

베이지안 네트워크와 규칙기반 병합 알고리즘을 이용한 자동 세포 분류 및 분할

정미라*, 고병철, 남재열
계명대학교 컴퓨터공학과
e-mail: {jungmr, niceko, jynam}@kmu.ac.kr

Automatic Cell Classification and Segmentation based on Bayesian Networks and Rule-based Merging Algorithm

Mi-Ra Jeong, ByoungChul Ko, Jae-Yeal Nam
Department of Computer Engineering, Keimyung University

요 약

본 논문에서는 세포영상을 분할하고 분류하는 알고리즘을 제안한다. 우선, 배경으로부터 세포를 분할한 후, 학습데이터로부터 얻은 Compactness, Smoothness, Moments와 같은 형태학적 특징을 추출한다. 전경세포들이 분할된 후에, 보다 정밀한 세포 분석을 위해서 군집세포(Overlapped Cell)와 독립세포(Isolated Cell)를 분류 할 수 있는 알고리즘의 개발이 필수적이다. 이를 위해서 본 논문에서는 베이지안 네트워크와 각 노드에 대한 3개의 확률밀도함수를 사용하여 각 세포 영역을 분류한다. 분류된 군집세포영역은 향후 정확한 세포 분석을 위해서 군집세포가 포함하는 독립세포의 수만큼 마커를 찾고, Watershed 알고리즘과 병합과정을 거쳐 하나의 독립세포를 분리하게 된다. 현미경으로부터 얻은 세포영상에 대한 실험 결과는 이전 논문들에서 제안한 방법들과 비교했을 때, 각 군집세포의 독립세포로의 분리 이전에 세포영역에 대한 분류과정을 먼저 수행하였기 때문에 분할 성능이 크게 향상되었음을 확인할 수 있다.

1. 서론

최근 디지털 사회의 변화에 맞춰 의료기관에서 또한 디지털화가 많이 진행되어 왔다. 그중에서 가장 대표적으로 관심을 받고 있는 것이 PACS(Picture Archiving Communication System)이다. 대용량의 의료 영상들이 디지털화 되고 하나의 데이터베이스에서 통합 관리 되면서 디지털 영상 관리 및 전송시스템인 PACS의 필요성은 더욱 증대되고 있다. PACS의 필요성 증대와 함께 이 시스템에서 관리되는 다양한 의료영상 중에서, 세포 및 분자 영상을 자동 분석하고 의사결정에 도움을 줄 수 있는 진단 보조 시스템의 중요성도 강조되고 있다. 병리학 영상의 자동 분석 보조 시스템의 개발에 앞서, 병리학 영상에는 여러 세포가 겹쳐져 있는 군집세포(Overlapped Cell)와 독립세포(Isolated Cell)가 혼재되어 있으므로 보다 신뢰성 있는 세포 진단을 위해서는 군집세포와 독립세포에 대한 정밀한 분류가 필수적이다. 기존에 제안된 방법들은 배경으로부터 세포를 분할함에 있어서 간편하고 기본적인 방법들을 적용하기 때문에 독립세포와 군집세포를 구분하지 못하고 동일한 알고리즘을 가져오는 문제점이 있다. 따라서, 좀 더 정확한 세포진단을 위해서는 군집세포들에 대한 정밀한 분류가 필요하다.

본 논문에서는 세포핵과 세포질을 포함하는 세포영역을 분할하고, 이후 추출된 세포 영역들에 대해 세포의 형태학적 특성들과 베이지안 네트워크를 이용하여 독립세포와 군집세포에 대한 분류방법과 군집세포로 분류된 영역을 독립세포로 분리하는 방법을 제안한다.

