
영역 확장 기법과 오류 역전파 알고리즘을 이용한 자궁경부 세포진 영역 분할 및 인식

김광백* · 김성신**

Nucleus Segmentation and Recognition of Uterine Cervical Pop-Smears using Region Growing Technique and Backpropagation Algorithm

Kwang-baek Kim* · Sungshin Kim**

요 약

자궁 경부 세포진 영상의 핵 영역 분할은 자궁 경부암 자동화 검색 시스템의 가장 어렵고도 중요한 분야로 알려져 있다. 자궁 경부 세포진 영상은 배경과 세포의 영역이 확실히 구분되지 않는 경우가 많기 때문에 이들을 확실하게 구분 하는 것이 매우 중요하다. 본 논문에서는 이러한 문제점을 해결하기 위해 자궁 경부 세포진 영상에서 Region growing 기법을 적용하여 세포 영상을 분할한다. Region growing 기법은 화소간의 유사도를 측정하여 영역을 확장하여 분할하는 방법이다. 세포와 배경이 분할된 영상을 일정 임계값을 이용하여 영상을 이진화 한 후, 8방향 윤곽선 추적 알고리즘을 이용해 세포 영역을 추출한다. 추출된 세포 영역을 원 영상인 RGB 컬러로 변환한 후에 K-means 알고리즘을 적용하여 각 세포 영역의 RGB 화소를 R, G, B 채널로 각각 분리하여 클러스터링한다. 클러스터링된 각각의 R, G, B 채널의 클러스터 값을 이용하여 HSI 모델로 변환시킨 후에 세포핵 영역의 Hue 정보를 추출한다. 추출된 세포핵의 특징을 오류 역전파 알고리즘을 적용하여 정상 세포와 비정상 세포를 분류하고 인식한다.

ABSTRACT

The classification of the background and cell areas is very important research area because of the ambiguous boundary. In this paper, the region of cell is extracted from an image of uterine cervical cytodiagnosis using the region growing method that increases the region of interest based on similarity between pixels. Segmented image from background and cell areas is binarized using a threshold value. And then 8-directional tracking algorithm for contour lines is applied to extract the cell area. First, the extracted nucleus is transformed to RGB color that is the original image. Second, the K-means clustering algorithm is employed to classify RGB pixels to the R, G, and B channels, respectively. Third, the Hue information of nucleus is extracted from the HSI models that is the transformation of the clustering values in R, G, and B channels. The backpropagation algorithm is employed to classify and identify the normal or abnormal nucleus.

키워드

Region Growing, K-means, 오류 역전파 알고리즘, 세포진 영역 분할

* 신라대학교 컴퓨터공학과

** 부산대학교 전기공학과

I. 서 론

자궁경부암은 한국 여성 성기의 악성종양 중 발생빈도가 가장 흔한 종양으로 조기진단 되면 치료가 가능하며 정복하지 않으면 안 될 중요한 질병중의 하나이다. 자궁경부암은 우리나라 전체 악성종양 중의 16.4%~49.6%를 차지하며, 여성 악성종양의 26.3%~68.2%를 점유하는 것으로 보고 된 바 있다[1][2]. 이런 자궁경부암으로의 이행을 차단하는 것이 자궁경부암을 완치하는 가장 확실한 방법이고, 그로 인해 지난 40년 동안 세포 진단 검사를 전체 혹은 부분적으로 자동화하는 많은 방법들이 연구 되었다 [3]. 의료 영상에서 관심 있는 영역에 대한 진단은 크게 영상의 영역 분할, 특징 추출, 특징 분석으로 구성된다. 영역 분할은 전문의가 원하는 이상 부위를 전체 영상에서 분리해내는 것이고, 특징 추출은 분리된 이상 부위에서 조직적인 정보 특징을 추출하는 것이다. 특징 분석은 추출된 특징을 판독하여 임상적인 정보와 비교 및 분석하여 전문의가 진단하는 보조 자료로 적용된다. 세포의 핵 영역 분할 방법은 대상이 되는 영상에 따라 달라지며, 크게 화소 중심 방법과 영역 중심 방법으로 나누어진다[4][5]. 화소 중심 방법은 각 화소마다 독립적으로 특정 기준에 따라 의미를 부여하게 된다. 화소 중심 방법은 전역적 특징을 이용할 수 있다[4]. 영역 중심 방법은 주변 화소와의 관계를 분석하여 의미를 부여하게 된다. 또한 영역 중심 방법은 계산량이 많아 상대적으로 시간이 많이 소요되지만, 지역적인 특성을 이용할 수 있다[5]. 본 논문에서는 획득된 자궁 세포진 영상을 그레이스케일 영상으로 변환한 후에 영역 확장 기법으로 세포진과 배경을 분할한다. 분할된 세포진 영상에서 Thresholding 기법으로 영상을 이진화한 후, 이진화된 영상에서 8 방향 윤곽선 추적 알고리즘을 이용해 세포핵 부분을 추출한다[6]. 그리고 추출된 세포핵 영역에서 HSI 모델의 Hue 정보를 오류 역전파 알고리즘을 적용하여 세포를 분류하고 인식한다.

