

자궁 세포진 핵 영역 분할에 관한 연구

김광백* · 송두현**

1. 서 론

오늘날 우리나라는 자궁경부암에 의한 사망률이 모든 여성 암 중 가장 높은 빈도를 보이고 있다. 자궁경부암은 세포진 검사와 조직 검사를 통해 정확히 조기진단 되면 치료될 수 있는 병이다. 하지만 실제 전문 의료원의 자궁경부암 진단에서 세포진 검사 소견과 치료 전 자궁경부 조직검사 소견이 정확히 일치한 경우가 약 75%였으며 세포진 검사가 조직검사 소견보다 진행된 병변의 경우로 판정된 경우가 약 7.7%였고 세포진 검사에서 조직검사 소견보다 경도의 병변으로 판정된 경우는 17.3%였다[1]. 자궁경부암의 조기 진단에 많이 사용되는 세포진단 검사법(Papanicolaou Smear)은 암의 조기 발견에 우수하지만 검사를 하는 병리학자가 하루에 25-100개의 표본을 검사하므로 오랜 작업의 지루함과 높은 집중력에 따른 피곤에 의해 진단의 위 음성률(비정상을 정상으로 진단하는 확률)이 5~50%정도로 크게 나타날 수 있는 문제점이 있다[2,3]. 그래서 지난 40년 동안 세포진단 검사를 전체 혹은 부분적으로 자동화하는

많은 시스템이 개발되었다[4].

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 진단하는데 자료로 적용된다. 세포의 핵 영역 분할 방법은 대상이 되는 영상에 따라 달라지며, 크게 화소 중심 방법과 영역 중심 방법으로 나뉘어진다. 화소 중심 방법은 각 화소마다 독립적으로 특정 기준에 따라 의미를 부여하게 된다. 화소 중심 방법은 전역적 특징을 이용할 수 있다. 영역 중심 방법은 주변 화소와의 관계를 계산하여 의미를 부여하게 된다. 또한 영역 중심 방법은 계산량이 많아 상대적으로 시간이 많이 소요되지만, 지역적인 특성을 인용할 수 있다.

본 논문에서는 명암값과 5*5 마스크 Closing 연산을 통해 잡음을 제거하고, 잡음이 제거된 영상에 대해 핵 영역을 분할한다. 분할된 핵 영역에서 형태학적 정보 13개와 명암에 관한 조직 정보 8개, 색상에 관한 정보 18개, 길잡에 관한 정보 1개와 wavelet 계수에 대한 정보 6개의 특징 정보를 이용하여 최적의 선택 기준을 정함으로써 추출률을 높일 수 있는 방법을 제안한다.

※ 교신저자(Corresponding Author): 김광백, 주소: 부산광역시 사상구 신라대학길 100번(617-736) 전화 :051)999-5052, FAX : 051)999-5657, E-mail : gbkim@silla.ac.kr

* 신라대학교 컴퓨터공학부 교수
 ** 용인송담대학 컴퓨터게임정보과 교수
 (E-mail : dsong@ysc.ac.kr)

2. 자궁 경부 세포진 핵 영역 분할

2.1 자궁 경부 세포진 영상의 특성과 검사 시스템 구성

자궁 경부 세포진 영상의 경우, 핵과 세포질로 이루어진 자궁 경부 세포, 백혈구 그리고 배경 등으로 구성된다. 주로 관심 있는 부분은 자궁 경부 세포의 핵 부분으로 암을 판별하는 많은 요소가 핵에 집중되어 있다. 하지만 자궁 세포진 영상에서의 핵이 나타나는 색은 슬라이드의 상태나 정상, 비정상에 따라 많은 차이를 보인다[5]. 그림 1은 자궁 경부 세포진 영상의 한 예이다.

본 논문에서 제안하는 세포 핵 영역 검출 과정은 그림 2와 같다. 먼저 전처리 과정을 통해 잡음(백혈구 포함)은 핵보다 크기가 작으므로 이들을 제거한다. 이 과정에서 크기가 작은 자궁 경부 세포핵도 제거되는데, 이렇게 세포핵의 크기가 아주 작은 세포는 정상 세포에 해당되므로 비정상 핵을 검출하는데 영향을 주지 않는다. 전처리 과정을 수행한 후에는 세포 영역에서 핵 영역을 분할한다[6].

