

자궁경부암 진단을 위한 3차원 세포핵 질감 특성값 유의성 평가에 관한 연구

최현주*, 김태윤**, Patrik Malm***, Ewert Bengtsson***, 최흥국*,****

Study on evaluating the significance of 3D nuclear texture features for diagnosis of cervical cancer

Hyun-Ju Choi*, Tae-Yun Kim**, Patrik Malm***, Ewert Bengtsson***, Heung-Kook Choi*,****

요약

본 연구의 목적은 세포핵의 3차원 염색질 질감 특성값이 암의 진행정도를 인식하는데 있어 유용한 특성값인지 평가하는데 있다. 특히, 제안한 방법이 악성이라고 진단된 세포진 도말 표본에서 정상으로 보이는 세포의 염색질 패턴에서의 미세한 차이를 인식할 수 있는지 살펴보고자 한다. 분류등급 정상(Normal), 저등급 편평 상피내 병변(LSIL, Low grade Squamous Intraepithelial Lesion), 고등급 편평 상피내 병변(HSIL, High grade Squamous Intraepithelial Lesion)에서 각각 100개씩의 세포 볼륨데이터로부터 3차원 GLCM(Gray Level Co occurrence Matrix)에 기반한 질감 특성값과 3차원 Wavelet 변환에 기반한 질감 특성값을 추출하고 분류기를 생성한 후 각 분류기에 대한 분류정확도를 비교하였으며, 2차원 세포진 영상에서의 세포핵 질감 특성값과 비교하기 위해 동일한 실험 볼륨데이터의 투영된 2차원 영상을 이용하여 같은 방법으로 2차원 세포핵 질감 특성값을 추출하고 분류기를 생성한 후 분류정확도를 비교하였다. 2차원 세포핵 질감 특성값과의 비교연구에서 3차원 세포핵 질감 특성값이 등급별 분류에 있어 보다 효율적인 것을 확인 할 수 있었으며 이는 3차원 염색질 질감 특성값이 자궁경부 세포의 정량화에 대한 정확성과 재현성을 개선할 수 있음을 의미한다.

▶ Keyword : 자궁경부암, 자궁경부 세포진 영상, 3차원 염색질 질감 특성값, 3차원 GLCM, 3차원 웨이블릿 변환

Abstract

The aim of this study is to evaluate whether 3D nuclear chromatin texture features are

• 제1저자 : 최현주 • 교신저자 : 최흥국

• 투고일 : 2011. 04. 28, 심사일 : 2011. 06. 09, 게재확정일 : 2011. 08. 04.

* 인제대학교 컴퓨터공학부(School of Computer Engineering, Inje University)

**** 인제대학교 유비쿼터스 헬스케어 연구 센터(Ubiquitous Healthcare Research Center, Inje University)

※ 이 논문은 2008년 정부(교육과학기술부)의 재원으로 학술진흥재단의 지원을 받아 수행된 연구임[KRF-2008-357-D00249]

significant in recognizing the progression of cervical cancer. In particular, we assessed that our method could detect subtle differences in the chromatin pattern of seemingly normal cells on specimens with malignancy. We extracted nuclear texture features based on 3D GLCM(Gray Level Co occurrence Matrix) and 3D Wavelet transform from 100 cell volume data for each group (Normal, LSIL and HSIL). To evaluate the feasibility of 3D chromatin texture analysis, we compared the correct classification rate for each of the classifiers using them. In addition to this, we compared the correct classification rates for the classifiers using the proposed 3D nuclear texture features and the 2D nuclear texture features which were extracted in the same way. The results showed that the classifier using the 3D nuclear texture features provided better results. This means our method could improve the accuracy and reproducibility of quantification of cervical cell.

▶ Keyword : Cervical Cancer, Pap Smear Images, 3D Chromatin Texture Feature, 3D GLCM, 3D Wavelet Transform

1. 서 론

자궁경부암은 세계적으로 유방암에 이어 여성에게 발생하는 악성종양 중 두 번째로 흔한 암이며 저개발국으로 갈수록 유병률과 사망률이 높은 실정이다[1].

자궁경부암 검진 방법으로는 자궁경부 세포진 검사(PAP smear), 질 확대경 검사(Colposcopy), 조직 검사(Biopsy), 초음파, 자기공명영상을 이용한 영상 검사, 인유두종바이러스 검사(HPV test) 등이 있으며, 이러한 검진 방법 중에서 1943년 George Papanicolaou에 의하여 처음으로 제시된 자궁경부 세포진 검사는 자궁경부암 조기진단을 위한 효과적인 방법으로 인정받아 왔으며 편리성, 저비용 등의 장점으로 가장 널리 사용되고 있다[2]. 하지만 자궁경부 세포진 검사의 높은 위음성률과 임상과의 혼란도에 따른 신뢰도의 차이가 문제로 제기되고 있다.

