

유사객체 분류에 의한 유사 의료영상의 검색

원정임⁰ 이덕형 송혜정 윤지희 김백섭
한림대학교 컴퓨터공학과

(jiwon⁰, hanuhm, hjsong, jhyoor, bskim)⁰@hallym.ac.kr

Retrieval of Similar Medical Image Objects using Conceptual Clustering Methods

Jung-Im Won⁰, Duck-Hyung Lee, Hyeo-Jung Song, Jee-Hee Yoon, Baek-Sop Kim
Dept. of Computer Engineering, Hallym University

요약

의료영상 처리시스템의 자동인식 결과 등과 함께 진단 중인 의료 영상과 유사한 영상객체를 임의로 검색하여, 부가정보로 활용할 수 있는 지능적 의료정보 시스템 구현에 대하여 논한다. 의료 영상객체간 유사도 계산을 위하여 객체로부터 추출된 특정 정보를 객체 속성으로 이용하며, 이들 특정 값들의 빈도와 관련 분포, 속성 간 관련성 등을 고려한 유사객체 분류방식을 사용한다. 이와 같이 얻어진 영상객체 간 유사도 정보는 지식베이스로 관리되어 자동 영상 인식에 사용될 뿐 아니라 유사 영상 검색 및 진단의 근거자료로 사용된다. 즉 전문의나 병리학자들은 진단 과정에서 유사영상의 판독 결과 등을 참조함으로써 영상의 정확한 판독 및 진단 확증의 객관적 근거 자료를 확보하는데 도움을 받을 수 있다. 구현된 시스템의 적용 예로 자궁경부 세포진 영상인식 시스템을 이용하여 그 유용성을 보인다.

1. 서론

의료 기술의 발전과 함께 의료 영상의 획득과 분석, 그리고 진단을 위한 기술도 많은 발전을 거듭하고 있다. X-Ray 뿐 아니라 CT(Computed Tomography), MR(Magnetic Resonance), 초음파 등 최첨단 의료영상 진단 장비를 통해 얻어진 의료 영상은 전문의들에게 환자의 이상 유무를 판별하고, 치료를 위한 계획과 방법 등을 결정하는 중요한 자료로 활용된다. 컴퓨터에 의한 의료 영상의 분석 및 진단정보 추출을 위한 영상 처리는 컴퓨터에 의한 시각 정보 획득과 처리로 정의 될 수 있으며, 그 중 인식 과정은 영상의 특징을 추출하여 얻어진 각종 정보를 사용하여 객체를 인식하는 과정으로 볼 수 있다. 이 분야의 특징은 대량의 데이터 생성과 복잡한 처리를 요하며, 대상 분야 전문가의 전문적 지식을 반드시 수반하여야 한다는 점을 들 수 있다. 이와 같은 특징들로 인하여 의료 영상 처리 시스템의 개발은 오류에 쉽게 노출되고, 많은 비용과 시간이 요구되므로 현재 개발된 대부분의 자동화 시스템은 영상 판독의 정확성 및 속도에 있어 시스템 성능이 그리 만족스럽지 못한 상태이다.

이와 같은 문제점을 해결하기 위한 시스템 구현 방식으로서, 자동화 시스템의 성능을 보완하여 전문가의 진료 및 최종 진단을 보조할 수 있는 지능적 의료 정보 시스템[1,4]의 구현 예를 보인다. 구현된 의료정보 시스템에서는 영상의 자동 인식/분석 결과 등과 함께 현재 진단 중인 의료 영상과 유사한 영상객체를 임의로 검색하여, 부가정보로 활용할 수 있는 유사 영상객체 검색 기능을 제공한다. 유사 영상객체 검색에 의한 부가 정보 검색 기능은 사용자의 편의에 따라 점진적으로 제공되며, 전문의나 병리학자들은 진단 과정에서 이미 판독된 유사 영상의 진단 결과 등을 필요에 따라 임의로 참조함으로써 영상의 정확한 판독 및 진단 확증의 객관적 근거 자료를 확보하는데 도움을 받을 수 있다.

유사 영상 객체 검색을 위한 유사도 계산에는 각 객체로부터 얻어진 특정 정보를 사용한다. 객체간의 유사도 계산 방식은 객체를 구성하는 각 속성의 데이터 타입에 의존적이다[4]. 본 시스템에서는 영상 객체 속성간의 관련성, 속성 값의 분포, 빈도 등을 고려한 유사도 계산을 위하여 속성값이 수치인 경우 적용 가능한 DISC, M-DISC 알고리즘[3]을 혼용 사용하였다.

