

# XMARS: XML 기반 멀티미디어 주석 및 검색 시스템

남윤영 · 황인준  
아주대학교 정보통신전문대학원 정보통신공학과

## XMARS: XML-based Multimedia Annotation And Retrieval System

Yunyoung Nam °, Eenjun Hwang  
Graduate School of Information and Communication, Ajou University  
E-mail : {youngman, ehwang}@ajou.ac.kr

### 요약

본 논문에서는 XML을 이용하여 멀티미디어 데이터를 구조적으로 표현하고 효율적으로 추출하기 위한 XML 기반 멀티미디어 주석 및 검색 시스템을 제안한다. 이 시스템은 멀티미디어 정보를 표현하기 위해 계층적 메타데이터 모델을 기반으로 하여 구현되었으며, 멀티미디어 데이터에 대한 주석, 검색, 브라우징 인터페이스를 제공한다. 멀티미디어에 대한 메타데이터는 MPEG-7에 정의되어 있는 서술 스키마를 기반으로 XML 스키마를 이용하여 작성하였다. 또한, 멀티미디어 데이터의 효율적인 인덱싱과 추출을 위하여 자막과 주석을 바탕으로 한 카테고라이징 기법을 사용하였다. 본 시스템의 목적은 멀티미디어 데이터에 대해 주석을 처리하고, 다양한 방법으로 검색과 그 결과를 브라우징하는 데 있다.

### 1. 서론

인터넷 보급의 확산과 대용량 저장 장치 가격의 하락은 많은 사람들이 멀티미디어 데이터를 저장하고 공유하는 것을 가능하게 만들었다. 또한 네트워크의 고속화와 멀티미디어의 압축기술이 발달로 인하여 멀티미디어의 사용이 늘어나게 되었다. 이러한 멀티미디어 정보의 사용량 증가는 VOD(Video-On-Demand)와 전자 도서관과 같은 서비스를 등장 시켰으나 많은 양의 멀티미디어 데이터를 효율적으로 조직화하고 추출하는 것이 필요하게 되었다. 멀티미디어 데이터 중 비디오 데이터는 가장 대표적인 원천 데이터이며, 텍스트, 이미지, 오디오, 객체 움직임 같은 풍부한 정보를 담고 있지만, 다양한 정보를 조직적으로 기술하는 것은 어렵다. 이러한 점을 해결하기 위해서 멀티미디어 정보를 조직화하는 정의 언어와 스키마가 필요하게 되었다. 이와 같은 요구에 따라 MPEG(Motion Picture Expert Group)은 멀티미디어 정보를 기술하고 풍부한 도구들을 제공하는 MPEG-7[1]을 개발하였다. MPEG-7은 주 요소인 Descriptors(D)와 Description Schemes(DS)를 이용하여 의미와 syntax를 정의하고 요소들간의 관계 의미와 구조를 정의하며, 멀티미디어 데이터를 다양하고 효율적으로 표현하기 위해 XML(eXtensible Markup Language)을 사용하였다. 또한, XML 스키마를 기반으로 하여

MPEG-7 서술 도구의 syntax를 정의하였다. 결과적으로 MPEG-7은 멀티미디어 데이터를 서술하기 위한 유동적이고 확장 가능한 프레임워크를 제공하고 있다.

이러한 MPEG-7을 기반으로 XML을 사용하여 멀티미디어 데이터에 대해 주석 처리하여 저장하고 검색하는 시스템을 구현하였다. 멀티미디어 데이터는 이미지와 비디오 데이터를 이용하였으며, 주석 처리하기 전에 이미지 데이터는 색상 히스토그램으로 분석하였고, 비디오 데이터는 샷 검출과 자막을 이용하여 비디오 분석을 하고 이를 바탕으로 주석을 XML 형태로 저장하는 방식을 사용하였다. 또한 메타데이터의 크기가 증가하여 검색하는데 소요되는 시간을 줄이고 유사도가 높은 데이터를 검색할 수 있도록 카테고라이징 기법을 사용하였다. 유연한 사용자 인터페이스를 제공하기 위해 웹을 기반으로 하여 개발되었으며, XML과 계층적 구조를 사용하여 메타데이터의 다양한 정보를 사용자가 쉽게 다룰 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 그 동안 연구 개발된 멀티미디어 검색시스템과 멀티미디어 데이터를 구조화와 관련된 연구들을 조사하였으며, 3장에서는 멀티미디어 데이터 조직화에 수반되는 메타데이터의 모델링 방법과 주석 처리, 카테고라이징 기법을 설명한다. 4장에서는 시스템의 전체적인 구조를 설명하고, 주석 처리 도구와 질의 및 추출 인터페이스를 설명한다. 5장에서는 실험 결과를 보여주며, 마지막으로 6장에서는 결론과 향후 계획에 대해서 논의한다.