2. 군집세포영역의 분류

2.1 적응적 지역 임계값(Adaptive Local Thresholding)과 Active Contour Model을 이용한 영상 분할

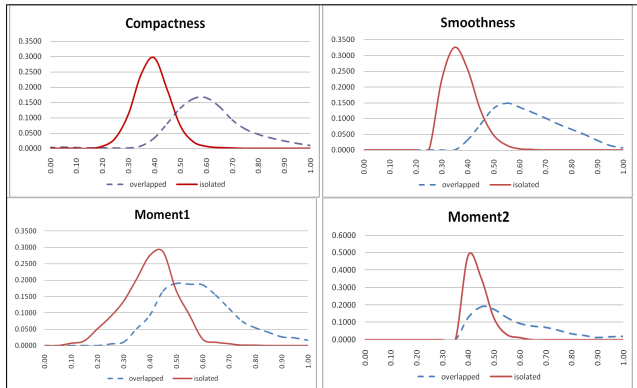
일반적으로 세포 영상을 분할하기 위해 전역적 임계값(Global Thresholding)분할 방법을 주로 사용하지만 본 논문의 실험에서는 분할 성능을 향상시키기 위해영상을 NxN 개의 부 영역으로 분할하고 각 영역으로부터 적응적 임계값(Adaptive Local Thresholding)을 적용하여 세포 영상을 분할하는 방법을 사용한다. 첫 번째 단계로, 각 지역마다 최적의 분할 임계값을 계산하기 위해 가우시안 혼합 모델(Gaussian Mixture Model)과 최대 우도 함수(Maximum Likelihood Function)[1]를 이용하여 배경과 분리된 세포 영역을 1차적으로 추출한다. 2차적으로, 지역 임계값으로 얻어진 세포 영역에 Active Contour Model(ACM)[2]을 적용함으로써 보다 정확한 세포 영역을 추출한다.

2.2 세포 영상 분류

세포 영상에서 추출된 세포영역은 군집세포영역과 독립 세포영역으로 분류할 수 있다. 두 영역 중 여러개의 겹쳐진 세포로 이루어진 군집세포영역은 이후의 정확한 세포 분석을 위해서 독립세포로 분리할 필요성이 있다. 두 타입의 세포영역을 분류하기 위해서 Chen등[3]은 세포의 형태학적 특징값인 Compactness와 몇 개의 경험적 규칙을 이용하여 세포를 분류하고 있지만, 이는 세포의 모양이 완전히 등근모양일 경우에만 좋은 결과를 가져오는 단점을 가지고 있다. 이 문제점을 해결하기 위해서 본 논문에서는 군집세포영역과 독립세포영역을 분류하기 위해서 각 영역의 6가지의 형태학적 특징들을 추출하여, 그 특징벡터를 베이지안 네트워크(Bayesian Networks)에 적용하여 세포 영역을 분류하는 방법을 제안한다.

2.2.1 형태학 특징 추출과 Parzen Window를 이용한 확률 밀도 추정

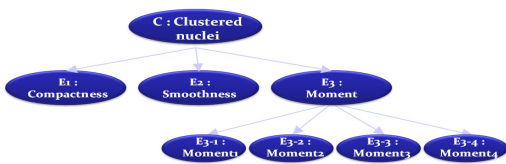
먼저, 학습데이터 군집으로 구성된 세포 영상에서 군집 세포영역과 독립세포영역의 형태학적 특징을 추출한다. 형태학적 특징은 객체의 모양이나 크기 등을 나타내기 위해서 객체의 원주, 면적, 둘레, Compactness, Smoothness, Convexity, 장축과 단축 등의 특징들이 사용된다. 독립세포의 형태는 원형뿐 아니라 타원형의 모양을 가지며, 그 크기는 일정하지 않다고 가정하고 군집세포영역과 독립세포영역의 두 영역을 가장 이상적으로 분류할 수 있는 형태학적 특징인 Compactness, Smoothness, Moment1,2,3,4를 선택한다. 이 3가지 특징들은 확률 모델을 만들기 위한 실험에서 독립 및 군집세포영역을 가장 잘 분리할 수 있도록 모델링 된 특징만을 선택한 것이다. 각 세포 영역별로 6개의 형태학 특징을 추출하여 확률 모델을 생성하게 된다. 이렇게 수집된 각 특징벡터는 데이터의 분포 형태가 일정하지 않고 일반적으로 많이 사용되는 가우시안 분포형태를 보인다고 할 수 없으므로, 본 논문에서는 각 데이터로부터 추출된 특징 벡터에서 확률분포를 추정하기 위해 비모수(Non-parametric)방법인 히스토그램을 구성하고, Parzen window중 가우시안 스무드 커널(Gaussian Smooth-Kernel)을 적용하여 보다 정밀한 확률 밀도를 추정한다. (그림 1)는 특징벡터를 스무드 커널을 이용해 얻어진 확률 밀도 추정 결과를 보여주고 있다.



