# Kapur 방법과 형태학적 특징을 이용한 자궁경부암 세포 추출 및 인식

김 광 백*

# Detection and Recognition of Uterine Cervical Carcinoma Cells in Pap Smear Using Kapur Method and Morphological Features

Kwang-baek Kim*


## 요 약

자궁 경부 세포진 영상의 효과적인 세포핵 영역 추출과 인식 및 분류를 위해서는 세포진 영상의 배경 그리고 세포핵과 세포질 영역의 정확한 구분이 중요하다. 본 논문에서는 자궁 경부 세포진 영상에서 세포핵 영역과 배경을 효과적으로 분할하기 위해 Median 필터를 적용하여 전체적인 영상의 명암값을 보정한 후, Gaussian 필터를 적용하여 그레이 영상에서 존재하는 잡음을 제거한다. Kapur 방법을 통해 배경과 세포의 엔트로피 누적 확률을 이용하여 영상을 이진화 한다. 자궁 경부진 영상에서는 군집화된 세포 영역이 빈번하게 나타난다. 군집화가 심화된 세포영역에서는 그 영역의 평균 명암도 값을 이용하여 세밀하게 영역을 재분할 한다. 그런 후, 미세잡음을 제거하기 위해 3×3 마스크를 적용하여 미세한 잡음을 제거한 후, 8 방향 윤곽선 추적 알고리즘을 적용하여 분할된 영역에서 세포들의 후보영역을 추출한다. 추출된 세포영역은 크기, 면적 비율, 핵 외곽의 방향성 정보를 이용하여 정상 세포와 암세포를 인식 및 분류한다. 실험 결과에서는 제안된 방법의 성능이 전문의와 소견과 비교적 근접한 것을 보여준다.

## ABSTRACT

It is important to obtain correct cytodiagnosis to classify background, cytoplasm, and nucleus from the diagnostic image. This study propose an algorithm that detects and classifies carcinoma cells of the uterine cervix in Pap smear using features of cervical cancer. It applies Median filter and Gaussian filter to get noise-removed nucleus area and also applies Kapur method in binarization of the resultant image. We apply 8-directional contour tracking algorithm and stretching technique to identify and revise clustered cells that often hinder to obtain correct analysis. The resulted nucleus area has distinguishable features such as cell size, integration rate, and directional coefficient from normal cells so that we can detect and classify carcinoma cells successfully. The experiment results show that the performance of the algorithm is competitive with human expert.



키워드

Cytodiagnosis(세포진단), Uterine Cervical Pap smears(자궁경부 세포진검사), Kapur Method, 8-directional Contour Tracking Algorithm(8 방향 윤곽선 추적 알고리즘)


* 신라대학교 컴퓨터정보공학부

접수일자 : 2007. 5. 21

## Ⅰ. 서 론

세계적으로 발병하는 대표적인 여성암인 자궁경부암은 전체 여성암의 10%를 점유하며 발병률 2위를 차지하고 있다[1]. 우리나라에서도 매년 6000여명의 새로운 환자가 발생하고 있으며 진단된 자궁암의 90~95%는 자궁경부암으로 나타나고 있다. 최근 세포학의 발달과 세포진 검사의 대중화로 자궁경부암의 조기 발견율이 향상되고는 있으나 여전히 우리나라 여성 사망원인의 주요 요인으로 지적되고 있는 실정이다[2].

자궁경부암의 검진법으로는 자궁경부 세포진 검사, 자궁경부 확대촬영검사, 질 확대경 검사, 조직검사 등이 진단 방법으로 시행되고 있다. 이러한 진단방법 중 특히 세포진 검사는 자궁 경부암의 조기진단에 일차적으로 적용하는 꼭 필요한 진단 방법으로서 정밀 검사를 받아보아야 하는 대상을 선별하는데 적합한 진단법이며, 대체로 자궁 경부암 진단에서는 75~80%의 정확도를 가지고 있다. 세포진 검사는 암 부분에서 채취한 세포나 암에서 떨어져 나온 것을 유리판에 도말시켜 현미경으로 관찰하여 진단하는 방법이다. 하지만 높은 암 발생 빈도에 비해 의사나 병리학자가 하루에 검진 할 수 있는 양은 제한되어 있다.