## 1. 관련 연구

이미지, 오디오, 비디오 데이터와 같은 멀티미디어 데이터는 다양하고 방대한 양의 의미를 포함하고 있어 효율적인 검색을 지원하기 위해서는 데이터를 기술하는 구조적이고 체계화된 형태의 메타데이터가 요구되었다. 멀티미디어 데이터가 보편화되고 웹 상에서 교류되고 있지만 멀티미디어 데이터에 대한 메타데이터 모델이나 스키마가 서로 다르고 교환의 표준이 정해지지 않아 통합하는데 어려움이 있다. 또한, 여러 가지 멀티미디어 검색 시스템이 제안되고 개발되고 있으나, 멀티미디어 데이터가 내포하고 있는 정보가 복합적이고 다양하기 때문에 시스템 개발에 어려움이 있다. 본 장에서는 멀티미디어 검색과 관련된 기존의 연구를 알아보고, 효율적인 메타데이터를 구축하는 방법들을 설명한다. 2.1절에서는 지금까지 연구된 멀티미디어 검색 시스템에 대해서 살펴보고, 2.2절에서는 멀티미디어 데이터에 대한 메타데이터 모델링에 관한 관련 연구에 대해 알아보도록 한다.

### 1.1. 기존의 멀티미디어 검색 시스템

최근까지, 멀티미디어 데이터베이스와 관련된 여러 가지 프로토타입 시스템이 제안되었으며 구현되었다. 멀티미디어 데이터의 검색은 크게 내용 기반 검색과 주석 기반 검색으로 나눌 수 있다. 내용 기반 검색은 객체의 색상, 질감, 형태, 위치, 동작 등 시각적인 정보와 청각적인 정보를 이용하여 이루어지며, 주석 기반 검색은 데이터의 의미나 내용 등에 대해 서술한 정보를 바탕으로 이루어진다.

우선, UC Berkeley에서 개발된 Chabot[2] 프로젝트는 POSTGRES라는 데이터베이스 시스템에 B트리와 R트리 기법을 이용하여 단순 속성, 객체에 대한 설명 정보를 이용한 텍스트 기반 검색과 색상 분석 기술을 이용한 내용 기반 이미지 검색 시스템이다.

Columbia 대학에서 개발된 VisualSEEK[3] 시스템은 색상과 공간 질의를 제공하기 위한 이미지 데이터베이스 시스템이다. 이미지의 구분은 색상 히스토그램과 같은 특성에 의하여 이루어지는데, 이미지 비교를 위해서 이미지의 영역과 색상, 크기, 공간적 위치들을 추가적으로 사용하는 방식이다.

QBIC[4]은 예제 이미지를 통한 유사도 질의를 하며 사용자 스케치에 의한 질의와 색상 및 질감 패턴에 대해 질의를 지원하는 시스템이며, 이미지 데이터베이스에 내용 정보를 저장하여 database population을 통해 추출하는 방식이다. QBIC은 이미지뿐만 아니라 비디오 데이터도 지원하는 시스템으로써 샷 검출, 각 샷에 대한 대표 프레임 생성과 객체의 움직임을 이용하여 database population을 구성한다.

SMOOTH Video DB[5]는 비디오 데이터를 데이터베이스화하여 질의, 브라우징, 주석처리 소프트웨어를 제안한 시스템이다. VIDEX모델을 기반으로 하여 데이터베이스를 구성하였으며 이벤트, 객체, 장소의 의미정보를 저장하여 텍스트 기반 질의를 지원한다.

VideoQ[6]은 객체들 간의 특성과 움직임을 정의하기 위해 스케치 드로잉 질의를 지원하는 웹 기반 비

디오 검색 시스템이다. 비디오 객체에 대해 세그먼트하고 추적하는 자동화 알고리즘을 개발하였으며 사용자 질의에 대해 실시간 비디오 편집 기법을 사용한다.

Informedia[7]는 자동화된 오디오-비디오 인덱싱과, 네비게이션, 검색, 추출 시스템이다. 이 시스템은 음성 인식과 이미지와 자연 언어 처리를 제공한다.

Virage[8]는 샷 경계, 키 프레임, 음성과 문자들을 자동적으로 탐색하는 비디오 인덱싱 도구를 지원한다. 또한 비디오 세그먼트에 대해서 주석 정보를 처리할 수 있도록 구현되었다.