(그림 1) Compactness, Smoothness, Moments의 확률 밀도 추정

2.2.2 베이지안 네트워크를 이용한 세포 영역의 분류

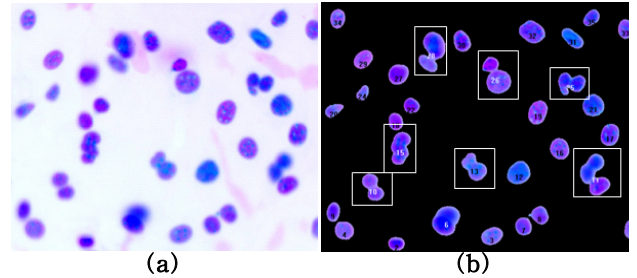
베이지안 네트워크는 기본적으로 이산 확률 변수들의 조건부 확률 관계를 방향성 비순환 그래프(DAG: Directed Acyclic Graph)와 조건부 확률 테이블로 가지고 있으며 이 확률 테이블은 파라미터간의 확률적 의존 관계를 나타내는 그래프와 각 파라미터별 조건부 확률로 구성된다. (그림 2)과 같이 베이지안 네트워크의 한 노드는 다른 노드가 가지는 값의 대한 조건부 확률을 계산하여 얻어진다.



(그림 2) 군집세포영역의 분류를 위한 베이지안 네트워크

(그림 2)에서 루트 노드 C는 3개의 자식 노드 E_1, E_2, E_3 를 가지며, 그 중 자식 노드 E_3 는 3개의 자식 노드로 구성된다. 베이지안 분류기에 의한 루트 노드의 확률 밀도인 세포영역이 군집세포영역으로 분류될 가능성은 (그림 2)와 같이 자식 노드의 조건부 확률로부터 추측되어진다.

(그림 3)은 베이지안 네트워크를 이용하여 군집세포영역을 분류한 결과를 보여준다.



(그림 3) 베이지안 네트워크를 이용한 군집세포영역 분류 결과 화면 : (a) 원영상, (b) 세포영역 분할 및 군집세포영역 분류 영상

3. 군집세포영역의 분할

세포 영역의 분류 후, 영역기반 접근을 사용하여 군집세포영역을 독립세포영역으로 분리한다. 이를 위해, 각 군집세포영역으로부터 수정된 Duchene[4]의 알고리즘을 사용하여 마커를 추출하고, Watershed 알고리즘과 병합 알고리즘을 적용하여 독립세포영역을 분리하게 된다.

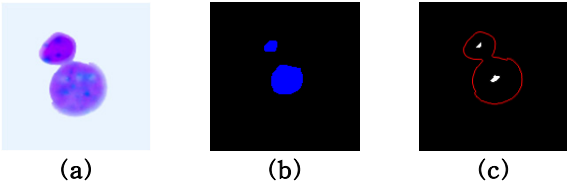
3.1 마커(Marker) 추출

성공적인 군집세포영역을 독립세포영역으로 분리하는데 있어서 정확한 마커의 추출은 필수적이다. 하나의 군집세포영역 안에는 하나 이상의 독립세포가 존재하므로 군집세포가 포함하는 독립세포 수만큼의 마커를 찾아야 한다. 정확한 마커 추출을 위해서 본 논문에서는 Ultimate 침식[4] 연산을 사용한다. Ultimate 침식 연산은 반복적인 침식연산 과정을 기반으로 마커를 추출하며, 그 과정은 다음과 같다.

$$\begin{aligned}
 1) & X_i = E(X_{i-1}) \\
 2) & Y_i = X_{i-1} - D(X_i) \\
 3) & U_i = U_{i-1} \cup (X_i - Y_i)
 \end{aligned} \tag{1}$$

$(X_0 = \text{input image}, i = 1, 2, \dots, n)$

X_i 는 입력 영상에서 침식연산을 적용한 영상이고, Y_i 는 X_{i-1} 과 X_i 의 Dilation연산 결과와의 차영상이다. 이는 군집세포영역에서 침식연산을 적용 후, Dilation연산을 적용한 객체영역과 침식연산이 적용되기 전의 객체영역을 비교하면 침식연산 후 사라진 영역이 추출된다. 본 논문에서는 이를 마커영역으로 설정한다. 객체영역의 픽셀이 모두 없어질때까지 과정1,2를 반복하여, 입력된 군집세포영상으로부터 최종 마커를 획득하게 된다. (그림 4)는 군집세포영역에서 최종 마커를 추출한 결과를 보여준다.