따라서 본 논문에서 제안하는 방법의 목적은 세포진 검사에 사용되는 자궁 경부 세포핵의 특징을 이용하여 세포진 보조 진단 도구로서 활용하여 효과적인 세포 검진과 세포 감별의 위음성률 및 위양성률을 낮추는데 있다. 의사와 병리학자들이 검사에 사용하는 도말된 세포진 영상은 보통 400~1000 배율의 영상이다. 본 논문에서는 400배율의 자궁경부 편평세포 암종 영상을 이용하며, 세포핵 추출과 세포의 형태학적 특징을 이용하여 인식 및 분류한다.

## Ⅱ. 세포진 영역 분할

본 논문에서 제안하는 자궁경부암 세포 영역 추출의 처리 과정은 그림 1과 같다.

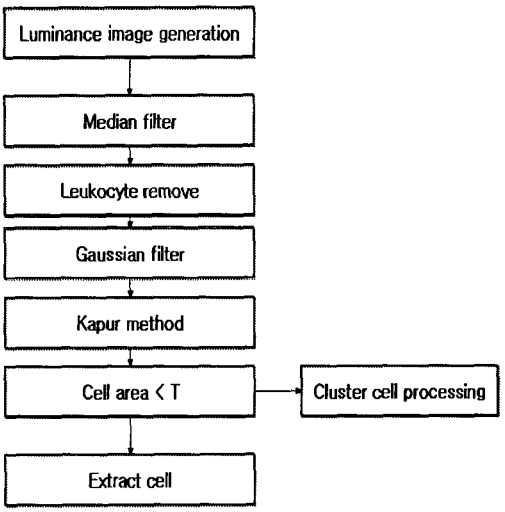



그림 1. 자궁 경부 암세포 추출 과정
Fig. 1. Extracting processes of cervical carcinoma cells from the Pap smear images

자궁 세포진 영상은 크게 배경, 세포질, 세포핵의 3부분의 영역으로 나눌 수 있다. 세포들을 정확히 분류하기 위해서는 전처리 과정에서 세포핵의 특징을 최대한 손상시키지 않아야한다. 따라서 세포진 영상이 컴퓨터로 처리되기 위해 디지털화됨에 따라 손실된 정보를 복원하기 위해 Median 필터를 적용하여 영상을 보정한다. 영상의 각 화소마다 주위 8 방향의 화소를 스캔하여 평균값을 대치시킴으로써 보다 부드러운 영상을 획득한다. 보정된 영상에서 1차적으로 암종 세포진 영상에서 빈번하게 나타나는 백혈구를 제거한다. 백혈구는 세포핵과 비슷한 명암도를 가지고 있지만 세포질을 가지고 있지 않으며, 크기도 세포핵 보다 작기 때문에 쉽게 분별되며 제거할 수 있다. 백혈구는 레이블링을 이용하여 영상에서 모든 객체를 추출한 후, 크기 정보를 이용하여 백혈구를 제거한다. 그림 2는 제안된 방법으로 백혈구를 제거한 결과 영상이다.

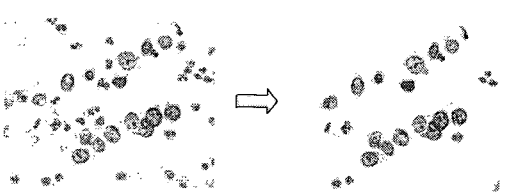


그림 2. 백혈구를 제거한 영상
Fig. 2. Removing leukocytes from the image

백혈구가 제거된 영상에서 배경과 세포영역을 1차적으로 분할하기 위해 Gaussian 필터를 적용한다. 세포핵에 비해 비교적 옅은 명암도를 가지는 배경과 세포질을 제거하기 위하여 그레이 레벨 영상에서 Gaussian 필터를 적용하여 세포질과 배경을 제거한다. 그림 3은 Gaussian 필터를 적용 전과 후의 영상이다.

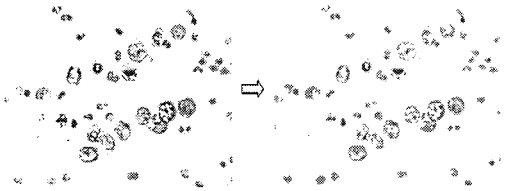


그림 3. 가우시안 필터가 적용된 영상
Fig. 3. Image after Gaussian Filtering

그림 3에서 도말된 세포진 영상에서 옅은 명암도를 가지는 배경과 세포질 부분이 제거된 것을 확인할 수 있다. 잡음이 제거된 영상은 Kapur 방법을 적용하여 영상을 이진화 한 후, 8 방향 윤곽선 추적알고리즘을 이용하여 세포핵을 추출한다.

