제22회 한국정보처리학회 추계학술발표대회 논문집 제11권 제2호 (2004. 11)

다중특징을 이용한 유방종양영상 내용기반검색 시스템 개발

김민경[a], 최흥국[a,b]
[a]인제대학교 의료영상과학대학원
[b]인제대학교 컴퓨터공학부
e-mail:mkkim@mitl.inje.ac.kr

Development of Content Based Breast Tumor Image Retrieval System Using Multi Features

[a]Min-Kyoung Kim, [a,b]Heong-Kook Choi
[a]Dept. of Medical Image Science, Inje University
[b]Dept. of Computer Science, Inje University

요 약

현재 병리 의사에 의해 주관적으로 이루어지고 있는 병리 영상의 진단에 도움을 주기 위해 병리영상에서 객관적으로 추출 가능한 정보를 이용하여 유방종양 검색 시스템을 개발하였다. 다중 특징을 이용한 내용 기반 검색 방법을 사용하였으며, 영상에서 자동으로 추출 가능한 다양한 특징을 검색의 파라미터로 이용하였다. 진단에 도움을 주기 위해 전체 영상 뿐만 아니라 관심 있는 영역의 부분영상도 추출하여 검색이 가능하게 설계하였으며 시스템의 평가를 위해 단일 특징을 이용하여 영상을 검색 하였을 때와 다중 특징을 이용하여 영상을 검색 하였을 때의 검색율을 비교하였다. 향후 이 시스템은 병리영상의 진단에 있어 객관적이고 높은 재현성을 가지게 하는 보조도구로 사용될 수 있을 것이다.

1. 서론

기존의 해부병리에서 이루어지고 있는 병리 진단은 대부분 병리 전문가에 의해 주관적으로 이루어지고 있다. 따라서 병리 영상의 판독과 진단에 있어 좀 더 객관성을 줄 수 있는 시스템이 요구되어지고 있다. 따라서 본 연구에서는 진단이 필요한 유방종양 병리 영상이 주어졌을 때 영상의 특징값을 추출하여 시스템 내의 데이터베이스의 특징값 데이터 중 가장 유사한 값을 가진 영상을 검색해 내는 내용기반검색 방법을 사용하여 시스템을 구축하였다. 유방종양세포를 이용한 내용기반 검색 시스템 구축을 위하여 영상에서 자동으로 추출 가능한 객관적인 특징인 색상 특징값, 질감 특징값, 웨이블렛 변환을 적용한 질감 특징값 등 총 22개의 특징값을 사용하여 영상을 검색할 수 있도록 하였다. Visual Basic을 이용하여 개발하였으며 영상에서 추출된 특징값들은 SQL을 이용하여 데이터베이스를 구축하였다.

본 논문의 구성은 다음과 같다. 2장에서는 유방종양영상에서 자동으로 추출할 수 있는 각 특징값에 대한 함수들을 설명하였고, 3장에서는 영상의 획득과 획득되어진 영상의 특징값 추출 및 본 시스템의 전체적인 구조에 대해 언급할 것이다. 4장에서는 시스템 구현과 성능평가결과에 대해 말하고 5장에서 결론 및 향후 연구과제에 대해 논의할 것이다.

2. 영상의 특징 추출

2.1 Color특징 추출

Color특징 추출을 위해 Markus의 유사도 측정벡터를 사용하여 색상분포의 유사도를 측정하였다. 이것은 밴드별 컬러 분포를 정량화 한 값에서 세 가지 밴드의 관계성을 평가해 주어 컬러 영상이 갖는 상대적인 분포 특성을 전체적으로 평가해 주는 것으로 아래의 식과 같이 나타낼 수 있다.

$$\begin{aligned} F_1 &= R-G && \text{(R = average of R band)} \\ F_2 &= 2XB-R-G && \text{(G = average of G band)} \\ F_3 &= R+G+B && \text{(B = average of B band)} \end{aligned} \quad (1)$$

