
자궁 경부진 핵 추출에 관한 연구

김광백*

A Study on Nucleus Extraction of Uterine Cervical Pap-Smears

Kwang-baek Kim*

요 약

자궁경부암은 다른 암과 달리 전암(前癌) 단계가 존재하므로 조기 발견할 경우에 생존율이 높다. 그러나 의사나 병리학자가 하루에 검진할 수 있는 양은 제한되어 있다. 따라서 본 논문에서는 세포 도말 검사에 사용되는 자궁 경부진 세포에서 핵을 추출하는 방법을 제안한다. 조기 자궁 경부 세포진 영상에서 핵의 추출은 영상의 배경 그리고 핵과 세포질 영역의 구분이 중요하기 때문에 Lighting Compensation을 적용하여 영상을 보정하고, 명암도의 분포가 가장 작은 B 채널에서 3×3 마스크를 이용하여 잡음을 제거한다. 잡음이 제거된 영상을 이진화하고 Grassfire 알고리즘을 적용하여 세포 객체를 추출한다. 추출된 세포 객체 중에서 군집화된 세포 영역에 대해서는 R 채널의 명암도 값을 반복 이진화에 적용하여 핵 영역을 추출한다. 실제 진단 세포학에서 사용하는 자궁경부 세포진 400 배율 영상을 대상으로 실험한 결과, 45개의 세포 영역 중에서 40개의 핵이 추출되었다.

ABSTRACT

If detected early enough, cervical cancer may have a good survival rate due to its preneoplastic state. However, the process is so time consuming that a medical expert can handle only a small amount of such examinations. In this paper, we propose a new nucleus extraction algorithm for uterine cervical pap smears in order to mitigate such burdens of medical experts. In the preneoplastic state cytodiagnosis images, it is important to differentiate three main areas - background, cytoplasm and nucleus. Thus, we apply lighting compensation and 3×3 mask of B channel in order to restore damaged image and remove noises respectively. The cell object is obtained from those clean binarized images with Grossfire algorithm. When there are clusters of cells, the target nucleus can be obtained with repetitive binarization of R channel brightness. In our experiment of using uterine cervical pap smears of 400 magnifications that is common in the diagnostic cytology, our method is able to extract 40 nucleus out of 45 population successfully.

키워드

cervical cancer, uterine cervical pap smears, cytoplasm, nucleus

* 신라대학교 컴퓨터정보공학부

접수일자 2009. 04. 15
심사완료일자 2009. 05. 11

I. 서 론

우리나라의 여성 자궁경부암 사망률은 매년 감소하고 있다. 하지만 자궁경부암 발생률은 세계에서 중간정도에 위치하고 있다. 사망률 감소의 가장 큰 원인은 조기 검진이 확대됐기 때문이다. 자궁경부암은 10~20년 동안 전암(前癌) 단계가 있으므로 조기 발견하여 치료하면 생존율이 높다.

자궁경부암에 대한 검진법으로는 세포 도말 검사(pap smear), 골반 검사, 질경 검사, 인유두종 바이러스 검사(HPV-PCR) 등이 있다[1]. 이 중 조기 전암 단계에서의 검진은 세포 도말 검사가 주로 쓰인다. 세포 도말 검사는 자궁 입구 주변의 여러 곳에서 세포조직 표본을 채취하여 염색한 후, 현미경으로 이상 징후를 검사하는 방법으로 상당히 정확한 검진 결과를 가진다. 하지만 높은 암 발생 빈도에 비해 의사나 병리학자가 하루에 검진할 수 있는 양은 제한되어 있다[2].

따라서 본 논문에서는 세포 도말 검사에 사용되는 자궁 경부 세포핵을 추출하는 연구를 통하여 효과적인 세포 검진을 할 수 있도록 하고 세포 감별의 위음성률 및 위양성률을 낮추는데 도움을 주고자 한다. 의사와 병리학자들이 검사에 사용하는 도말된 세포 영상은 보통 400~1000 배율이다[3].