자궁경부 세포진 검사 자동화 시스템[5,6]은 슬라이드로부터 비정상 세포(abnormal cell)라고 여겨지는 부분을 자동 검색하여 보여줌으로써 전문가의 최종 진단을 지원 하는 시스템으로서, 반복되는 진단과정을 자동화하여 수

련된 전문가의 작업량을 줄여줄 뿐 아니라, 객관적 데이터를 제공함으로써 진단의 주관적 판단으로 인한 오진률 감소 등을 목적으로 하는 시스템이다. 본 논문에서는 자궁경부 세포진 인식 시스템[5,6]을 그 적용 예로 하여, 시스템 기능을 설명한다. 우선, 자궁경부 세포진 인식 시스템 개요 및 시스템 구성에 대하여 설명하고, 각 영상 내 세포에 대하여 추출된 특정 값에 근거한 유사객체 식별, 분류 방식에 대하여 논한다. 다음, 유사객체 분류 정보 활용 방식과 사용자 인터페이스 설계 및 구현 방식에 대하여 설명하고, 시스템의 유용성에 대하여 논한다.

2. 자궁경부 세포진 인식 시스템

2.1 자궁경부 세포진 자동 인식

자궁경부 세포진 영상은 슬라이드로부터 현미경을 통해 디지털 카메라로 얻어지는데 자궁 경부 세포 자체가 추출된 부위나 세포의 종류에 따라 형태나 크기 등의 특징들이 많이 다르며 각각이 따로 추출되는 것이 아니라 한 영상에 군집되어 나타난다. 이러한 일정한 유형을 보이지 않는 영상에서 영역을 정확하게 분할하는 문제는 간단한 일이 아니다. 이 외에도 염색약의 종류와 농도, 환자의 나이와 호르몬의 영향, 건강 상태, 생리주기, 심지어는 병리학자의 습관이나 병원 환경에 따라서도 영상이 많이 달라진다. 또한 디지털 카메라로 얻어진 자궁경부 세포진 영상에는 핵, 세포질, 배경, 백혈구, 염증 세포 등의 기본적인 분포 이외에 슬라이드를 만들 때 생긴 기포나 현미경 자체의 먼지 등이 같이 얻어지는 경우도 있어 영상의 처리를 더욱 어렵게 만든다. [5,6]에서는 다양한 실험을 통하여 자궁경부 세포진 영상 인식에 적합한 방법을 개발하고 구현하였다.

2.2 시스템 구성

본 시스템은 그림 1과 같이 시스템 모듈, 데이터베이스, 지식베이스, 사용자 인터페이스로 구성된다. 주요 시스템 모듈의 기능 및 구현 방식은 다음과 같다.

[1] 의료영상 처리 모듈[5,6]

의료영상 처리 모듈은 영상 획득부와 영상 인식부로 구성된다. 영상 획득부에서는 CCD 카메라가 부착된 현미경을 통해 세포가 도말된 슬라이드로부터 저배율(100X) 영상을 획득한다. 이 때 획득된 영상으로부터 ROI(Region of Interest) 검색에 의하여 세포가 존재하는 영상만이 선별되어 영상 인식부로 전달된다. ROI는 전문의의 수작업으로 설정될 수 있으나, 규칙과 지식베이스를 활용한 알고리즘을 이용하여 자동 추출될 수 있

다. 영상 인식부는 영상 획득부로부터 세포가 존재하는 영상을 입력받아 세포영역분할, 특징추출, 인식 등의 단계를 거쳐 영상 안에서 비정상 세포를 검출한다. 세포영역 분할에서는 자궁경부 세포진 영상으로부터 암 판별에 영향을 미치지 않는 염중세포(주로 백혈구), 먼지, 파편(debris)을 우선 제거하고 핵과 세포질 영역을 추출한다. 특징 추출부에서는 분할된 결과를 이용하여 핵 특징과 영상 특징을 추출한다. 인식부에서는 추출된 특징을 이용하여 정상/비정상 여부를 판별한다. 판별을 위하여는 3장에서 보이는 유사객체 분류방식을 사용한다.

