XML 기반 멀티미디어 검색시스템

윤미희 (충북과학대학)·최동선 (우송공업대학)·최병갑 (목원대학교)

차 례

Ⅰ. 서론
Ⅱ. 메타데이터
Ⅲ. 메타데이터 모델링을 위한 기법
Ⅳ. 메타데이터 모델링
Ⅳ. XML 모델링
Ⅳ. 멀티미디어 검색 시스템

Ⅰ. 서론

인터넷의 보급과 더불어 대용량 저장장치의 가격이 하락하면서 멀티미디어 데이터의 저장과 공유가 보편화되었다. 또한 멀티미디어 데이터를 활용하는 지리정보시스템, 멀티미디어 의료 정보시스템, 전자신문, 전자도서관, 홈쇼핑, VOD등의 새로운 응용 분야들이 각광을 받고 있다. 이러한 환경적인 요인은 대용량의 멀티미디어 데이터를 효율적으로 추출, 저장, 검색하기 위한 멀티미디어 데이터 처리기술이 요구된다.[1,2]

비디오는 멀티미디어 데이터의 가장 대표적인 형태로, 텍스트나, 이미지, 오디오와 객체의 움직임 같은 풍부한 정보를 담고 있다. 이러한 멀티미디어의 다양한 정보를 조직적으로 기술하기 위한 정의 언어와 스키마가 필요함에 따라 MPEG(Motion Picture Expert Group)은 멀티미디어 정보를 기술하고 다양한 도구들을 제공하는 MPEG-7[3]을 발표하였다. 비정형의 멀티미디어 데이터를 다양하고 효율적으로 표현하기 위해서는 XML(eXtensible Markup Language)[4]을 사용하여 저장하고 검색하는 멀티미디어 검색시스템이 필수적이다.

일반적으로 XML 문서의 검색은 요소(Element)단위의 구조검색으로 이루어지며 검색결과의 현태는 조건에 맞는 요소들의 집합으로 표현된다. XML 문서의 특성을 고려한 다양한 구조 검색 방법이 이루지고 있지만 문서 구조에 대한 사전지식이 필요하다. 그러므로 멀티미디어 데이터에 대한 검색을 위해서는 멀티미디어 데이터의 내용을 구조적으로 설명하는 메타데이터가 필요하고 이 메타데이터를 XML을 사용하여 표현하며 저장하고 검색하기 위한 멀티미디어 검색시스

템이 요구된다. 본 논문의 구성은 2장에서는 메타데이터를 구축하기 위한 방법들을 살펴보고 3장에서는 구성된 메타데이터에 대한 모델링 방법을 설명하고 4장에서는 메타데이터에 모델링 기법으로 UML을 이용한 모델링 기법을 설명하고 5장에서는 4장에서 설명한 메타데이터 모델을 XML 모델로 표현하고 6장에서는 XML 기반의 멀티미디어 검색 시스템을 설명한다.

II. 메타데이터

메타데이터의 구성은 컴퓨터에 의해 비디오의 특징들을 자동적으로 추출하여 데이터베이스를 구축하는 특징기반 색인기법과 사람에 의해 입력하여 데이터베이스를 구축하는 주석기반 색인으로 분류할 수 있다. 이미지의 경우 색상 질감, 형태 위치 등 시각적인 정보를 히스토그램이나 윤곽선 추출(edge detection) 알고리즘을 이용하여 사람의 주관적인 판단에 따라 객체의 특성을 텍스트로 기술하는 방식으로 메타데이터를 기술하였다. 히스토그램이나 윤곽선 검출을 이용한 특징기반 검색은 비용이 많이 들고 정확도가 낮은 단점이 있고, 이미지를 보고 사람이 주관적으로 텍스트로 기술하는 방식의 주석기반 검색은 데이터가 커지고 종류가 다양해짐에 다라 주석 처리에 소용되는 시간이 많이 걸리고 사람의 주관적인 판단에 따라 서술됨으로써 객관성의 문제가 있다.