### 2.1 Kapur method를 이용한 영상 이진화

본 논문에서는 thresholding 기법으로 잡음이 제거된 영상에 Kapur 방법을 적용한다[3]. Kapur 방법은 그레이 레벨 영상에서 히스토그램을 획득 후, 0~255사이의 임의의 명암값 k를 정하여 두 구간으로 나눈 후, 객체와 배경의 엔트로피 누적확률 $C_0$와 $C_1$의 합이 최대가 되는 지점 k를 임계치로 설정하여 영상을 이진화 하는 방법이다. 원하는 객체의 엔트로피 $C_0$을 구하는 식은 (1)과 같으며, 배경의 엔트로피 $C_1$을 구하는 식은 (2)와 같다. 여기서 $p_i$는 $i$번째 명암도 값을 의미하며, $w_k$는 0부터 $k$까지의 누적률이다.

$$Class\ C_0 : \frac{p_0}{w_k}, \frac{p_1}{w_k}, \cdots, \frac{p_k}{w_k} \tag{1}$$

$$Class\ C_1 : \frac{p_{k+1}}{L-w_k}, \frac{p_{k+2}}{L-w_k}, \cdots, \frac{p_{L-1}}{L-w_k} \tag{2}$$

객체의 엔트로피 Class $C_0$의 $w_k$는 $w_k = \sum_{i=0}^{k} p_i$ 와 같이 정의되며, 배경의 Class $C_1$의 $L-w_k$는 $L-w_k = \sum_{i=k+1}^{L-1} p_i$와 같이 정의된다.

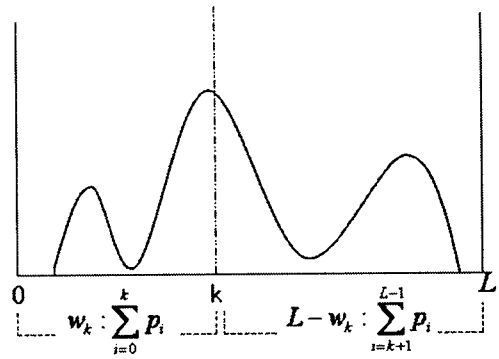



그림 4. Kapur 방법의 히스트그램
Fig. 4. Histogram of Kapur method

그림 4에서 임의의 임계치 k에서 두 구간으로 나눈 다음 각 구간의 누적확률을 계산한 후, 식 (3)과 (4)를 이용하여 두 구간의 누적 엔트로피를 계산한다.

$$H(C_0) = -\sum_{i=0}^{k} \frac{p_i}{w_k} ln \frac{p_i}{w_k} \tag{3}$$

$$H(C_1) = -\sum_{i=k+1}^{L-1} \frac{p_i}{L-w_k} ln \frac{p_i}{L-w_k} \tag{4}$$

두 누적 엔트로피의 합 $\Psi(k)$는 식(5)와 같다.

$$\Psi(k) = H(C_0) + H(C_1) \tag{5}$$

두 구간 $C_0$와 $C_1$의 합이 최대가 되는 지점 k를 임계치로 설정하여 이진화 한다. 그림 5는 Kapur 방법을 적용하여 이진화된 영상이다.

