

퍼지 c-Means 클러스터링 알고리즘을 이용한 자궁 세포진 핵 인식에 관한 연구

허정민 · 김정민 · 김광백
신라대학교 컴퓨터공학과

A Study on Nucleus Recognition of Uterine Cervical Pap-Smears using Fuzzy c-Means Clustering Algorithm

Jung-Min Heo · Jung-Min Kim · Kwang-Baek Kim
Dept. of Computer Engineering, Silla University

요 약

자궁 경부 세포진 영상의 핵 영역 분할은 자궁 경부암 자동화 검색 시스템의 가장 어렵고도 중요한 분야로 알려져 있다. 본 논문에서는 자궁 경부 세포진 영상에서 HSI모델을 이용하여 세포진 핵 영역을 추출한다. 추출된 세포진 핵 영역은 형태학적 정보(morphometric feature)와 명암 정보(densitometric feature), 색상 정보(colorimetric feature), 질감 정보(textural features)를 분석하여 핵의 특징을 추출한다. 또한 Bethesda System에서의 분류 기준에 따라 핵의 분류 기준을 정하고 추출된 핵의 특징들을 퍼지 c-Means 클러스터링 알고리즘에 적용하여 실험한 결과, 제안된 방법이 자궁 세포진 핵 추출과 인식에 있어서 효율적임을 확인하였다.

키워드

자궁 경부 세포진, 핵, 분할, 특징 추출, 질감 정보, 퍼지 c-Means

1. 서 론

자궁경부암은 한국 여성 성기의 악성종양 중 발병빈도가 가장 흔한 종양으로 조기진단 되면 치료 가능성이 높으며 정복하지 않으면 안 될 중요한 질병 중의 하나이다. 자궁경부암은 우리나라 전체 악성종양의 26.3%~68.2%를 점유하는 것으로 보고된 바 있다[1][2]. 이런 자궁경부암으로의 이행을 차단하는 것이 자궁경부암을 완치하는 가장 확실한 방법이고, 그로인해 지난 40년 동안 세포진단 검사를 전체 혹은 부분적으로 자동화하는 많은 시스템이 개발되었다[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 진단하는데 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 특징을 인용할 수 있다[5].

본 논문에서는 획득된 자궁세포진 영상을 그레이 스케일 영상으로 변환 후, 영상의 밝기 값을 이용하여 잡음을 제거하고 5*5 마스크 형태의 퍼지 그레이 모폴로지 연산자를 이용하여 영상을 단순화시켜서 기존의 영상에서 보다 쉽게 관심을 갖는 세포핵을 찾을 수 있게 한다. 단순화된 영상에서 핵 영역을 분할하기 위해 어두운 영역에 대해 핵을 분할하고, 분할된 핵 영역에서 형태학적 정보 13개와 명암에 관한 조직 정보 8개, 색상에 관한 정보 18개, 질감에 관한 정보 1개의 특징 정보를 추출한다. 추출된 정보들은 퍼지 c-Means 클러스터링 알고리즘을 적용하여 각 핵의 비정상 정도를 4등급으로 분류하고 인식한다.

II. 자궁 경부 세포진 핵 영역 분할

본 논문에서 제안하는 자궁 세포진 핵 추출

과정은 그림 1과 같다.

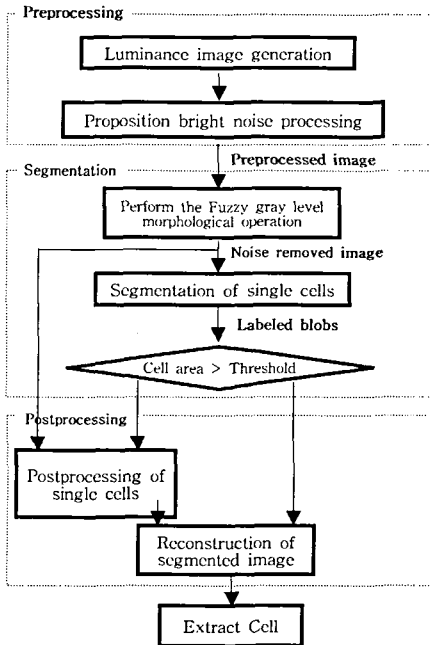


그림 1. 자궁 경부 세포 핵 영역 추출 과정

본 논문에서는 전처리 과정을 통해 잡음(백혈구 등)은 핵의 크기보다 적다는 정보를 이용하여 잡음을 제거한다. 이 과정에서 크기가 작은 자궁 경부 세포핵도 제거되는데, 이렇게 세포핵의 크기가 아주 작은 세포는 정상 세포에 해당되므로 비정상 핵을 검출하는데 영향을 주지 않는다[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 1개의 특징 정보를 추출한다. 추출된 핵의 특징에 대한 분류 기준은 Bethesda System에서 제안한 방법을 적용 하고 정해진 기준에 따라 추출된 핵에 대해서 악성 정도를 분류하기 위하여 퍼지 c-Means 클러스터링 알고리즘을 적용한다.

