XMARS : XML 기반 멀티미디어 주석 및 검색 시스템

남 윤 영† · 황 인 준††

요 약

본 논문에서는 XML을 이용하여 멀티미디어 데이터를 구조적으로 표현하고 효율적으로 추출하기 위한 XML 기반 멀티미디어 주석 및 검색 시스템을 제안한다. 이 시스템은 멀티미디어 데이터에 대한 주석이나 검색, 브라우징을 위한 그래픽 인터페이스를 제공하며 멀티미디어 정보를 표현하기 위해 계층적 메타데이터 모델을 기반으로 하여 구현되었다. 비디오에 대한 메타데이터는 MPEG-7 표준에 정의되어 있는 멀티미디어 서술 스키마를 기반으로 XML 스키마를 사용하여 작성하였다. 또한, 멀티미디어 데이터의 효율적인 인덱싱과 추출을 위하여 자막과 주석을 바탕으로 한 카테고라이징 기법을 사용한다.

XMARS : XML-based Multimedia Annotation and Retrieval System

Yunyoung Nam† · Eenjun Hwang††

ABSTRACT

This paper proposes an XML based Multimedia Annotation and Retrieval System, which can represent and retrieve video data efficiently using XML. The system provides a graphical user interface for annotating, searching, and browsing multimedia data. It is implemented based on the hierarchical metadata model to represent multimedia information. The metadata about video is organized based on multimedia description schema using XML Schema that basically conforms to the MPEG-7 standard. Also, for the effective indexing and retrieval of multimedia data, video segments are annotated and categorized using the closed caption.

키워드 : 멀티미디어(Multimedia), 주석(annotation), XML, MPEG-7, 메타데이터(metadata)

1. 서 론

인터넷의 보급과 더불어 대용량 저장 장치의 가격이 하락하면서, 멀티미디어 데이터의 저장과 공유가 상당히 보편화되었다. 또한 고속 네트워크가 보급되고 압축기술이 발달하면서 멀티미디어 데이터의 사용이 크게 늘어났다. 이러한 추세에 발맞추어 VOD나 전자 도서관과 같은 새로운 서비스가 등장하였으며 여기에는 많은 양의 멀티미디어 데이터를 효율적으로 관리하고 추출하는 것이 중요한 이슈로 등장하게 되었다. 멀티미디어 데이터 중 비디오 데이터는 가장 대표적인 원천 데이터로써, 텍스트나 이미지, 오디오와 객체 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 풍부한 도구들을 제공하는 MPEG-7[1]을 발표하였다. MPEG-7은 주 요소인 Descriptors(D)와 Description Schemes(DS)를 이용하여, 데이터 특성의 의미와 구조를 정의하고 구성들간의 관계 의미와 구조를 정의한다. 또한, 멀티미디어 데이터를 다양하고 효율적으로 표현하기 위해 XML(eXtensible Markup Language)을 사용하였으며, XML 스키마를 기반으로 하여 MPEG-7 서술 도구의 구조를 정의하였다. 결과적으로 MPEG-7은 멀티미디어 데이터를 서술하기 위한 유동적이고 확장 가능한 프레임워크를 제공하고 있다.

컨텐츠 정보로부터 질의를 생성하는 것은 멀티미디어 검색 시스템에서 중요한 부분이다. 그러나 MPEG-7에서는 검색과 추출에 대한 충분한 해결책을 제시하고 있지 않다. 본 시스템에서는 XPath와 XQuery를 사용하여 XML로 제작된 메타데이터에 대한 질의를 생성한다. XPath는 XML 문서 구조를 접근하는데 사용되며, XQuery는 좀더 정확한 질의 생성을 위해 사용된다.

본 논문에서는 이러한 MPEG-7을 기반으로 멀티미디어 데이터를 주석 처리하고 XML로 표현하여 저장하고 검색하는 시스템을 구현하였다. 멀티미디어 데이터는 이미지와 비디오 데이터를 주석 처리하기전에 이미지 데이터는 색상 히스토그램으로 분석하였고 비디오 데이터는 샷 검출과 자막

※ 본 연구는 한국과학재단 목적기초연구 R05-2002-000-01224-0 지원으로 수행되었음
† 준 회 원 : 아주대학교 정보통신전문대학원
†† 종신회원 : 아주대학교 정보통신전문대학원 교수
논문접수 : 2002년 7월 29일, 심사완료 : 2002년 10월 29일

을 이용하여 비디오 분석을 하고 이를 바탕으로 주석을 XML 형태로 저장하는 방식을 사용하였다. 또한, 메타데이터의 증가에 따른 검색 소요 시간을 줄이고 유사도가 높은 데이터를 검색하기 위해 카테고라이징 기법을 사용하였다. 웹을 기반으로 하여 유연한 사용자 인터페이스를 제공하며 XML과 메타데이터의 계층적 구조를 사용하여 메타데이터의 다양한 정보를 사용자가 쉽게 다룰 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 멀티미디어 데이터의 검색과 조직화와 관련된 연구들을 살피고, 3장에서는 멀티미디어 데이터의 조직화에 수반되는 메타데이터 모델링 방법과 주석 처리, 카테고라이징 기법을 설명한다. 4장에서는 전체적인 시스템 구조를 설명하고 주석 처리 도구와 질의 및 추출 인터페이스를 설명한다. 5장에서는 시스템의 실험 결과를 보이고, 마지막으로 6장에서는 결론과 향후 계획에 대해서 논의한다.

