반 암세포 전이 운동 모니터링 기술이 암 전이 억제 치료 효과 검증에 효과적인 방법으로 고려됨에 따라 관련 연구들이 활발하게 진행되고 있다[2]. 특히 이 과정에서 세포의 형태를 분할하는 기술과 함께 그 대상의 형태를 활성 및 비활성 세포로 분류하는 연구가 행동 패턴 분석에 필수적인 기술로 인식되고 있다.

현미경 영상 내 세포의 형태학적 분류를 위한 연구는 크게 세 과정(세포 분할, 형상특징 추출, 분류기 구축)을 동반한다. 첫째, 분류 대상 세포를 분할하는 연구에서 보편적으로 사용되는 분할 방법은 크게 watershed 방법과 level-set 방법으로 구분할 수 있다 [3,4]. 둘째, 형상특징 추출 방법은 Haralick, Zernike moments 등의 단순한 영상 내 통계정보에 기반하는 방법[5,6]과 다양한 feature 레벨의 정보를 혼합하여 구성하는 방법으로 구분할 수 있다. 후자의 대표적인 연구로는 형태구조적 특성, 에지 특성, 텍스처 특성을 혼합하여 사용하는 SLF sets 방법을 들 수 있다[7]. 마지막, 분류기 구축을 위한 기계학습 방법에는 주로

neural network 와 SVMs 등이 이용되고 있다[8,9].

본 논문에서는 뇌종양 중에서도 가장 흔하고 치명적인 다형성교아종(glioblastoma multiforme)의 특성을 갖는 U87 줄기세포를 대상으로 효과적인 형태학적 세포 분류 기술을 개발한다. U87 는 동일 세포라인일 지라도 그 세포의 형태, 크기, 방향성이 달라 정확한 세포 형태 분류를 어렵게 한다. 본 논문에서는 U87 세포 분류기 구축의 프레임워크로서 해당 세포를 크게 활성(원추형) 및 비활성(원형) 세포로 구분하는 분류기 구축을 연구 목표로 한다. 제시한 U87 세포 분류의 어려움을 해결하기 위해 본 논문에서는 분류 대상이 되는 세포의 형태, 크기, 방향성에 무관한 세포 고유 형태 특성을 제공하는 Fourier descriptor 를 분류기 구축의 입력 정보로 사용한 SVM 기반 이진 분류기 모델 구축을 통해 영상 내 세포의 형태학적 분류 성능 평가를 수행한다.

본 논문의 구성은 다음과 같다. 2 장에서는 U87 세포의 정확한 형태학적 분류를 위한 전처리, 세포형상 표현 추출, SVM 기반 모델 구축 방법을 설명한다. 3 장에서는 실험 및 결과를 제시하고, 마지막으로 4 장에서는 결론 및 향후 연구 방향을 기술한다.

* 교신저자

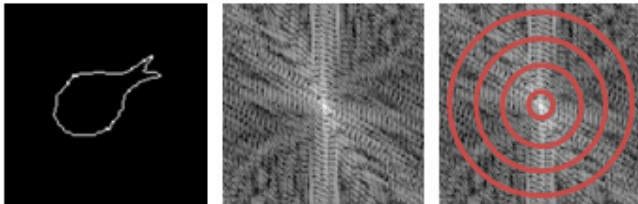
2. 세포의 형태학적 분류 과정

2.1 전처리

본 연구에서는 암세포의 형태학적 분류를 위한 전처리 과정으로써 우선 time-lapse 위상차 현미경을 통해 얻어진 U87 세포 영상으로부터 단일 세포를 포함한 블록 영상화를 통해 분류기 학습 및 테스트를 위한 샘플 데이터를 수집한다. 그런 후 각각의 샘플 영상으로부터 세포 오브젝트를 수동 분할하고 주요 영상 처리 방법을 이용하여 세포 contour 를 추출한다. 그림 1 은 수동 분할된 세포 오브젝트(왼쪽)와 해당 세포의 contour 을 추출한 결과의 예를 보여 준다. 이때 세포 오브젝트의 contour 는 4 neighborhood 연결성을 유지하여 1 픽셀 너비로 추출된다.



(그림 1) 수동 세포 분할(왼쪽) 및 contour(오른쪽) 추출 결과의 예



(그림 2) contour 영상 및 주파수 영역에서의 영상; (왼쪽)contour 영상 (가운데) 주파수 스펙트럼 (오른쪽) 동심원으로 나눈 주파수 스펙트럼

2.2 세포 형상 표현 추출

본 장에서는 영상 내 분류 대상이 되는 세포의 방향성과 크기 및 위치에 무관하게 고유의 형태 특성을 표현해주는 Fourier descriptor 추출 과정을 설명한다.