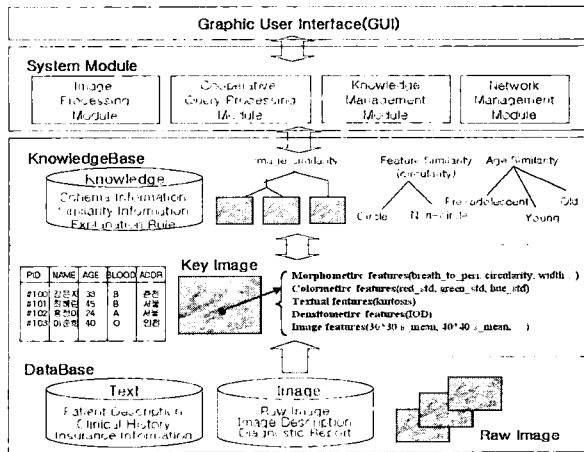


그림 1. 시스템 구성도

[2] 데이터베이스/지식베이스 구성 및 지식 관리 모듈
환자의 병력, 처방 및 약제 정보, 의료영상 정보를 효율적으로 관리, 운영하기 위한 데이터 모델로 객체관계형 데이터모델을 이용한다. 그림 2에 데이터베이스 구조의 일부 예를 보인다. 데이터베이스 내의 각 객체는 환자 이름, 성별, 혈액형, 증상 등과 같이 단순 문자열을 합으로 하는 속성과 키, 몸무게, 나이 등과 같이 숫자 값을 갖는 속성 그리고 검사이미지 등과 같이 OID로 표현된 복합객체를 속성 값으로 하는 속성들로 구성되어 있다. 따라서 객체간 유사도 정보를 얻기 위한 유사도 계산은 객체가 갖는 데이터 타입에 따라 서로 다른 방식을 적용하여야 한다[3,4].

지식관리 모듈은 의료 데이터베이스 내의 각 객체간 계산된 유사도 정보를 내부 클래스 형태의 지식베이스로 저장, 관리하는 모듈이다. 이들 자동 추출된 지식들은 영상인식 과정과 협력질의 과정에서 활용되며 전문가에 의해 편집되어 사용될 수 있다.

[3] 협력질의 처리 모듈
협력질의 처리 모듈은 사용자의 질의 수행 결과나 시스템이 자동으로 판별한 영상의 인식 결과에 대해 그와 관련된 유사 정보를 부가적으로 제공하는 모듈이다. 이는 사용자의 진단 과정에 있어, 진단의 객관적인 근거 자료 및 병명이 확정된 다른 환자와의 사례 비교를 통해 질병의 시간적 경과에 따른 변이를 추정하고 그에 따른 처방을 내릴 수 있도록 보조하는 기능을 갖는다.

[4] 네트워크 관리 모듈
네트워크 관리 모듈은 질의 수행 결과 검색된 환자의 신상, 병력, 처방 정보와 같은 문자 정보 그리고 의료 영상 및 영상 인식의 결과 등을 타 병원 및 병원 내 다른 진료과들로 전송하기 위한 모듈이다. 이때 전송 표준으로 XML을 이용한다[7].

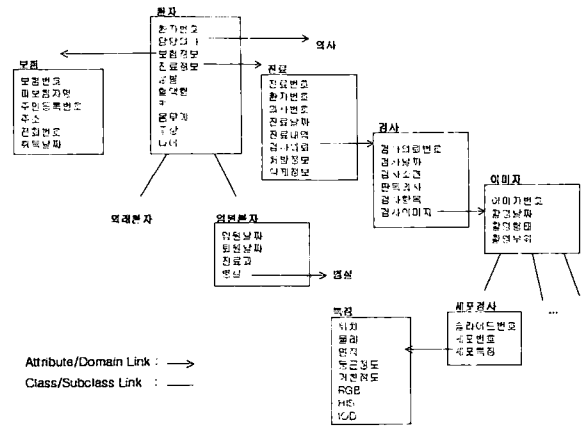


그림 2. 예제 데이터베이스

3. 유사 객체 분류

3.1. 영상 객체간 유사도

객체는 단순 문자열이나 숫자, OID 등을 속성 값으로 하는 다수의 속성들이 복합된 형태를 갖는다. 객체간의 유사도 계산은 데이터 타입에 의존한다. 데이터 타입이 단순 문자열인 속성 값을 근거로 객체를 유사 분류하기 위해서는 이를 제외한 나머지 속성 값을 이용한 유사도 계산 방식을 채택한다. 그러나 객체가 숫자를 속성 값으로 하는 경우 객체간 유사도는 값의 분포와 빈도 등을 고려하여 계산하는 것이 효율적이다[3,4].