2. 관련 연구

이미지나 오디오, 비디오와 같은 멀티미디어 데이터는 다양하고 방대한 양의 정보를 포함하고 있어 효율적인 검색을 위해서는 데이터를 기술하는 구조적이고 체계화된 형태의 메타데이터가 요구된다. 멀티미디어 데이터가 보편화되고 웹 상에서 빈번히 사용되고 있지만 멀티미디어 데이터에 대한 메타데이터 모델이나 스키마가 서로 다르고 교환의 표준이 정해지지 않아 통합하는데 어려움이 있다. 뿐만 아니라, 이제까지 제안된 여러가지 멀티미디어 검색 시스템은 멀티미디어 데이터가 내포하는 다양하고 복합적인 정보를 효율적으로 처리하는 데 제약점이 많다. 본 장에서는 멀티미디어 검색과 관련된 기존의 연구를 알아보고, 효율적인 메타데이터를 구축하는 방법들을 설명한다.

2.1 기존의 멀티미디어 검색 시스템

최근까지 멀티미디어 데이터베이스와 관련된 여러가지 프로토타입 시스템이 제안되었다. 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 기반으로 이루어지며, 주석 기반 검색은 데이터의 의미나 내용 등에 대해 서술한 정보를 바탕으로 이루어진다.

우선, UC Berkeley에서 개발된 Chabot[2]는 POSTGRES라는 데이터베이스 시스템에 B트리와 R트리 기법을 이용하여 단순 속성, 객체에 대한 설명 정보를 이용한 텍스트 기반 검색과 색상 분석을 이용한 내용 기반 이미지 검색을 지원하는 시스템이다.

Columbia 대학에서 개발된 VisualSEEK[3]는 색상과 공간 질의를 지원하는 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 저장하고 추출하는 방식이다. QBIC은 이미지 뿐만 아니라 비디오 데이터도 지원하는 시스템으로써 샷의 검출, 샷에 대한 대표 프레임 생성과 객체의 움직임을 이용하여 데이터베이스를 구성한다.

SMOOTH Video DB[5]는 비디오 데이터를 데이터베이스에 저장하고 질의하며 브라우징 및 주석 처리를 지원하는 시스템이다. VIDEX 모델을 기반으로 하여 데이터베이스를 구성하였으며 이벤트, 객체, 장소의 의미정보를 저장하며 텍스트 기반의 질의를 지원한다.

VideoQ[6]는 객체들간의 특성과 움직임을 정의하기 위해 스케치 드로잉 질의를 지원하는 웹 기반 비디오 검색 시스템이다. 비디오 객체에 대해 세그먼트를 수행하고 추적하는 자동화 알고리즘을 개발하였으며 사용자 질의에 대해 실시간 비디오 편집 기법을 사용한다.

Informedia[7]는 자동화된 오디오-비디오 인덱싱과 네비게이션 및 검색을 지원하는 시스템으로써 음성 인식과 이미지와 자연 언어 처리를 제공한다.

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

2.2 멀티미디어 데이터의 구조화

기존의 연구에서 이미지의 경우 색상, 질감, 형태, 위치 등 시각적인 정보를 히스토그램이나 윤곽선 검출(edge detection) 알고리즘 등을 사용하거나 사람의 주관적인 판단에 따라 객체들의 특성을 텍스트로 기술하는 방식으로 메타데이터를 구축하였다. 히스토그램이나 윤곽선 검출을 이용한 내용 기반 검색은 비용이 많이 들고 정확도가 낮은 단점이 있으며, 텍스트로 서술하는 방식의 주석 기반 검색은 데이터가 커지고 종류가 다양해짐에 따라 주석 처리에 소요되는 시간이 증가하고 사람의 주관적인 판단에 따라 기술됨으로써 서술의 객관성에 문제점이 있다.

비디오는 프레임, 샷, 장면 등의 계층적인 구조로 표현하며, 이를 위해서 비디오 분석을 통해 그 경계를 검출하는 것이 필요하다. 하지만, 카메라 기법이나 장면의 변환이 거의 없는 경우 샷의 경계를 정확하게 검출하는 것은 상당히 어려우며, 단일 스토리를 구성하는 장면의 경계 검출은 더욱 어렵다. 많은 연구에서 비디오 프레임에 대한 색상 히스토그램과 대표 색상을 기반으로 장면과 샷의 경계 검출이 이루어

져 왔지만, 최근 연구에서는 영상정보 뿐만 아니라 오디오 정보를 이용하여 좀더 효과적인 검출을 할 수 있는 기법을 보이고 있다. 그러나 비디오 분석은 여전히 어렵고 발전시켜야 할 과제로 남아 있다. 자동화된 경계 검출 연구가 과거 몇 년 동안 이루어져 왔지만, 비디오 데이터에 대해 쉽고 빠르게 의미 정보를 해석하는 것은 사람이며, 사람에 의한 주석 처리는 비디오 인덱싱에서 필수적이다.