1. HSI 정보를 이용한 잡음 제거

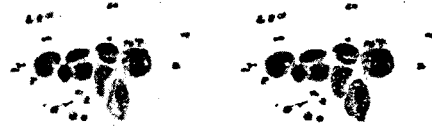
전처리 과정에서는 칼라 영상을 명암 영상으로 변환하고 획득된 자궁암 세포 영상의 질을 높이기 위해 P-tile 스트레칭을 수행한 영상에 명암 변화량을 이용하여 잡음을 제거한다. 그리고 잡음이 제거된 영상을 8*8 블록으로 나눈 후, 각 블록에 대해 식 (1)을 적용한다.

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

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

고 밝기 값 * 30%)이고 b는 최고 밝기 값 * 80% 이다. a'와 b'는 각각 0과 255이다.

본 논문에 적용한 자궁 경부 세포진 영상은 그림 2(a)와 같고 제안된 전처리 과정에서 잡음이 제거된 그레이 영상은 그림 2(b)와 같다.



(a) 자궁 경부 세포진 영상 (b) 잡음이 제거된 영상
그림 2. 잡음이 제거된 자궁 경부 세포진 영상

2. 퍼지를 이용한 모폴로지 연산

제안된 전처리 과정에 의해 잡음이 제거되면 정상 세포의 핵과 암 세포 핵의 일부 정보가 손실되어 핵을 정확히 추출하기가 어렵다. 따라서 이러한 문제점을 개선하기 위하여 형태학적 연산인 5*5 마스크 형태의 퍼지 그레이 모폴로지 연산을 적용하여 정상 세포와 비정상 세포의 핵을 정확히 추출할 수 있도록 한다. 퍼지 모폴로지 연산은 식(2)와 식(3)과 같다. 5*5 마스크 형태의 퍼지 그레이 Erosion · Dilation 형태학적 모폴로지 연산을 적용한 결과 영상은 그림 3과 같다.

$$A \ominus B = \{(x, \mu_{A \ominus B}(x)) \mid x \in E^N\}$$

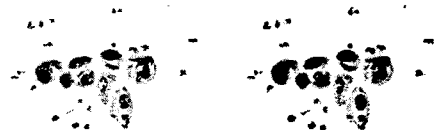
$$\mu_{A \ominus B}(x) = \inf_{z \in E^N} \min[1, MRF(\mu_A(z), \mu_{(B:z)}(z))]$$

$$MRF(a, b) = \begin{cases} 0 & \text{if } b = 0 \\ \frac{a}{b} & \text{otherwise} \end{cases} \quad (2)$$

$$A \oplus B = \{(x, \mu_{A \oplus B}(x)) \mid x \in E^N\}$$

$$\mu_{A \oplus B}(x) = \sup_{z \in E^N} [\mu_A(z) \times \mu_{(B:z)}(z)] \quad (3)$$

여기서 a는 원 영상이고 b는 5*5 마스크이다.



a. Erosion b. Dilation
그림 3. Closing 결과 영상

3. 반복 임계값 선정 기법을 이용한 세포핵의 영역 분할

제안된 전처리 방법에 의해서 단순화된 영상의 히스토그램 45%~100% 구간에 대해서 그림 4와 같은 반복 임계값 선정 기법을 이용하여 임계값을 선정한다. 그림 5은 반복 임계값 선정 알고리즘을 이용하여 세포핵 영역을 분할한 결과 영상이다.