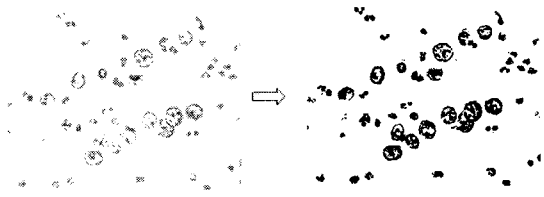


그림 5. Kapur 방법을 적용한 이진화
Fig. 5. Binarization with Kapur method

그림 5는 잡음을 처리한 그레이 레벨 영상에서 남아 있던 약간의 세포질과 배경 부분이 Kapur 방법을 적용한 후에 제거된 것을 보여준다. 하지만 세포진 영상의 특성상 전처리 과정을 거친 영상에서도 미세한 잡음이 존재하며, 이런 미세한 잡음은 프로세스의 전체 처리 속도를 느리게 하는 요인이 될 수 있다. 따라서 본 논문에서는 남아있는 미세한 잡음을 제거하기 위해 3×3 마스크 연산을 이용하여 잡음을 제거한다. 그림 6과 같이 3×3 마스크를 이용하여 중심 화소의 주위 8 방향을 스캔 한 후, 흰 색 픽셀이 검정색 픽셀보다 많이 존재할 경우 중심 화소를 흰색으로 대치함으로써 영상에 존재하는 미세잡음을 제거한다.

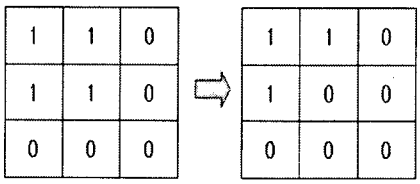



그림 6. 3X3 마스크 연산
Fig. 6. 3x3 mask operation

### 2.2 군집화된 세포 영역 분할

미세한 잡음을 제거 후, 8 방향 윤곽선 추적 알고리즘을 적용하여 이진화 된 영상에서 세포핵 영역을 추출한다. 암종 세포진 영상에서는 암세포 전이에 의해 군집화된 세포가 빈번하게 나타난다. 군집된 세포 영역이 객체로 추출되는 과정에서 여러 개의 세포가 하나의 단일 세포로 추출 및 인식되므로 정확히 암종 세포를 진단할 수 없다. 따라서 본 논문에서는 그림 7과 같이 군집화된 세포 영역에 대해서 그림 8과 같은 처리 과정을 거친다.

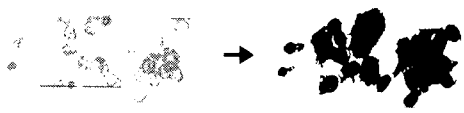


그림 7. 군집화된 세포 영상
Fig. 7. Clustered cell image

군집화 된 세포영역의 전체적인 처리 과정은 그림 8과 같다.

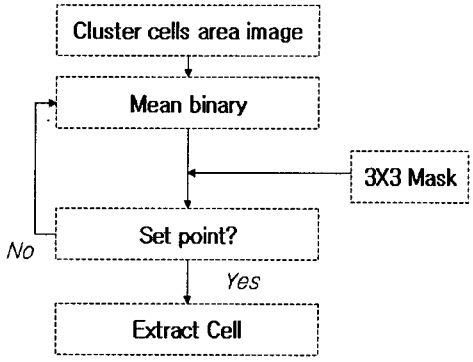



그림 8. 군집화된 세포 영역 처리 과정
Fig. 8. Manipulation processes of clustered cell

군집화 된 세포 영역에서 평균 이진화를 적용하여 영상을 이진화 한 후, 3×3 마스크를 이용하여 미세한 잡음을 제거한다. 그런 후, 잡음이 제거 된 영상에서 8 방향 윤곽선 추적 알고리즘을 적용하여 단일 세포로 추출한다. 여기서 오추출을 방지하기 위해 영상에서 상위 10%에 해당하는 명암도를 가진 화소는 핵의 영역으로 가정한 후, 이를 정점으로 두고 추출된 영역 속에 정점이 속해 있는지를 검사하여 최종적으로 핵 영역을 추출한다. 그림 9는 군집화된 세포 영역에서 제안된 방법을 적용하여 단일 세포로 분할한 결과이다.

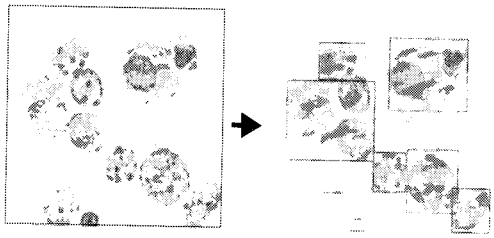
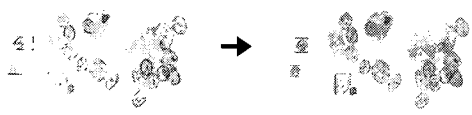


그림 9. Cluster cells 처리 결과
Fig. 9. Manipulating clustered Cells

다양한 명암도 분포와 복잡성을 가지고 있는 세포진 영상은 세포핵 영역만을 정확히 분할하기 힘들며, 분할하는 과정에서 영상에 존재하는 세포핵의 특징이 손실 될 가능성이 있다. 그렇게 되면 정확한 세포의 인식과 분류가 어려워진다.