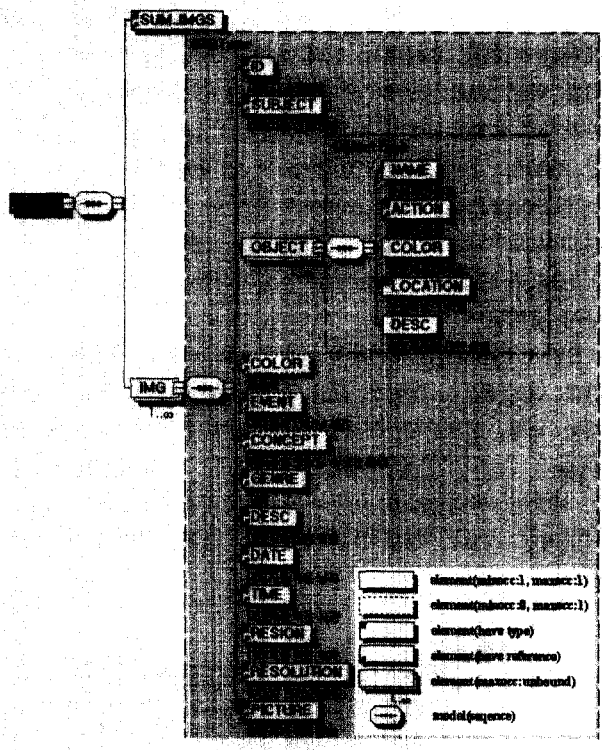
주석 기반의 검색은 데이터에 대한 일관되고 정확한 표현이 어렵기 때문에, 색상 히스토그램 등을 이용하여 멀티미디어 데이터로부터 자동으로 추출하여 인덱스를 생성 관리하는 내용 기반의 검색이 병행되어야 하며, 주석 기반의 검색이 멀티미디어 데이터에 대한 모든 정보를 제공할 수 없더라도 객체들간의 관계를 일관성 있게 해석할 수 있는 모델을 정립하여 의미 정보를 체계적으로 표현하기 위한 스키마를 구축한다면 효과적인 멀티미디어 검색이 될 수 있다.

3. 메타데이터 모델링

본 장에서는 멀티미디어 검색에 필요한 메타데이터 모델링과 주석을 처리하는 방법에 대해서 설명한다. 검색 시스템에 사용된 멀티미디어 데이터는 이미지와 비디오이며, 데이터의 본질적인 특성이 상이하기 때문에 각 특성에 맞는 메타데이터를 구축하였다.

3.1 이미지 메타데이터

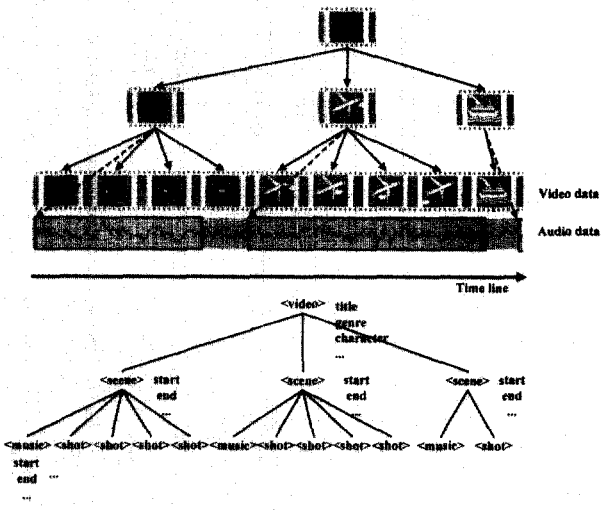
(그림 1)은 이미지 메타데이터를 위한 XML 스키마의 다이어그램이다. 루트 엘리먼트 IMGS는 SUM_IMGS와 IMG라는 엘리먼트를 가지며, SUM_IMGS 엘리먼트는 데이터베이스에 저장되어 있는 이미지의 총 개수를 값으로 갖는다. 각 이미지는 IMGType이라는 complex type을 참조하며, 이미지 타입은 이미지의 고유번호, 대표 이름, 객체, 색상, 사건, 개념, 장르, 서술, 촬영 날짜, 시간, 장소, 해상도, 파일의 위치 정보를 포함한다. 객체는 ObjectType이라는 complex type을 참조하며 객체 타입은 객체의 이름, 동작, 색상, 이미지 상에서 객체의 위치, 설명을 포함한다. 색상과 객체의 위치, 장르, 촬영 날짜, 시간과 같이 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. 실선 상자는 나타날 수 있는 최소 횟수가 1이며 점선 상자는 나타날 수 있는 최소 횟수가 0임을 뜻한다. 상자의 왼쪽 하단에 있는 화살표는 이미 정해진 타입을 참조함을 의미하고 왼쪽 상단에 있는 세 개의 줄은 엘리먼트의 형태가 정해진 것이며 줄이 없는 엘리먼트는 아직 정해지지 않은 것을 뜻한다.



(그림 1) 이미지 메타데이터를 위한 XML 스키마

3.2 비디오 메타데이터

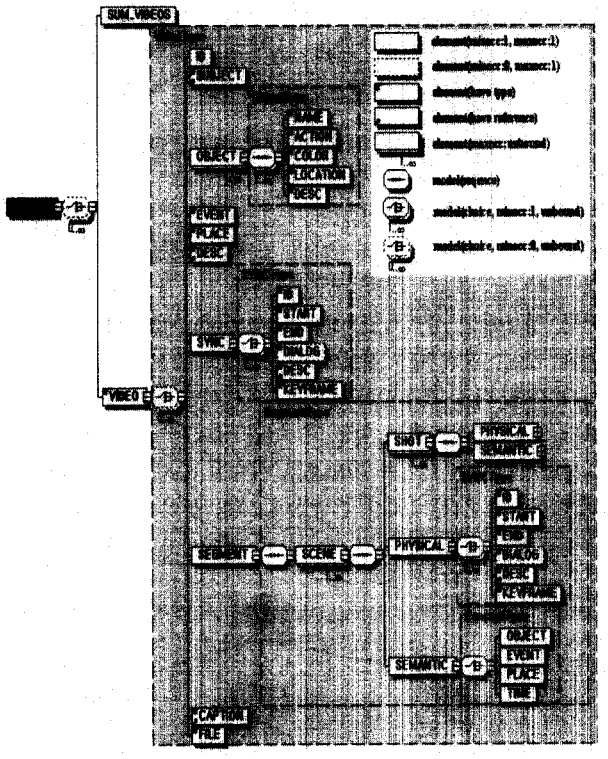
비디오의 계층적 구조는 트리 형태로 표현할 수 있으며 이 경우 각 노드에 는 샷과 장면에 대한 정보를 포함시킨다. 이러한 정보에는 물리적인 정보와 의미적인 정보가 포함된다. 예를 들어, 물리적인 정보에는 세그먼트의 시작과 끝 시간, 샷에 대한 대표 프레임이 속하며, 의미적인 정보에는 영상 정보나 오디오 정보에서 얻을 수 있는 내용이 포함된다. 본 시스템에서는 MPEG-7을 기반으로 서술된 주석 정보를 말하며, MPEG-7 표준 문서에서 컨텐츠에 대한 설명은 이벤트, 객체, 개념(concept), 장소, 시간, 요약으로 정의하고 있다.