이러한 문제들을 해결하기 위해 육안에 의한 진단을 보조할 수 있는 컴퓨터를 이용한 자궁경부 세포진 영상 분석 방법이 연구되어지고 있으며 주로 세포진 영상에서의 세포핵 분할과 특징 추출, 인식을 위한 알고리즘 개발에 관한 연구가 진행되고 있다. 은 등은 자궁경부 세포진 영상에서 암세포핵 인식 방법으로 에지 정보를 이용하여 세포핵을 분할한 후 세포핵의 형태학적 특징인 핵의 크기, 핵과 세포질 비율, 원형성을 추출하고 이를 바탕으로 정상 세포핵과 비정상 세포핵을 인식하였으며[3], 김 등은 자궁경부 세포진 영상에서 퍼지 그레이 모폴로지 연산을 이용하여 핵을 분할하고 형태학적 정보와 명암 정보, 색상 정보 및 질감 정보를 분석하여 핵의 특징

을 추출하고 개선된 퍼지 ART 알고리즘을 적용하여 분류하였다[4]. Plissiti 등은 특히 자궁경부암 세포핵 분류에 있어 FCM(Fuzzy C-means) 방법과 SVM(Support Vector Machine) 방법을 사용하여 성능을 비교하였으며[5], 이미 AutoPap(Neopath Inc.)과 같은 자동화된 자궁경부암 진단 시스템이 상용화되고 보조 진단 도구로서 실제 임상에서 사용되어지고 있다.

그러나 대부분의 연구가 2차원 세포진 영상에서의 세포핵 분할과 분할된 세포핵의 형태학적 특징을 기반으로 암의 유무 및 진행 정도를 판단하는데 중점을 두고 있어 본질적으로 3차원 구조를 가지는 세포핵의 정확한 특성을 정량화하기에는 한계를 가지고 있다. 또한 병리진단학적인 관점에서 암의 진행에 따른 세포핵 내부의 변화는 형태학적인 변화와 함께 진단 및 예후 추정을 위한 중요한 요소로 평가받고 있음에도 불구하고 세포핵 내부의 3차원 염색질(Chromatin) 변화를 분석하기 위한 방법에 대한 연구는 미비한 실정이다. 그리고 자동화된 자궁경부암 진단 시스템 역시 자궁경부 세포진 슬라이드의 초기 선별 과정에만 사용되고 있어 병리의사의 2차적 선별 및 진단 과정이 반드시 필요하며 특히 고가의 장비로 자궁경부암의 유병률과 사망률이 높은 저개발국에서 사용하기에는 많은 어려움이 있다.

본 연구팀은 현재 스웨덴 욥살라대학교(Uppsala University, Sweden), 인도 암센터(Regional Cancer Center, India)와 공동으로 저개발국에서도 낮은 비용으로 쉽게 사용할 수 있고 높은 처리율과 정확성을 가진 자궁경부암 진단 시스템 개발을 위해 연구하고 있다[6,7]. 이 프로젝트에서 본 연구팀은 자궁경부암 진단을 위한 세포핵 내부의 3차원 염색질 변화를 정량화 할 수 있는 분석 기술 개발에 관한 연구를 수행하

고 있으며 본 논문에서는 자궁경부 세포핵의 3차원 염색질 질감 특성값이 암의 진행 정도를 인식하는데 있어 유용한 특성값 인지를 평가하고자 한다. 특히, 악성이라고 진단된 세포진 도말 표본에서 정상으로 보이는 세포가 주변의 악성 세포의 영향으로 정상 세포의 형태를 가지고 있으나 세포핵 내부의 염색질 분포에는 정상 세포와 미세한 차이가 있다는 연구들을 바탕으로[8-10] 형태는 정상 세포처럼 보이지만 세포핵 내부의 염색질 분포나 패턴에서는 차이를 보이는지를 살펴보고자 한다.

II. 재료 및 방법

1. 재료 및 영상 획득

본 연구에서 사용한 실험영상은 인도 암센터에서 제공한 자궁경부 세포진 도말 표본 중 편평상피세포(Squamous cell) 영상으로, 병리전문의가 베데스다 시스템(Bethesda system)을 바탕으로 정상(Normal), 저등급 편평 상피내 병변(LSIL, Low grade Squamous Intraepithelial Lesion), 고등급 편평 상피내 병변(HSIL, High grade Squamous Intraepithelial Lesion)으로 분류한 슬라이드에서 형태학적으로는 정상 세포(NILM, Negative for Intraepithelial Lesion and Malignancy)로 보인다고 판단한 세포 영상이다. 즉, 서론에서 언급한 대로 악성이라고 진단된 세포진 도말 표본에서 형태는 정상 세포처럼 보이지만 주변의 악성 세포의 영향으로 세포핵 내부의 염색질 분포나 패턴에서 차이를 보이는지를 살펴보기 위해, Normal, LSIL, HSIL로 진단된 슬라이드에서 0.4 μ m 간격으로 연속적인 2차원 광학 절편 영상 40장으로 구성된 1344x1024 크기의 볼륨데이터를 40배율에서 현미경에 부착된 CCD 카메라를 통해 8비트 그레이 영상으로 디지털화하였다. 3개의 분류등급별로 얻은 볼륨데이터에서 각각 100개씩 정상으로 보이는 세포를 선택하고 수동으로 잘라내어 총 300개의 세포 볼륨데이터를 실험영상으로 사용하였다. 그림 1은 3개의 분류등급 Normal, LSIL, HSIL 각각에 대해 사용한 볼륨데이터의 한 예이다.