본 논문에서는 400 배율의 자궁경부 편평세포 암종 영상을 이용하여, 암종 세포핵의 형태학적 특징을 이용하여 세포를 추출한다.

II. 세포 영역 추출

본 논문에서 제안하는 자궁 경부진 핵 영역 추출의 전체 과정은 그림 1과 같다.

2.1 Lighting compensation에 의한 영상 보정

세포 영상 획득 시 빛의 영향을 받아 잘못 분류 될 수 있는 영역의 오류를 보정하기 위해 Lighting Compensation 방법이 제안되었다[4].

따라서 본 논문에서는 Lighting Compensation을 적용하여 세포 영상 R, G, B 각 채널에 대해 명암도 분포를 조사하여, 각 채널에 대한 상위 5%에 해당하는 영역을 보정한다.

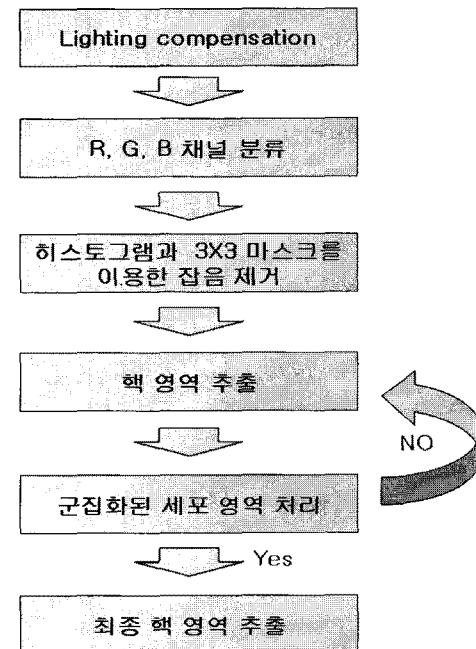


그림 1. 자궁 경부진 핵 영역 추출 처리 과정
Fig. 1 Overall nucleus extraction process

각 채널의 5% 포함되는 명암도의 평균 R_m, G_m, B_m 의 값을 계산하고, 식 (1)과 같이 $R_{rate}, G_{rate}, B_{rate}$ 의 비율을 구한다.

$$R_{rate} = \frac{255}{R_m}, G_{rate} = \frac{255}{G_m}, B_{rate} = \frac{255}{B_m} \quad (1)$$

계산된 각각의 비율을 식 (2)와 같이 계산하여 세포 영상에 보정한다.

$$\begin{aligned} R_{result} &= R_{original} \times R_{rate} \\ G_{result} &= G_{original} \times G_{rate} \\ B_{result} &= B_{original} \times B_{rate} \end{aligned} \quad (2)$$

Lighting Compensation을 적용하여 세포 영상을 보정한 결과는 그림 2와 같다.

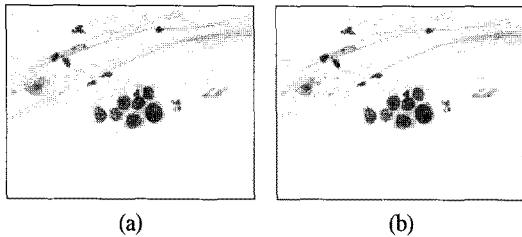


그림 2. Lighting Compensation 적용하여 보정된 효과
(a) 세포 영상 (b) 보정된 영상

Fig. 2 The effect of Lighting Compensation
(a) Original cell image (b) After compensation

2.2 R, G, B 채널 분류

그림 3과 같이 세포 영상은 R, G, B 채널에 따라 명암도의 분포도가 많이 다르다. R 채널에서는 명암도의 분포가 0~255사이에 어두운 명암과 밝은 명암이 모두 존재 한다. 그러나 B 채널에서는 명암도의 분포가 어두운 명암이 거의 존재하지 않는다. 즉 B 채널에서는 R 채널에 서 보다 분포의 영역이 좁은 것을 확인할 수 있다. 분포의 영역이 좁게 되면, 세포 영상에서 세포의 핵과 세포질, 그리고 백혈구의 명암도가 비슷하게 나타난다. 따라서 본 논문에서는 분포의 영역이 좁은 B 채널 영상에서