비디오의 경우 프레임, 샷, 장면 시컨스등의 계층구조로 표현하며, 이를 위해 비디오 파싱을 통해 자동적으로 카메라의 샷의 분할점을 찾아내야한다. 즉 하나의 카메라 샷에서 그 다음 샷으로 전이되는 샷전이 검출해야한다. 그러나, 카메라 기법이나 장면의 변환이 거의 없는 경우 이러한 샷전이를 이용하여 샷의 경계를 검출하는 것은 상당히 어렵다. 또한 단일 스토리를 구성하는 경우 장면의 경계를 검출하는 것은 더욱 어려운 문제이다. 많은 연구에서 비디오 프레임에 대한 색상 히스토그램과 대표 색상을 기반으로 장면과 샷의 경계 검출이 이루어져 왔지만 최근에는 영상 정보뿐만 아니라 오디오 정보를 이용하여 더 효율적으로 경계를 검출할 수 있는 기법들이 연구되고 있다.

그러나 비디오의 파싱은 여전히 난해한 과제이며 연구되어야 할 과제이다. 자동으로 비디오에 대한 정보를 추출하기 위한 연구가 활발히 진행되고는 있지만 비디오 데이터에 대해 빠르고 효율적으로 의미 정보를 해석하는 것은 사람이며, 사람에 의한 주석기반 메타데이터 구축은 비디오 파싱에서 필수적이다.

Ⅲ. 메타데이터 모델링을 위한 기법

멀티미디어 데이터의 대표적인 형태인 이미지와 비디오 데이터는 본질적으로 특성이 다르기 때문에 각각의 특성에 맞게 메타데이터를 구축하여야한다.

이미지를 위한 메타데이터의 XML 스키마는 그림 1과 같이 나타낼 수 있다. 이미지는 비디오 데이터의 샷으로 매핑될 수 있다.

이미지 데이터는 complex type으로 이미지 번호인 id와 이미지의 이름 이미지에 나타난 객체, 이미지의 전체 색상, 이미지가 저장된 MPEG화일의 위치정보를 포함하고 있다. 객

체는 complex type으로 객체의 고유 번호인 id와 객체의 이름, 색상, 이미지에서의 위치 정보를 포함하고 있다.

