

칼라와 질감특징을 이용한 유방종양영상 내용 기반 검색

김민경^a, 조미정^a, 황해길^b, 최홍국^{ab}
인제대학교 의료영상대학원^a, 인제대학교 컴퓨터공학부^b

Content Based Image Retrieval in Breast Tumor Images Using Color and Texture Features

Min-Kyoung Kim^a, Mi-Jung Cho^a, Hea-Gil Hwang^b, Heong-Kook Choi^{ab}
^aDept. of Medical Image Science, Inje University
^bDept. of Computer Science, Inje University

요 약

기존의 병리 영상을 판독하고 저장, 관리 하는 시스템이 수작업으로 이루어져 발생하는 문제점들을 보완하는 방안으로 유방종양 영상을 사용하여 세포영상 내용기반 검색 시스템을 구축하고자 한다. 유방암 세포를 사용하여 효율적인 내용기반 영상 검색 시스템을 구축하기 위해서는 유방암 영상에서 검색에 가장 적합한 영상의 질감, 칼라, 형태학적 특징값의 조합이 필요하다. 따라서 본 논문에서는 세포영상의 분류에 많이 사용되는 질감 특징과 칼라값을 사용하여 내용기반 검색 시스템을 구축 하였으며, 칼라값과 질감특징값을 사용하여 검색했을 때의 효율성을 비교하였다. 향후 이러한 실험들을 통하여 세포영상검색에서 가장 최적의 파라미터들을 조합한 내용기반 검색 시스템을 구축하고자 한다.

1. 서론

최근 멀티미디어 기술의 발전으로 인해 영상 데이터베이스 시스템이 현실화되고 있다. 영상 데이터베이스 시스템은 많은 영상들로 이루어져 있기 때문에 사용자가 원하는 영상을 찾는 데에는 상당한 어려움이 있다. 따라서 효율적인 검색 방법이 요구되어지고 있다. 이에 내용 기반 영상 검색 방법이 다양하게 연구되어지고 있는데 이러한 내용 기반 영상 검색에서 갖추어야 할 조건들을 살펴보면 첫 번째로 물체의 회전, 이동, 크기 변화에 민감하지 않아야 하며, 두 번째로 많은 수의 영상을 위해 데이터베이스를 구축하므로 영상 특성에 대한 파라미터의 개수가 적어야 한다는 것이다. 현재 이러한 조건을 만족시키는 방법들이 제안되고 있는데 제안되고 있는 방법들을 살펴보면, 영상 특성 추출을 위해 영상의 색(Color), 물체의 모양(Shape), 전반적인 레이아웃(Layout), 질감(Texture) 등을 사용하고 있다 [1][2].

이러한 내용기반 영상검색은 일반 멀티미디어 분야 뿐만 아니라 의료영상검색에도 응용되어질 수 있다.

본 연구에서는 기존의 병리 영상을 저장하고 검색하는 시스템이 수작업으로 이루어져 발생하는 문제점을 보완하기 위한 방안으로 유방 종양 영상을 사용하여 내용기반 영상 검색 시스템을 구축하였다. 영상이 가진 정보 중에서 자동으로 추출이 가능한 색상, 모양, 질감 정보 중에서 우선적으로 색상 정보와 질감 정보를 사용한 검색 시스템을 구축하였다. 영상에서 추출한 색상과 질감 특징값은 유클리디안 거리함수(Euclidean Distance)를 사용하여 데이터베이스 내에 있는 영상들과의 유사도를 측정하였으며, 검색효율의 측정을 위한 평가 기준은 재현율(Recall)과 적합율(Precision)을 사용하였다.

논문의 구성은 다음과 같다. 2장에서는 유방종양영상에서 칼라와 질감특징을 어떻게 추출했는지에 관하여 살펴볼 것이며, 3장에서는 검색 시스템의 구성과 질의 영상과 데이터 베이스 내의 영상 사이의 유사도 측정을 위한 방법에 대하여 기술하였다. 4장에서는 이 검색 시스템에 대한 성능 평가를 하였으며 5장에서는 성능 평가 결과 및 향후 연구과제에 대하여 논의한다.

2. 특징벡터의 추출

내용기반 영상 검색 방법은 텍스트 기반 검색 기법들처럼 주관적인 판단이 들어가지 않고 이미지 자체에서 정보를 추출하여 이를 검색에 이용함으로써 효율적인 검색이 가능하게 한다. 이렇게 추출된 정보는 벡터화하여 데이터 베이스에 저장되며 이를 특징벡터(feature vector)라고 한다.