### 1.2. 멀티미디어 데이터의 구조화

기존의 연구에서 이미지의 경우 색상, 질감, 형태, 위치 등 시각적인 정보를 히스토그램이나 Edge Detection 알고리즘 등을 사용하여 메타데이터를 구성하거나, 사람의 주관적인 판단에 의해 객체들의 특성을 텍스트로 서술해 놓는 방식으로 메타데이터를 구축하였다. 히스토그램이나 Edge Detection을 사용한 내용 기반 검색은 그 비용이 높고 사람의 판단보다 정확도가 낮은 단점이 있으며, 텍스트로 서술하는 방식의 주석 기반 검색은 데이터의 크기가 크고 종류가 다양화됨에 따라 주석 처리에 소요되는 시간이 증가하는 문제와 사람의 주관적인 판단에 따라 기술된다 는 점에서 서술 내용의 객관성에 문제점이 있었다.

비디오는 프레임, 샷, 장면과 같은 계층적인 구조로 표현할 수 있으며, 이러한 계층적 구조로 나타내기 위해서는 비디오 분석을 통해 그 경계를 검출하는 것이 필요하다. 하지만, 카메라 기법이나 장면의 변환에 거의 없는 경우의 비디오는 샷의 경계를 검출하는 데 있어 상당한 어려움이 있으며, 단일 스토리를 구성하는 장면 경계 검출 경우에는 더욱 어렵다. 많은 연구에서 비디오 프레임에 대한 색상 히스토그램과 대표 색상을 기반으로 장면과 샷의 경계 검출이 이루어져 왔지만, 최근 연구에서는 영상 정보뿐만 아니라 오디오 정보를 검출에 이용할 경우 검출율을 높일 수 있다고 보고하고 있다. 그러나 비디오 분석은 여전히 어렵고 발전시켜야 할 과제로 남아있다. 자동화된 경계 검출 연구가 과거 몇 년 동안 이루어져 왔지만, 비디오 데이터에 대해 쉽고 빠르게 의미 있는 정보를 해석하는 것은 사람이며, 사람에 의한 주석 처리는 비디오 인덱싱에서 필수적이다.

따라서, 주석 기반 검색은 데이터에 대한 일관되고 정확하게 표현하기 어렵다는 문제점이 있어, 색상 히스토그램 등을 이용하여 멀티미디어 데이터로부터 자동으로 추출하여 인덱스를 유지 관리하는 내용 기반의 검색이 병행되어야 하며, 주석 기반 검색이 멀티미디어 데이터에 대한 모든 정보를 제공할 수 없더라도 객체들간의 관계를 일관성 있게 해석할 수 있는 모델을 정립하여 의미 정보를 체계적으로 표현하기 위한 스키마를 구축한다면 효과적인 멀티미디어 검색이 될 수 있다.

## 2. 메타데이터 모델링

본 장에서는 멀티미디어 검색 시스템 구축에 필요

한 메타데이터 모델링과 주석을 처리하는 방법에 대해서 설명한다. 검색에 사용된 멀티미디어 데이터는 이미지와 비디오 데이터이며, 데이터의 본질적인 특성이 상이하기 때문에 각 특성에 맞는 메타데이터를 구축하였다.

3.1절에서는 이미지 메타데이터 구축에 사용된 스키마를 설명하고 3.2절에서는 비디오 영상 정보 분석 방법과 비디오 메타데이터 구축에 사용된 스키마를 설명하고 구조적인 표현 방법에 대해서 알아본다. 3.3절에서는 비디오 데이터에 대한 주석 처리 방법에 대해 알아보고, 끝으로 3.4절에서는 주석 정보와 자막을 이용한 카테고리징 기술에 대해서 설명하도록 하겠다.