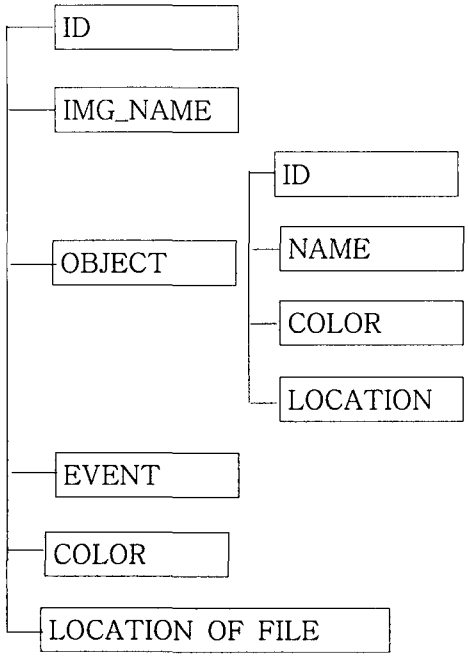

그림 1. 이미지 데이터를 위한 XML스키마

비디오다큐먼트는 필름 이론(film theory) [5]에 의해 시퀀스(sequence), 장면(scene), 화면(shot), 프레임(frame)으로 구성된다. 그러므로 비디오는 계층적 구조의 트리 형태로 표현할 수 있고 그림 2와 같다.하나의 비디오는 각각의 노드에 시퀀스 장면, 화면에 대한 정보를 가질 수 있고 이러한 정보에는 물리적 정보와 내용 정보를 갖는다. 물리정보란 시작과 끝 프레임, 장면에서의 대표 화면, 화면에서의 대표 프레임을 포함하고, 내용정보에는 영상정보나 오디오 정보에서 추출할 수 있는 내용으로 파싱에 의해 자동적으로 추출될 수 있는 정보와 사람에 의해 주관적으로 기술될 수 있는 모든 정보를 포함한다. 이러한 정보는 대체로 MPEG-7을 기반으로 서술된 정보로 MPEG-7 표준 문서에서 컨텐츠에 대한 설명은 이벤트, 객체, 개념(concept), 장소, 시간, 요약으로 정의하고 있다.

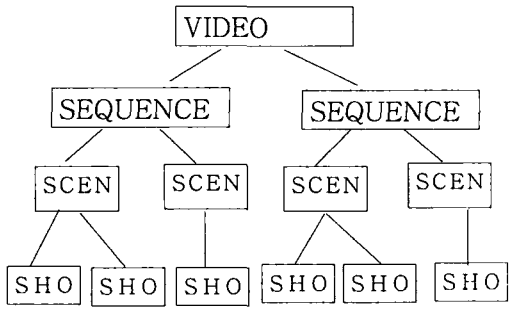

그림 2. 비디오의 계층적 구조

계층적 구조를 가진 비디오 메타데이터는 물리적인 정보를 나타내는 타입과 내용의 정보를 나타내는 두개의 타입으로 구성될 수 있으며 이 두가지 타입은 시퀀스, 신, 샷을 구성하는 단위이다. 물리적인 정보는 자신의 고유번호와 비디오의 시작과 끝 프레임 번호 키 프레임 번호를 가지면 내용의 정보는 내용 독립적인 정보와 내용 기반 정보로 구성될 수 있다. 내용 독립적인 정보는 비디오의 제목이나, 제작사 방송년월일등과 같은 내용에 독립적인 정보이고, "누가 언제 어디서 무엇을 어떻게"와 같은 비디오의 의미를 설명하는 내용기반 정보들을 갖는다.

비디오는 샷의 경계 검출과 자막데이터를 기반으로 세그먼테이션되며, 이미지는 자동적으로 컴퓨터에 의해 특성을 추출하여 그 특성을 기반으로 주석을 처리한다. 자막데이터의 단어는 스토리 세그먼테이션에 의해 처리되어 추출된다. 자막데이터는 영화나, 뉴

스, 스포츠에서 사용되며, 내용뿐만 아니라 샷의 경계 추출이나, 장면의 경계 추출에도 유용하게 활용될 수 있는 정보를 포함하고 있다. 그러나 대부분의 경우 비디오 데이터의 내용 추출을 하기에는 충분하지 않으므로 사람에 의한 주석처리는 반드시 필요하다.

비디오의 주석처리는 사람에 의해 샷이나 장면에 나오는 객체, 시간, 사건, 장소와 같은 의미 있는 정보를 기술한다. 주석을 다는 사람은 의미를 추출함과 동시에 세그먼트를 구분하고, 설명과 키워드를 삽입한다. 이러한 주석정보는 XML데이터베이스에 정장된다. 그러나 정보의 양이 증가함에 따라 검색에 소요되는 시간이 증가될 뿐 아니라 주석을 다는 사람의 주관적인 정보가 늘어나 검색 효율이 떨어질 가능성이 높아진다. 그러므로 이러한 문제를 해결하기 위한 적절한 기법이 포함되어야 검색의 효율을 높일 수 있다.