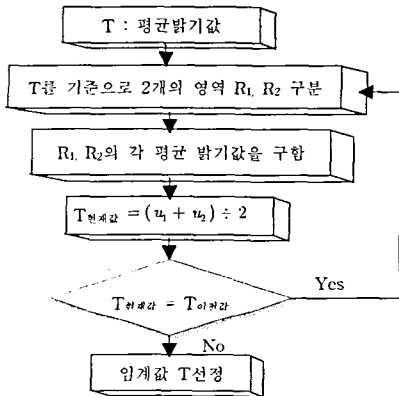


그림 4. 반복적 임계값 선정 알고리즘



그림 5. 세포핵 영역 분할 영상

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 모양을 가지게 된다. 또한 비정상 세포핵은 염색되는 정도가 다르므로 핵이 어둡게 나타나며 세포핵 안의 염색질(chromatin)이 균일하게 분포하지 않고 거칠게 나타난다[7][8].

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 사각영역의 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 픽셀 수를 추출한다. 색상에 관한 정보는 Red, Green, Blue, Hue, Saturation, Intensity 성분의 평균, 표준편차를 구한다. 질감에 관한 정보는 HVS로 나누어진 각 채널에서 에너지 벡터와 에너지 편차를 구한다. 에너지를 사용한 질감 특징 벡터는 다음과 같이 계산한다.

$$q_{mn} = C_{mn} \sum_w \sum_{\phi} [P_{\phi}(w)]^2 \quad (4)$$

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

여기서 $P_{\phi}(w)$ 는 각 채널의 주파수 공간에서의 값을 의미한다. 여기서 C_{mn} 은 정규화 값을 구하기 위한 상수이다. 그리고 에너지 편차를 이용한 질감 특징 벡터를 계산하는 식은 다음과 같다[9].

$$r_{mn} = \sqrt{D_{mn} \sum_w [(f_w^n(w))' - p_{mn}]^2} \quad (6)$$

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

여기서 D_{mn} 은 정규화 값을 구하기 위한 상수이다. 식 (4)에서부터 식(7)까지 구한 값과 한 영상의 평균값과 표준 편차를 이용하여 핵의 질감 정보를 나타내기 위한 질감 기술자는 식 (8)로 표현된다.

$$Descriptor_{i,j,k,l,m} = [\begin{matrix} dc & std & e_{00} & e_{01} & \dots \\ e_{45} & d_{00} & d_{01} & \dots & d_{45} \end{matrix}] \quad (8)$$

5. FCM 클러스터링 알고리즘을 이용한 핵 분류 및 인식

기본적으로는 퍼지-c-Means 클러스터링 알고리즘은 데이터 집합의 유클리디안 거리를 이용하여 각각의 클러스터를 분할하는 K-means 클러스터링 알고리즘과 유사하다[10]. 다만, 차이점은 데이터 개체들이 각 클러스터에 소속하는지(혹은 1)에 대한 정보가 아니라, 각 클러스터에 소속하는 정도(0과 1사이의 실수)를 분할 행렬(U)로 표시한다는 점과 초기 중심을 지정할 때 임의로 만들어진 초기분할 행렬을 가지고, 전체 데이터 분포의 중간 정도에 초기 중심을 설정한다는 점이다. 본 논문에서 적용한 퍼지 c-Means 클러스터링 알고리즘은 그림 6과 같다.

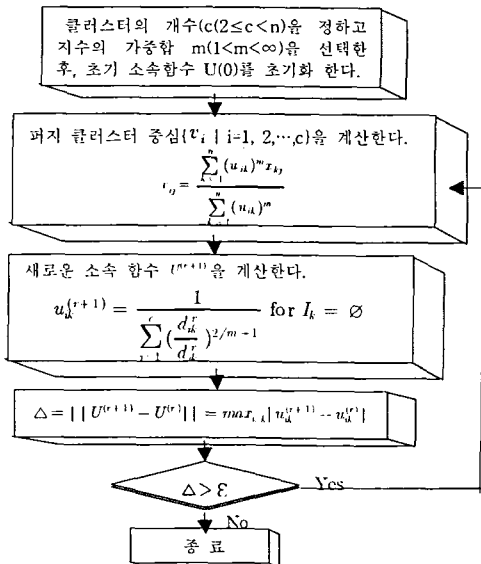
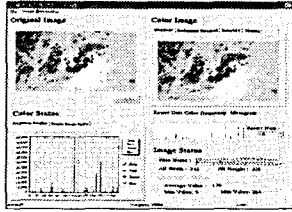


그림 6. 퍼지 c-Means 클러스터링 알고리즘

7. 실험 및 결과분석

실험 환경은 IBM 호환 기종의 Pentium-IV PC 상에서 Visual C++ 6.0와 C++ Builder 6.0로 구현하였다. 실험 표본은 부산대학교 부속 병원에서 획득한 640*480 크기의 자궁 경부 세포 영상 20장을 대상으로 실험하였다. 제안된 자궁암 경부 세포진 영상 분석 결과는 그림 7(a)와 같고, 자궁암 경부 세포진 핵 추출 결과는 그림 7(b)와 같다.