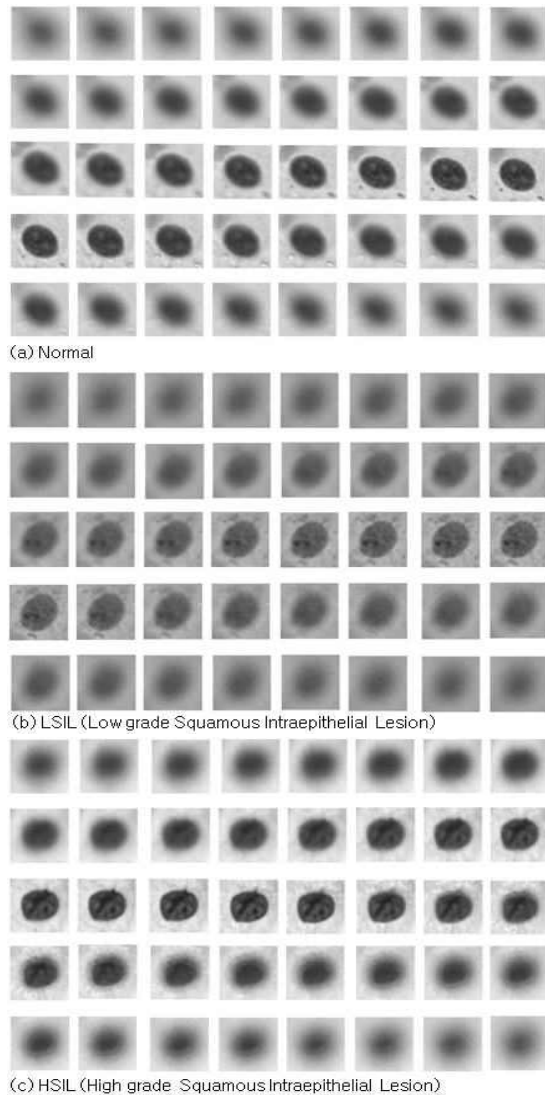


그림 1. 각 분류등급별 실험영상의 예
Fig. 1. Examples of representative images for each group

2. 세포핵 분할

공초점 현미경을 이용하여 획득한 3차원 볼륨데이터와는 달리 일반 광학현미경을 이용하여 얻은 3차원 볼륨데이터는 2차원 광학 절편 영상의 볼륨데이터 내에서의 위치가 z축의 가운데에서 멀어질수록 초점이 흐려지고 잡음이 많은 문제점이 있다. 그러므로 관심영역을 분할하기 위해서는 볼륨데이터를 구성하고 있는 각각의 2차원 광학 절편 영상에 맞는 서로 다른 임계치나 적절한 분할 방법을 찾아야하며 찾은 방법에

대해서도 볼륨데이터를 구성하고 있는 2차원 광학 절편 영상의 수만큼 반복 수행해야하는 단점이 있다.

이러한 문제점을 개선하기 위하여 세포핵을 분할하기 위한 전처리 단계로, 3차원 볼륨데이터에 최소 밝기값 투영(MIP, Minimum Intensity Projection) 방법을 적용하여 신호 대 잡음비(SNR, Signal to Noise ratio)가 높은 투영된 2차원 영상을 생성하고 그 결과 영상에 캐니 에지 추출 방법(Canny Edge Detection Method)을 사용하여 세포핵의 윤곽선을 추출하였다. 다음으로 세포핵의 윤곽선이 얻어진 이진 영상에서 침식과 팽창, 그리고 구멍을 메우는 형태학적 연산을 수행하여 세포핵 영역이 분할된 이진 영상을 얻었다.

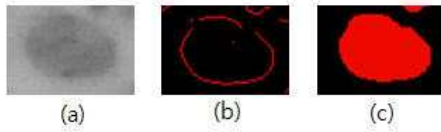


그림 2. 세포핵 분할 과정 (a) 최소 밝기값 투영 영상, (b) 캐니 에지 결과, (c) 형태학적 연산 수행 결과
 Fig. 2. Process for segmenting a cell nucleus (a) minimum intensity projection image, (b) Canny edge result, (c) final segmentation result achieved through filling of closed areas and opening

마지막으로 세포핵이 분할된 이진 영상을 바탕으로 원영상과 같은 개수를 가진 3차원 이진 볼륨데이터를 만들고 원래의 볼륨데이터와 AND 연산을 수행하여 세포핵 내부의 밝기값은 그대로 유지하면서 배경부분의 밝기값은 0인 세포핵 영역이 분할 된 새로운 볼륨데이터를 생성하였으며 이 결과 영상을 세포핵의 3차원 질감 특성값을 계산하기 위한 입력 영상으로 사용하였다.

3. 3차원 질감 특성값 정의 및 추출

3.1 3차원 GLCM(Gray Level Cooccurrence Matrix)에 기반한 질감 특성값 추출

영상에서의 질감 분석이란 밝기값 분포의 공간적 상호의존성에 따른 부드러움과 거침 정도를 분석하여 객관적 수치로 정량화하는 것이다. 질감 분석 방법은 기본적으로 통계적 방법, 구조적 방법이 있는데 세포 영상에 대한 질감 분석은 영상의 불규칙성과 다양성 때문에 규칙적으로 반복되는 일정한 패턴의 위치 규칙을 정의함으로써 분석하는 구조적 방법보다는 통계적 방법을 사용하는 것이 더 적절하다.