위의 식(1)의 F_1, F_2, F_3은 컬러밴드의 차이값 으로 구성된 특성 함수로 컬러 영역에 대한 분포를 인덱스 하는 함수이다. 또한 영상의 컬러 분포는 확률분포로 해석될 때 확률적인 평균값에 의해 특징화 될 수 있으므로, 영상의 평균과 분산, 표준편차를 이용한 분포도는 영상해석에 대한 자료를 제공한다. 아래의 식(2)는 확률분포적인 영상 특징값이며, m×n은 영상의 전체 화소수를 말한다. 아래의 식에서 I_{ij} 는 영상에서의 gray level intensity이며, E_I는 영상의 평균값을, σ_I는 영상의 표준편차값을, S_I는 영상의 3차 모멘트로서 왜도(skewness)를 나타낸다[1][2].

$$E_I = \frac{1}{\square \times \square} \sum_{\square=0}^{\square-1} \sum_{\square=0}^{\square-1} I_{\square}$$

$$\sigma_I = \left(\frac{1}{\square \times \square} \sum_{\square=0}^{\square-1} \sum_{\square=0}^{\square-1} (I_{\square} - E_I)^2 \right)^{\frac{1}{2}} \quad (2)$$

$$S_I = \left(\frac{1}{\square \times \square} \sum_{\square=0}^{\square-1} \sum_{\square=0}^{\square-1} (I_{\square} - E_I)^3 \right)^{\frac{1}{3}}$$

2.2 질감특징 추출

유방세포영상의 분석에서는 염색질 패턴의 내부적 구조 변화를 분석할 수 있는 질감 분석이 필요하다. 본 논문에서는 사용된 질감 특징 추출 방법은 세포 조직 영상이 거의 불규칙적인 패턴을 보이므로 통계학적 방법 중의 하나인 GLCM (Gray Level Co-occurrence Matrix)을 사용하였다. GLCM 방법은 한 영상을 대표하는 통계 값을 얻기 위하여 각 화소들간의 공간적인 상호 관계를 분석하는 방법이다. MN크기의 영상에서 거리 d와 방향 θ에 의하여 그레이 레벨 i,j를 갖는 해당 화소쌍이 얼마나 자주 나타나는가를 표시하는 관련 빈도수 행렬 P[i,j]이다. 텍스쳐 패턴의 영향을 많이 받는 특징인 Contrast와 Energy, 그레이 레벨의 차이에 영향을 많이 받는 Contrast와 Unifromity를 특징값으로 사용하였으며 추출한 질감 특징의 식은 아래의 식(3)과 같다.(단, PM은 co- occurrence matrix이다)[3][4].

$$\text{- Entropy} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (PM) \log (Px(i)Py(j))$$

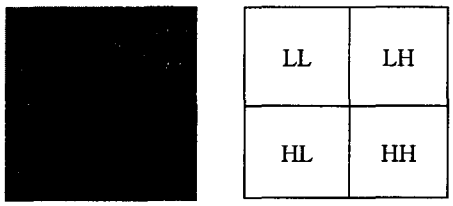
$$\text{- Energy} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (PM)^2 \quad (3)$$

$$\text{- Contrast} = \sum_{n=0}^{N-1} n^2 \left(\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} PM \right) \quad (\text{단}, |i-j| = n)$$

$$\text{- Uniform} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p[i,j]}{1+n} \quad (\text{단}, |i-j| = n)$$