(그림 4) Ultimate 침식 이용한 마커 추출 결과 화면: (a) 원영상, (b) Ultimate 침식 적용 전 영상, (c) 마커 추출 결과

3.2 규칙기반(Rule-based) 영역 병합

마커 추출 후, 마커를 이용해서 군집세포영역을 독립세포영역으로 분리하기 위해서 Watershed 알고리즘과 병합 알고리즘을 사용한다. 하지만 일반적인 Watershed 알고리즘은 오버-세그멘테이션의 문제점을 가진다. 우리는 이를 해결하기 위해서 추출된 마커와 규칙기반 병합 전략을 사용해서 오버-세그멘테이션 결과를 병합시킨다.

군집세포영역의 그레이 영상에 Watershed 알고리즘을 적용한 후, 마커를 이용하여 오버-세그멘트된 영역 중 중심영역을 정하고 중심영역을 중심으로 다음 몇 가지 조건을 고려하면서 병합시켜 나간다. 병합단계와 병합조건은 다음과 같다.

- (1) 앞서 추출된 마커 영역의 좌표와 대응되는 세그멘트된 영역을 초기의 중심영역 S_k 으로 정의한다.

$$S_k \text{ is a seed region, } k = 1, 2, \dots, n$$

- (2) 임의의 세그먼트 영역 R_i 는 다음 조건을 만족하는 중심영역 S_k 를 선택하여 병합하게 된다.

$$\begin{aligned} S'_k &= S_k \cup R_i, \quad S_k \leftarrow \min(d_i) \\ \text{if } N_{i,k} > 1, \quad S_k &\leftarrow \max(Ob_{i,k}) \\ N_{i,k} &= Nm(R_i \cap S_k), \quad Ob_{i,k} = \text{card}(R_i \cap S_k) \end{aligned} \quad (2)$$

여기서 $N_{i,k}$ 는 영역 R_i 를 기준으로 하나 이상의 맞닿은 중심영역 S_k 의 수를 나타내고, $Ob_{i,k}$ 는 R_i 와 S_k 가 맞닿은 픽셀들의 집합을 나타낸다. d_i 는 영역 R_i 로부터 중심영역 S_k 까지의 거리이다. 영역의 병합은 임의의 영역 R_i 에서 최소한의 d_i 를 가지는 중심영역 S_k 와 병합된다. 하지만, 만약 두 개 이상의 중심영역 S_k 와 맞닿게 된다면, 영역 R_i 는 가장 많은 경계영역이 맞닿은 중심 영역 S_k 와 병합영역으로 결정된다. 결정된 중심영역 S_k 는 영역 R_i 와 병합되고, 중심영역 S_k 은 새롭게 정의된다.

- (3) 더 이상 영역이 존재하지 않을 때까지 과정 (2)를 반복한다.

(그림 5)는 Watershed 알고리즘과 제안된 영역 병합 과정을 통한 독립세포영역으로 분리되는 결과를 보여준다.



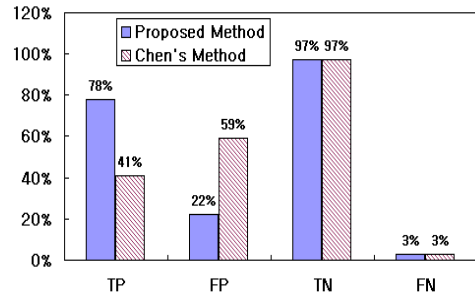
(그림 5) 제안된 병합 알고리즘을 이용한 세포 분할: (a) Watershed 세그멘테이션 결과, (b) 최종 결과

3. 실험 결과

시스템의 성능평가를 위해서 Visual C++ 6.0을 이용하여 개발하였으며, 실험은 개인용 펜티엄 PC(CPU 3.2GHz)와 운영체제는 Window를 사용하여 수행하였다.