II. 자궁 경부 세포 영역 분할

본 논문에서 제안하는 자궁 세포 영역의 추출과정은 그림 1과 같다.

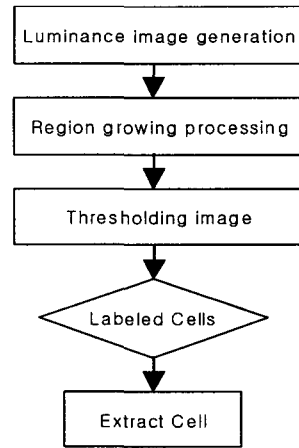


그림 1. 자궁 경부 세포 영역 추출 과정
Fig. 2. Process to extract nucleus of cervix uteri

2.1 Region growing 기법을 이용한 영상 분할

Region growing 기법은 화소간의 유사도를 측정하여 영역을 확장하여 분할하는 방법이다. 우선 하나의 중심 영역을 설정하여 인접 화소의 유사도를 측정하고 중심 영역에 속하는지를 판단한다. 두 인접 화소 A, B에 대한 유사도를 검증하는 식은 다음과 같다.

$$G(A) - G(B) < T \tag{1}$$

여기서 $G(A)$ 와 $G(B)$ 는 Gray level의 명암도 값이고 T는 임계치이다. 즉 두 Gray level 차인 유사도가 임계치 T보다 적으면 같은 영역으로 확장하고 그렇지 않으면 확장을 중단하여 영역을 분할한다. 그림 2의 (a)는 영역 확장 기법을 적용한 후의 이진화된 영상이고 (b) 영역 확장 기법을 적용하지 않고 이진화된 영상이다.

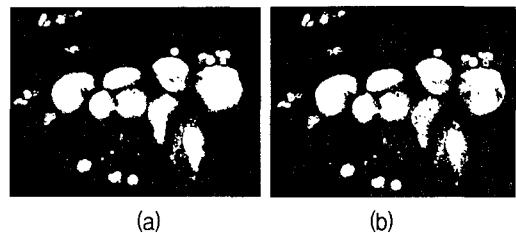


그림 2. 세포 이진화 결과: (a) 영역 확장이 적용된 이진화된 영상, (b) 이진화된 영상

Fig. 2. Binarized images of the cervix uteri:
(a) result of the region growing method
(b) Binarized image

그림 2에서 (a)는 세포의 영역이 선명하게 이진화된 것을 확인할 수 있고 (b)는 세포 영역이 많이 손상되어 있음을 확인할 수 있다.