Fourier descriptor 을 추출하기 위해서는 우선 Fourier transform 을 이용하여 공간 도메인에서 표현된 대상 세포의 contour 정보(그림 2 왼쪽)를 주파수 도메인으로 변환한다(그림 2 가운데). 이렇게 변환된 주파수 공간 내 표현으로부터 phase 성분을 제외하여 amplitude 값만을 취한다. 그런 후 얻어진 amplitude 값들로부터 DC 성분을 제거함으로써 분류 대상의 상대적 크기에 영향을 받지 않고 형태학적으로 분류가 가능한 형상 표현을 얻는다[10]. 이렇게 만들어진 Fourier descriptor 를 concentric function 을 이용하여 주파수 공간을 균일 간격의 동심원으로 나눈다(그림 2 오른쪽). 마지막으로 각 동심원 영역에서의 amplitude 의 합을 구한 후 모든 구역의 amplitude 합을 히스토

그램으로 표현하여 최종 세포 형상 표현을 생성한다. 이렇게 계산된 Fourier descriptor 는 세포 형태 분류를 위한 feature vector 로 사용된다.

2.3 세포 이진 형태 분류기 구축

본 장에서는 각 세포 샘플로부터 추출한 Fourier descriptor 를 이용하여 세포 형태를 원추형과 원형 세포로 구분하는 분류기 구축에 대하여 설명한다. 본 연구에서는 적은 샘플 개수에서도 over-fitting 문제에도 강인하면서 정확한 이진 분류 성능을 제공하고 SVM(Support Vector Machine)을 사용한다[9].

주어진 학습 데이터가 $\{(x_k, y_k), 1 \leq k \leq n\}$ 일 때, 분류 대상은 x_k , 분류 그룹은 y_k 에 해당한다. 그리고 Kernel function $K: \square^n \times \square^n \mapsto \square$ 이다. 이때 SVM 분류 함수는 수식 (1)로 정의된다.

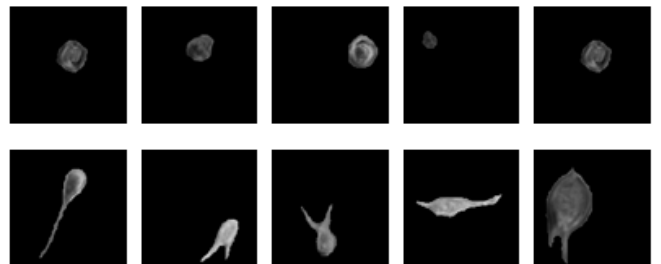
$$y_k(x) = \sum_{k=1}^n a_k y_k K(x, x_k) + b \quad (1)$$

수식 (1)에서 계수 a_k 와 b 는 각각 두 클래스 집단간 최대화된 margin 을 결정하기 위하여 quadratic 최적화 문제를 풀어 산출된다. 본 방법에서는 비선형적 분류의 경우를 처리하기 위하여 polynomial, RBF, sigmoid 커널 함수를 채택하여 분류기 학습 및 테스트를 수행한다. 또한 구축된 분류기의 정확도를 평가하고 비선형적 케이스에 대한 최적 매개변수 환경을 결정하기 위하여 k-fold cross validation 기법을 적용한다. 분류 성능 평가 기준으로는 error(학습오류), recall(재현율), precision(정밀도)을 계산하여 실험 결과로 제시한다[11].

3. 실험 및 결과

본 장에서는 U87 세포의 이진 세포 형태 분류 성능을 평가하기 위하여 세포 형상 표현(Fourier descriptor)의 feature 해상도 및 SVM 에 적용된 커널 함수의 종류에 따른 분류 성능을 평가하였다.

본 실험에서는 SVM 이용 이진 세포 분류 학습 및 그 성능을 평가하기 위하여 총 66 개의 샘플 데이터를 구성하였다(원추형 33 개, 원형 33 개)(그림 3 참조).



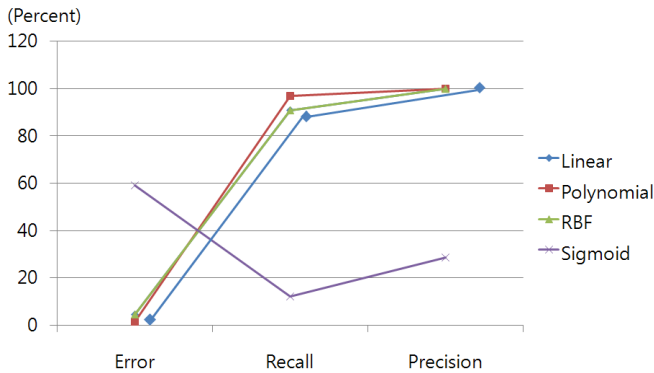
(그림 3) 실험 데이터 샘플

아래 표 1 는 feature 해상도에 따른 분류 결과를 나타낸다. 본 결과는 linear kernel function 을 사용하였고, 2-fold cross validation 방법으로 실험을 한 것이다.