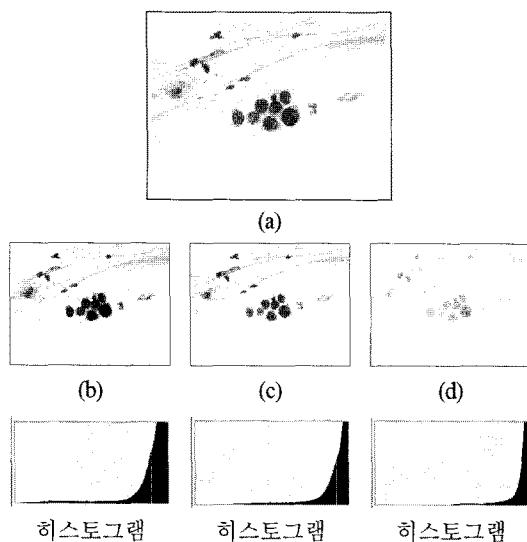


그림 3. 각 채널에 대한 세포 영상과 히스토그램
(a) 보정된 영상 (b) R 채널 (c) G 채널 (d) B 채널

Fig. 3 Cell image and Histogram of each channel
(a) Image after compensation (b) R channel
(c) G channel (d) B channel

세포의 핵과 세포질을 그룹화하고 세포의 형태학적 정보를 이용하여 백혈구를 제거한다.

2.3 히스토그램과 3×3 마스크를 이용한 잡음 제거
그림 4와 같이 각 채널의 히스토그램에서 상위 80%는 세포핵과 세포질, 백혈구를 제외한 잡음이 포함되어 있다. 따라서 명암도 80% 이상의 영역에 대해서는 배경 잡음으로 간주하여 제거한다. 하지만 배경 잡음을 사이의 미세한 잡음이 남게 된다.

따라서 미세한 잡음을 제거하기 위해 그림 5와 같은 3×3 마스크를 이용하여 잡음을 제거한다. 중심 화소의 주위 8 방향을 스캔 한 후, 흰색 픽셀이 검정색 픽셀보다 많이 존재할 경우, 중심 화소를 흰색으로 대체함으로써 영상에 존재하는 미세 잡음을 제거한다.

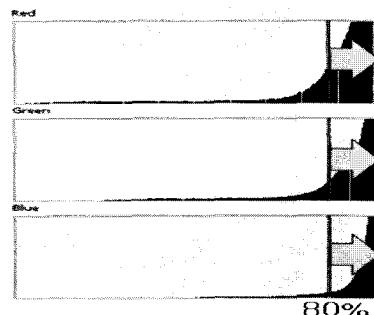


그림 4. 배경의 잡음 영역
Fig. 4 Noisy area of background

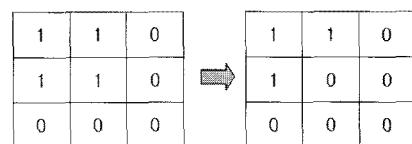


그림 5. 3×3 마스크 구조
Fig. 5 3×3 mask structure

2.4 세포 영역 추출

본 논문에서는 잡음이 제거된 세포 영상에서 세포 영역을 추출한다. 명암 분포의 영역이 좁은 B 채널에서 식 (3)을 적용하여 이진화를 한다.

```
if (brightness != 255) Result = 0;
else Result = 255; (3)
```

이진화된 영상에서 *Grassfire* 알고리즘[5]을 적용한 후에 표 1과 같은 세포의 형태학적 특징[6]을 이용하여 객체를 추출한 결과는 그림 6과 같다.