2.3 Wavelet 변환을 적용한 질감특징 추출

유방세포영상의 질감이 가지는 2차원의 공간적인 성질 때문에 단순한 1차원적인 방법으로 질감 특징을 분석하는 것과 다해상도 방법으로 영상을 분석하는 것에 있어 차이가 있다. 따라서 공간영역과 주파수영역에서 그 지역적 특징을 잘 나타내는 특성으로 질감정보를 얻고자 할 때 많이 쓰이는 wavelet변환을 적용하여 질감특징을 분석해 보았다. 빠른 계산과 효율적인 특징 추출을 위해 영상처리 응용분야에서 많이 쓰이는 Harr wavelet 변환을 사용하였다. 웨이블렛 변환을 사용하여 생성되어진 하위 영상들에 대한 각각의 질감 특징값을 특징벡터로 사용하였다. 그러나 LL 하위 영상은 저해상도 영상으로 원래 영상을 blurring함으로써 생성되는 영상인 반면 나머지 하위 영상은 원영상이 blurring됨으로 인해 손실되는 정보를 성분별로 나타내는 영상이므로 LL하위 영상을 제외한 나머지 3개의 하위영상으로부터 질감 특징을 추출하였다[5][6].

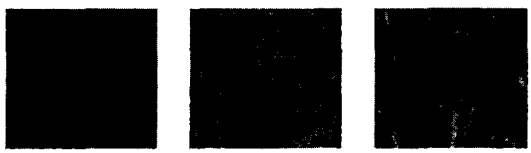


[그림1] 1단계 Harr wavelet 변환

3. 시스템의 흐름 및 구성

3.1 유방종양영상의 획득

디지털 카메라가 연결되어 있는 광학 현미경을 통해 병리 전문의에 의해 병변을 대표하는 부위를 100배율에서 획득한 후 256×256크기의 칼라 영상으로 저장하였다. 또한 주요병변이 있는 부분을 128×128 크기로 선택하여 한 영상에서 5개 이상의 부분 영상이 저장되도록 하였다. FCC(fibrocystic change-섬유낭성변화), DCIS(ductal carcinoma in situ - 관상피 내암종), CA(carcinoma, invasive Ductal Carcinoma - 침윤성 관암종)로 유방종양 영상을 세 종류로 분류한 후 각각 100개씩 총 300개의 영상을 데이터베이스에 저장하였다.

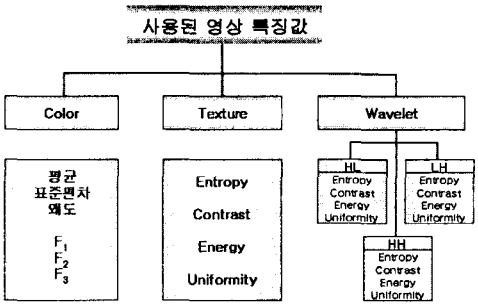


(a)FCC 10×　(b) DCIS 10×　(c) CA 10×

[그림2] 질환별 유방종양영상

3.2 영상 특징값 분류

본 시스템은 다양한 특징값을 사용하여 세포영상을 검색하는 시스템으로 아래의 그림3과 같이 22개의 특징값을 검색의 파라미터로 사용하였다.

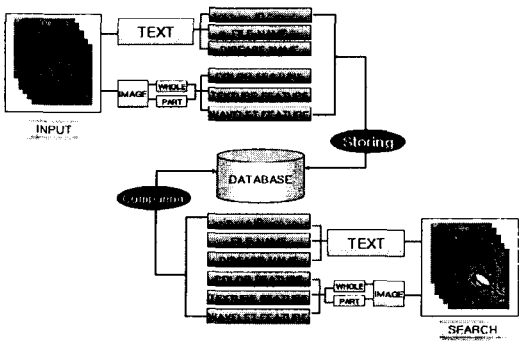


[그림3] 영상 특징값의 분류

3.3 시스템의 전체적인 구조

본 시스템의 전체적인 구조는 크게 입력된 이미지에서 영상의 특징 벡터를 추출하여 이를 데이터베이스에 저장하는 부분과 질의 이미지를 입력하여 특징 벡터가 유사한 이미지를 검색하는 두 부분으로 나뉜다. 아래의 그림[4]는 본 시스템의 전체적인 구조를 나타내고 있다.