1) Color 정보 추출 기법

영상에서 색상정보는 대상을 아주 단순화 시켜주고 대략적인 구분 정보를 표현할 수 있는 특징이 있다. 색상 정보는 이미지의 화소(pixel)로부터 직접 정보를 추출한다. 색상 정보는 보통 RGB 형태로 표시되며, 경우에 따라서 HSI, CIE 등의 다른 컬러 공간을 이용할 수도 있다.

색상정보의 분석은 히스토그램을 통한 분포 특성을 이용하거나 확률 밀도 함수를 통한 분포 특성을 이용하는 방법이 있는데 본 연구에서는 Swain에 의해 제안된 히스토그램을 사용하여 영상에서 칼라 특징벡터를 추출하였다[3].

히스토그램은 색상 정보를 이용하는 검색 기법 중에서 주로 이용되는 기법이다. 유사한 영상들은 유사한 색 분포를 갖는다는 가설하에 제안되었다. 히스토그램의 경우 주로 RGB 컬러 공간을 사용하며 하나의 화소는 RGB (0, 0, 0)부터 RGB (255, 255, 255)까지의 256^3 개의 조합 가능한 값을 갖게 된다. 따라서 이미지의 모든 화소에 대하여 각각의 값이 몇 번 나오는지 계산하여 그 빈도수를 특징벡터로 사용하게 된다. 이 방법의 경우 계산이 쉽고 효율적이며 영상의 회전과 작은 이동에도 민감하지 않다는 장점이 있는 반면에, 색상 정보의 빈도수를 계산하는 것이므로 전혀 다른 이미지가 우연히 색상 정보와 일치하여 검색결과로 나오는 문제가 발생할 수 있는 단점이 있다[4].

2) 질감정보 추출 기법

영상에서의 질감 분석은 영상이 가지고 있는 질감의 상호 대비성, 질감의 방향성, 질감 무늬의 규칙성을 분석하여 객관적 수치로 나타내는 것이다.

질감 분석 방법은 크게 영상 원형의 규칙적인 공간상의 배열을 다루는 구조적인 방법과 영상내 각 화소 사이의 상호 관계를 분석하는 통계적인 방법으로 나눌 수 있다[5]. 구조적인 방법은 질감의 구조가 규칙적으로 반복되는 기하학적인

도형의 배열을 분석하는 데 이용된다. 이 방법은 영상의 구조적인 원형(Primitive)과 그들의 위치 규칙을 정의함으로써 질감 정보를 표현한다. 하지만, 영상 내 원형의 구조가 크고, 일정한 규칙성을 가지는 영상만이 분석 가능하다는 단점을 가진다. 그리고, 통계적인 방법은 질감 성질을 묘사하기 위한 가장 단순한 접근법으로, 한 영상을 대표하는 통계 값을 얻기 위하여 일반적으로 화소들간의 공간적인 상호 관계를 분석한다. 본 논문에서는 사용된 질감 특징 추출 방법은 세포 조직 영상이 거의 불규칙적인 패턴을 보이므로 통계학적 방법중의 하나인 GLCM (Gray Level Co-occurrence Matrix)을 사용하였다. GLCM은 원 영상에서 해당 그레이 레벨쌍의 출현 횟수를 기본으로 한다. 즉, GLCM은 MN크기의 영상에서 거리 d 와 방향 θ 에 의하여 그레이 레벨 ij 를 갖는 해당 화소쌍이 얼마나 자주 나타나는가를 표시하는 관련 빈도수 행렬 $P[i,j]$ 이다.

우선 칼라로 획득되어 저장된 RGB 칼라 영상을 삼색 스펙트럼에서의 빛의 분포를 감안하여 아래의 식(1)을 사용하여 8비트 그레이 레벨 영상으로 전환하였다[6].

$$Gray = Red \times 0.35 + Green \times 0.58 + Blue \times 0.07 \quad (1)$$

Joint Probability Density 함수 $f(i, j, d, \theta)$ 를 통해 Gray level의 Co-occurrence Matrix(GLCM)을 생성한 후, 두 화소간의 거리 d 값은 1에 대해, 각 θ 은 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 에 대해 계산하였다. 원 영상의 256 gray level에 대해 모든 두 화소 간의 밝기값의 분포를 계산 할 경우, 크기도 클 뿐만 아니라 계산량도 너무 많이 요구되므로, 16 gray level로 축소하여 16×16 GLCM을 생성하여 질감 특징을 추출하였다[7]. Entropy, Energy, Contrast, Uniformity의 4개의 질감 특징을 사용하여 시스템을 구축하였다[8].