자궁경부 세포에서 핵이 검출되면 핵의 변화 및 특징을 분석하기 위하여 핵의 형태학적 정보 13개와 명암에 관한 조직 정보 8개, 색상에 관한 정보 18개, 질감에 관한 정보 1개와 wavelet 계수에 대한 정보 6개의 특징 정보를 추출한다. 추출된 특징

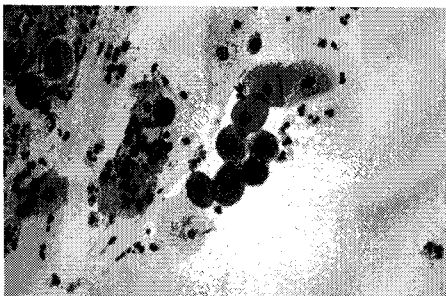


그림 1. 자궁 경부 세포진 영상

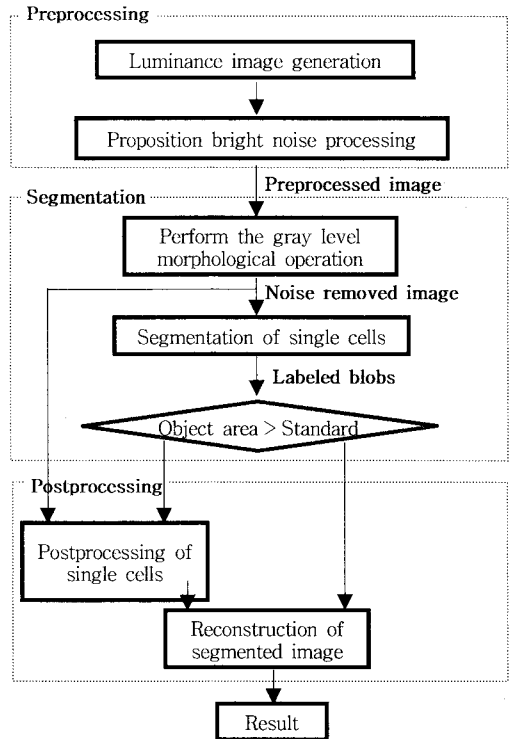


그림 2. 자궁 경부 세포 핵 영역 추출 과정

들의 선택기준은 제안된 평가 함수를 적용하여 각각의 기준을 정하고 정해진 기준에 따라서 암세포와 정상 세포에 대한 핵의 특징을 결정한다.

2.2 전처리

전처리 과정에서는 칼라 영상에서 명암 영상으로 변환하고 획득된 자궁암 세포 영상의 질을 높이기 위해 P-tile 스트레칭을 수행한 영상에 명암 변화량을 이용하여 잡음을 제거한다. 그리고 잡음이 제거된 영상을 8*8 블록으로 나눈 후, 각 블록에 대해 식 (1)을 적용한다.

$$z' = \frac{b' - a'}{b - a} \times (z - a) + a' \quad (1)$$

식 (1)에서 a, b는 원 영상에서 밝기 값이고 z는 $a \leq z \leq b$ 이다. 여기서 a는 최저 밝기 값 + (최고



그림 3. 잡음이 제거된 영상

밝기 값 * 30%)이고 b는 최고 밝기 값 * 80% 이다. $a' = 0$, $b' = 255$ 이다.

제안된 전처리 과정에서 잡음이 제거된 그레이 영상은 그림 3과 같다.

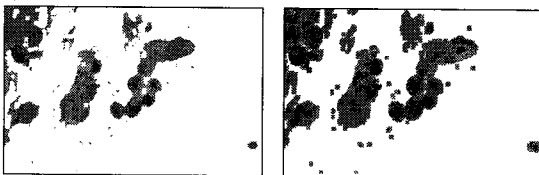
제안된 전처리가 과정을 수행하더라도 자궁 경부 세포진 영상에는 바이러스 등과 같은 잡음이 존재한다. 따라서 이러한 잡음을 제거하기 위하여 형태학적 연산인 5*5 마스크 형태의 Closing 기법을 사용하여 잡음들을 제거한다. Closing 연산은 다음과 같고 Closing 연산 후의 결과 영상은 그림 4와 같다.