그림 3은 멀티미디어 데이터를 처리하기 위한 방법으로 이미지 데이터의 특징 추출과 비디오 데이터의 특징 추출과 사람에 의한 주석 처리에 의해 XML베타데이터베이스를 구성하는 단계를 보여주고 있다. 파싱단계에서는 비디오 스트림을 일반적인 샷으로 분할하는 일을 수행한다. 이들 샷은 데터베이스 내에서 인덱싱되어지는 가장 작은 단위이다. 인덱싱 단계에서는 각각의 샷에 내용(content)을 다는 일을 수행하는데 여가에서의 내용은 이미지가 가지고 있는 의미를 분류화(classification)할 수 있는 지식에 근거한 정보를 포함해야한다. 이런 과정을 통해 얻은 특징과 사람에 의해 미디오의 의미를 설명하는 정보들이 XML데이터베이스에 저장되어 검색에 활용될 수 있다. 이미지 처리에서는 컬러히스토그램등과 같은 기법을 이용하여 이미지의 특징을 추출하게되며 자동적으로 검출된 특징들은 비디오데이터와 같은 인덱싱 단계와 주석을 다는 단계를 거쳐 정보를 표현한다.

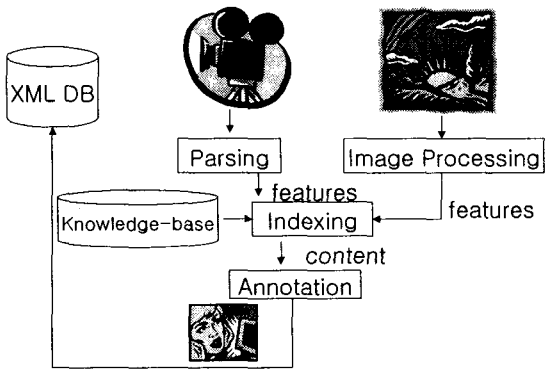

그림 3. 멀티미디어 메타데이터 처리

Ⅳ. 메타데이터 모델링

UML은 소프트웨어 시스템의 구조물을 명세화 구조화 시각화 문서화할 수 있는 언어이다. 또한 XML 프로그램이나 XML을 이용한 B2B 시스템 구축과 같은 XML 프로젝트에서는 객체지향적 설계언어인 UML기법이 효율적이다. UML을 이용해 만든 메타데이터는 쉽게 XML 스키마로 변환할 수 있다.

UML의 정적 구조도는 시스템내의객체, 개체관의 관계, 속성 연산을 보임으로써 시스템의 정적 구조를 기술하는 다이어그램의 전반을 지칭하여 내부에 클래스(class), 인터페이스(interface), 관계(relationship)등의 요소로 구성되어 있다. 객체는 하나의 개념으로 추상 또는 문제의 정의에서서로 구분되는 것

이다. 클래스는 유사한 성질, 공통적인 행위, 다른 객체에 대한 공통적인 관계와 의미를 가지는 객체의 집합을 의미하며, 직사각형으로 표기하고 직사각형안의 영역을 세부분으로 나누어 가장 윗부분은 클래스의 이름을 중간 부분은 속성을 하단은 연산을 기입한다. 정적 구조도는 추상적인 모델링과 실질적인 프로그램의 성계에 유용하게 사용된다. 속성은 한 클래스의 인스턴스인 객체가 가져야 할 데이터 값을 의미하며, 연산은 클래스 내의 객체들의 행위를 나타낸다.

관계는 객체의 인스턴스들 사이의 개념적이거나 물리적인 연결을 의미하며, Dependency, Association, Generalization, Refinement로 구성되고, Association은 다시 Composition과 Aggregation으로 나뉜다. Composition은 Aggregation의 한 형태로 단순한 Aggregation이 경우 부분이 여러개의 전체에 공유될 수 있는 반면 Composition은 전체에 대한 부분이 강한 소속감을 가지고 동일한 생명기간을 가질 때를 나타내며 Generalization은 일반적인 것과 상세화된 것 사이의 관계를 나타낼 때 사용하며, 상속의 의미를 나타낸다. 즉 일반적인 것은 부모가 되고 상세화된 것은 자식이 된다. 그러므로 상세화 된 것은 부모의 모든 속성과 연산을 가진다. Generalization(IS-A관계)은 개념적인 모델링이나 구현에 유용한 구조를 제공하고 상속성은 일반성 관계를 이용하여 공유되는 속성과 동작을 나타낸다.