2.2 Thresholding 기법을 이용한 영상 이진화 및 세포 영역 추출

Thresholding 기법은 일정 임계치에 의해 영상을 이진화하는 방법으로, 간단하고 속도가 빠르지만 복잡한 영상에 대해 정확한 영역 분할이 어렵다. 그러나 세포 영상의 명암 히스토그램이 세포핵, 세포질과 배경과 같이 3부분으로 나타나므로 thresholding 기법에 의해 세포핵 영역의 일부분을 추출하여 중심 좌표를 얻을 수 있다. 세포 영상 전체에서 그림 3과 같은 히스토그램 분포를 획득하고, 히스토그램 분포에서 최소값에서부터 0.5%, 최대값에서부터 0.5%의 명암도를 가진 화소를 잡음으로 간주하고, 각각의 최소 및 최대값에서 0.5%의 지점까지 화소를 제거한다. 제거된 지점을 각각 새로운 최소, 최대값으로 설정하고 두 지점의 50% 지점을 영상의 임계치로 결정하여 영상을 이진화한다. 이진화된 영상에서 8방향 윤곽선 추적 알고리즘을 적용하여 세포 영역을 추출한다. 추출된 세포는 1차적으로 정상과 비정상 세포로 분류하기 위해 세포의 면적(Area)과 세포의 둥근 정도(Circularity)를 이용한다. 그림 3은 세포 영상에 대한 히스토그램 분포와 임계치를 구하는 과정이다.

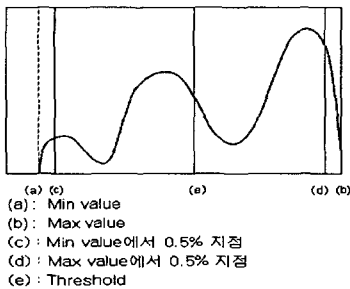


그림 3. 히스토그램 분포도와 임계치
 Fig. 3. Histogram distribution and threshold

세포의 둥근 정도를 구하는 식은 다음과 같고 결과는 그림 4와 같다.

$$Circularity = \frac{Perimeter^2}{4\pi Area} Z$$

(단, $perimeter = 2\pi r$) (2)

여기서, Perimeter는 핵의 둘레이다.

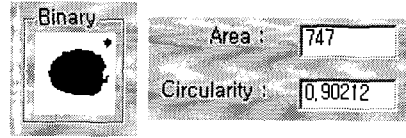


그림 4. 단일 세포의 둥근 정도 결과
 Fig. 4. The circularity of a single cell

추출된 세포는 세포핵과 세포질, 배경과 같이 크게 3부분으로 구분할 수 있으며, 추출된 세포 영상은 그림 5와 같다.

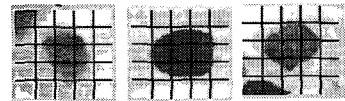


그림 5. 추출된 세포 영상
 Fig. 5. Extracted cell images

자궁 경부 세포에서 추출되는 특징은 형태학적 정보, 명암 정보, 색상 정보, 질감 정보 등이 있다.

본 논문에서는 추출된 세포 영역의 원 영상을 K-means 알고리즘을 적용하여 RGB 컬러 화소를 각각 R, G, B로 채널로 분리하여 클러스터링 한다. 클러스터링된 각각의 R, G, B 채널의 클러스터 값을 이용하여 HSI 모델로 변환한다. 변환된 HSI 모델에서 Hue 정보를 정상 세포와 비정상 세포를 분류하고 인식하기 위한 또 하나의 특징으로 적용한다.

2.3 K-means를 이용한 핵 특징 추출

K-means 알고리즘은 가장 널리 알려진 클러스터링 기법으로 각 입력 값에 대해 유사도를 결정한 후 유사하다고 판단되는 값들을 K개의 그룹으로 분류하는 기법이다.

본 논문에서는 세포 영역의 영상을 블록화 한 후, 각 블록의 RGB 화소를 R, G, B 채널로 각각 분리하여 K-means 알고리즘을 적용하여 그림 6과 같이 클러스터링 한다. 한 블록에 대해 K-means 알고리즘을 적용하여 각각의 R, G, B 채널에서 10개의 클러스터가 각각 생성되고 25개의 블록에서는 세포핵 부분에 해당되는 중앙 9개 블록의 code book이 생성된다.

생성된 코드북의 클러스터의 값을 식(3)에 의해 HSI 모델로 변환시킨 후에 Hue 값을 오류 역전과 알고리즘의 입력 패턴으로 적용하여 핵을 분류하고 인식한다. 그림 7

은 세포 영역에서 한 블록에 대한 클러스터링 과정이고 그림 8은 각각의 R, G, B 채널에 대한 클러스터링 결과 값을 HSI 모델로 변환시킨 후에 Hue 정보를 추출하는 과정이다.