이들 방식을 확대, 적용하여 세포 간 유사도 계산에 활용한다. 이를 위해 영상처리 단계에서 자동 추출된 핵의 형태, 색상, 질감, 명암 그리고 핵을 중심으로 부영상을 추출하여 얻어진 세포 영상 특징을 사용한다. 추출된 특징은 전문의와 시스템 설계자의 협의 하에 정보량이 많은 우수한 특징들만 선택되어 사용될 수 있다. 그림 2의 데이터베이스 스키마의 자궁 경부 세포진 영상에서의 세포 특징의 도메인인 특징 클래스에 대한 구체적인 인스턴스 예를 표 1에 보인다. 여기서 객체는 영상(슬라이드 영상) 내에 출현하는 각 세포를 의미하며 특징 추출을 위한 기하학적 계산 방식에 의해 산출된 일련의 숫자들을 속성 값으로 갖는다. 따라서 자궁경부 세포진 영상에서의 세포 간 유사도 계산에는 각 특징 값(숫자)들의 분포와 빈도 등을 고려한 계산 방식이 적용되어야 한다. 객체간 유사도 계산 방식으로 특징을 나타내는 m 개의 속성 중 각각의 독립적인 단일 속성 값을 근거로 유사도를 계산하여 혼합, 적용할 수 있으나, 객체를 구성하는 m 개의 속성들은 상호 의미적 연관성을 가지므로 m 개의 복수 속성 값의 분포, 빈도를 동시에 고려한 유사도 방식을 채택하는 것이 더욱 정확한 정보를 얻을 수 있다. 즉 자궁경부 세포진 영상에서의 세포의 정상/비정상 여부의 판별을 위해서는 단일 특징이 아닌 핵의 크기, 모양, 핵 내부의 염색 정도, 염색질의 거친 정도 등 다수의 특징들을 모두 고려해야 한다.

본 시스템에서는 객체의 유사분류 방식으로서 DISC, MDISC 알고리즘[3]을 사용하였다. 이 방식은 주어진 객체 집합(클러스터) $C = \{t_1, t_2, \dots, t_n\}$ 를 다음 식의 $RE(Relaxation Error)(c)$ 가 최소화되도록 클러스터를 순환적으로 분류해 나가는 방식이다.

$$RE(C) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m P(t_i)P(t_j)D(t_i, t_j) \quad (1)$$

여기서, RE 는 그 클러스터에 포함된 객체를 동일 객체

로 인식하는 경우의 오차 범위를 나타내는 것으로 유사도를 측정하는 기준으로 사용된다. 이 때 $P(t_i)$ 는 분류된 집합에서 m 개의 속성을 갖는 객체 t_i 가 나타날 확률이며, $D(t_i, t_j)$ 는 객체 t_i 와 t_j 의 거리를 나타낸다. 그림 3에 표 1의 자궁경부 세포진 영상에서 암 판별의 주요한 요인이 되는 핵과 핵 영역을 중심으로 자동 추출된 특징 값을 근거로 계산된 세포 간 유사분류 형태를 보인다. 이와 같은 유사분류 정보는 인덱스형태의 지식베이스로 등록, 관리되어 영상인식에 사용된다.

표 1. 특징 클래스의 인스턴스 예

OID	breath_to_peri	circularity	log_of_h_to_w	width
OID100	30.0000	1.3513	-0.0512	9.0000
OID101	70.0000	1.6664	0.0969	16.0000
OID102	102.0000	2.2437	-0.1091	27.0000
OID103	68.0000	1.4837	0.0000	18.0000
OID104	46.0000	1.4392	0.1047	11.0000
OID105	48.0000	1.4437	-0.0669	14.0000
OID106	44.0000	1.3397	-0.0726	13.0000
OID107	68.0000	1.7195	0.0483	17.0000
OID108	50.0000	1.4109	0.1627	11.0000
OID109	68.0000	1.5929	-0.1249	20.0000

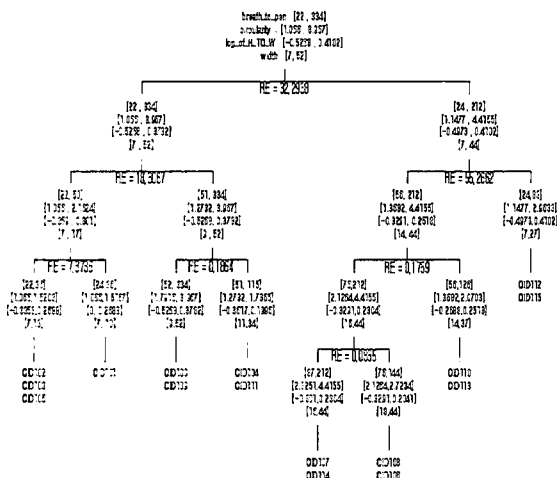


그림 3. 영상 객체간 유사도

3.2. 유사 분류에 의한 인식

자궁경부 세포진 영상에서 세포의 정상/비정상 여부를 판별하기 위한 여러 가지 영상 인식 방법들이 소개되어져 왔다[5,6]. 본 시스템에서는 계산된 객체간 유사도를 기반으로 객체의 정상/비정상 여부를 판단하는 인식 방식을 도입하였다. 영상 인식기는 임의의 세포 객체를 입력받아 각 특징을 추출한 후, 유사 분류 인덱스를 이용하여 대상 객체와 가장 유사한 객체들이 포함된 최근접 클러스터를 검색한다. 그 후 검색된 클러스터내의 정상 객체와 비정상 객체의 출현 빈도를 조사하여 이들 객체의 출현 빈도를 이용하여 세포의 정상/비정상 여부를 판별한다.