UML기법을 이용한 비디오메디터의 객체들 사이의 관계는 그림 4와 같다. 물리적인 프레임 스트림을 나타내는 비디오는 원시 데이터를 공유할 수 있으며 메타데이터도 재사용할 수 있다.

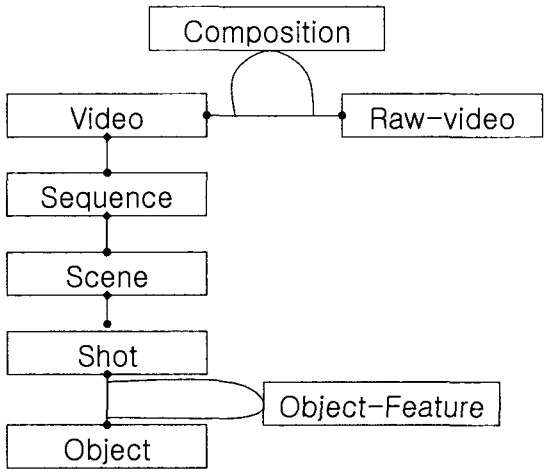

그림 4. 메타데이터 모델링

Raw-Video는 디지털형식으로 저장소에 저장되어 있는 비디오 스트림을 표현하는 객체로 비디오 형식(MPEG, AVI), 사영시간의 물리적 속성으로 정의된다. Video는 하나의 제목을 가지는 한편의 비디오로 하나이상의 시퀀스로 구성되며 비디오를 구성하는 물리적인 비디오 스트림은 원시비디오 객체와 "Composition"관계를 갖는다.

Composition 객체는 Video 객체와 Raw-Video 객체 사이의 합성관계를 나타내며 한편의 Video는 하나이상의 Raw-Video에 의해 생성되고 Raw-Video는 여러 Video에 의해 공유될 수 있다. Sequence 객체는 비디오 내에서 연결 가능한 이야기, 또는 완전한 기사나 비디오에 나타난 의미 있는 사건을 나타낸다. 대부분의 경우 Sequence 객체는 사건을 기반으로 하나의 사건이 시퀀스로 구성된다. 시퀀스는 비디오 객체의 속성을 상속받고 하나이상의 Scene으로 구성된

다. Scene 객체는 동일시간 동일 장소에서 발생한 내용으로 구성된 데이터이다. Scene은 여러개의 Shot으로 구성될 수 있다. Shot은 처리의 기본단위로서 Scene을 구성하는 이미지이다. Object 객체는 한 Shot에 나타나는 관심 있는 Object를 표현한다. Object는 사람, 사물, 배경등 한 Shot 내에서 사용자가 관심 있는 대상을 의미한다. Object- feature 객체는 Shot과 Object사이의 관계를 표현하기 위한 객체로 shot에 나타나는 관심 있는 객체의 색상, 시간적 관계, 공간적 관계의 특징들을 표현한다. 이것은 영상처리에 의해 자동으로 추출될 수 있다. 정밀도에 따라 다르지만 하나의 이미지에서 하아의 특징을 추출하기 위해서 1초가 소요된다고 할때, 질의처리를 위한 시간은 데이터베이스에 100개의 이미지가 있으면 약 17분 이상의 시간이 요구된다.

V. XML 모델링

XML 스키마의 표기법은 제안한 업체마다 조금씩 다르기는 하지만 데이터의 형식 및 반복 횟수의 지정 등 그 자체가 XML 구조로 되어 있다는 것은 공통점이다[6].