$$A \ominus B = x \mid (B)_x \subseteq A \quad (2)$$

$$A \oplus B = x \mid (B)_x \supseteq A \quad (3)$$

2.3 반복 임계값 선정 기법을 이용한 세포핵의 영역 분할

세포핵을 분할하기 위해서 영상을 8*8 블록으



(a) 침식, (b) 팽창

그림 4. Closing 결과 영상. (a) 침식, (b) 팽창

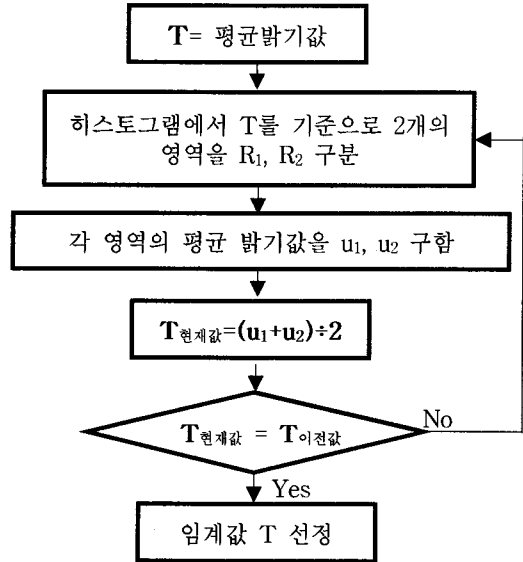


그림 5. 반복적 임계값 선정 알고리즘

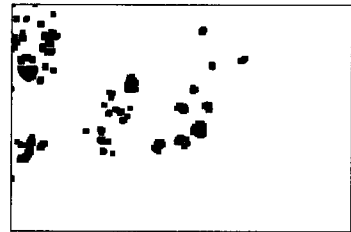


그림 6. 세포핵 영역 분할 영상

로 나눈 후, 잡음이 제거된 영상의 히스토그램 45%~100% 구간에 대해서 그림 5와 같이 반복 임계값 선정 기법을 이용하여 임계값을 선정한다. 그림 6은 반복 임계값 선정 알고리즘을 이용하여 세포핵 영역을 분할한 결과 영상이다.

2.4 암세포 인식을 위한 핵 특징 추출

자궁경부암 세포진 검사에서 나타나는 정상 세포는 세포핵이 작고 진하지 않은 색상을 가지며, 핵과 세포질의 비율이 작다. 이에 반해 비정상 세포는 정상세포에 비하여 핵의 크기가 크거나 핵의 모양이 길쭉하거나 원형에서 벗어난 불규칙한 모

양을 가지게 된다. 또한 비정상 세포핵은 염색되는 정도가 다르므로 핵이 어둡게 나타나며 세포핵 안의 염색질(chromatin)이 균일하게 분포하지 않고 거칠게 나타난다[6].

본 논문에서는 이러한 특징들을 분류하기 위하여 크게 핵 특징과 세포영상 특징을 추출한다. 먼저 핵 특징은 형태에 관한 정보인 핵 면적, 핵 둘레, 핵 둘레와 사각 둘레와의 비율, 핵 모양의 둥근 정도, 핵 모양의 둥근 정도의 역수, 높이와 폭의 비율과 최소사각 영역에서의 $\log_{10}(\text{height}/\text{width})$, 수평·수직방향의 가장 긴 내선과 핵 면적과 사각 영역의 면적 비율을 추출하고 convex영역에서 핵을 감싸는 convex_hull의 면적, 핵 면적을 제외한 면적을 구한다. 명암에 관한 정보에서는 밝기의 평균, 표준편차, 분산, histogram의 최대 명암값, 중앙값, 최소 명암값, 명암값이 60 이상인 pixel수를 추출한다. 색상에 관한 정보는 Red, Green, Blue, Hue, Saturation, Intensity성분의 평균, 표준편차를 구하게 된다. 질감에 관한 정보는 HVS로 나누어진 각 채널에서 에너지 벡터와 에너지 편차를 구한다. 에너지를 사용한 질감 특징 벡터는 다음과 같이 계산된다.

$$q_{mn} = C_{mn} \sum_w \sum_{\theta} [p_{\theta}(w)]^2 \tag{4}$$

$$e_{mn} = \log(1 + p_{mn}) \tag{5}$$

여기서 $P_{\theta}(w)$ 는 각 채널의 주파수 공간에서의 값을 의미한다.