Contrast와 Energy는 텍스처 패턴의 영향을 많이 받는 특징이고, Contrast와 Uniformity는 그레이 레벨의 차이에 영향을 많이 받는 특징이다. 추출한 질감 특징의 식은 아래와 같다.(단, PM은 co-occurrence matrix이다)

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

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

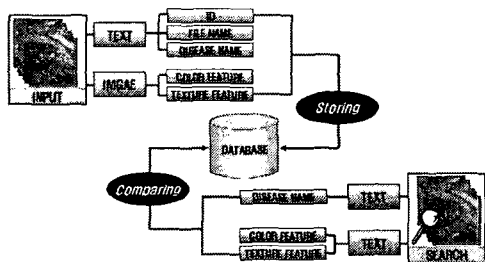
$$- \text{Contrast} = \sum_{i=0}^{N-1} n^2 \left(\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} PM \right) \quad (\text{단, } |i-j| = n)$$

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

3. 내용기반 이미지 검색 시스템의 구현

1) 시스템의 구성

본 시스템은 구축한 영상 데이터베이스에서 질의 영상을 뽑아 유사한 영상들을 검색하는 query by example방식을 이용한다. 시스템의 전체적인 구성은 크게 새로운 이미지를 입력하여 칼라와 질감 특징 벡터를 추출하여 이를 데이터베이스에 저장하는 부분과 질의 이미지를 입력하여 특징 벡터가 유사한 이미지를 검색하는 두 부분으로 나뉜다. 아래의 <그림1>은 본 시스템의 전체적인 구조를 나타내고 있다.



<그림 1> 유방종양영상 검색 시스템의 구조

2) 유사도 측정

각 영상에서 추출된 대표 특징 벡터들은 데이터베이스내에 저장되어 영상 검색 과정에 이용된다. 특징 벡터들을 이용하여 사용자가 질의한 영상과 유사한 영상들을 검색하는 과정은 다음과 같다. 우선 질의 영상의 대표 특징 벡터를 구한다. 다음 DB내 각 영상의 대표 특징 벡터와의 차이를 이용해 유사도 계산을 한 후 유사도 값이 큰 영상부터 출력한다. 유사도 계산은 유클리디언 거리를 이용하였고, 평준화 과정을 거쳐 유사도 값은 0과 1사이에 표현된다[9].

$$S_D = \sum_{K=0}^m (F_{Q,K} - F_{D,K})^2$$

Q: 질의 영상에 관한 인덱스, D: DB내 영상에 관한 색인

m: 영상의 특징 벡터 빈(bin) 수

4. 성능평가

유방종양영상을 사용한 내용기반 검색 시스템의 성능을 평가하기 위해 디지털 카메라가 연결되어 있는 Olympus광학 현미경을 통해 병리 전문의에 의해 획득되어진 병변을 대표하는 부위를 100배율에서

2048*1536크기로 획득한 후 512*512크기의 칼라 영상으로 저장하였다. Benign, DCIS, CA로 유방영상을 세종류로 분류한 후 각각 30개씩 총 90개의 영상으로 실험하였다.

내용기반 영상 데이터 검색의 효율성을 분석하기 위하여, 일반적으로 Recall과 Precision의 두 가지 성능 평가 척도를 이용한다. Recall은 영상 데이터베이스 내에서 질의와 관련된 영상 중 검색된 영상의 비율을 말한다. 그리고, Precision은 검색된 영상 중에서 질의와 관련된 영상의 비율을 나타낸다[10][11].

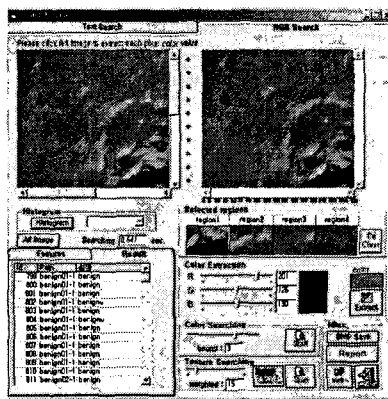
$$\text{Recall} = \frac{R_r}{T} \quad \text{Precision} = \frac{R_r}{T_r} \quad (2)$$

식(2)에서, R_r은 검색된 항목 중에서 질의와 관련된 항목의 수를 나타내고, T는 검색 대상 중에서 질의와 관련된 항목의 총수를 말하고, T_r은 검색된 항목의 총수를 나타낸다. 아래의 <표1>은 칼라와 질감특징의 검색결과를 보인 것이다.