## 2.1. 이미지 메타데이터

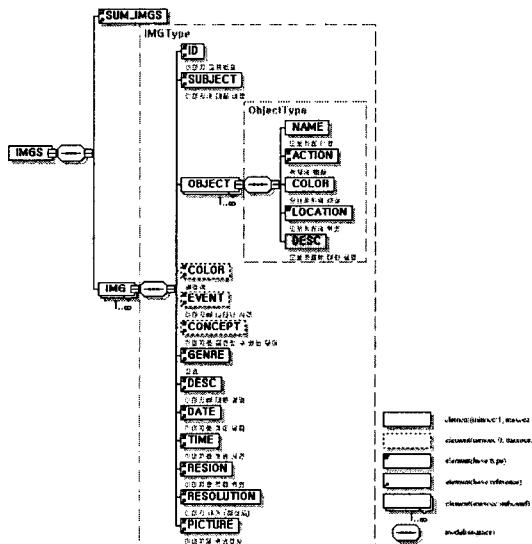


그림 1 이미지 메타데이터를 위한 XML 스키마

그림 1은 이미지 메타데이터를 위한 XML 스키마의 다이어그램이다. 루트 엘리먼트 **IMGS**와 **IMG**라는 엘리먼트를 가지며, **SUM\_IMGS** 엘리먼트는 데이터베이스에 저장되어 있는 이미지의 총 개수를 값으로 갖는다. 각 이미지는 **IMGType**이라는 complex type을 참조하며, 이미지 타입은 이미지의 고유번호, 대표이름, 객체, 객체 색상, 사건, 컨셉, 장르, 서술, 이미지를 활용한 일자·시간·장소, 해상도, 파일의 위치 정보를 포함한다. 객체는 **ObjectType**이라는 complex type을 참조하며 객체 타입은 객체 이름, 동작, 색상, 이미지 상에서 객체 위치, 설명을 포함한다. 색상과 객체의 위치, 장르, 활용 날짜, 시간과 같은 엘리먼트의 값이 한정되어 있는 경우 XML 스키마에서 단어 사용을 제한할 수 있다. 다이어그램에서 상자는 엘리먼트를 뜻하며, 두 개의 상자가 겹친 형태의 엘리먼트는 나타날 수 있는 최고 횟수가 무한대임을 뜻한다. 실선상자는 나타날 수 있는 최소 횟수가 1이며 점선상자는 나타날 수 있는 최

소 횟수가 0을 뜻한다. 상자의 왼쪽 하단에 화살표는 이미 정해진 타입을 참조함을 의미한다. 상자의 왼쪽 상단에 세 개의 줄은 엘리먼트의 형태가 정해진 것이며 줄이 없는 엘리먼트는 정해지지 않은 것을 뜻한다.

## 2.2. 비디오 메타데이터

비디오의 계층적 구조를 이용해 트리로 표현할 수 있으며, 트리로 표현하면, 노드들은 샷과 장면에 대한 정보를 포함시킬 수 있다. 정보에는 물리적인 정보와 의미적인 정보가 포함되는데, 예를 들어, 물리적인 정보 비디오 세그먼트에 대한 시작과 끝 시간 정보나 샷에 대한 대표 프레임이 이에 속하며, 의미적인 정보는 영상 정보와 오디오 정보에서 얻을 수 있는 내용 정보를 말한다. 본 시스템에서는 MPEG-7을 기반으로 서술된 주석 정보를 말하며, MPEG-7 표준 문서에서 컨텐츠에 대한 설명은 이벤트, 객체, 컨셉(concept), 장소, 시간, 요약으로 정의하고 있다.

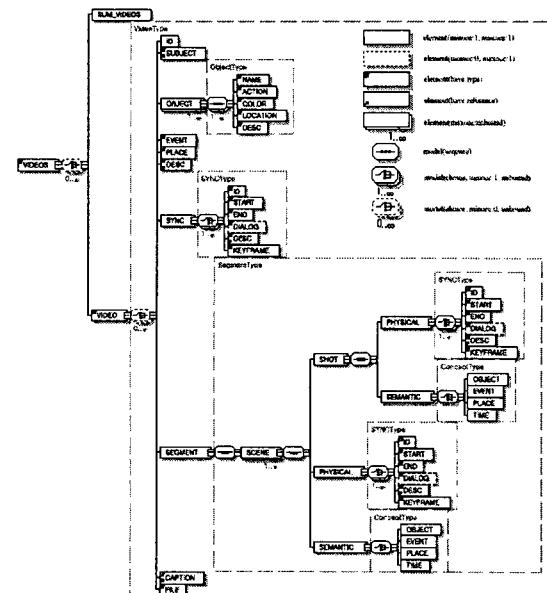


그림 2 비디오 메타데이터를 위한 XML 스키마

그림 2은 비디오 메타데이터를 위한 XML 스키마의 다이어그램이다. 이미지와 동일하게 그림 1에서 보았던 **ObjectType**이 있으며, 뿐만 아니라 물리적인 정보를 나타내는 **SYNCTYPE**과 의미적인 정보를 나타내는 **ConceptType**이라는 complex 타입을 정의한다. 두 가지 타입은 **SCENE** 또는 **SHOT**을 구성하는 단위가 되며, **SCENE**과 **SHOT**을 모아 **SegmentType**이라는 complex 타입을 다시 정의한다.

**SYNCTYPE**은 고유번호와 비디오 세그먼트의 시작과 끝 시간을 밀리세컨드 단위로 값을 갖는 엘리먼트와 객체들 간에 대화 내용, 설명, 키 프레임을 엘리먼트로 갖는다. **ConceptType**은 ‘누가’, ‘언제’, ‘어디서’, ‘무엇’을 예 해당되는 내용을 엘리먼트로 이루어진다.

이러한 물리적인 정보와 의미적인 정보는 장면이나 샷이라는 상위 엘리먼트에 묶여 있다.