객체들 사이의 관계는 4가지(복합, 집합, 상속, 연관)의 관계성을 가지고 있다. 이것은 XML 스키마에 의해 표현될 수 있다[7].
복합관계는 하나의 컨테이너 안에 한 객체의 단순 중첩을 말한다. XML에서 이것은 하나의 element 자식과 같이 element 또는 attribute를 정의하는 것으로 변환한다. 집합관계는 element를 포함하는 복합관계의 내부에 독립적으로 존재하는 객체를 정의한다. 상속관계는 오직 Type에 적용한다. XML 스키마에서 상속 구조는 매우 간단하다고 할 수 있다. XML 스키마는 2개의 기본적인 형식이 있는데 simple type과 complex type이다. 연관관계는 2개의 객체 또는 속성들이 서로 상호간에 관련된 곳에서 이들 두개의 항목들의 함께 묶도록 XML instance document안에 링크를 생성하는 것이 가능하다. 이것은 집합관계에서 보여준 동일한 방법을 사용하며 관련된 항목들에 관개서는 key/keyref를 pair로 생성해서 처리한다.

3장에서 제시한 UML객체를 이용한 XML 모델링 과정은 그림 5와 같다.

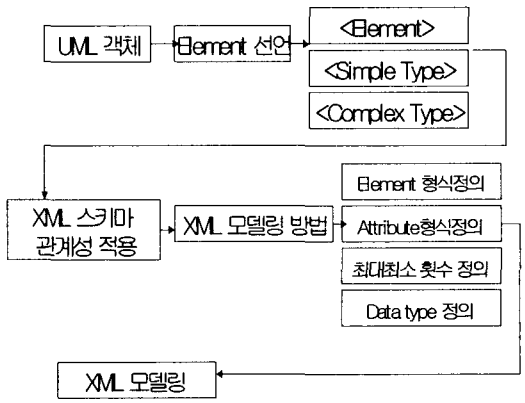

그림 5. XML 모델링

VI. 멀티미디어 검색 시스템

멀티미디어 검색 시스템은 생성된 메타데이터의 XML 데이터베이스를 이용하여 사용자가 원하는 이미지나 비디오의 장면 또는 샷을 검색한다. 검색은 웹서버와 XML 엔진 비디오 엔진, 데이터베이스가 담당한다. XML 엔진은 XSLT 프로세서, 파일시스템,

DOM, XML 데이터베이스를 위한 API등으로 구성된다. XSLT 프로세서는 XML 문서를 HTML 문서로 변환할때 사용되고, 저장된 XML 문서의 갱신이나 질의 및 XSLT 컴포넌트는 W3C XML 표준 인터페이스를 이용한다[8].

사용자의 질의는 XML 문서에 스타일 언어인 XSLT를 적용시켜 폼을 구성하여 사용자의 입력을 받고 폼을 통해 입력 받은 값은 Xquery를 사용하여 XML 데이터베이스에 대한 질의를 통하여 결과 값을 산출한다. 산출된 결과 값은 사용자가 원하는 형태로 보여주기 위해 적당한 XSLT 문서를 적용하여 HTML 문서로 만들어져 미디어 플레이어나 웹브라우저를 통해 브라우징된다. 비디오의 경우 계층적인 구조로 구성되었기 때문에 시퀀스, 장면, 샷, 프레임기반으로 검색하고 추출할 수 있다. 멀티미디어 검색시스템을 위한 구조도는 그림 6과 같다. XML DB는 컴퓨터에 의해 자동 추출된 이미지와 비디오의 특징과 사람에 의해 기술된 주석에 의해 전처리 과정에서 생성된 데이테베이스이다. 사용자 질의가 입력되면 XML 엔진이 XML 데이터베이스와 연동하여 해당하는 이미지나 비디오를 추출하고 추출된 결과를 사용자의 웹브라우저나 동영상 플레이어를 이용하여 보여주게된다. 비디오 데이터베이스와 이미지데이터베이스는 원시 데이터가 저장된 곳으로 추출된 결과에 따라 해당 이미지가 있는 파일이나, 비디오의 해당 프레임을 보여주기 위해 엑세스하고 그 결과를 미디어 처리과정을 거쳐 웹서버를 통해 사용자의 컴퓨터에 보여주게된다..