세포영상 실험을 위해 사용된 영상은 카네기 멜론 대학의 PSLID(Protein Subcellular Localization Image Database) 데이터베이스를 사용하였다. PSLID영상에는 640x480 사이즈의 2D/3D 단백질 분자 영상들을 포함하고 있다.

본 논문에서 제안한 세포영상 분류 알고리즘의 성능과 효율성을 확인하기 위해 Chen 등의[3] 알고리즘과 비교하였다. 영상 분류 시스템의 성능 비교는 True Positive(TP), False Positive(FP), True Negative(TN), False Negative(FN)의 4가지 분류성능 측정방법을 이용하여 테스트 하였다. (그림 6)은 Chen[3]의 알고리즘과 본 논문에서 제안한 시스템을 비교한 결과이다.



(그림 6) 제안된 방법과 Xiaowei Chen의 방법 비교

표에서 볼 수 있듯이, TP와 FP에서 본 논문에서 제안한 알고리즘이 전체적으로 우수한 수행 결과를 나타냄을 확인할 수 있다. 특히 제안한 알고리즘은 베이지안 추론을 사용하여 효과적으로 false positive를 줄일 수 있었음을 알 수 있다. 또한, 두 알고리즘은 TN의 결과에서 보듯이 독립세포영역을 분류하는 성능은 좋은 결과를 보여주지만, 군집세포영역을 정확하게 분류하는 성능(TP)은 본 연구에서 제안된 알고리즘이 훨씬 정확한 결과를 보여준다는 것을 알 수 있다. 결과적으로, 독립세포영역보다는 군집세포영역의 추출이 더욱 중요시 되므로 본 연구의 결과의 성능이 우수하다고 말할 수 있다.

또한 우리는 세포영역의 분리과정의 성능을 평가하기 위해서 군집세포영역을 분류한 후 독립세포영역으로 분리할 때 (<표 2>)와 군집세포영역을 분류과정을 거치지 않을 경우 (<표 3>)를 비교하였다. <표 2>와 <표 3>에서 'Over-segmentation'은 하나의 독립세포영역에서 두 개 이상의 마커를 찾을 경우를 나타내고, 'Fusion'은 군집세포영역에서 실제로 존재하는 독립세포의 수보다 작은 수의 마커를 찾을 경우를 나타낸다. 또한 'Ignored'는 마커를 찾지 못했을 경우를 나타낸다[11]. <표 2>와 <표 3>을 비교하였을 때 'Correctly segmented', 'Fused', 'Ignored'는 비슷한 성능을 나타냈지만, 'Over-segment'에서 <표 3>보다 <표 2>가 훨씬 낮은 결과를 보여주었고, 실행 시간도 3배 이상의 실행 시간 감소를 보여주었다. 본 실험결과를 통하여 본 논문에서 제안한 알고리즘이 독립세포영역의 추출 정확성을 잃어

버리지 않으면서 훨씬 빠른 실행속도와 더 나은 세포영역의 분할 결과를 확인할 수 있다.

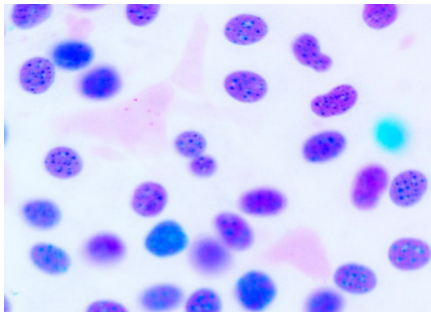
<표 2> 세포 분류 과정을 적용한 규칙기반 영역 병합

Number of nuclei	Isolated nuclei	Overlapped nuclei	All cases
Analyzed	2237	247	2484
Correctly segmented	2218(99.2%)	189(76.5%)	2407(96.9%)
Over-segmented	19(0.8%)	25(10.1%)	44(1.8%)
Fused	0(0%)	33(13.4%)	33(1.3%)
Ignored	0(0%)	0(0%)	0(0%)
Segmentation Time	34.92 sec		