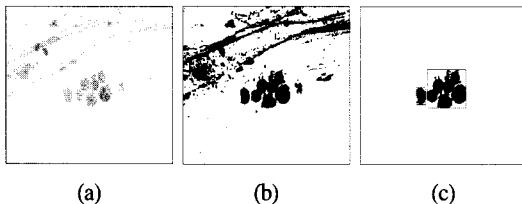


그림 6. 세포 추출 결과 (a) B 채널 영상 (b) 이진화 영상 (c) 추출된 세포 객체 (b) After binarization
Fig. 6 Result of cell extraction (a) B channel image
(c) Cell extracted

표 1. 세포의 형태학적 특징
Table 1. Morphological feature of cell

세포의 특징	
1. 핵 크기 / 세포의 크기가 각각의 세포마다 거의 유사하다.	
2. 핵의 위치가 세포의 중간에서 거의 벗어나지 않는다.	
3. 핵 크기가 세포질의 크기에 비해 매우 작은 비율을 차지 한다.	
암종 세포의 특징	
암 세포는 크게 여섯 가지의 형태를 가지며, 세포가 다음 여섯 가지 중에 포함되면 암세포로 분류한다.	
1. 암종세포 1 : 등근 핵의 핵막은 두껍고 농염되고, 염색질이 엉켜서 불규칙하고 굽은 과립상이다.	
2. 암종세포 2 : 올챙이 세포핵 - 핵은 항상 올챙이의 머리부분에 위치한다.	
3. 암종세포 3 : 다형성 세포이다. 크고 작은, 둥글고 길쭉한 암세포들이 매우 농염된 핵을 가진다. 또한 세포질의 염색성도 다양하다.	
4. 암종세포 4 : 암세포들이 두 개 이상 모여서 암진주를 형성 한다.	
5. 암종세포 5 : 무핵성 세포질을 가진다.	
6. 암종세포 6 : 무핵성 세포질로 변해가는 도중의 암세포는 핵이 농축되고 한쪽으로 편재된다.	

2.5 핵 영역 추출

세포 객체를 추출하는 과정에서 그림 7과 같이 군집화된 세포 영역에서 핵들이 겹쳐지는 경우가 발생한다. 따라서 본 논문에서는 핵을 정확히 추출하기 위해

서 그림 8과 같은 처리 과정을 수행하여 단일 핵을 추출 한다.

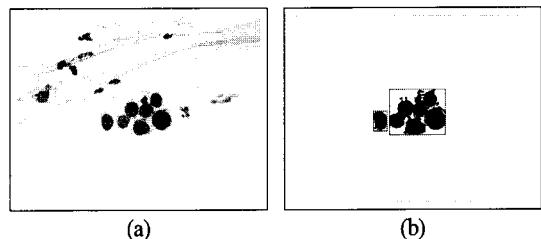


그림 7. 군집화된 세포 영역 (a) 세포 영상
(b) 군집화된 세포
Fig. 7 Clustered cell area (a) Cell image
(b) Clustered cells