세포영상 검색을 위해 데이터베이스에 영상을 입력할 때 문자부분과 영상부분을 나누어 저장하였으며, 문자부분에서는 ID, 파일이름, 질병명 등의 텍스트 정보만으로 검색이 가능하게 하였다. 영상부분은 영상에 대한 정보가 전체영상과 부분영상으로 나뉘어 저장되는데, 전체 영상은 256×256 크기의 입력영상 그대로에서 해당되는 특징값을 추출하여 저장하였고, 부분영상은 한 영상에서 128×128크기의 샘플을 5개 이상 뽑아 그 각각의 특징값을 추출하여 데이터베이스에 저장하였다[7].

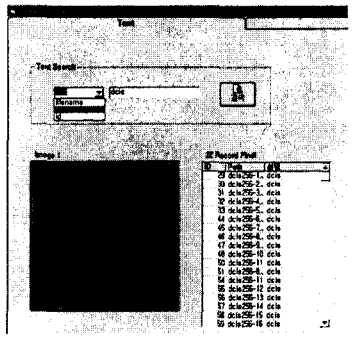


[그림4] 유방종양 검색시스템의 전체적인 구조

4. 시스템 구현 및 성능평가

4.1 문자 검색 부분

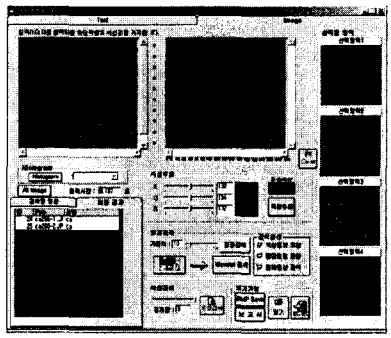
병리 영상에 대한 기본적인 텍스트 정보를 저장하여 텍스트를 이용하여 검색이 가능하도록 설계하였다. 문자검색부분에서는 아래의 그림[5]와 같이 ID, 파일이름, 질병명 등으로 검색이 가능하다.



[그림5] 텍스트 기반 영상 검색 결과

4.2 영상 검색부분

유방종양세포영상은 한 영상에서도 여러 가지 병변이 발견될 수 있기 때문에 전체 영상에 대한 정보뿐 아니라 관심영역에 대한 정보 또한 검색에 필요하다. 따라서 아래 그림[6]에서 보듯이 전체 영상과 부분영상을 필요에 따라 검색 가능하도록 시스템을 설계하였다.



[그림6] 내용기반 영상 검색 결과

4.3 성능평가 및 결과

부분영상과 전체영상 각각에서 단일특징의 검색률 및 두개 이상의 특징을 결합한 다중특징의 검색률을 테스트하여 비교해 보았다. 성능평가척도는 일반적으로 정확한 매칭이 아닌 유사매칭을 수행하는 내용기반검색에서 많이 사용되고 있는 Recall과 Precision을 사용하였다. 다음의 표[1]과 표[2]는 각각 부분영상과 전체 영상에 대해 검색되어진 결과에

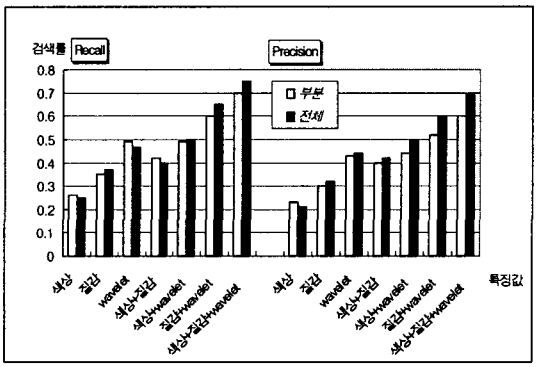
대한 Recall과 Precision을 각각 나타낸 것이다. 실험 결과에서 알 수 있듯이 부분 영상과 전체 영상의 검색에서는 전체 영상이 검색률이 높았고, 세 가지 특징 모두를 결합한 다중 특징값으로 검색했을 때 검색률이 단일 특징으로 검색했을 때보다 월등히 높음을 알 수 있었다.