(그림 2) 비디오의 계층적 구조와 트리 형태의 XML 구조

(그림 2)는 비디오에 대한 계층적인 구조와 이러한 구조를 표현한 트리를 보여주고 있다. (그림 3)은 비디오 메타데이터를 위한 XML 스키마의 다이어그램이다. 이미지에서와 마찬가지로 (그림 1)에서 보았던 ObjectType이 있으며, 물리적인 정보를 나타내는 SYNCType과 의미적인 정보를 나타내는 ConceptType이라는 complex 타입을 정의한다. 이 두 가지 타입은 SCENE 또는 SHOT을 구성하는 단위가 되며, SCENE과 SHOT은 SegmentType이라는 complex 타입으로 다시 정의된다.

SYNCType은 고유번호와 비디오 세그먼트의 시작과 끝 시간을 밀리세컨드 단위로 값을 갖는 엘리먼트와 객체들 간에 대화 내용, 설명, 키 프레임을 엘리먼트로 갖는다. ConceptType은 '누가', '언제', '어디서', '무엇을', '어떻게' 에 해당되는 내용을 엘리먼트로 이루어진다. 이러한 물리적인 정보와 의미적인 정보는 장면이나 샷이라는 상위 엘리먼트에 묶여 있다.

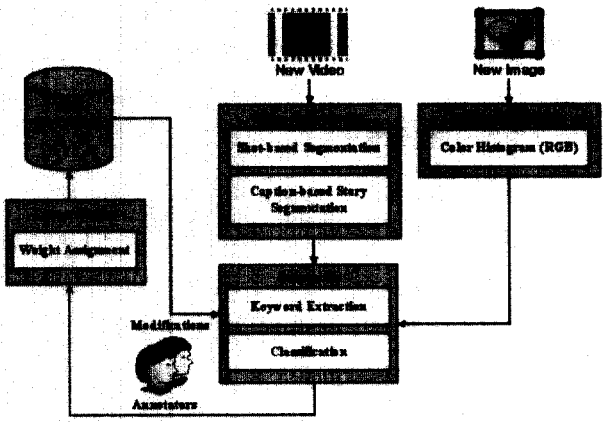


(그림 3) 비디오 메타데이터를 위한 XML 스키마

3.3 주 석

(그림 4)는 비디오 조직화, 주석, 카테고라이징하는 처리 과정을 보여주고 있다. 우선 비디오는 샷 경계 검출과 자막 데이터를 기반 세그멘테이션 처리 과정을 거치며, 이미지는 색상 히스토그램을 통해 특성을 추출하며 이 특성을 바탕으로 주석을 처리한다. 자막 데이터의 단어는 스토리 세그멘테이션 처리기에서 추출되는데, 자막 데이터는 영화, TV 뉴스, 스포츠 등에서 사용되는 자료이며, 대화 내용 뿐만 아니라 화자와 대화 시작·지속시간 정보를 담고 있다. 그러나 비디오 데이터를 인덱싱하는데 충분하지 않기 때문에, 미디어의 세그먼테이션의 주석 처리 작업이 필요하다. 세그먼테이션에 대한 주석은 샷 검출과 스토리 세그멘테이션을 기반으로 하여 수행된다.

의미적인 정보의 스키마는 오디오-영상 컨텐츠를 표현하기 위한 이야기식의 서술방법을 나타내며 객체, 사건, 장소, 시간과 같은 의미적인 형태를 서술한다. 주석 정보는 이 모델을 바탕으로 만들어진다. 주석자는 세그먼트를 구분할 뿐만 아니라 적당한 설명과 키워드를 지정한다. 그 후에 세그먼트와 관련된 주석 정보로부터 키워드의 빈도수를 바탕으로 계산하는 카테고라이징을 하게 된다. 마지막으로 주석 정보와 빈도수는 XML 데이터베이스에 저장된다.



(그림 4) 멀티미디어의 메타데이터 처리 단계

3.4 카테고라이징

정보의 양이 증가함에 따라서 검색에 소요되는 시간은 더욱 증가하게 된다. 이러한 문제를 해결하고 효율적인 정보의 인덱싱과 추출을 위해 주석 정보에 대한 카테고라이징을 수행한다.

시스템에서 문서의 총 개수를 N, 인덱스 단어를 k_i는 k_i가 나타나는 문서의 개수라 하고, 문서 d_j에서의 단어 k_i의 빈도수를 $freq_{i,j}$라고 하면, 문서 d_j에서 단어 k_i의 정규화된 빈도수 $f_{i,j}$는 식 (1)과 같이 주어진다[11].