4. 사용자 인터페이스

그림 4에 구현된 사용자 인터페이스의 사용 예를 보

인다. 인식 시스템에 의해 처리된 저배율(100X) 영상의 결과와 환자의 병력 및 임상정보를 전문의나 병리학자에게 보여준다. 최종 진단은 전문의가 직접 입력하거나 또는 제공된 형태(악성도를 나타내는 세포의 종류, 유형, 진행여부 등)를 선택하여 입력 할 수 있도록 한다. 이때 결과 영상에서 비정상 판별된 세포에는 해당 위치에 '✓'표시가 된다. 이 표시를 클릭하면 선택된 세포를 중심으로 고배율(400X)로 확대된 영상과 함께 비정상 세포로 판별하게 된 근거 자료인 영상 내에 출현하는 비정상 세포들에 대한 특징 정보 및 계산된 유사도, 판별에 대한 확신도를 제시한다. 또한 선택된 세포를 토대로 같은 유형으로 유사 분류된 세포들에 대한 영상과 특징 정보를 함께 제공하여 다른 환자들과의 사례를 비교할 수 있도록 한다. 표시된 영상에 대해서는 기본적으로 Zoom In/Out기능, 기하학적 계산 기능, 메모 기능, 포맷 변환 기능 등이 제공된다.

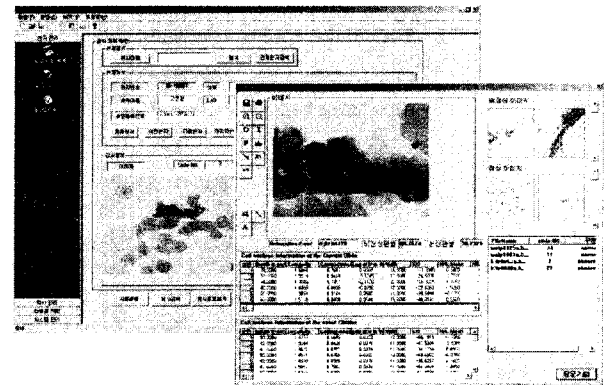


그림 4. 구현된 시스템 예

5. 결론

본 논문에서는 자궁경부 세포진 영상을 예로 전문의의 진단 과정을 보조할 수 있는 유사 영상 객체 검색 기능을 갖는 지능적 의료 정보 시스템 설계 및 구현에 대하여 논하였다. Windows 2000상에서 의료영상 처리기, 협력질의 처리기, 지식 관리기, 네트워크 관리기 등의 모듈별 프로그램을 Visual C++를 사용하여 개발하였다.

참고 문헌

[1] Chu, W.W., Cardenas, A.F., Taira, R.K., "KMED : A Knowledge-Based Multimedia Medical Distributed Database System", Information Systems, Vol.19, No.4, pp. 33-55, 1994.
 [2] Motro,A., "FLEX : A Tolerant and Cooperative User Interface to Databases", IEEE Trans. on Knowledge and Data Engineering, Vol. 2, No. 2, pp. 231-246, 1990.
 [3] Chu, W. W., Chiang, K., Hsu, C. C., Yau, H., "An Error-based Conceptual Clustering Method for Providing Approximate Query Answers", Communication of ACM, 1996.
 [4] 원정임, 박형주, 안상원, 윤지희, "유사 객체 검색을 지원하는 협력 의료 정보 시스템 설계", 춘계 정보과학회 학술대회 논문집, pp. 119-121, 2000.
 [5] 광치경, 김백섭, 김호성, "형태학적 연산을 이용한 자궁 경부 세포진 영상의 핵 영역 분할", 24회 대한의생명체공학회 춘계 학술대회, pp. 58-59, 2001.
 [6] 송선아, 송해정, 김백섭, "자궁 경부 세포진 검사 시스템의 구현에 관한 연구", 한국정보과학회 컴퓨터비전 및 패턴인식 연구회, pp. 209-210, 2001.
 [7] 양정욱, 홍동완, 윤지희, 주한규, "HMS을 기반으로 한 웹상의 병원정보 통합 및 검색", 추계 정보과학회 학술대회 논문집, pp. 76-79, 2001.