따라서 본 논문에서는 영상에서 전역적으로 명암값이 균등하도록 스트레칭 기법을 적용하여 각 세포핵과 세포질의 경계 부분을 향상시켜 세포핵의 윤곽을 뚜렷하게 보정한다. 보정된 영상에서 세포핵을 추출한 후, 세포의 특징 정보를 이용하여 인식과 분류를 수행한다.



(a) 보정전 세포핵 (b) 보정후 세포핵
그림 10. 보정된 세포핵
Fig. 10. Revised nucleus

그림 10에서 영역 보정 전의 영상(a)보다 영역 보정 후의 영상(b)가 세포핵 영역이 비교적 정확하게 추출되는 것을 확인할 수 있다.

## Ⅲ. 세포의 형태학적 특징을 이용한 분류 및 인식

일반적으로 암종 세포의 핵은 정상 세포의 핵에 비하여 크고, 다양한 형태를 보이며, 뚜렷한 핵소체를 가진다. 또한 암종 세포핵은 이염색질(heterochromatin)의 증가 때문에 거친 염색질을 보인다. 실제로 암종 세포핵은 이와 같은 특징 때문에 자궁 경부 암종 등의 세포학적 진단이 가능하다[4].

본 논문에서는 이러한 세포핵의 형태학적 특징을 이용하여 핵 외곽의 방향성 정보와 면적 비율, 그리고 핵의 크기 정보를 이용하여 세포를 인식하고 분류한다. 그림 11은 본 논문에서 제안하는 세포의 인식 및 분류 과정이다.

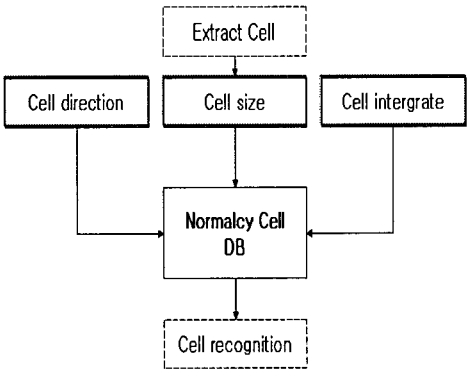
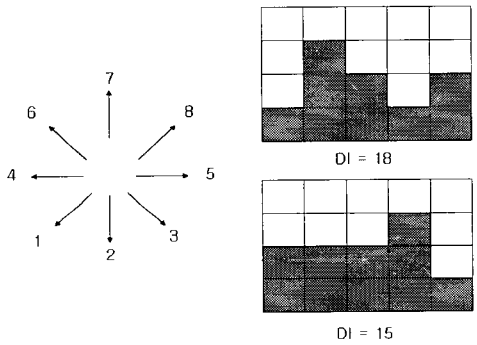



그림 11. 세포의 분류 및 인식 과정
Fig. 11. Processes of classification and recognition

### 3.1 세포핵 외곽의 방향성 정보

본 논문에서는 1차적으로 8 방향 윤곽선 추적 알고리즘의 진행 방향에 따른 가중치를 적용하여, 이전 방향의 가중치와 다음 방향의 가중치의 차를 계산하여 계산된 값을 모두 더한 수치를 세포핵 인식을 하기 위한 첫 번째 특징 정보로 적용한다. 8 방향의 각각의 가중치는 그림 12와 같다. 그림 12의 (a)는 각 방향에 따른 가중치이고, (b)는 각 영상에서 외곽선을 추적하면서 가중치를 모두 합한 수치인 DI의 한 부분이다. 정상 세포보다 세포핵의 윤곽이 거친 암종 세포의 경우 DI의 수치가 상대적으로 높게 나타나며, 이런 특징을 이용하여 1차적인 세포 분류를 위한 특징 정보로 사용한다.

암종 세포핵의 윤곽은 비교적 정상 세포보다 거친 면을 보이므로 정상과 암종 세포를 분류할 때 방향성 정보는 중요한 특징이다[5].