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

식 (1)에서 $\max_l$은 문서 d_j에서 나타나는 모든 단어의 개수를 말한다. 만약 단어 k_i가 문서 d_j에 나타나지 않으면 $f_{i,j} = 0$이다. k_i에 대한 빈도수의 역을 이용하여 단어의 가중치 $w_{i,j}$를 식 (2)와 같이 정의한다.

$$w_{i,j} = f_{i,j} \times \log \frac{N}{n_i} \quad (2)$$

특정화된 단어의 키워드의 가중치를 $K_{i,j}$라 하고, $K_{i,j}$ $w_{i,j}$에 대한 상대적 중요도 $\alpha(0 \le \alpha \le 1)$를 사용하여 본 논문에서 사용하는 단어에 대한 가중치는 식 (3)과 같다.

$$W_{i,j} = \alpha \times w_{i,j} + (1 - \alpha) \times K_{i,j} \quad (3)$$

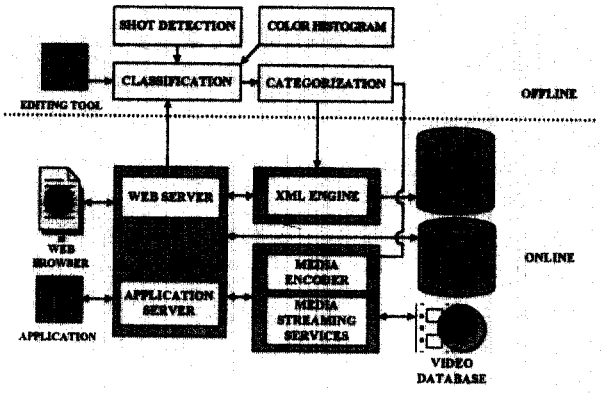
실험에서는 카테고라이징에서 도출된 단어의 가중치와 키워드의 가중치에 대한 중요도을 같은 비율로 두어 α를 0.5로 정했다.

4. XMARS system

4.1 구 조

(그림 5)는 XMARS(XML-based Multimedia Annotation and Retrieval System)의 개략적인 구조를 보여주고 있다. 오프라인 과정에서 샷 검출과 색상 히스토그램을 통해 비디오와 이미지 분석을 하고 주석 처리와 카테고라이징을 수행한다. 한편, 온라인 처리는 웹 서버와 XML/미디어 엔진, 그리고 데이터베이스가 담당한다. XML 엔진은 XSLT 프로세서, 파일 시스템, DOM, XML 데이터베이스를 위한 API 등으로 이루어진다. XSLT 프로세서는 XML 문서를 HTML 문서로 변환할때 사용되고, 저장된 XML 문서의 갱신이나 질의 및 XSLT 컴포넌트(component)는 W3C XML 표준 인터페이스에 맞춰 구현되었다. 미디어 엔진은 비디오 데이터베이스에 적당한 포맷으로 저장하기 위한 미디어 인코더와 전체 비디오 데이터에서 원하는 부분만 스트리밍해 주는 미디어 스트리밍 서비스로 이루어져 있다.

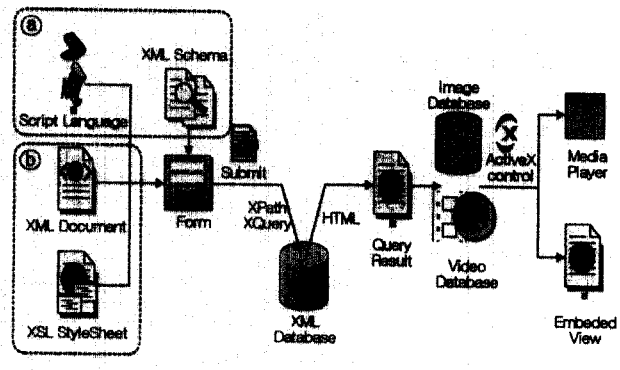
XML 데이터베이스로는 엑셀론(eXcelon)을 사용하였고, 동적이고 확장 가능한 서버 구성을 위해 톰캣(Tomcat)과 액티브엑스 컨트롤(ActiveX control)을 사용하였다. 서버와 클라이언트 모듈들은 CORBA 환경에서 서로 통신을 하므로 방화벽(firewall)이 설치되어 있는 네트워크 내에서도 동작이 가능하다.



(그림 5) XMARS의 구조

4.2 전체 작업 흐름도

(그림 6)은 웹 상에서 XMARS의 전체적인 프로세스 흐름을 보여주며, 크게 멀티미디어 데이터에 대한 주석 정보 입력과 질의로 구분 지을 수 있다. ⓐ는 주석 정보의 입력에 필요한 요소이며, ⓑ는 질의에 필요한 요소들이다. ⓐ에서는 XML 스키마에서 엘리먼트와 속성의 이름, 나타나는 횟수 등을 분석하여 스크립트 언어를 통해 폼을 구성하며, 폼을 통해 입력받은 데이터는 XML 문서로 변환되어 XML 데이터베이스에 저장한다. 이때 변환된 XML 문서는 XML 스키마에 대해 유효성 검사(validation check)를 하여 유효한 문서만 저장시킬 수 있으며, 유효하지 않은 문서는 수정하여 다시 유효성 검사를 거쳐야 한다.