GLCM(Gray Level Co occurrence Matrix)은 통계적 질감 분석을 위한 대표적 방법으로 두 픽셀간의 밝기값 변

화 패턴이 얼마나 빈번히 나타나는가를 결합 확률 밀도(Joint probability density) 함수 $f(i, j : d, \theta)$ 로 표현한 것이다. 여기서 i, j 는 인접한 두 픽셀을 의미하여 d 는 두 픽셀간의 거리, θ 는 두 픽셀이 이루는 각을 나타내며 d 와 θ 값을 변화시킴으로써 여러 가지 GLCM를 생성할 수 있다. 2차원 영상 분석을 위한 GLCM이 하나의 평면상의 픽셀들의 밝기값 분포에 대한 공간적 의존성을 고려하는 반면에, 3차원 영상 분석을 위한 GLCM은 볼륨데이터를 구성하는 기본 단위인 복셀 밝기값 분포의 3차원 공간적 의존성을 고려해야한다. 한 복셀을 중심으로 인접한 복셀에 대해 26개 변위에 대한 GLCM이 생성되어야 하나 중복성으로 인해 13개 변위 벡터에 의한 GLCM을 생성하고 2차원 GLCM을 3차원 GLCM으로 확장하였다. 표 1은 볼륨데이터의 GLCM을 계산하기 위한 13개 변위 벡터와 그에 대응하는 중복 벡터를 나타내고 있으며 D는 거리를 의미한다.

표 1. 3D GLCM의 변위 벡터와 중복벡터
 Table 1. Displacement vector and duplicate displacement vector of 3D GLCM

Direction(ϕ, θ)	Displacement Vector	Duplicate Displacement Vector
(0°, 45°)	(D, 0, D)	(-D, 0, -D)
(0°, 90°)	(D, 0, 0)	(-D, 0, 0)
(0°, 135°)	(D, 0, -D)	(-D, 0, D)
(45°, 45°)	(D, D, D)	(-D, -D, -D)
(45°, 90°)	(D, D, 0)	(-D, -D, 0)
(45°, 135°)	(D, D, -D)	(-D, -D, D)
(90°, 45°)	(0, D, D)	(0, -D, -D)
(90°, 90°)	(0, D, 0)	(0, -D, 0)
(90°, 135°)	(0, D, -D)	(-D, 0, -D)
(135°, 45°)	(-D, D, D)	(0, -D, D)
(135°, 90°)	(-D, D, 0)	(D, -D, -D)
(135°, 135°)	(-D, D, -D)	(D, -D, 0)
(-, 0°)	(0, 0, D)	(0, 0, -D)

또한 256 그레이 레벨에 대한 모든 두 복셀 간의 밝기값 분포에 대한 GLCM을 생성할 경우 GLCM의 크기가 크고 많은 계산량이 요구되는 단점이 있으므로 원영상의 256 그레이

레벨을 64 그레이 레벨로 양자화하고 두 복셀 간의 거리 1에 대해 64x64 크기의 GLCM을 생성하였다. 생성된 13개의 GLCM으로부터 Haralik에 의해 정의된 질감 특성값 Energy, Entropy, Contrast, Homogeneity를 계산하였으며 그 식은 다음과 같다[11].

$$Energy = \sum_{i=1}^N \sum_{j=1}^N PM^2 \quad (1)$$

$$Entropy = - \sum_{i=1}^N \sum_{j=1}^N (PM) \log(PM) \quad (2)$$

$$Contrast = \sum_{i=1}^N \sum_{j=1}^N (i-j)^2 PM \quad (3)$$

$$Homogeneity = \sum_{i=1}^N \sum_{j=1}^N (1+(i-j)^2)^{-1} PM \quad (4)$$

3.2 3차원 Wavelet 변환에 기반한 질감 특성값 추출

Wavelet 변환 방법은 공간 영역과 주파수 영역에서 그 지역적 특징을 잘 나타내고 다해상도 분석이 가능하므로, 서로 다른 해상도로 영상을 분할하여 개별적으로 부영상(Sub band)을 분석할 수 있다는 특징으로 인해 질감 정보를 얻고자 할 때 많이 쓰이고 있다. 이 특성은 Wavelet이라고 불리는 기저함수를 이동하고 확장함으로써 저대역 부밴드와 고대역 부밴드로 주파수 영역에 따른 다해상도를 갖게 된다. Wavelet 기저함수는 아래 식과 같으며 여기서 a는 척도 변수(Scale Variable)이고 b는 이동 변수(Translation Variable)이다[12].

$$CWT_x^a(a,b) = \frac{1}{\sqrt{a}} \int x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (5)$$

2차원 영상의 Wavelet 변환은 필터 뱅크 구조를 구성하여 x축 방향과 y축 방향으로 1차원 Wavelet 변환을 반복적으로 수행해서 얻어진다. 그러므로 3차원 볼륨데이터에 대한 Wavelet 변환을 적용하기 위해 1차원 Wavelet 변환을 x축, y축, z축으로 반복 수행하였다. 사용한 Wavelet 기저함수는 Haar와 Daubechies 2이며 방향별로 고주파 대역 통과 필터와 저주파 대역 통과 필터를 적용하고 주파수 성분에 따라 Wavelet 계수로 이루어진 8개의 3차원 부영상(LLL, LLH, LHL, HLL, LHH, HLH, HHL, HHH)을 얻었다. 그림 3은 3차원 Wavelet 변환의 볼륨데이터 분할과정이다.