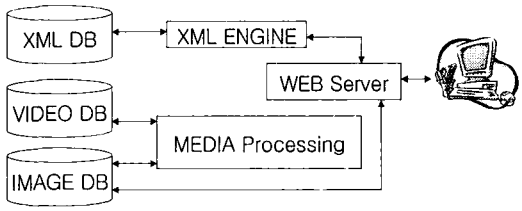

그림 6. 멀티미디어 검색시스템의 구조

효율적인 멀티미디어 검색시스템을 위해서는 자동화된 멀티미디어 분석기법과 효율적적인 검색을 위한 인덱싱 기법과 자동적인 지식 발견에 의한 지식베이스 구성 기법을 개발하는 것이 요구된다. 끝으로 본 연구는 2003년도 IT학과 장비지원사업에 의한 연구결과임을 부기하는바이다.

참고문헌

[1] R.Hjesvold and R.Midstraum, "Modeling and Querying sequence Data", In Proc of the 20th VLDB conference, Sandiago, Chile, Sep. 1994.

[2] P. Senthil Kumar, G. Phanendra Babu, "Intelligent multimedia data : data indices + interface", Multimedia systems, Vol 6, P. 395-407, 1998.

[3] Moving Picture Experts Group, "Overview of MPEG-7 standara(Version 6.0)," ISO/IEC JTIC1/SC29/WG11 N4509, Pattaya, Thailand, December, 2001.

[4] "Extensible Markup Language (XML) 1.0", World Wide Web Consortium Recommendation. (http://www.w3.org/TR/REC-xml)

[5] J. Monaco, "How to Read a Film. The Art, Technology, Language, History and Theory of Film and Media", Oxford University Press,

1981.

[6] 김채미, 최학열, 김심석 공저, "전문가와 함께가는 XML Camp", 마이트 Press, 2001.

[7] Duckett Jon, Ozu Nik, williams Kevin, Mohr Stephen, Stephen, Carl Jurt, Griffin Oliver, Norton Francis, Stokes Rees Ian, and Tennison Jeni, Professional XML Schemas, Wrox Pr Inc, 2001.

[8] 남윤영, 황인준, " XMARS : XML 기반 멀티미디어 주석 및 검색 시스템", 정보처리학회 논문지 B 제9-B권 제5호, 2002.10.

저자소개

● 윤미희

1990년 : 숙명여대 전산학과 (이학사)
1992년 : 숙명여대 대학원 전산학과 (이학석사)
1999년 : 숙명여대 대학원 전산학과 (이학박사)
1998~1999년 : 숙명여대 자연과학연구소 선임연구원
2000~현 재 : 도립 충북과학대학 컴퓨터정보과학과 교수
<관심분야> : 멀티미디어, XML, 영상처리, 지식베이스

● 최동선

1972년 : 항공대학교 항공전자공학과 (공학사)
1980년 : 숭실대학교 대학원 전자공학과 (공학석사)
1997년 : 청주대학교 대학원 전자공학과 (공학박사)
현재 : 우송공업대학 디지털전자정보계열 교수
<관심분야> : 영상공학, 디지털 지적재산권 보호, 디지털 콘텐츠 공학, 영상 처리 및 이해

● 최병갑

1976년 : 한양대학교 전자공학과 (공학사)
1982년 : 한양대학교 대학원 전자계산학과 (공학석사)
1994년 : 청주대학교 대학원 전자공학과 (공학박사)
현재 : 목원대학교 컴퓨터공학과 교수
<관심분야> : 영상공학, 데이터 베이스, 멀티미디어 데이터 베이스, 디지털 콘텐츠 공학