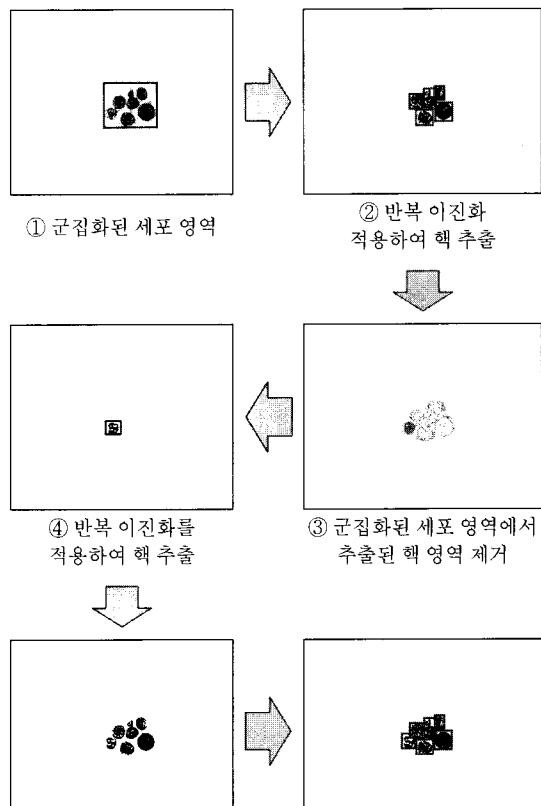


그림 8. 단일 핵 영역 추출 과정
Fig. 8 Process of singular cell extraction

군집화된 영역에서 단일 핵 영역을 추출하는 과정은 다음과 같다. 세포 영상에서 가장 넓은 명암도 분포를 가지는 *R* 채널의 명암도 값을 이용하여 군집화된 영역에서 반복 이진화 적용하여 영상을 이진화 한다. 이진화된 영상에 *Grassfire* 알고리즘을 적용하여 핵 영역을 추출한다. 추출한 세포 핵 영역이 단일 핵의 크기 정보보다 큰 경우에는 군집화된 영역으로 판단하여 그림 8의 과정을 반복 수행하여 단일 세포핵을 추출한다. 그림 8의 ⑥은 군집화된 영역에서 제안된 기법을 통해 추출된 세포 핵 영역이다.

III. 실험 및 결과 분석

본 논문에서 제안한 방법을 Intel Pentium-IV 2.80GHz CPU와 1.00GB RAM이 장착된 IBM 호환 PC상에서 VC++ 6.0으로 구현하여 실험하였다. 실험표본은 640×480 크기의 자궁경부 세포진 400 배율 영상을 대상으로 실험하였다. 제안된 자궁 경부진 핵 추출 화면은 그림 9와 같다.

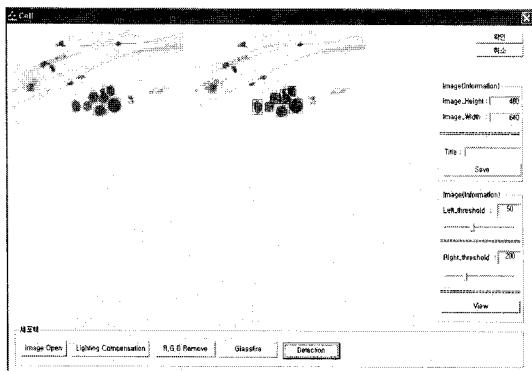


그림 9. 제안된 자궁 경부진 핵 추출 화면
Fig. 9 Interface of nucleus extraction algorithm for uterine cervical pap smears

그림 10은 핵을 추출한 결과 영상이다. 표 2는 3개의 영상에서 암세포의 핵을 추출한 개수를 나타내었다. 총 45개의 암세포에 대해 40개의 핵이 추출되었다.

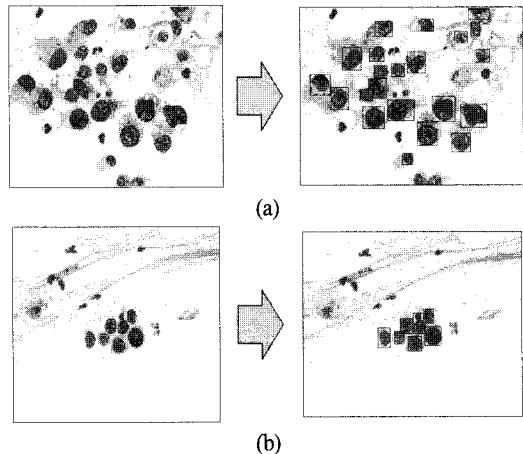


그림 10. 핵 추출 결과 영상 (a) 투명세포선암종
(b) 소세포성편평세포암종1

Fig. 10 Nucleus extraction result image (a) Clear cell adenocarcinoma (b) Small cell carcinoma 1

표 2. 핵 추출 개수
Table 2. Number of nucleus extraction

암종 세포명	추출된 핵의 수	총 암세포의 핵의 수
소세포성편평세포암종1	7	8
소세포성편평세포암종2	13	15
투명세포선암종	20	22

그림 11은 단일 핵을 추출하는 과정에서 일부 세포 영역에서 단일 핵 추출에 실패한 경우를 나타내었다. 그림 11에서와 같이 군집화된 영역에서 세포들이 겹쳐져 있는 정도가 높은 경우에는 단일 핵 추출에 실패하였다.