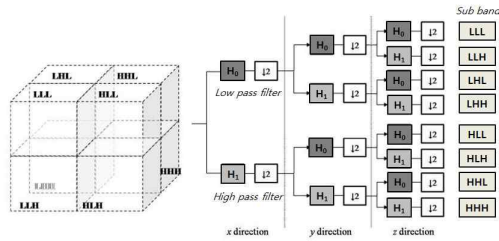


그림 3. 3차원 wavelet 변환의 볼륨데이터 분할과정
Fig. 3. Decomposition of a volume data by 3D wavelet transform

Wavelet 변환 후 얻어진 부영상에서 바로 질감 특성값을 계산하는 기존의 방법과는 달리 부영상의 각 Wavelet 계수값을 바탕으로 CM(Co-occurrence Matrix)을 생성하고 WCCM(Wavelet Coefficient Co-occurrence Matrix)이라고 정의하였다. 생성방법은 3차원 GLCM과 같으며 각 부영상에 대한 Wavelet 계수값의 최대값과 최소값을 기준으로 64개의 계수값을 가지도록 양자화 한 후 64x64 크기의 WCCM를 13개 변위 벡터에 대해 생성하였다. 생성된 WCCM으로부터 GLCM과 동일한 질감 특성값 Energy, Entropy, Contrast, Homogeneity를 추출하였다.

4. 추출한 특성값에 대한 유의성 검증

추출한 3차원 세포핵 질감 특성값 중에서 암의 진행 정도에 따른 각 그룹간의 차이를 의미있게 설명해 줄 수 있는 특성값들을 찾고 독립변수의 차원을 줄이면서도 분류정확도를 높이기 위해 순차적 선택 방법(Stepwise Selection Method)을 사용하였다. 순차적 선택 방법은 전진 선택 방법(Forward Selection Method)과 후진 선택 방법(Backward Selection Method)의 절충적인 형태로, 전진 선택 방법에 의해 추출한 특성값들 중에서 분류에 가장 큰 기여를 하는 특성값을 택하는 동시에 각 단계에서 후진 선택 방법을 같이 사용하여 유의적으로 기여하지 못하는 특성값을 제거하는 방법이다[13].

표 2, 3, 4는 유의수준 0.05%로 한 순차적 선택 방법의 결과로, 3차원 GLCM에 기반하여 추출한 3차원 세포핵 질감 특성값 52개(13개의 GLCM x 4개의 질감 특성값) 중 7개의 특성값이 각 그룹으로 분류하는데 있어 유의적인 영향을 주는 것으로 나타났으며 Haar 기저 함수를 사용한 3차원 Wavelet 변환에 기반한 질감 특성값 416개(13개의 WCCM x 8개의 부영상 x 4개의 질감 특성값) 중에서는 29개의 특성값이, Daubechies 2 기저 함수를 사용했을 때 추출한 질감

특성값 416개(13개의 WCCM x 8개의 부영상 x 4개의 질감 특성값) 중에서는 25개의 특성값이 선택되었다. 윌크스 람다 (Wilks' lambda)값은 집단 내 변동/총 변동(집단 내 변동+ 집단 간 변동)의 비율로 이 값이 작을수록 판별함수의 판별력이 높다고 해석되어지며 D1에서 D13은 3차원 GLCM과 WCCM 계산을 위한 13개의 변위 벡터를 나타내며 LLL, LLH, LHL, HLL, LHH, HLH, HHL, HHH는 3차원 Wavelet 변환 후 생성된 8개의 3차원 부영상을 의미하며 Energy, Entropy, Contrast, Homogeneity는 추출한 질감 특성값이다. 그러므로 예를 들어 선택된 특성값 D13_LLL_entropy는 Wavelet 변환을 적용한 후 얻어진 LLL 부영상에서 13번째 변위 벡터를 고려하여 계산한 WCCM으로부터 추출한 Entropy값을 뜻하며 표 2, 3, 4에서 최종 선택 된 특성값들은 회색 음영으로 표시하였다.

표 2. 3차원 GLCM에 기반하여 추출한 3차원 세포핵 질감 특성값 중 순차적 선택 방법에 의해 선택된 특성값 벡터
Table 2. Stepwise selection summary of cell nuclear texture features extracted from 3D GLCM

Step	Entered	Removed	Wilks' lambda
1	D5_homogeneity		0.890
2	D6_energy		0.810
3	D11_homogeneity		0.760
4	D10_homogeneity		0.724
5	D6_entropy		0.657
6	D4_entropy		0.504
7	D10_contrast		0.463
8		D10_homogeneity	0.464
9	D2_energy		0.430
10	D5_contrast		0.394
11		D6_energy	0.401

표 3. Haar 기저 함수를 사용한 3차원 wavelet에 기반하여 추출한 3차원 세포핵 질감 특성값 중 순차적 선택 방법에 의해 선택된 특성값 벡터
Table 3. Stepwise selection summary of cell nuclear texture features extracted from 3D WCCM based on haar basis function