(a) (b)
그림 12. (a) 방향 (b) 윤곽선에 대한 가중치 합
Fig. 12. (a) Directions (b)Computing sum of weights through contour

### 3.2 세포핵 면적 비율

정상과 암종 세포를 분류하기 위한 두 번째 특징은 핵의 면적 비율이다. 암종 세포핵은 핵의 파괴와 염색질이 엉켜서 불규칙한 형태를 가짐으로써 핵 내에 비어있는 부분이 많다[6]. 본 논문에서는 이러한 암종 세포의 특징을 면적 비율이라 정의하며, 인식을 위한 특징으로 적용한다.

### 3.3 세포핵 크기 정보

세 번째 특징으로는 핵의 크기 정보를 이용하여 세포의 특징으로 적용한다. 암종 세포의 경우 정상 세포의 경

우보다 크기가 크며, 크기 정보 역시 인식을 위한 중요한 특징으로 이용한다. 추출된 세포의 인식 및 분류를 위해서 정상 세포핵의 형태학적 특징 정보와 암종 세포의 특징정보를 비교하여 각 특징별로 가중치를 적용시켜 정상 세포와 암종 세포를 인식하고 분류한다.

그림 13은 추출된 정상 세포의 특징정보이다. 정상 세포의 경우 방향성 정보의 수치는 평균 210.2로 나타났으며, 평균 크기 수치는 898.6, 면적 비율은 93.47%로 나타났다.

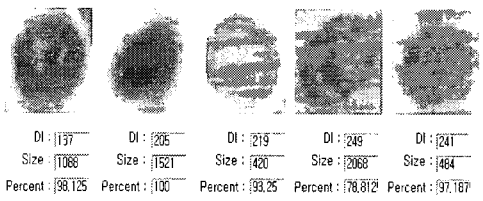



그림 13. 정상 세포의 형태학적 특징
Fig. 13. Morphological features of normal cells

## Ⅳ. 실험 및 결과 분석

실험 환경은 IBM 호환 기종의 Pentium-IV PC상에서 Visual C++ 6.0으로 구현하였다. 실험 표본은 640×480 크기의 자궁경부 편평 세포진 400 배율 영상 10장을 대상으로 실험하였다. 자궁 경부 세포진 영상에서 추출된 세포의 총 개수는 171개였으며, 표 1은 각 암종 영상에 대한 암세포 인식 결과를 나타내었다.

표 1. 암세포 인식 개수
Table 1. Recognition result of cancer cells

| 암종 세포명 | 인식된 암세포 개수 | 총 암세포 개수 |
|---|---|---|
| 각화성 편평세포 암종1 | 15 | 16 |
| 각화성 편평세포 암종2 | 17 | 19 |
| 경관 상피내 선암종 | 13 | 13 |
| 대세포 편평세포 암종1 | 20 | 23 |
| 대세포 편평세포 암종2 | 6 | 6 |
| 소세포 편평세포 암종1 | 16 | 16 |
| 소세포 편평세포 암종2 | 8 | 10 |
| 투명 세포 선암종 | 21 | 24 |
| 미소침윤성 편평세포 암종1 | 16 | 21 |
| 미소침윤성 편평세포 암종2 | 17 | 23 |

표 2는 각 영상에 대한 대표적인 암종 세포의 특징 파라미터를 나타내었다. 표 2에서 DI는 방향성 정보를 의미하며, Size는 세포핵의 크기, Percent는 면적 비율이다. 표 2에서 암종 세포들은 대체적으로 정상 세포의 파라미터 평균치에 비해 파라미터의 수치가 높은 것을 알 수 있다.