### 2.3. 주석

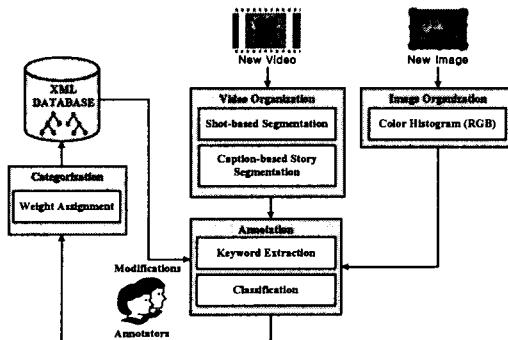


그림 3는 비디오 조직화, 주석, 카테고라이징하는 처리 과정을 보여주고 있다. 우선 비디오는 샷 경계 검출과 자막 데이터를 기반 세그먼테이션 처리 과정을 거치며, 이미지는 색상 히스토그램을 통해 특성을 추출하며 이 특성을 바탕으로 주석을 처리한다. 자막 데이터의 단어는 스토리 세그먼테이션 처리기에서 추출되는데, 자막 데이터는 영화, TV 뉴스, 스포츠 등에서 사용되는 자료이며, 대화 내용뿐만 아니라 화자와 시작 시간, 대화 지속시간 정보를 담고 있다. 그러나 비디오 데이터를 인덱싱하는 데 충분하지 않기 때문에, 미디어의 세그먼테이션은 주석 처리 작업이 필요하다. 세그먼테이션에 대한 주석은 샷 검출과 스토리 세그먼테이션을 기반으로 하여 수행된다.

의미적인 정보의 스키마는 오디오-영상 컨텐츠를 표현하기 위한 이야기식의 서술방법을 나타낸다. 이것은 객체, 사건, 장소, 시간과 같은 의미적인 형태를 서술한다. 주석 정보는 이 모델을 바탕으로 만들어진다. 주석자는 세그먼트에 대해 구분 지을 뿐만 아니라 알맞은 설명과 키워드를 정해야 한다. 이것은 세그먼트와 관련된 주석정보에서의 키워드의 빈도수를 바탕으로 하여 계산하는 카테고라이징을하게 된다. 마지막으로 주석 정보와 빈도수는 XML 데이터베이스에 저장이 된다. 다음절에서는 카테고라이징에 대한 자세히 설명한다.

### 2.4. 카테고라이징

정보의 양이 증가함에 따라서 검색에 필요한 소요 시간은 더욱 증가하게 된다. 이러한 문제를 해결하고 효율적으로 인덱싱과 추출하는데 사용하기 위해서 주석 정보에 대한 카테고라이징 기법을 사용한다.

시스템에서 문서의 총 개수가  $N$ 이라고 하고,  $n_i$ 를 인덱스 단어,  $k_i$ 가 나타나는 문서의 개수라 하고, 문서에서의 단어  $k_i$ 가 빈도수를  $freq_{i,j}$ 라고 한다. 이 때, 문서  $k_i$ 에서 단어  $k_i$ 의 빈도수는 다음과 같이 주어진다.[11]

여기서 최대값은 문서에서 나타난 텍스트들의 모든 단어를 계산한 것이다. 만약 단어  $k_i$ 가 문서에서 나타나지 않으면  $freq_{i,i}=0$ 이다.  $k_i$ 의 빈도수의 역을 구하고 나서 단어의 가중치를 아래와 같이 정의한다.

$$f_{i,j} = \frac{freq_{i,j}}{\max_i freq_{i,j}} \quad (1)$$

$K_{i,j}$ 는 문서의 키워드 가중치이고  $\alpha$ 는 단어의 상대적 중요성이다. 실험에서는  $\alpha$ 는 0.5로 정했다.

### 3. XMARS(XML-based Multimedia Annotation and Retrieval System)

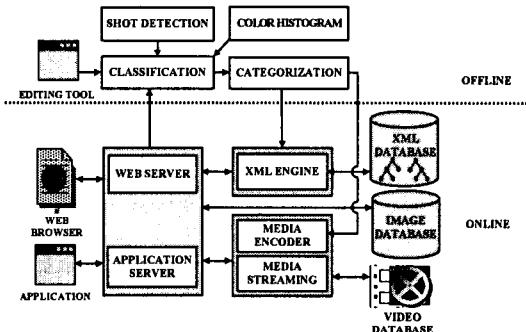


그림 4 XMARS의 구조

그림 4는 XMARS 시스템의 구조를 개략적으로 보여주고 있다. 오프라인에서는 샷 검출과 색상 히스토그램을 통해 비디오와 이미지 분석을 하고 나서 주석 처리와 카테고라이징을 한다. 온라인에서는 세 개의 계층(서버, XML/미디어 엔진, 데이터베이스)으로 구성되어 있다. XML 엔진은 XSLT 프로세서, 파일 시스템, DOM, 업데이트(update), 질의, XML 데이터베이스를 위한 API 등으로 이루어진다. XSLT 프로세서는 XML 문서를 HTML 문서로 변환할 때 사용되고, 업데이트는 저장된 XML 문서를 수정할 때 사용된다. 업데이트, 질의, XSLT 컴포넌트(component)는 W3C XML 표준 인터페이스에 맞춰 구현되었다. 미디어 엔진은 비디오 데이터베이스에 적당한 포맷으로 저장하기 위한 미디어 인코더와 용량이 큰 비디오 데이터에서 원하는 부분만 스트리밍해 주는 미디어 스트리밍으로 이루어져 있다.