Step	Entered	Removed	Wilks' lambda
1	D13_LLL_entropy		0.671
2	D4_LLH_homogeneity		0.565
3	D3_LLL_contrast		0.483
4	D9_HHH_contrast		0.439
5	D9_HLH_contrast		0.394
6	D5_LLL_energy		0.361
7	D9_LLH_contrast		0.337
8	D8_HLH_contrast		0.308
9	D4_LHL_contrast		0.289

10	D1_HLH_homogeneity		0.273
11	D3_HHL_energy		0.257
12	D6_HLL_entropy		0.244
13	D6_HHH_entropy		0.231
14	D4_HHL_energy		0.218
15	D3_LLL_entropy		0.209
16	D3_LHH_contrast		0.200
17	D6_LHL_energy		0.193
18	D6_HLL_energy		0.186
19	D4_HHH_energy		0.181
20	D2_LHH_entropy		0.175
21	D1_LHL_energy		0.165
22		D4_HHH_energy	0.168
23	D2_LHH_contrast		0.162
24		D9_HHH_contrast	0.164
25	D9_LHH_contrast		0.157
26	D1_HHL_entropy		0.152
27		D3_LLL_contrast	0.154
28	D8_LHH_energy		0.148
29	D2_LHL_entropy		0.140
30	D2_LLL_entropy		0.134
31	D1_HHH_energy		0.128
32	D1_LHH_entropy		0.122
33		D3_LHH_contrast	0.125
34	D1_LHH_energy		0.118
35	D4_LHH_energy		0.113
36	D7_LLH_energy		0.109
37		D6_HLL_energy	0.109
38	D8_LLL_entropy		0.106
39		D8_HLH_contrast	0.108
40	D12_LLL_entropy		0.103
41	D8_HHH_energy		0.100

선택된 질감특성값들을 바탕으로 판별분석(Discriminant Analysis)에 의한 3종류의 분류기를 생성하였다. 판별분석에 의한 분류기란 선형결합으로 이루어진 판별식으로, 분류하고자 하는 각 객체의 특성값들을 대입해서 각 객체들이 속하는 집단을 찾아내는 판별함수이다. 생성된 분류기가 그룹간의 차이를 잘 반영하는지에 대한 유의성 검증은 Leave one out cross validation에 의한 분류정확도를 보고 판단하였다

표 4. Daubechies 2 기저 함수를 사용한 3차원 wavelet에 기반하여 추출한 3차원 세포핵 질감 특성값 중 순차적 선택 방법에 의해 선택된 특성값 벡터
Table 4. Stepwise selection summary of cell nuclear texture features extracted from 3D WCCM based on Daubechies 2 basis function

Step	Entered	Removed	Wilks' lambda
1	D3_LLL_contrast		0.761
2	D5_HHL_energy		0.636
3	D9_HLH_entropy		0.552
4	D1_LLH_entropy		0.488

5	D2_HLL_contrast		0.418
6	D1_HHL_energy		0.343
7	D2_HHH_contrast		0.315
8	D1_LHL_energy		0.287
9	D6_HLL_entropy		0.268
10	D6_HLH_entropy		0.250
11	D8_LLL_contrast		0.229
12	D1_HHH_energy		0.216
13	D3_LLH_contrast		0.206
14	D3_HHL_contrast		0.192
15	D9_HLH_contrast		0.183
16	D3_LLL_homogeneity		0.175
17	D7_LHH_contrast		0.168
18	D2_LHH_contrast		0.161
19	D2_LHH_energy		0.153
20	D6_LLH_entropy		0.144
21		D5_HHL_energy	0.145
22	D3_LLH_contrast		0.140
23	D4_LHH_entropy		0.133
24	D2_HHLcontrast		0.127
25	D2_HLH_contrast		0.123
26	D11_HHH_contrast		0.119
27	D2_LHL_contrast		0.114
28		D2_HLL_contrast	0.116
29	D9_LLH_energy		0.112
30		D6_HLH_entropy	0.114
31		D6_HLL_entropy	0.116
32	D9_LLH_energy		0.112
33	D1_HLH_contrast		0.108

III. 실험 및 결과 분석

표 5, 6, 7은 분류등급 Normal, LSIL, HSIL 각각 100 개씩 총 300개의 세포 불륨데이터에 대하여 생성된 3개의 분류기에 대한 Leave one out cross validation에 의한 분류 정확도 비교 결과이다.

표 5. 3차원 GLCM에 기반한 질감 특성값들로 생성된 분류기에 대한 분류행렬
Table 5. Confusion matrix for the classifier using texture features based on 3D GLCM

	Normal	LSIL	HSIL	Total	%
Normal	67	16	17	100	67
LSIL	14	77	9	100	77
HSIL	9	16	75	100	75
Total	90	109	101	300	73

표 6. 3차원 wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기에 대한 분류행렬 (Haar)
Table 6. Confusion matrix for the classifier using texture features based on 3D wavelet transform (Haar)

	Normal	LSIL	HSIL	Total	%
Normal	92	3	5	100	92
LSIL	3	92	5	100	92
HSIL	8	8	84	100	84
Total	103	103	94	300	89.3

표 7. 3차원 wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기에 대한 분류행렬 (Daubechies 2)
Table 7. Confusion matrix for the classifier using texture features based on 3D wavelet transform (Daubechies 2)

	Normal	LSIL	HSIL	Total	%
Normal	91	3	6	100	91
LSIL	6	85	9	100	85
HSIL	4	8	88	100	88
Total	101	96	103	300	88

3차원 GLCM에 기반한 질감 특성값들로 생성된 분류기의 분류정확도는 73%, Haar 기저 함수를 사용한 3차원 Wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기의 분류정확도는 89.3%, Daubechies 2 기저 함수를 사용한 3차원 Wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기의 분류정확도는 88%로 3차원 GLCM에 기반한 질감 특성값들로 생성된 분류기보다는 3차원 Wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기의 분류정확도가 높게 나타났으며 서로 다른 Wavelet 기저 함수에 대해서는 큰 차이를 보이지 않았지만 Haar 기저함수를 사용한 분류기의 분류정확도가 가장 높게 나타났다.