표 2. 암종 세포의 형태학적 특징
Table 2. Morphological features of carcinoma cells

| 암종 세포명 | 특징 정보 | |
|---|---|---|
| 각화성 편평세포 암종1 | DI | 227 |
| | Size | 4212 |
| | Percent | 85.54% |
| 각화성 편평세포 암종2 | DI | 221 |
| | Size | 5767 |
| | Percent | 92.15% |
| 경관 상피내 선암종 | DI | 243 |
| | Size | 2107 |
| | Percent | 79.25% |
| 대세포 편평세포 암종1 | DI | 257 |
| | Size | 3009 |
| | Percent | 71.42% |
| 대세포 편평세포 암종2 | DI | 543 |
| | Size | 3968 |
| | Percent | 62.40% |
| 소세포 편평세포 암종1 | DI | 243 |
| | Size | 2205 |
| | Percent | 77.56% |
| 소세포 편평세포 암종2 | DI | 315 |
| | Size | 3150 |
| | Percent | 82.76% |
| 투명 세포 선암종 | DI | 175 |
| | Size | 5658 |
| | Percent | 75.56% |
| 미소침윤성 편평세포 암종1 | DI | 251 |
| | Size | 1404 |
| | Percent | 77.62% |
| 미소침윤성 편평세포 암종2 | DI | 283 |
| | Size | 4290 |
| | Percent | 85.76% |

## V. 결 론

본 논문에서는 세포 도말 검진에 사용되는 자궁경부 세포진 영상에서 세포들의 후보 영역을 추출하기 위해 그레이 레벨 영상에서 Filter 기법을 적용하여 잡음을 제거하였고 Kapur 방법을 적용하여 최종적으로 세포 영역을 분할하였다. 분할된 세포 영역에서 8 방향 윤곽선 추적 알고리즘을 적용하여 세포핵을 추출하였으며, 핵의 형태학적 특징 정보를 이용하여 암세포를 분류하고 인식하였다. 암세포를 분류하고 판별에 적용된 특징들은 100개 정도이다. 하지만 그 중 암 세포의 특징들을 선택하기 어렵고 암세포의 특징들을 모두 추출하기 위해서는 오랜 시간이 걸린다[7]. 따라서 본 논문에서는 세포핵의 여러 특징 정보 중 3가지의 특징 정보를 추출하여 암종 세포의 판별에 적용하였다. 본 논문에서는 10개의 세포진 영상에서 171개의 세포가 추출되었다. 추출된 171개의 세포 중에서 149개의 세포가 인식되었고, 암세포의 인식률은 87.13%로 나타났다.

향후 연구 과제로는 세포진 영상에서 배경과 세포 영역을 분할하는 전처리 과정에서 세포핵의 정보가 손실되는 부분을 개선할 것이다. 또한 제안된 방법에서는 세포핵의 여러 특징 정보들 중에서 3가지의 특징 정보만을 적용하였지만, 향후에는 더 많은 특징 정보를 적용하여 세포를 분류하고 암세포의 인식률을 개선할 것이다.

## 참고문헌

[1] 이선경, 유희석, 류기성, "HPV 예방백신 시대의 자궁경부암 선별검사," 대한부인종양 콜포스코피학회지, pp.71-76, 2005.

[2] 박관규, 권선영, 제갈문영, 한상미, 장은순, 차순도, 한인숙, "자궁경부암종에서 암 전이의 조기 발견을 위한 면역조직화학 연구," 대한병리학회지, 35권, 5호, pp.391-400, 2001.

[3] J. N. Kapur, P. K. Sahoo, A. K. C. Wong, "A New method for Gray Level Picture Thresholding Using the Entropy of the Histogram," Computer Vision, Graphics, and Image Processing, Vol.29, pp.273-285, 1985.

[4] 김미숙, 정상우, "진염색질-이염색질 경계면에서의 mRNA 합성: 경계면 이론의 제안," 대한병리학회지, 35권, 2호, pp.93-97, 2001.

[5] 김광백, 윤홍원, "퍼지 신경망을 이용한 기관지 편평암 인식에 관한 연구," 대한의료정보학회지, 5권, 1호, pp.77-87, 1999.

[6] 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.

[7] 강미영, 하진영, 김백섭, 이헌길, "자궁경부세포 인식을 위한 가변 블록 기반 특징 추출 기법," 정보통신논문지, 제5집, pp.66-71, 2001.

### 저자소개

**김 광 백(Kwang-Baek Kim)**

1999년 부산대학교 전자계산학과 (이학박사)
1997년~현재 신라대학교 컴퓨터정보공학부 부교수
2005년~현재 : 한국해양정보통신학회 이사 및 논문지 편집 부위원장
2005년~현재 : 한국멀티미디어학회 이사 및 논문지 편집위원
※관심분야: Neural Networks, Image Processing, Fuzzy Logic, Medical Imaging and Biomedical System, Support Vector Machines