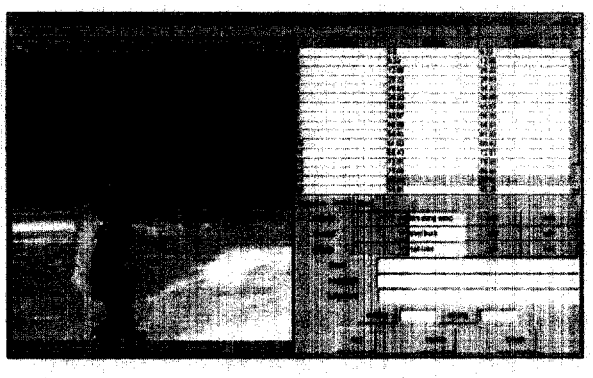


(그림 6) 웹상에서 XMARS의 작업 흐름도

ⓑ에서는 XML 문서에 스타일언어인 XSLT를 적용시켜 폼을 구성하며, 사용자의 입력을 받는다. 폼을 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 생성된 HTML 문서를 통해 사용자가 원하는 비디오의 세그먼트가 미디어 플레이어나 웹 브라우저를 통해 브라우징 되며, 이때 전체 비디오 데이터로부터 추출된 시간 정보를 이용한 부분 재생을 위해 액티브엑스 컨트롤을 사용한다.

4.3 주석 입력 도구

(그림 7)은 비디오 세그먼트에 대해 주석을 입력하기 위한 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 정보의 추가, 삭제, 수정 등을 수행하며, 샷의 리스트를 이용하여 비디오의 특정 시간 위치에 대해 풍부한 정보를 서술할 수 있다. 비디오에 대한 주석 입력 단계 후에 비디오 세그먼트의 서술내용은 카테고라이징 단계를 거쳐 생성된 요약 문서와 함께 XML 데이터베이스에 저장이 되고, 비디오 스트림은 미디어 인코더를 거쳐 변환되어 비디오 데이터베이스에 저장이 된다.



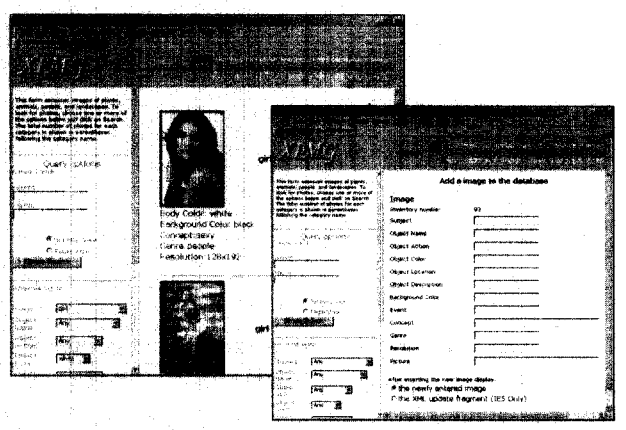
(그림 7) 비디오를 위한 주석 처리 도구

4.4 질의와 브라우징

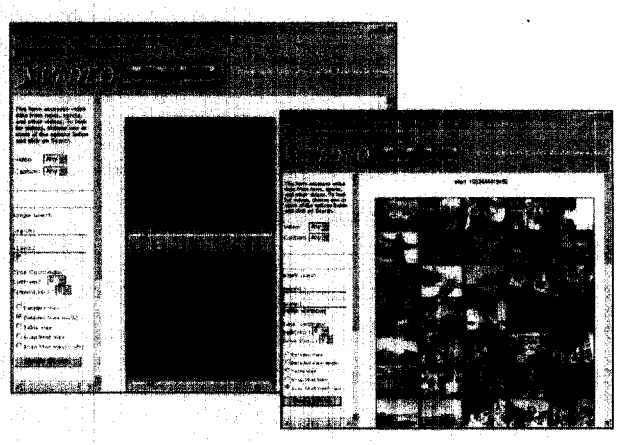
사용자가 원하는 이미지나 비디오 세그먼트를 효율적으로 검색하고 추출하기 위해서, 검색 시스템은 다양한 방법으로 질의를 생성하고 그 결과를 브라우징할 수 있게 하여야 한다. XMARS는 멀티미디어 데이터에 대한 다양한 질의를 가능케 하는 기본 검색과 고급 검색의 두 가지 모드를 제공한다. 기본 검색은 카테고라이징 단계에서 생성된 문서에 대해 검색하는 것으로, 이것은 트리 형태의 주석 문서에 존재하는 모든 노드들을 액세스하는 것이다. 기본 검색은 스키마에 익숙하지 않은 사용자에게 유용하지만, 낮은 정확률(precision)과 높은 재현율(recall)의 검색 결과를 보여준다. 고급 검색은 사용자가 스키마에 대해 잘 알고 정확한 검색을 하고자 할 때 사용되는 방법이다. 이와같은 검색은 XPath와 XQuery를 이용하여 사용자의 입력 값과 주석의 엘리먼트 값을 비교하는 방법으로 수행한다.

(그림 8)은 XMARS의 질의와 브라우징 인터페이스를 보여주고 있다. 이미지와 비디오 데이터의 이질적인 특성으로 인한 분산 처리 환경을 제공하기 위해 이미지 검색 시스템인 XIMG와 비디오 검색 시스템인 XVIDEO로 나누어 개발하였다. (a)와 (b)는 XIMG의 인터페이스 화면이며, (c)와 (d) 질의는 XVIDEO의 인터페이스 화면이다. 질의는 왼쪽 프레임에서 생성하고 질의 결과는 오른쪽 프레임에서 보여진다. 사용자는 질의 결과를 다양한 방법으로 브라우징할 수 있으며 구체적으로 테이블이나 스냅 샷, 내장 뷰 중 하나를 선택할 수 있다. 테이블 뷰는 테이블 형태로 데이터에 대해 저장되어 있는 주석 정보를 보여주고, 스냅 샷은XIMG의 경우 주석 정보와 함께 이미지를 보여주며, XVIDEO의 경우는 비디오 세그먼트의 키 프레임을 보여준다. 내장 뷰는 비디오 세그먼트을 웹 브라우저 상에서 재생하고자 할때 사용된다. 특히, 비디오 데이터에 대한 메타데이터가 계층적 구조를 기반으로 작성되어 있으므로, 장면-샷-프레임 기반으로 검색하고 추출할 수 있다.