표 8, 9, 10은 2차원 세포진 영상에서의 세포핵 질감 특성값들과 비교하기 위해 동일한 실험 불륨데이터의 투영된 2차원 영상을 이용하여 2차원 GLCM에 기반한 질감 특성값, Haar 기저 함수를 사용한 2차원 Wavelet 변환에 기반한 질감 특성값, Daubechies 2 기저 함수를 사용한 2차원 Wavelet 변환에 기반한 질감 특성값들을 추출하고 같은 방법에 의해 분류기를 생성한 후 분류정확도를 비교한 결과이다. 2차원 GLCM에 기반한 질감 특성값들로 생성된 분류기의 분류정확도는 52.3%, Haar 기저 함수를 사용한 2차원 Wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기의 분류정확도는 51%, Daubechies 2 기저 함수를 사용한 2차원 Wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기의 분류정확도는 49%로 2차원 영상에서는 GLCM에 기반한 질감 특성값들로 생성된 분류기와 Wavelet 변환에 기반한 질감특성값들에 의해 생성된 분류기의 분류정

확도가 크게 차이를 보이지 않았다. 그러나 3차원 세포핵 질감 특성값들과 비교했을 때 2차원 세포핵 질감 특성값들보다는 3차원 세포핵 질감 특성값들이 등급별 분류에 있어 효율적인 것을 확인 할 수 있었으며, 그림4는 본 연구에서 생성한 6개의 분류기에 대한 분류정확도이다.

표 8. 2차원 GLCM에 기반한 질감 특성값들로 생성된 분류기에 대한 분류행렬
Table 8. Confusion matrix for the classifier using texture features based on 2D GLCM

	Normal	LSIL	HSIL	Total	%
Normal	55	23	22	100	55
LSIL	46	29	25	100	29
HSIL	10	17	73	100	73
Total	111	69	120	300	52.3

표 9. 2차원 wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기에 대한 분류행렬 (Haar)
Table 9. Confusion matrix for the classifier using features based on 2D wavelet transform (Haar)

	Normal	LSIL	HSIL	Total	%
Normal	55	24	21	100	55
LSIL	39	29	32	100	29
HSIL	13	18	69	100	69
Total	107	71	122	300	51

표 10. 2차원 wavelet 변환에 기반한 질감 특성값들에 의해 생성된 분류기에 대한 분류행렬 (Daubechies 2)
Table 10. Confusion matrix for the classifier using features based on 2D wavelet transform (Daubechies 2)

	Normal	LSIL	HSIL	Total	%
Normal	53	27	20	100	53
LSIL	51	27	22	100	27
HSIL	11	22	67	100	67
Total	115	76	109	300	49

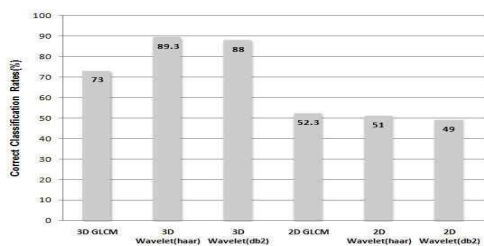


그림 4. 6개 분류기에 대한 분류정확도
Fig. 4. Correct classification rates of 6 classifiers

IV. 결론

본 논문에서는 자궁경부 세포핵의 3차원 염색질 질감 특성값이 암의 진행정도를 인식하는데 있어 유용한 특성값인지를 평가하기 위해 분류등급 Normal, LSIL, HSIL 각각 100개씩 총 300개의 세포 볼륨데이터에 대하여 3차원 GLCM에 기반한 질감 특성값과 3차원 Wavelet 변환에 기반한 질감 특성값으로 이루어진 분류기에 대한 분류정확도를 분석하였다. 실험한 분류기에 대한 분류정확도가 모두 70% 이상 분류정확도를 나타내었으며 3차원 Wavelet 변환에 기반한 질감 특성값으로 이루어진 분류기에 대한 분류정확도는 88% 이상으로 3차원 GLCM에 기반한 질감 특성값보다는 3차원 Wavelet 변환에 기반한 질감 특성값이 자궁경부암의 진행정도를 인식하는데 있어서 보다 유용한 특성값으로 나타났다. 그리고 2차원 세포핵 질감 특성값과의 비교연구에서 3차원 세포핵 질감 특성값들이 등급별 분류에 있어 보다 효율적인 것을 확인 할 수 있었다. 이는 2차원 세포핵 질감 특성값만으로는 세포핵 내부의 염색질 분포에서의 미세한 변화를 인식하기 어렵다는 것을 의미하며 이에 비해 본 논문에서 제안한 3차원 세포핵 질감 특성값은 이러한 미세한 차이를 인식하는데 있어서 효과적이며 특히 악성이라고 진단된 세포진 도말 표본에서 정상 세포의 형태는 가지고 있으나 주변악성 세포의 영향으로 세포핵 내부의 염색질 분포에서 보이는 미세한 차이를 인식하고 분류할 수 있다는 점에서 장점을 가진다.