<표 1> feature 해상도 변화에 따른 분류 성능

단위(%)	Error	Recall	Precision
Feature 7 개	34.85	51.52	70.83
Feature 4 개	18.18	63.64	100.00
Feature 3 개	4.55	90.91	100.00

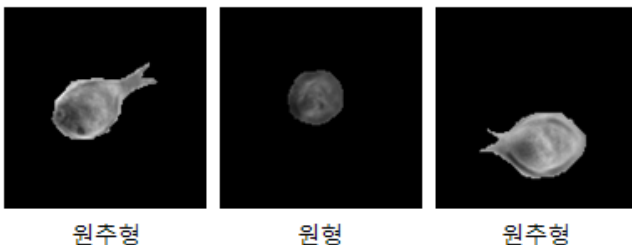
본 실험에서는 SVM 의 Kernel function 중 대표적으로 사용되는 Linear, Polynomial, Radial basis function(RBF), Sigmoid function 을 사용하여 얻어진 error 와 recall 그리고 precision 값을 비교하여 polynomial function 이 본 실험에서 가장 월등한 성능을 보였음을 확인하였다.(그림 4 참조) 본 실험에서 사용된 Polynomial 의 parameter 는 gamma: 1, degree: 3, coefficient: 0 이다.



단위:(%)	Error	Recall	Precision
Linear	4.55	90.91	100.00
Polynomial	1.52	96.97	100.00
RBF	4.55	90.91	100.00
Sigmoid	59.09	12.12	28.57

(그림 4) 적용 커널함수에 따른 분류 성능 비교

그림 5 는 본 연구에서 구축한 이진 세포 분류기를 이용하여 분류 결과를 보여준다.



(그림 5) 분류 결과

4. 결론 및 향후 연구 과제

본 연구에서 제안한 Fourier descriptor 와 SVM 을 이용한 U87 세포의 형태학적 분류기 구축 방법은 분류 대상의 다양한 방향성과 크기에 무관하게 형태적 정보만을 이용하여 정확한 분류 성능을 제공한다. 또한

Fourier descriptor 를 주파수에 따라 분류하여 SVM 기계학습 방법을 사용하여 학습시킴으로써 자동적으로 분류가 가능하다.

향후 연구 과제로는 본 연구에서는 두 종류의 형태만을 분류하였으나 원 영상에는 원형과 원주형 이외의 형태를 띠고 있는 세포들이 존재하고 있다. 다양한 형태 중에서 대칭성을 띠고 있는 세포들이 많이 존재한다. 그러한 대칭적인 특성을 지니고 있는 세포들도 함께 분류함으로써 멀티 분류가 가능하게 할 것이다. 또한 이에 대한 생물학적 전문가를 통한 검증 과정을 통하여 실제 분류결과가 생물학적 기준에 대한 적합성을 평가할 것이다.

감사의 글

이 연구는 한국과학재단(과학기술부)지원 도약연구 지원사업에 의해 수행되었습니다.

참고문헌

- [1] S. Wadler, H. Zhang, M. Cammer, and X.P. Hu. Quantification of Ribonucleotide Reductase Expression in Wild-Type and Hydroxyurea-Resistant Cell Lines Employing in Situ Reverse Transcriptase Polymerase Chain Reaction and a Computerized Image Analysis System, *Anal. Biochem.* 267 (1999), pp. 24–29.
- [2] Y. Hamamoto, S. Uchimura and S. Tomita. On the behavior of artificial neural network classifiers in high-dimensional spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 5 (1996), pp. 571–575.
- [3] Kachouie, N., P. Fieguth, et al. (2008). "Watershed deconvolution for cell segmentation", *Engineering in Medicine and Biology Society*, 2008.
- [4] Yayun Zhou, Arjan Kuijper, Lin He, Multiphase level set method and its application in cell segmentation, *Proceedings of the Fifth IASTED International Conference on Signal Processing, Pattern Recognition and Applications*, February 13-15, 2008
- [5] Huang, K. and Murphy, R.F., Automated classification of subcellular patterns in multicell images without segmentation into single cells. In: *IEEE international symposium on biomedical imaging: nano to macro*, pp. 1139-1142.
- [6] Hamilton, N., Pantelic, R., Hanson, K. and Teasdale, R.D., Fast automated cell phenotype classification. *BMC Bioinform.* v8. 110
- [7] Chen, X. and Murphy, R.F., Objective clustering of proteins based on subcellular location patterns. *J Biomed Biotech.* v2. 87-95.
- [8] Arbib, Michael A. (Ed.) (1995). *The Handbook of Brain Theory and Neural Networks*.
- [9] Vapnik, V. N.: *The Nature of Statistical Learning Theory*, Springer, (1995)
- [10] Kazhdan, M., Funkhouser, T., Rusinkiewicz, S. 2003 Rotation invariant spherical harmonic representation of 3D shape descriptors. In Kobbelt, S., Schröder, S., Hoppe, S. (Eds.). *Eurographics Symposium on Geometry Processing EG Digital Library*
- [11] Olson, David L.; Delen, Dursun "Advanced Data Mining Techniques" Springer; 1 edition (February 1, 2008), page 138, ISBN 3540769161