그리고 에너지 편차를 이용한 질감 특징 벡터를 계산하는 식은 다음과 같다.

$$q_{mn} = \sqrt{D_{mn} \sum_w \sum_{\theta} [(P_{\theta}(w))^2 - p_{mn}]^2} \tag{6}$$

$$d_{mn} = \log(1 + q_{mn}) \tag{7}$$

여기서 C_{mn} 와 D_{MN} 는 정규화 값을 구하기 위한

상수이다. 이러한 값과 한 영상의 평균값과 표준편차를 사용하여 질감 기술자는 다음 식 (8)로 표현된다.

$$Descriptor_{texture} = \begin{bmatrix} dc & std & e_{00} & e_{01} & \dots \\ e_{45} & d_{00} & d_{01} & \dots & d_{45} \end{bmatrix} \tag{8}$$

세포 영상에서 특징을 추출하기 위해 wavelet 계수를 이용하여 LL1과 LL2의 평균, 분산, 에너지 값의 특징을 추출한다. 본 논문에서 적용한 웨이블릿 변환 구조는 그림 7과 같고 저주파 대역과 고주파 대역 필터는 식 (9)와 같다.

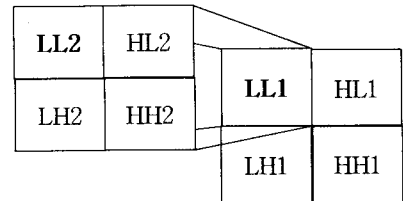


그림 7. 적용한 웨이블릿 변환 구조

$$\begin{bmatrix} 1 + \sqrt{3} & 3 + \sqrt{3} \\ 3 - \sqrt{3} & 1 - \sqrt{3} \end{bmatrix} \times 4\sqrt{2} \tag{9}$$

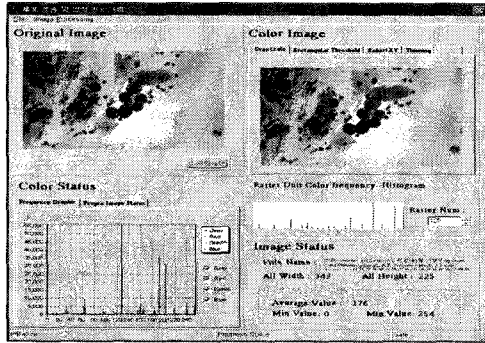
$$\begin{bmatrix} 1 - \sqrt{3} & \sqrt{3} - 3 \\ 3 + \sqrt{3} & -1 - \sqrt{3} \end{bmatrix} \times 4\sqrt{2}$$

2.5 실험 및 결과분석

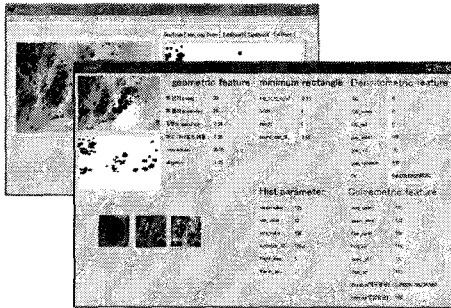
실험 환경은 IBM 호환 기종의 Pentium-IV PC 상에서 C++ 빌더로 구현하였다. 실험 표본은 부산대학교 부속 병원에서 획득한 640*480 크기의 자궁 경부 세포 영상을 사용하였다.

자궁암 경부 세포진 영상에서 제안된 세포핵 분할 추출 화면은 그림 8(a)과 같고, 자궁암 경부 세포진에서 추출한 핵 대한 특징 값은 그림 8(b)와 같다.

진단 시스템에서 비정상적인 세포가 정상 세포로 판정되는 것은 큰 문제가 되지만 정상인 세포



(a)



(b)

그림 8. 분할 화면과 특징 정보. (a) 제안된 핵 영역 분할 화면, (b) 추출된 핵에 대한 특징 정보

가 비정상 세포로 판정되는 것은 큰 문제가 되지 않는다. 하지만 영상에 나타나는 모든 핵을 분할하는 것보다 관심의 대상이 되는 세포핵을 분할하는 것이 인식 단계에서 더욱 효율적이다[7].

전문의의 임상적인 실험 정보와 세포 진단학 정보를 기반으로 제안된 방법에 의해 추출된 핵에 대한 특징 정보 중에서 일부는 다음과 같다.

일반적으로 자궁 경부 세포진에서 핵을 이용하여 정상 세포와 비정상 세포로 분류한다. 비정상 세포는 악성도에 따라 4 등급으로 분류하며 악성도가 매우 높은 세포는 암세포가 된다. 따라서 본 논문에서는 자궁암 세포를 판별하기 위한 전 단계로 자궁암 경부 세포진에서 정상 세포와 비정상 세포에서 추출한 핵의 개수를 표 2로 나타내었다.