그러나 임상적 유용성을 갖기 위해서는 보다 많은 영상에 대한 테스트가 필요하며 다양한 Wavelet 기저함수 차이에 대한 비교와 2차원 세포핵 질감 특성값과의 비교 연구에서도 투영된 2차원 영상이 아닌 볼륨 데이터 중에서 최적의 초점을 가지는 2차원 영상을 생성하여 특성값을 추출하고 비교하는 연구가 필요하다.

참고문헌

- [1] P. Pisani, F. Bray and D.M. Parkin, "Estimation of the Worldwide Prevalence of Cancer for 25 Sites in the Adult Population," International Journal of Cancer Vol. 97 pp. 72-81, 2002.
- [2] Y.T. Kim, "Causes and Diagnoses of Cervical Cancer,"

- Journal of the Korean Medical Association, Vol. 50, No. 19, pp. 769-777, 2007.
- [3] S.K. Uh, C.M. Park, H.C. Park, S.Y. Yoon, M.S. Cho, S.Y. Cho and S.S. Kim, "A Study on Automatic Detection of Uterine Cervical Pap-Smears by Image Processing," Korean Journal of Cytopathology, Vol. 5, No. 1, pp. 15-22, 1994.
- [4] K.B. Kim, "Nucleus Segmentation and Recognition of Uterine Cervical Pap-Smear using Enhanced Fuzzy ART Algorithm," Journal of Fuzzy Logic and Intelligent Systems, Vol. 16, No. 5, pp. 519-524, 2006.
- [5] M.E. Plissiti and C. Nikou, "Automated Detection of Cell Nuclei in Pap Smear Images using Morphological Reconstruction and Clustering," IEEE Transaction on Information Technology in Biomedicine, Vol. 15, No. 2, pp. 233-241, 2011.
- [6] P. Malm and A. Brun, "Closing Curves with Riemannian Dilation: Application to Segmentation in Automated Cervical Cancer Screening," Lecture Notes in Computer Science, Vol. 5875, pp. 337-346, 2009.
- [7] P. Malm, A. Brun and E. Bengtsson, "PAPSYNTH: Simulated Bright Field Images of Cervical Smears," Proceedings of IEEE International Conference on Biomedical Imaging, pp. 117-120, Rotterdam, Netherlands, 2010.
- [8] G. Burger, U. Jutting and K. Rodenacker, "Changes in Benign Cell-Populations in Cases of Cervical Cancer and Its Precursors," Analytical and Quantitative Cytology and Histology, Vol. 3, No. 4 pp. 261-271, 1981.
- [9] M. Guillaud, A. Doudkine, D. Gamer, C. MacAulay and B. Palcic, "Malignancy Associated Changes in Cervical Smears: Systematic Changes in Cytometric Features with the Grade of Dysplasia," Analytical Cellular Pathology, Vol. 9, No. 3 pp. 191-204, 1995.
- [10] R.A. Kemp, C. MacAulay, D. Gamer and B. Palcic, "Detection of Malignancy Associated Changes in Cervical Cell Nuclei using Feed-Forward Neural Networks," Analytical Cellular Pathology, Vol. 14, pp. 31-40, 1997.
- [11] R.M. Haralick, K. Shanmugam and I. Dinstein, "Texture Feature for Image Classification," IEEE Transaction on Systems, Machines and Cybernetics, Vol. 3, No. 6, pp. 610-621, 1973.
- [12] C.S. Burrus, R.A. Gopinath and H. Guo, "Introduction to Wavelets and Wavelet Transform" Prentice-Hall, 1998.
- [13] A.J. Richard and K.S. Jeffrey, "Applied Multi variate Statistical Analysis" Prentice-Hall, 1997.

저 자 소개



최 현 주

2005 : 인제대학교 전산학과
전산학박사.
현 재 : 인제대학교 컴퓨터공학부
연구교수.
관심분야: 의료영상 처리 및 분석.
Email : hyunjuchi77@gmail.com



김 태 운

2011 : 인제대학교 전산학과
전산학박사.
현 재 : 국립암센터 연구원.
관심분야: 의료영상 처리 및 분석.
Email : liminus@paran.com



Patrik Malm

현 재 : 웁살라대학교(스웨덴)컴
퓨터공학 박사과정.
관심분야: 영상처리.
Email : patrik@cb.uu.se



Ewert Bengtsson

현 재 : 웁살라대학교(스웨덴)영
상분석센터 센터장.
관심분야: 영상분석.
Email : ewert@cb.uu.se



최 흥 국

1996 : 울산대학교(스웨덴)컴퓨터공학 공학박사.

현 재 : 인제대학교 컴퓨터공학부 교수.유비쿼터스 헬스케어 연구 센터센터장

관심분야 : 의료영상 처리 및 분석, 컴퓨터 그래픽스, 유비쿼터스 헬스케어.

Email : cschk@inje.ac.kr