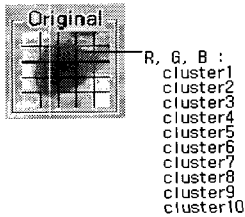


그림 6. K-means을 이용한 한 블록에 대한 클러스터링 과정
Fig. 6. Clustering process of a block by means of K-means

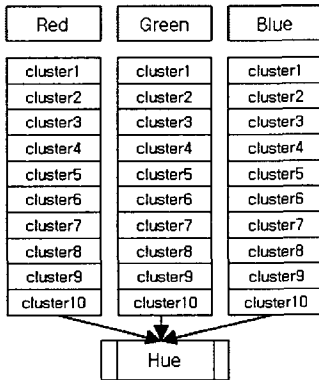


그림 7. 클러스터링 구조도
Fig. 7. Clustering structure

$$H = \cos^{-1} \left[\frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right] \quad (3)$$

2.4 오류 역전파 알고리즘을 이용한 핵 분류 및 인식

본 논문에서 세포핵의 Hue 값을 이용하여 정상 세포와 비정상 세포를 분류하기 위해 오류 역전파 학습 알고리즘을 적용한다. 오류 역전파 알고리즘은 비선형 판별 문제를 해결할 수 있어 최근에 가장 보편적으로 사용되고 있다. 그러나 이 학습 알고리즘은 학습 시간이 오래 걸리고, 새로운 패턴을 추가하여 학습할 때 기존의 학습한 패턴에 영향을 주며, 필요한 정보가 모든 연결선에 고르게 분포되지 않아서 정보가 있는 연결선으로 들어온 오류의 중복으로 인하여 신경회로망 전체의 효율이 떨어지는 문제점을 가지고 있다[7].

따라서 본 논문에서는 일반화된 델타 학습 방법에

Delta-bar-Delta 알고리즘을 적용하여 학습률을 동적으로 조정하는 개선된 오류 역전파 알고리즘을 적용하여 정상 세포와 비정상 세포를 인식한다.

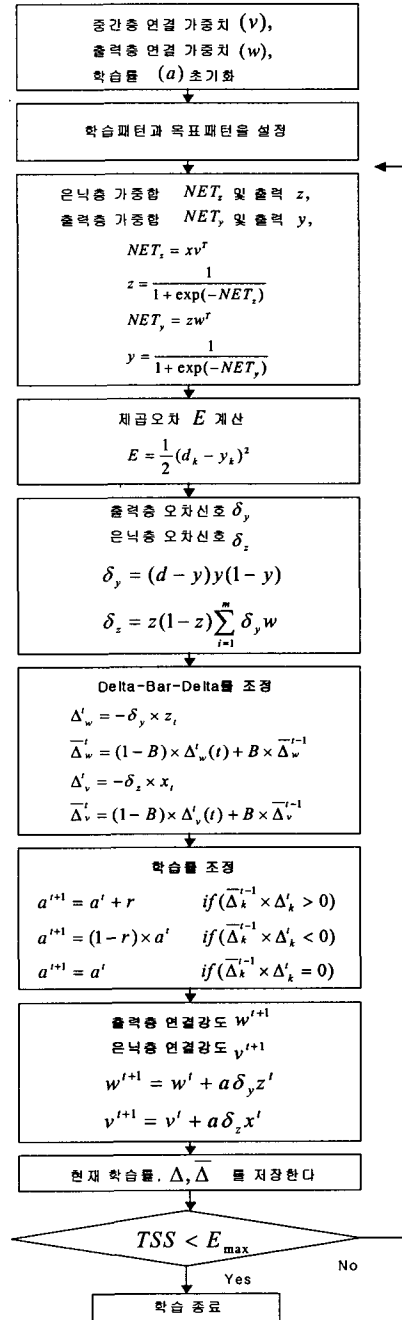


그림 8. 오류 역전파 알고리즘
Fig. 8. Backpropagation algorithm

개선된 오류 역전과 알고리즘은 그림 8과 같다. 그리고 추출된 세포핵의 특징인 Hue값을 오류 역전과 알고리즘의 학습과 인식을 위해 0과 1사이의 값으로 정규화한다.