(a). 제안된 자궁암 경부 세포진 영상 분석 결과



(b) 자궁암 경부 세포진 핵 추출 결과
그림 7. 자궁암 경부 세포진 핵 분석 결과

실제 영상 20장에서 제안된 핵 영역 분할 방법의 성능을 평가하기 위해 전문의가 진단한 결과와 제안된 핵 추출 방법의 결과를 비교하였다. 20장의 자궁 경부 세포 영상에서 전문의가 추출한 자궁암 경부 세포진의 핵은 316개이고, 본 연구를 통해 추출된 핵은 278개이다. 표 1은 제안된 방법에 의해 추출된 핵의 수를 나타내었다.

표 1. 핵 추출 결과

	전문 의	제안된 방법	제안된 방법의 추출률
핵 추출	316	278	87.9%

표 1에서 알 수 있듯이 본 연구를 통해 추출된 핵의 추출률은 87.9%의 정확도를 가졌다. 2개 이상 세포가 겹쳐서 있는 클러스터 세포인 경우에는 정확히 추출되지 않았다.

세포핵은 정상 세포의 핵과 비정상 세포의 핵 및 암세포 핵으로 분류된다. 따라서 본 논문에서는 핵의 분류를 Bethesda System에서 분류한 WNL, ACUS, LSIL, HSIL, SCC의 5 등급으로 분류하였다. 여기서 WNL은 정상세포이고 뒤로 갈수록 악성도가 높은 비정상 세포이고, SCC로 분류되는 세포는 암 세포를 의미한다.

그림 8은 본 논문에서 추출한 자궁암 경부 세포진 핵의 특징 정보들을 나타내었다.

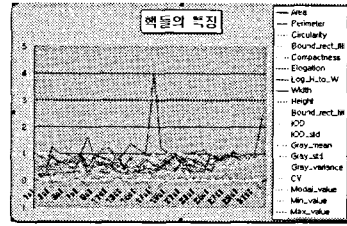


그림 8. 핵의 특징 정보 결과

그림 8과 같은 핵의 특징 정보들을 이용하여 Bethesda System에서 사용한 분류 기준에 따라 핵의 특징 정보들을 분류할 기준에 대한 정보는 그림 9와 같다.

그림 9와 같이 핵의 분류 기준 정보를 통해 정상 세포와 비정상 세포의 등급 및 암세포를 판별할 수 있는 정보로 분류를 함으로써 영상에 나타나는 모든 핵을 분할하는 것보다 관심의 대상이 되는 세포핵을 분류하여 암 세포 진단의 정확성을 향상시킬 수 있다.

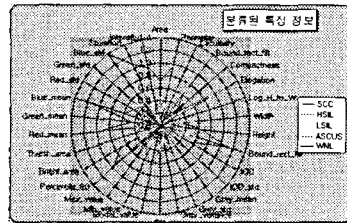


그림 9. Bethesda System에 의한 핵의 분류 기준 정보

그림 8와 같이 추출한 핵의 특징 정보들을 퍼지 c-Means 클러스터링 알고리즘을 적용하여 정상 세포와 비정상 세포 및 암세포를 분류하고 판별하였다. 최종 FCM 클러스터링 알고리즘을 적용하여 세포의 상태를 분류 및 판별한 결과는 표 2와 같다.

표 2. FCM 클러스터링 알고리즘에 의한 세포의 상태 분류 및 인식 결과

전문 의		
정상 세포(WNL)		98
비정상 세포	ASCUS	88
	LSIL	40
	HSIL	61
암세포(SCC)		29
제안된 방법		
정상 세포(WNL)		78
비정상 세포	ASCUS	100
	LSIL	45
	HSIL	64
암세포(SCC)		29