(그림 8)에서 (d)는 기본 검색에서 'Korea worldcup'라는 문자열을 검색한 것이며, (c)는 축구경기에서 슛 장면을 검색한 것이다. 슛 장면은 아나운서가 '슛'이라고 말한 비디오 세그먼트를 추출하는 것이며, 추출한 비디오 세그먼트의 시작과 끝 시간은 'Time coordinator'를 통해서 조정할 수 있다. 내장 뷰의 경우, 네트워크 트래픽(traffic)과 클라이언트의 시스템 성능에 따라 비디오 스트리밍 개수를 조절할 수 있다.



8-(a) 8-(b)



8-(c) 8-(d)

(그림 8) 멀티미디어 데이터에 대한 질의와 브라우징 인터페이스

5. 실험 및 분석

XMARS는 펜티엄 4 프로세서와 윈도우 2000을 운영체제로 하여 개발하였으며, 메타데이터는 XML 데이베이스인 엑셀론에 저장하였다. 비디오 데이터는 데이터 손실을 줄이고 전송률을 높이기 위해 RAID(Redundant Array of Inexpensive Disks)에 저장하였다.

실험을 위해 이미지 데이터로는 코렐 이미지 1000 개를 블랍(BLOB)의 형태로 엑셀론에 저장하였으며, 비디오 데이터로는 영화 10편, 스포츠 5편, 뉴스 10편을 RAID에 저장하였다. 이들에 대한 메타데이터를 구축하여 총 633,000개의 노드로 구성하였고, 실험 데이터에 대한 자세한 내용은 <표 1>과 같다.

〈표 1〉 실험 데이터

구 분		저장형태	개수	사이즈	메타데이터의 노드수
이미지	코렐 이미지	JPEG/BLOB	1000	15 MB	17,000
비디오	영 화	MPEG-4	10	6.87 GB	34,850
	스포츠	MPEG-1	5	7.51 GB	5,250
	뉴 스	ASF	10	660 MB	6,200

<표 2>과 <표 3>은 이미지와 비디오에 대해 각각 세 가지 질의를 수행하여 얻은 검색 결과이다. 이미지에 대한 질의는 "녹색 바탕에 노란색 객체가 있는 이미지를 찾아라", "초록색 바탕에 표범이 걷고 있는 이미지를 찾아라", "어둠 속에서 검은 사람이 미소짓고 있는 이미지를 찾아라" 이며, 비디오에 대한 질의는 "뉴스에서 어선이 나오는 장면을 찾아라", "축구 경기에서 슛을 하는 장면을 찾아라", "영화에서 '사랑'이라는 말을 하는 장면을 찾아라" 이다.

〈표 2〉 이미지에 대한 질의 결과

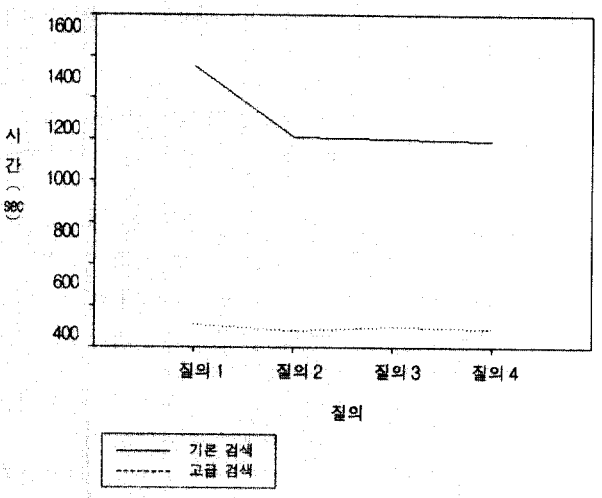
질 의	검 색	retrieved	relevant	relevant images in DB	precision	recall
질의 1	기본검색	102	18	21	17.64 %	85.71%
	고급검색	24	18	21	75.00%	
질의 2	기본검색	15	13	14	86.66%	92.85%
	고급검색	14	13	14	92.85%	
질의 3	기본검색	15	12	15	80.00%	80.00%
	고급검색	13	12	15	92.30%	

〈표 3〉 비디오에 대한 질의 결과

질 의	검 색	retrieved	relevant	relevant shots in DB	precision	recall
질의 1	기본검색	36	7	9	19.44 %	77.77%
	고급검색	8	7	9	87.50%	
질의 2	기본검색	102	94	94	92.15%	100.00%
	고급검색	97	94	94	96.90%	
질의 3	기본검색	188	42	45	22.34%	93.33%
	고급검색	56	42	45	75.00%	

위의 표에서 알 수 있듯이 대부분의 질의에서 기본 검색보다는 고급 검색이 더 정확하다는 것을 알 수 있다. 현재 연구되고 있는 멀티미디어 검색 시스템은 각각의 메타데이터의 스키마가 상이하며 내용 기반의 검색이 아닌 주석 기반의 검색은 주석자의 판단과 사용 어휘에 따라 검색률이 다르게 나올 수 있다. 위의 실험에서 비디오에 대한 질의1의 경우 주석자가 '어선'이 아니라 '배'라고 기술하였다면 검색에 실패할 수 있다. 그러나, 질의2에서 축구 경기의 '슛'을 하는 장면을 검색하는 경우, 축구 경기의 용어는 한정되어 있기 때문에 검색률이 높게 나온다. 여러가지 질의에 대한 실험결과를 분석해 본 결과, 자막을 이용한 영화의 검색보다는 스포츠나 뉴스의 경우가 더 높은 정밀도를 얻을 수 있었다. 이것은 영화 자막이 연기자간의 대화가 주를 이루고, 스포츠나 뉴스의 자막은 객체와 이벤트에 대한 설명이 주를 이루기 때문이다.