본 시스템에서 XML 데이터베이스로는 엑셀론(eXcelon)을 사용하였으며, 동적이고 확장 가능한 서버 구성을 위해 톰캣(Tomcat)과 액티브엑스 컨트롤(ActiveX control)을 사용하였다. 서버와 클라이언트 모듈들은 CORBA 환경에서 서로 통신을 하므로 방화벽(firewall)이 설치되어 있는 네트워크 내에서도 동작이 가능하다.

### 3.1. 전체 흐름도

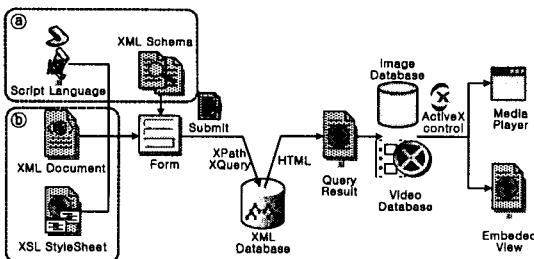


그림 5 웹상에서 XMARS의 프로세스 흐름도

그림 5 은 웹상에서 XMARS의 프로세스 흐름을 보여주고 있다. 크게 멀티미디어 데이터에 대한 주석 정보를 입력하는 품과 질의하는 품으로 구분 짓을 수 있다. ④는 주석 정보를 입력하는 품 구성 시 필요한 요소이며, ⑤는 질의하는 품 구성 시 필요한 요소들이다. ④에서는 XML 스키마에서 엘리먼트와 속성의 이름과 나타나는 횟수를 분석하여 스크립트 언어를 통해 품을 구성하며, 품을 통해 입력받은 데이터는 XML 형태로 변환되어 XML 데이터베이스에 저장된다. 이 때 변환된 XML 문서는 XML 스키마에 대해 유효성 검사(validation check)를 하여 유효한 문서만 저장시킬 수 있으며, 유효하지 않은 문서는 수정하여 다시 유효성 검사를 하여야 한다.

⑤에서는 XML 문서에 스타일언어인 XSLT를 적용시켜 품을 구성하며, 사용자의 입력을 받는다. 품을 통해 입력받은 값은 XPath와 XQuery를 사용하여 XML 데이터베이스에 대한 질의를 통하여 결과 값을 추출한다. 추출된 결과 값은 사용자가 원하는 형태로 보이기 위해 적당한 XSLT 문서를 적용하여 HTML 문서로 만들어진다. 사용자 질의에 의해 생산된 HTML 문서를 통해 사용자가 원하는 비디오의 세그먼트는 미디어 플레이어 또는 웹 브라우저를 통해 브라우징 되며, 이 때 전체 비디오 데이터로부터 추출된 시간정보를 이용하여 부분 재생을 위해 액티브엑스 컨트롤을 사용한다.

### 3.2. 주석 입력 도구

그림 6은 비디오 세그먼트에 대해 주석을 서술할 수 있도록 하는 인터페이스 화면이다. 비디오 데이터는 왼쪽 화면에서 재생이 되고, 오른쪽 상단 화면에서는 샷의 시작과 끝 시간을 테이블 형태로 보여주고 있다. 오른쪽 하단에서는 비디오 장면, 샷에 대한 정보를 추가, 삭제, 수정을 할 수 있으며, 그들은 샷의 리스트로부터 선택하여 비디오의 특정 시간 위치에 대해 풍부한 정보를 서술할 수 있다. 비디오에 대한 주석 입력 단계 후에 비디오 세그먼트의 서술내용은 카테고리화 단계를 거쳐 생성된 요약 문서와 함께 XML 데이터베이스에 저장이 되고, 비디오 스트림은 미디어 인코더를 거쳐 변환되어 비디오 데이터베이스에 저장이 된다.