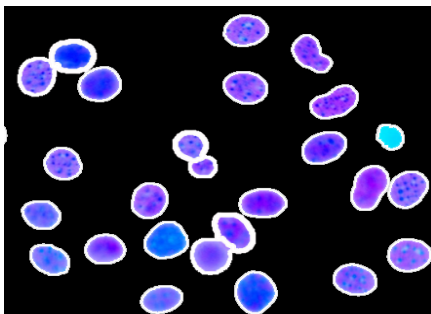
<표 3> 세포 분류 과정을 적용하지 않은 규칙기반 영역 병합

Number of nuclei	Isolated nuclei	Overlapped nuclei	All cases
Analyzed	2237	247	2484
Correctly segmented	2107(94.2%)	189(76.5%)	2296(92.4%)
Over-segmented	124(5.5%)	27(10.9%)	151(6.1%)
Fused	0(0%)	31(12.6%)	31(1.2%)
Ignored	6(0.3%)	0(0%)	6(0.2%)
Segmentation Time	117.78 sec		

(그림 7)은 제안된 알고리즘의 독립세포의 최종 분할 결과를 보여준다.



(a) 원영상



(b) 최종 분할 결과 영상

(그림 7) 세포 분할의 결과

4. 결론

본 논문에서 제안된 알고리즘은 세포 영상을 전처리 과

정을 통하여 정확하게 분할하고, 그 분할된 영역들 중에서 독립세포영역과 군집세포영역으로 분류하고, 군집세포영역은 독립세포영역으로 분할하는 알고리즘을 제안하였다. 세포 분류를 위해, 세포의 형태학적 특징을 추출하고 그 특징 값들의 확률 밀도를 이용하여 군집세포영역을 분류하였다. 또한 군집세포영역의 분할을 위해서는 마커를 추출하고 그 마커를 중심으로 Watershed 세그먼트 영역을 규칙기반 병합 알고리즘을 사용하여 군집세포영역을 독립세포영역으로 분할하도록 본 논문은 제안하고 있다. 본 논문은 세포영상을 독립세포영역으로 분할하는데 있어, 1차적으로 군집세포영역을 분류함으로써 높은 분할성능의 향상을 가져왔다.

향후, 군집세포영역의 분할 정확성을 좀 더 개선시키고, 최종적으로 정확하게 분할된 독립세포영역을 기반으로 비정상 세포의 유무를 진단하는 연구를 추가적으로 진행할 예정이다.

감사의 글

본 연구는 산업자원부의 지방기술혁신사업(RT104-01-01)지원으로 수행되었습니다.

참고문헌

- [1] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. On Systems, Man and Cybernetics* 9, 62-66 (1979)
- [2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models." *International Journal of Computer Vision*. v. 1, n. 4, pp. 321-331, (1987)
- [3] Xiaowei Chen, xiaobo Zhou, and Stephen T.C. Wong, "Automated Segmentation, Classification, and Tracking of Cancer Cell Nuclei in Time-lapse Microscopy." *IEEE Trans on Biomedical Engineering*, January, (2006)
- [4] Duchene and Lewis, 1996 P. Duchene and D. Lewis, *Visilog 5 Documentation*, Noesis Vision, Quebec, Canada (1996).
- [5] Gang Lin, Umesh Adiga, Kathy Olson, John Guzowski, Carol Barnes, and Badrinath Roysam. A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry*, 56A(1):23 - -36, November 2003.
- [6] H. Refai, L. Li, T. Teague, and R. Naukam, "Automatic count of hepatocytes in microscopic images," *IEEE International Conference on Image Processing*, (2003)
- [7] C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," *Rensselaer Polytechnic Institute, CS, Tech. Rep. TR-05-09*, (2005)
- [8] N. Malpica, C. Ortiz de Solorzano, J. J. Vaquero et al. "Applying watershed algorithms to the segmentation of clustered nuclei." *Cytometry* 28, pp. 289 - 297, (1997).
- [9] N. Malpica, C. Solorzano, J.J. Vaquero, A. Santos, I. Vallcorba, J. G. Sagredo, and F. Pozol, "Applying Watershed Algorithms to the Segmentation of Clustered Nuclei," *Cytometry*, 28, 289-297 (1997)
- [10] C. Restif and W. F. Clocksin, "Comparison of Segmentation Methods for Cytometric Assay," *International Conference on Medical Image Understanding and Analysis*, 153-156 (2004)