성능평가 특징벡터	Recall	Precision
Color	0.37	0.35
Texture	0.51	0.42

<표1> 성능평가 결과

<표1>에서 보는바와 같이 유방종양 세포에서 칼라값과 질감특징값을 사용하여 검색했을 때의 효율성은 질감특징값을 사용했을 때 보다 칼라값을 사용했을 때 더 낮게 나타났다. 이는 세포영상에서의 칼라값이 같은 시약으로 염색되어 질환별로 색상에 있어서 크게 차이가 나타나지 않아서이다.



<그림 2> 유방종양영상 검색 결과

<그림2>는 본 시스템에서 유방종양 영상을 검색했을 때 검색되어진 결과를 나타낸 것이며 아래의 <그림3>과<그림4>는 각각 칼라 특징값과 질감 특징값으로 유방 종양 영상을 검색 했을때 나온 결과를 나타내고 있다.

Query image	검색결과		
Benign	DCIS	CA	Benign

<그림3> 칼라특징값으로 유방종양영상 검색 결과

Query image	검색결과		
CA	CA	CA	DCIS

<그림4> 질감특징값으로 유방종양영상 검색 결과

5. 결론 및 향후과제

본 연구에서는 기존의 병리 영상을 저장하고 검색하는 시스템이 수작업으로 이루어져 발생하는 문제점을 보완하기 위한 방안으로 유방 종양 영상을 사용하여 내용기반 영상 검색 시스템을 구축하였다. 칼라와 질감 특징을 검색시스템의 파라미터로 이용하였으며, 이 두 특징값을 사용해 유방종양영상을 검색했을때, 칼라값보다 질감특징값이 검색의 효율성은 좋았으나 다른 내용기반 검색 시스템의 효율성과 비교해보았을 때는 낮은 검색율을 나타내어 좀 더 정확하고 검색율이 높은 시스템이 요구되어지고 있다. 따라서 칼라와 질감특징 뿐만이 아닌 모양(Shape)과 레이아웃(Layout)등을 파라미터를 사용한 검색율이 높은 유방종양영상 검색 시스템으로 발전해 나가야 할 것이다. 이러한 유방암영상 내용기반 검색 시스템은 향후 유방 종양의 진단 및 병리 영상의 학습에 이르기까지 다양하게 활용이 가능할 것이며 향후 다른 종류의 질병에까지 확대하여 포괄적인 세포검색시스템을 구축하여야 할 것이다.

[참고문헌]

- [1] 송준규, 최황규, "모양정보의 회귀 추정에 의한 내용기반 이미지 검색 기법, 정보통신 논문지 Vol 5, 2001
- [2] 이은주, 정성환, "Color N * M-grams를 이용한 영상 분류, 정보과학회논문지 Vol 25, No2, Oct, 1998
- [3] M. J. Swain and D. H. Ballard, "The role and use of color in a general vision system," Proceedings of Image Understanding Workshop, Vol. 2, pp 599-613, February, 1987
- [4] Barolo B., Gagliardi I., Schettini R., "An effective strategy for querying image databases by color distribtuion", Computer and the History of Art Journal, Special issue on Electronic Imaging and the Visual Arts, vol. 7(1), pp. 3-14, 1997.
- [5] 김희승, "영상인식-영상처리, 컴퓨터 비전, 패턴인식, 신경망", 생능출판사, 1994
- [6] Ramesh Jain. Rangachar Kasturi. Brian G, Schunck, "Machine Vision", ISBN 0-07-032018-7, pp.234-248, 1995
- [7] 김재륜, 하진영, 김백섭, 김호성, "자궁경부암 세포 영상 분할을 위한 Threshoding 기법", 한국정보과학회 학술발표논문집, Vol. 26, No. 2, pp.419-421, 1999
- [8] 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
- [9] Y. Rui, Thomas S. Huang, Michael Ortega, and Sharad Mehrotra, "Relevance Feedback : A Power Tool in Interactive Content-Based Image Retrieval", IEEE Tran on Circuits and Systems for Video Technology, Vol. 8(5), pp. 644-655, Sept. 1998.
- [10] Gerard Salton, Chris Buckley, "Improving Retrieval Performance by Relevance Feedback," IST 83-16166 and IRI 87-02, 88-898, February,1988
- [11] 최현섭, 김철원, 김성동, 최기호, "텍스처패턴과 윤곽점 기율기 성분을 이용한 내용기반 화상 검색 시스템의 설계 및 구현," 한국 정보과학회 논문지, 제4권, 제1호, 1997.1