III. 실험 및 결과분석

실험 환경은 IBM 호환 기종의 Pentium-IV PC상에서 Visual C++ 6.0으로 구현하였다. 실험 표본은 640*480 크기의 자궁 경부 세포 영상 20장을 대상으로 실험하였다. 제안된 자궁 경부 세포진 영상 추출 결과는 그림 9와 같다.

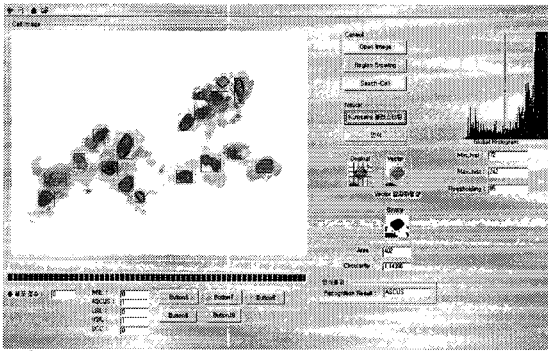


그림 9. 자궁 경부 세포진 추출 결과
Fig. 9. nucleus extraction result of the cervix uteri cytodiagnosis

실제 영상 20장에 대해서 제안된 핵 영역 분할 방법의 핵 추출 결과는 표 1과 같다.

표 1. 세포핵 추출률
Table 1. Nucleus extraction result

핵 추출	총 세포갯수	제안된 방법	추출률
	316	269	85%

표 1에서 알 수 있듯이 본 연구를 통해 추출된 핵의 추출률은 85%의 정확도를 가졌다. 그리고 추출된 단일 세포에 대해 오류 역전과 알고리즘의 학습 데이터와 실제 테스트 데이터는 표 2와 같다. 세포는 정상 세포와 비정상 세포 및 암세포로 분류된다. 따라서 본 논문에서는 세포의 분류를 The Bethesda System에서 분류한 WNL, ASCUS, LSIL, HSIL, SCC의 5등급으로 분류하였다. 여기

서 WNL은 정상세포이고 뒤로 갈수록 악성도가 높은 비정상 세포이고, SCC로 분류되는 세포는 암세포를 의미한다. 따라서 본 논문에서는 각 자궁 세포진 영상에서 WNL, ASCUS, LSIL에 대해서는 각각 60개, HSIL에 대해서는 20개, SCC에서는 10개를 학습데이터로 사용하였다. 따라서 학습 데이터는 총 210개이고 테스트 데이터는 269개이다.

표 2. 실험에 사용된 학습 및 테스트 데이터
Table 2. Training and testing data sets

학습 데이터		테스트 데이터	
WNL	60개	WNL	94개
ASCUS	60개	ASCUS	78개
LSIL	60개	LSIL	65개
HSIL	20개	HSIL	25개
SCC	10개	SCC	7개
Total	210개	Total	269개

표 2에서와 같이 총 210개의 단일 세포에서 추출된 핵의 Hue 정보를 오류 역전과 알고리즘을 적용하여 학습하였다. 신경망의 구성은 입력층의 입력 노드 수는 90개로 설정하였고 은닉층은 노드 수는 실험을 통하여 20개로 설정하는 것이 최적으로 나타났다. 그리고 출력층의 노드 수는 3개로 설정하였다. 학습의 반복 횟수는 1416번으로 나타났다. 그리고 학습률과 모멘텀은 각각 0.7과 0.3로 설정하였다. 학습 데이터와 테스트 데이터에 대한 정상 세포와 비정상 세포의 판별 결과는 표 3과 같다. 표 3에서는 The Bethesda System에서 분류한 WNL, ASCUS, LSIL, HSIL, SCC의 5등급으로 분류하여 인식한 결과이다.

표 3. 정상 및 비정상 세포 인식 결과
Table 3. Recognition rates for the normal and abnormal cervix uteri.