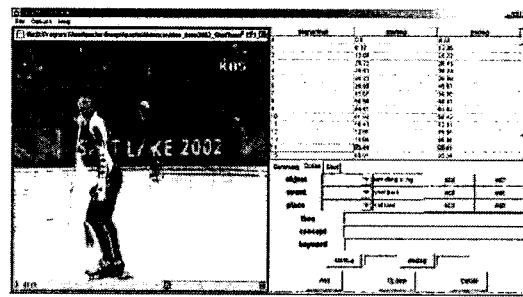


그림 6 비디오를 위한 주석 처리 도구

### 3.3. 질의와 브라우징

사용자가 원하는 이미지나 비디오 세그먼트를 효율적으로 검색 및 추출하기 위해서, 검색 시스템은 다양한 방법으로 질의를 생성하고 브라우징 할 수 있도록 지원되어야 한다. XMARS는 멀티미디어 데이터에 대한 다양한 질의를 가능케 하도록 기본 검색과 고급 검색의 두 가지 모듈을 제공한다. 기본 검색은 카테고리화 단계에서 생성된 문서에 대해 검색하는 것으로, 이것은 트리 형태의 주석 문서에서의 모든 노드들을 액세스하는 것이다. 기본 검색은 사용자가 스키마에 대해 모를 때 효율적이지만, 검색 결과가 낮은 정확률(Precision)과 높은 재현율(Recall)을 발생한다. 고급 검색은 사용자가 스키마에 대해 잘 알고 정확한 검색을 하고자 할 때 사용되는 방법이다. 이와 같은 검색은 사용자의 입력 값과 주석의 엘리먼트의 값을 XPath와 XQuery를 이용하여 비교하는 방법으로 수행한다.

그림 7은 XMARS의 질의와 브라우징 인터페이스를 보여주고 있다. 이미지와 비디오 데이터의 이질적인 특성과 분산 처리 환경을 위해 이미지 검색 시스템인 XIMG와 비디오 검색 시스템인 XVIDEO로 나누어 개발하였다. ④, ⑤는 XIMG 인터페이스의 화면이며, ⑥, ⑦는 XVIDEO의 인터페이스 화면이다. 질의는 왼쪽 프레임에서 생성하고 질의 결과는 오른쪽 프레임에서 보여진다. 사용자는 질의 결과를 원하는 브라우징 방법으로 볼 수 있는데, 다음의 세 가지 모듈 (table · snapshot · embedded view) 중에 하나를 선택하여 수행시킬 수 있다. table view는 테이블 형태로 데이터에 대해 서술된 주석 정보를 보여주고, snapshot view는 XIMG에서는 주석 정보와 함께 이미지를 보여주며, XVIDEO에서는 비디오 세그먼트의 키 프레임을 보여준다. embedded view는 비디오 세그먼트들을 웹 브라우저 내에서 재생을 하고자 할 때 사용된다. 특히, 비디오 데이터에 대한 메타데이터가 계층적 구조를 기반으로 작성되어, 장면-샷-프레임을 기반으로 검색 및 추출을 할 수 있도록 지원한다.

그림 7에서 ⑦는 기본 검색에서 'worldcup 2002'라는 문자열을 검색한 것이며, ⑧는 축구경기에서 슛 장면을 검색한 것이다. 슛 장면은 아나운서가 '슛'이라고 말한 비디오 세그먼트를 추출하는 것이며, 추출한 비

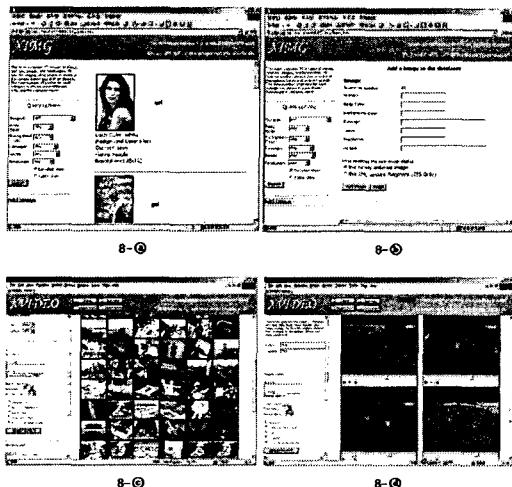


그림 7 멀티미디어 데이터에 대한 질의와  
브라우징 인터페이스

디오 세그먼트의 시작과 끝 시간은 'Time coordinator'를 통해서 조정할 수 있다. embedded view의 경우, 네트워크 트래픽(traffic)과 클라이언트의 시스템 성능에 따라 비디오 스트리밍 개수를 조절 할 수 있다.

#### 4. 실험 및 분석

이 시스템은 펜티엄 3 프로세서와 윈도우 2000을 운영체제로 하여 구축하였으며, 비디오 데이터는 데이터 손실을 줄이고 전송률을 높이기 위해 RAID(redundant array of inexpensive) 스토리지 시스템에 저장하였다.