표 1. 추출된 핵의 일부 특징 정보

핵의 특징 정보	핵의 특징 값
핵 둘레와 사각둘레와의 비율	0.5791
핵 모양의 등근 정도	0.8571
핵 면적과 사각영역의 면적 비율	0.6457
핵을 감싸는 convex_hull의 면적	0.5424
밝기의 표준편차	0.8721
밝기의 분산	0.7452
wavelet LL2 평균	0.7611
wavelet LL2 분산	0.7562

표 2. 정상 세포와 비정상 세포에서의 핵 추출 결과

전문의			실험 결과		
정상 세포 핵	비정상 세포		정상 세포 핵	비정상 세포 핵	
	암세포 핵	바이러스		암세포 핵	바이러스
10	10	15	10	9	15

표 2에서 전문의가 판단한 암세포에 대해 제안된 방법에서는 1개의 암세포 핵이 추출되지 않았다. 그 이유는 2개의 암세포가 모여서 암진주를 형성하여 추출되지 않았다. 그러나 정상 세포의 핵은 모두 추출된 것을 확인할 수 있다.

3. 결 론

자궁 경부 세포진은 그 종류가 다양하고 복잡하여 기존의 영상 처리로는 효과적인 세포의 핵 분할이 어려웠으며, 시간과 비용이 많다는 단점이 있었다. 이러한 부분을 개선하기 위하여 본 논문에서는 원 영상에서 바로 핵 영역의 특징을 추출함으로써 시간과 비용을 최소화하였다. 그리고 추출률을 높이기 위하여 세포 핵 영역의 특징 정보를 이용하여 최적의 선택 기준을 만듦으로써 비정상 핵 영역 추정치에 편차를 줄일 수 있었다.

향후 과제로는 cluster cells 분류에 대해 세포

의 형태적 특징과 핵의 색상 변화를 분석하여 암 세포의 특징을 정확히 추출할 것이고, 이를 기반으로 새로운 암세포 분류를 위한 클러스터링 기법을 연구할 것이다.

참 고 문 헌

[1] 서채원, 최석진, 홍민기, 이호영, 정우길, “자궁경 부암의 진단에 관한 역학적 고찰 및 비정상 세포진 검사 결과에 대한 조직검사와의 비교연구,” 가정의학회지, 제17권, 제1호, pp. 76-82, 1996.

[2] 은성경, 박찬모, 박화순, 윤소영, 조민선, 조수연, 김성숙, “영상처리를 이용한 자궁경부 세포진의 자동탐색 방법에 관한 연구,” 대한세포병리학회지, 제5권, 제1호, pp. 15-22, 1994.

[3] Heinz K. Grohs, O.A. Nasseem Husain, “Automated Cervical Cancer Screening,” Igaku-shoin, 1994.

[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] 김호영, 김선아, 최예찬, 김백섭, 김희숙, 남궁성은, “다단계 영역 분할 방법을 이용한 자궁 경부 세포진 핵 영역 분할에 관한 연구,” 대한의료정보학회지, 제 5권, 제1호, pp. 89-95, 1999.

[6] 남궁성은, “자궁경부 세포진 검사의 자동화 시스템에 대한 연구”, 보건복지부.

[7] Kittler J. and L.Lingworth J. “Threshold Selection Based on a Simple Image Statistic.”, Computer Vision, Graphics, and Image Processing, 1985. Vol.30. pp. 125-147.



김 광 백

- 1999년 부산대학교 전자계산학과 (이학박사)
- 1997년~현재 신라대학교 컴퓨터정보공학부 교수
- 1999년~2000년 Biomedical Fuzzy Systems Association Associate Editors (Japan)
- 2005년~현재 한국해양정보통신학회 이사 및 논문지 편집부위원장
- 2005년~현재 한국멀티미디어학회 이사 및 논문지 편집분과위원장
- 관심분야 : Neural Networks, Image Processing, Fuzzy Logic, Medical Imaging and Biomedical System, Support Vector Machines



송 두 현

- 1981년 서울대교 자연대학 계산통계학과(이학사)
- 1983년 KAIST 전산학과(석사)
- 1994년 University of California at Irvine 컴퓨터과학과 (이학박사 수료)
- 1997년~현재 용인송담대학 컴퓨터게임정보과 교수
- 관심분야 : 인공지능, 기계학습, 비전, 게임 소프트웨어공학, 인지과학 등