(그림 9)는 기본 검색과 고급 검색을 사용하여 비디오 데이터에 대한 질의를 처리하는데 소요되는 시간을 나타낸다. 그림에서 알 수 있듯이, 질의 처리시에 기본 검색이 고급 검색보다 더 많은 시간이 소요되었는데, 이것은 고급 검색의 경우 XML 문서에서 특정 노드의 값만 비교하는데 반해, 기본 검색은 한정된 부분이 아닌 모든 노드의 값을 비교하기 때문이다.



(그림 9) 검색별 질의 처리시간

6. 결론 및 향후 계획

본 논문에서는 XML 기반의 멀티미디어 주석 및 검색 시스템을 구현하였다. 이 시스템은 멀티미디어 데이터에 대한 분석이나 주석처리, 검색, 브라우징을 위한 인터페이스를 제공하고 있으며, 애플리케이션과 웹을 통해 검색이 이루어 질 수 있도록 하였다. 비디오 데이터의 효율적인 표현과 추출을 위해 계층적인 메타데이터 모델을 적용하였으며, MPEG-7 표준을 기반으로 XML 스키마로 작성하여 유연하고 확장 가능한 특성을 제공할 수 있었다. 비디오 세그먼트의 효율적인 인덱싱과 질의를 위해 자막 기반의 스토리 세그먼테이션 분석을 하였고, 카테고라이징 기법과 가중치 부여를 통해 통합적인 질의 검색을 지원할 수 있도록 하였다. 사용자의 질의 생성은 기본 검색과 고급 검색을 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 데이터 검색을 효율적으로 지원하기 위해 시소러스(thesaurus)와 온톨로지(ontology)와 관련하여 연구 개발하는 것이며, 이로써 좀 더 유연한 표현 방법에 의한 정보 표현이 가능하도록 하는 것이다. 또한, 자동화된 멀티미디어 데이터 분석 기법과 효율적인 멀티미디어 데이터의 인덱싱 방법을 개발하는 것이다.

참 고 문 헌

[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, Vol.28, No.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, LA, USA, pp.495-496, 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, Seattle, USA, pp. 313-324, 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), Tokyo, Japan, pp.1036-1039, August, 2001.

[13] D. Bargeron, 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, Toronto, Canada, pp.61-75, May, 1999.

[14] A. Berger and J. Lafferty, "Information Retrieval as Statistical Translation," ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99), California, USA, pp.222-229, August, 1999.

[15] M. Christel, D. Winkler, and R. Taylor, "Improving Access to a Digital Video Library," The 6th IFIP Conference on Human Computer Interaction, Sydney, Australia, pp.524-531, July, 1997.

[16] B. Emond, M. Brooks, and A. Smith, "A Broadband Web-based Application for Video Sharing and Annotation," ACM Multimedia 2001, Ottawa, Canada, September-October, 2001.

[17] H. Rehatschek, and H. Muller, "A Generic Annotation Model for Video Databases," The 3rd International Conference on VISUAL '99, Amsterdam, Netherlands, pp.383-390, June, 1999.

[18] M. V. Setten and E. Oltmans, "Demonstration of A Distributed MPEG-7 Video Search and Retrieval Application in the Educational Domain," ACM Multimedia 2001, Ottawa, Canada, September-October, 2001.

[19] L. Wilcox and J. Boreczky, "Annotation and Segmentation for Multimedia Indexing and Retrieval," The 31st Annual Hawaii International Conference on System Sciences, Vol. II, Wailea, Hawaii, pp.259-267, January, 1998.

[20] 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), Tokyo, Japan, pp.1036-1039, August, 2001.

[21] M. Carrer, L. Ligresti, G. Ahanger, and T. D. C. Little, "An Annotation Engine for Supporting Video Database Population," Multimedia Tools and Applications 5, pp.233-258, November, 1997.

[22] D. Bargeron, 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, Toronto, Canada, pp.61-75, May, 1999.

남 윤 영

e-mail : youngman@ajou.ac.kr

2001년 아주대학교 정보 및 컴퓨터공학과 (학사)

2001년~현재 아주대학교 정보통신전문대학원 석사과정

관심분야 : 데이터베이스, 멀티미디어 시스템, 정보 통합, XML 응용

황 인 준

e-mail : ehwang@ajou.ac.kr

1988년 서울대학교 컴퓨터공학과(학사)

1990년 서울대학교 컴퓨터공학과(석사)

1998년 Univ. of Maryland at College Park 전산학과(박사)

1998년~1999년 Bowie State Univ., Assistant Professor

1999년~1999년 Hughes Research Lab. 연구교수

1999년~현재 아주대학교 정보통신전문대학원 조교수

관심분야 : 데이터베이스, 멀티미디어 시스템, 정보 통합, 전자상거래, XML 응용