	WNL	ASCUS	LSIL	HSIL	SCC	각각의 악성 세포 등급에 따른 인식률	정상과 비정상에 대한 인식률
WNL	73	9	10	2	0	77.7%	77.6%
ASCUS	21	43	5	9	0	55.1%	73.0%
LSIL	13	5	39	7	1	60.0%	80.0%
HSIL	1	2	2	17	3	68.0%	96.0%
SCC	0	0	0	1	6	85.7%	100%

표 3에서와 같이 각각의 악성 세포 등급에 따른 인식은 평균 69.7%의 인식률로 나타났고, 정상 및 비정상 세포의 인식률은 비교적 높은 평균 85.32%의 인식률을 보였다.

IV. 결 론

자궁 경부 세포진 영상은 그 종류가 다양하고 영상이 복잡해 일반적인 영상 처리로는 정확한 영역 분할이 어려우며, 세포의 정확한 추출을 위한 명확한 기준을 세우기 힘들다. 따라서 본 논문에서는 세포 영역을 효율적으로 분할하기 위해 영역 확장 기법을 적용하였으며, 추출된 세포 영역에서 세포핵의 Hue값을 추출하여 오류 역전과 알고리즘을 적용하여 정상 세포와 비정상 세포를 분류하고 인식하였다.

향후 연구 과제는 세포의 형태학적 특징, 즉 구조적 변화와 비정상 세포의 색조 변화 특징을 모두 분석하여 세포 영역에서 세포핵을 분할하는 방법을 개선하여 임상적인 조직 검사와 비교 분석할 것이다.

참고문헌

- [1] 박종택, 노재운, 조경병, "자궁경부암의 임상 및 병리 조직학적연구," 대한산부회지, Vol. 21, pp.545-57, 1978.
- [2] 사운진, 최영철, 계정용, "자궁암의 역학적 연구," 대한산부회지, Vol.19, pp.181~190, 1976.
- [3] 서채원, 최석진, 홍민기, 이호영, 정우길, "자궁경부암의 진단에 관한 역학적 고찰 및 비정상 세포진 검사 결과에 대한 조직 검사와의 비교연구," 가정의학회지, 제17권, 제1호, pp.76-82, 1996
- [4] Hugo Banda-Gamboa, Ian Ricketts, Alistair Cairns, Kudair Hussein, James H. Tucker and Nasseem Husain, "Automation in cervical cytology : an overview," Analytical Cellular Pathology, 4, pp. 25-48, 1992.
- [5] 남궁성은, 자궁경부 세포진 검사의 자동화 시스템에 대한 연구, 보건복지부
- [6] 강미영, 하진영, 김백섭, 이현길, "자궁경부세포 인식을 위한 가변블록기반 특징 추출 기법," 정보통신논문지, 제5집, pp.66~71, 2001.

- [7] K. B. Kim, C. S. Park, "An Enhanced Fuzzy Multilayer Perceptron," Lecture Notes in Computer Science, LNCS 3316, Springer, pp.977-982, 2004.

저자 소개



김 광 백(Kwang-Baek Kim)

1999년 부산대학교 전자계산학과 (이학박사)

1996년~1997년 동의공업대학 사무자동화과 전임강사

1997년~현재 신라대학교 컴퓨터공학과 부교수

1999년~2000년 Biomedical Fuzzy Systems Association Associate Editors (Japan)

2005년~현재 한국멀티미디어학회 이사 및 논문지 편집위원

2005년~현재 한국해양정보통신학회 이사 및 논문지 편집위원

※ 관심분야 : Fuzzy Neural Networks, Image Processing, Fuzzy Logic, Medical Imaging and Biomedical System, Support Vector Machines



김 성 신(Sungshin Kim)

1984년 2월 연세대학교 전기공학과 졸업(공학사)

1986년 2월 연세대학교 전기공학과 졸업(공학석사)

1996년 8월 Georgia Institute of Technology, School of Electrical and Computer Eng. 졸업 (공학박사)

1998년~현재 부산대학교 전기공학과 부교수

2005년~현재 한국공학교육학회 편집위원

2006년~현재 한국퍼지및지능시스템학회 영문지 편집위원장

※ 관심분야: 지능시스템, 데이터마이닝, 기계학습