	부분영상	전체영상
색상	0.26	0.25
질감	0.35	0.37
wavelet	0.49	0.47
색상+질감	0.42	0.40
색상+wavelet	0.49	0.50
질감+wavelet	0.60	0.65
색상+질감+wavelet	0.70	0.75

[표1] Recall에 대한 성능평가 결과

	부분영상	전체영상
색상	0.23	0.21
질감	0.30	0.32
wavelet	0.43	0.44
색상+질감	0.40	0.42
색상+wavelet	0.44	0.50
질감+wavelet	0.52	0.60
색상+질감+wavelet	0.60	0.69

[표2] Precision에 대한 성능평가 결과



[그림7] 성능평가 결과 비교

5. 결론 및 향후 연구 과제

본 연구에서는 기존의 병리과에서 이루어지고 있는 진단이 대부분 병리 전문가에 의해 주관적으로 이루어짐에 따라 요구되어지는 객관적인 판단에 도움을 줄 수 있는 시스템을 구축하였다. 이 시스템은 진단이 필요한 병리 영상이 주어졌을 때 영상의 특징값을 추출하여 시스템 내의 데이터베이스의 특징값 데이터 중 가장 유사한 값을 가진 영상을 검색해내는 내용기반검색 방법을 사용하였다. 검색에 사용된 특징값으로는 칼라, 질감, wavelet 영역에서의 질감 등에서 각각 추출되어진 총 22개의 특징값을 검색의 파라미터로 이용하였다. 세포영상을 검색하는 시스템의 특징상 부분 영상과 전체 영상으로 나누어 검색할 수 있게 하였다. 칼라, 질감, wavelet영역의 질감 각각의 검색률과 결합된 특징의 검색률을 구하여 비교하였으며 실험결과 색상, 질감, wavelet의 모든 특징값을 파라미터로 사용했을 때 Recall과 Precision이 가장 좋았다. 이러한 다중 특징을 이용한 유방종양영상 내용기반 검색 시스템은 유방 종양의 진단에 있어 좀 더 객관적이고 높은 재현성을 가지도록 할 수 있다. 따라서 본 시스템은 진단의 보조도구로 사용될 뿐 아니라 저장되어진 병리 영상 데이터베이스를 활용하여 병리 영상의 학습 등에 활용 가능할 것이다. 향후 검색률 향상을 위해 세포의 형태학적 특징값을 검색의 파라미터로 추가해야할 필요가 있을 것이며, 검색의 효율성을 위한 속도 개선방안도 추가적으로 연구되어져야 할 것이다.

참고문헌

[1] 이병일, 손병환, 홍성욱, 손성건, 최홍국, "자동차 전조등 검색을 위한 다중지식기반의 영상검색 기법",신호처리시스템학회 논문지 제 3권3호, pp27-34

[2] R. Jain, R. Kasturi, B. G.Schunck, Machine vision, McGraw-Hill Inc, International edition, 1995

[3] Robert M. Haralick, K. Shanmugam, Its' Hak Dinstein, "Textural Features for Image Classification", IEEE Trans. On System, Man, and Cybernetics, Vol. SMC-3, No.6, pp. 610- 624, 1973

[4] T. Randen and J. H. Husoy, "Image Content Search by Color and Texture Properties".*ICIP '97*,Vol 1. pp580- 583, 1997

[5] 황해길, "유방 종양의 세포 및 조직 영상의 분류", 인제대학교 대학원 졸업논문, 2001

[6] B. Jahne, "Digital Image Processing : concepts, algorithm, and scientific applications" , Springer-Verlaf Berlin, 1995

[7] M. Flicker et al., "Query by image and video content: The QBIC system," IEEE Computer, vol. 28, pp. 23-32, Sept. 1995.