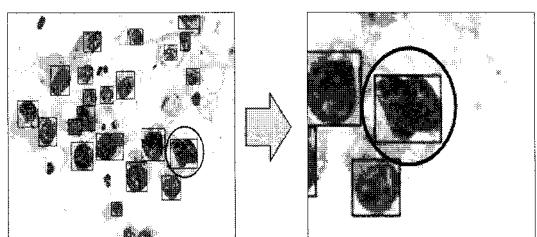


그림 11. 추출에 실패한 부분
Fig. 11 Failed extraction examples

IV. 결론

본 논문에서는 세포 도말 검진에 사용되는 자궁경부 세포진 영상에서 세포들의 영역을 추출하기 위해 Lighting Compensation을 적용하여 세포 영상 획득 시 빛의 영향을 받아 잘못 분류될 수 있는 영역의 오류를 보정하였고, 명암도의 분포가 가장 작은 *B* 채널을 이용한 히스토그램과 3×3 마스크를 이용하여 잡음을 제거하였다. 잡음이 제거된 영상을 이진화하고 Grassfire 알고리즘을 적용하여 세포 객체를 추출하였다. 추출된 세포 객체 중에서 군집화된 세포 영역에 대해서는 *R* 채널의 명암도 값을 반복 이진화에 적용하여 핵 영역을 추출하였다. 그러나 군집화된 영역에서 세포들이 겹쳐져 있는 정도가 높은 경우에는 단일 핵 추출에 실패하였다.

향후 연구 과제는 퍼지 기법을 적용하여 겹쳐져 있는 정도가 높은 세포진 영역을 세밀하게 분류하고 핵을 추출할 수 있는 방법에 대해 연구할 것이다.

참고문헌

- [1] 박관규, 권선영, 제갈문영, 한상미, 장은숙, 차순도, 한인숙, “자궁경부암종에서 암 전이의 조기 발견을 위한 면역조직화학 연구,” 대한병리학회지, Vol.35, pp.391-400, 2001.
- [2] 김미숙, 정상우, “진염색질-이염색질 경계면에서의 mRNA 합성: 경계면 이론의 제안,” 대한병리학회지, Vol.35, pp.93-97, 2001.
- [3] K. B. Kim, S. Kim, K. B. Sim, “Nucleus Classification and Recognition of Uterine Cervical Pap-Smears Using Fuzzy ART Algorithm,” Lecture Notes in Computer Science, LNCS 4247, Springer, pp.560-567, 2006.
- [4] K. B. Kim, D. H. Song, Y. W. Woo, “Nucleus Segmentation and Recognition of Uterine Cervical Pap-Smears,” Lecture Notes in Artificial Intelligence, LNAI 4482, Springer, pp.153-160, 2007.
- [5] R. C. Gonzalez, R. E. Woods, Digital Image Processing, Person Education, Inc., 2008.

- [6] K. B. Kim, H. W. Yun, “A Study on Recognition of Bronchogenic Cancer Cell Image using a New Physiological Fuzzy Neural Networks,” Japanese Journal of Medical Electronics and Biological Engineering, Vol.13, No.5, pp.39-43, 1999.

저자소개



김광백(Kwang-Baek Kim)

1999년 부산대학교
전자계산학과(이학박사)
1997년 ~ 현재 신라대학교
컴퓨터정보공학부 부교수

2005년 ~ 현재 한국해양정보통신학회 학술상임이사
및 영문지 편집위원

2007년 ~ 현재 Scientific Journals International (USA)
Editor

※ 관심분야: 퍼지 논리, 영상 처리, 유전자 알고리즘, 의료정보시스템, 생물정보학