표 2에서 Bethesda System에서 분류한 WNL, ACUS, LSIL, HSIL, SCC의 5 등급으로 분류하기 위하여 클러스터의 수를 5개로 설정하여 분류하였다. 표 2에서 알 수 있듯이 전문의가 진단한 결과보다 제안된 방법이 정상 세포를 비정상 세포로 분류한 경우가 발생하였다. 그 이유는 추출된 정상 세포의 핵이 농염되어서 비정상 세포로 분류된 경우이다. 그러나 표 2와 같이 전문의가 진단한 비정상 세포와 암세포에 대해 어느 정도의 정확성을 가졌으나 비정상 세포의 등급 분류에서는 차이가 있는 것을 확인할 수 있다. 그 이유는 비정상 세포핵 정보 중에 한 부분인 핵의 농염 정도를 FCM 클러스터링 알고리즘이 정확히 분류하지 못한 경우이다. 그러나 제안된 방법이 비교적 비정상 세포와 암세포 인식에 있어서 효율적인 것을 확인할 수 있고 전문의의 진단에 도움을 줄 수 있다는 것을 표 2를 통해서 알 수 있다.

III 결론

자궁 경부 세포진은 그 종류가 다양하고 복잡하여 기존의 영상 처리로는 효과적인 세포의 핵 추출 및 인식이 어렵다.

본 논문에서는 핵의 분류를 Bethesda System에서 분류한 WNL, ACUS, LSIL, HSIL, SCC의 5 등급 기준 정보를 이용하여 자궁 경부 세포진 핵 추출 및 인식 방법을 제안하였다. 제안된 방법은 획득된 자궁세포진 영상을 그레이스케일 영상으로 변환 후, 영상의 밝기 값을 이용하여 잡음을 제거하였고 5*5 마스크 형태의 퍼지 그레이 모폴로지 연산자를 이용하여 영상을 단순화시켜서 기존의 영상에서 보다 쉽게 관심을 갖는 세포핵을 찾을 수 있게 하였다. 단순화된 영상에서 핵 영역을 분할하기 위해 어두운 영역에 대해 핵을 분할하였고, 분할된 핵 영역에서 형태학적 정보 13개와 명암에 관한 조직 정보 8개, 색상에 관한 정보 18개, 질감에 관한 정보 1개의 특징 정보를 추출하였다. 추출된 정보들은 퍼지 c-Means 클러스터링 알고리즘을 적용하여 각 핵에 대해 정상 세포와 비정상 세포 3등급 및 암세포로 분류하여 인식하였다.

자궁 경부 세포 영상 20장을 대상으로 실험한 결과, 제안된 방법이 비교적 비정상 세포와 암세포 인식에 있어서 효율적인 것을 확인하였고 전문의가 자궁 경부 세포진 진단에 추정치에 오차가 적음을 확인하였다.

향후 과제로는 클러스터 세포의 분류에 대해 세포의 형태적 특징과 핵의 색상 변화를 분석하여 핵의 특징 정보를 정확히 추출할 것이고, 핵의 분류에 있어 추정치 편차를 줄이도록 여러 특징 정보를 추출하여 분류 기준을 만들도록 연구할 것이다.

참고문헌

- [1] 박종택, 노재윤, 조정병, 자궁경부암의 임상 및 병리조직학적연구. 대한산부회지, Vol.21, pp.545-575, 1978.
- [2] 사운진, 최영철, 계정용, 자궁암의 역학적연구, 대한산부회지, Vol.19, pp.181-190, 1976.
- [3] 서채원, 최석진, 홍민기, 이호영, 정우길, "자궁경부암의 진단에 관한 역학적 고찰 및 비정상 세포진 검사 결과에 대한 조직검사와의 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, Vol.4, pp.25-48, 1992.
- [5] 남궁성은, "자궁경부 세포진 검사의 자동화 시스템에 대한 연구", 보건보지부.
- [6] 김호영, 김선아, 최예찬, 김백섭, 김희숙, 남궁성은, "다단계 영역 분할 방법을 이용한 자궁 경부 세포진 핵 영역 분할에 관한 연구," 대한의료정보학회지, 제5권, 제1호, pp.89-95, 1999.
- [7] 김광백, 윤홍원, "퍼지 신경망을 이용한 기관지 편평암 인식에 관한 연구," 대한 의료정보학회지, 제5권, 제1호, pp.77-87, 1999.
- [8] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
- [9] B. S. Manjunath, W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.18, No.8, 1996.
- [10] George J. Klir, Bo Yuan, Fuzzy Sets and Fuzzy Logic Theory and Applications, Prentice Hall PTR, 1995.