실험을 위해 코렐 이미지와 영화, 스포츠, 뉴스 등에 비디오를 사용하였으며, 이러한 데이터에 대한 메타데이터는 34,850개의 노드를 구성하였다. 실험에서 질의 처리하는 시간이 고급 검색보다 기본 검색이 더 오래 소요되었는데, 이것은 고급 검색의 경우 XML 문서에서 특정 노드만 값을 비교를 하는데 반해, 기본 검색은 한정된 부분이 아닌 모든 노드를 비교하기 때문이다. 자막을 이용하여 영화를 대상으로 한 검색은 낮은 정밀도를 얻었지만, 스포츠나 뉴스를 대상으로 했을 때는 높은 정밀도를 얻을 수 있었다. 영화의 자막은 연기자간의 대화가 주를 이루고, 스포츠나 뉴스의 자막은 객체와 이벤트에 대한 설명이 주를 이루기 때문이다.

#### 5. 결론 및 향후 계획

본 논문에서는 XML 기반 멀티미디어 주석 및 검색 시스템을 구현하였다. 이 시스템은 멀티미디어 데이터에 대해 분석, 주석처리, 검색, 브라우징 인터페이스를 제공하고 있으며, 애플리케이션과 웹을 통해 검색이 이루어 질 수 있도록 하였다. 비디오 데이터의 효율적인 표현과 추출을 위해 계층적인 메타데이터

모델을 적용하였으며, 비디오 세그먼트의 효율적인 인덱싱과 질의를 위해 자막 기반의 스토리 세그먼테이션 분석을 하였다. 또한, 카테고라이징 기법과 가중치 부여를 통해 통합적인 질의 검색을 지원할 수 있도록 하였다. 메타데이터는 MPEG-7 표준을 기반으로 하여 XML 스키마로 작성하여 유연하고 확장 가능한 특성을 제공한다. 사용자의 질의 생성은 기본 검색과 고급 검색 통해 이루어지며, 멀티미디어 데이터와 주석을 다양하게 브라우징 할 수 있도록 세 가지 브라우징 모드로 질의 결과를 보여준다. 특히, 비디오 데이터의 구조적이고 계층적인 특성을 이용하여 질의 결과를 보여줄 수 있도록 하였다.

향후 계획은 자동화된 멀티미디어 데이터 분석 기법과 효율적인 멀티미디어 데이터의 인덱싱 방법을 개발하는 것이다.

#### [참고문헌]

- [1] Moving Picture Experts Group, "Overview of the MPEG-7 standard (version 6.0)," ISO/IEC JTC1/SC29/WG11 N4509, Pattaya, Thailand, December 2001.
- [2] Virginia E. Ogle and Michael Stonebraker, "Chabot: Retrieval from a Relational Database of Images," IEEE Computer, Vol. 28, No. 9, pp.40-48, September 1995.
- [3] J. R. Smith and S.-F. Chang, "VisualSEEK: a fully automated content-based image query system," ACM Multimedia, Boston, May 1996.
- [4] Flickner, Myron, et. al., "Query by Image and Video Content: The QBIC System," IEEE Computer, Volume 28, Number 9, September 1995.
- [5] H. Kosch, R. Tusch, L. Boszormenyi, A. Bachlechner, B. Doflinger, C. Hofbauer, and C. Riedler, "The SMOOTH Video DB - Demonstration of an integrated generic indexing approach," ACM Multimedia Conference, pp. 495-496, LA, USA, October-November 2000.
- [6] S.F. Chang, W. Chen, H. Meng, H. Sudaram, and D. Zhong, "VideoQ: An Automated Content Based Video Search System Using Visual Cues," ACM Multimedia Conference, pp. 313-324, Seattle, USA, November 1997.
- [7] Informedia Project at Carnegie Mellon University, <http://www.informedia.cs.cmu.edu/>
- [8] Virage. Inc, <http://www.virage.com/>
- [9] World Wide Web Consortium (W3C) Recommendations, <http://www.w3.org/>
- [10] Moving Picture Experts Group, "DDL Working Draft 4.0," ISO/IEC JTC1/SC29/WG11 N3575, Beijing, China, July 2000.
- [11] B.Y. Ricardo and R.N. Berthier, Modern Information Retrieval, ACM press, 1999.
- [12] W. Zhu, C. Toklu, and S.P. Liou, "Automatic News Video Segmentation and Categorization Based on Closed-Captioned Text," 2001 IEEE International Conference on Multimedia and Expo (ICME2001), pp.1036-1039, Tokyo, Japan, August 2001.
- [13] D. Barger, A. Gupta, J. Grudin, and E. Sanocki, "Annotations for Streaming Video on the Web: System Design and Usage Studies", The Eighth International World Wide Web Conference, pp. 61-75 Toronto, Canada, May, 1999.