

# 멀티미디어 데이터 검색을 위한 스키마 설계 및 시스템 구현

노승민<sup>0</sup> 황인준

아주대학교 정보통신전문대학원 정보통신공학과  
{anycall, ehwang}@ajou.ac.kr

## Schema Design and Implementation for Multimedia Data Retrieval

Seungmin Rho<sup>0</sup>

Eenjun Hwang

The Graduate School of Information and Communication, Ajou University

### 요약

본 논문에서는 효과적인 멀티미디어 데이터 검색을 위하여 기존의 주석과 내용 기반 검색 기법을 보완한 새로운 멀티미디어 데이터 검색 기법과 멀티미디어 데이터 모델을 사용한 통합 멀티미디어 검색 및 주석 시스템을 제안한다. 데이터 모델로는 MPEG-7 표준에 정의되어 있는 멀티미디어 기술 구조(MDS)와 기술 정의 언어인 XML Schema를 사용하였다. 이러한 모델을 기반으로 멀티미디어 데이터를 XML의 계층구조를 이용하여 주석 처리하고 다양한 방법을 통하여 검색을 할 수 있는 멀티미디어 검색 시스템을 구현하여, 실제 실험을 통하여 성능을 평가하였다.

### 1. 서론

최근 웹(WWW)의 급속한 발전으로 인한 웹 콘텐츠의 증가와 네트워크의 고속화와 함께 각종 멀티미디어의 압축기술과 같은 멀티미디어 기술의 발전에 따라 멀티미디어 정보의 사용량이 급증하게 되었다. 이러한 멀티미디어 정보들은 이미지, 비디오, 오디오, 음성 등 다양한 형태로 표현되고, 사용자들의 멀티미디어 정보 생성, 가공, 교환 및 필요한 정보를 검색하는 등의 요구가 늘어남에 따라 점차 질의의 효율성과 편리성에 중점을 두는 시스템이 개발되었다. 사용자는 기존의 이미지의 색상 히스토그램이나 질감 등의 특성을 이용하거나 예제에 의한 질의(Query by example) 방식을 가지고 검색을 하던 방식에서 색상 히스토그램이나 질감 등의 포괄적인 특성 외에 이미지내의 각 물체의 색, 모양(Shape)이나 오브젝트의 배치 등 다양한 이미지의 특성을 검색에 이용한 새로운 형태의 질의 방식인 스케치에 의한 질의(Query by sketch)가 제안되었다. 그러나 이것은 검색하고자 하는 목표 이미지와는 전혀 다른 검색 결과가 나올 수 있는 단점이 있다. 이러한 문제점을 해결하기 위해 이미지 오브젝트간의 공간 관계를 이용한 공간 유사도 기반 이미지 검색(Spatial Similarity-based Image Retrieval) 기법[8]을 사용한다.

멀티미디어 데이터 검색은 사용되는 특징에 따라 주석 기반 검색과 내용 기반 검색으로 나누어진다. 내용 기반 검색(Content-based retrieval)은 멀티미디어 데이터로부터 모양, 색깔, 질감, 움직임 등의 특징적인 내용을 표현하는 특정 데이터를 자동으로 추출하여 저장하고 이를 바탕으로 이미지나, 비디오 클립 등을 질의로 입력하여 검색하는 방법이다. 따라서 사용자의 주관이 개입되지 않아 객관적인 검색이 가능하고, 특정 도메인에서 좋은 결과를 나타내지만 그 데이터가

정확한 내용을 표현한다는 것이 어렵다. 그러나, 주석 기반 검색(Annotation-based retrieval)은 멀티미디어 데이터에 대한 의미정보를 사용자가 미리 파악한 뒤에 이를 주석으로 작성하여 저장한 후에 데이터들을 기술한 질의에 나타난 내용들을 주석과 비교하여 관련된 데이터를 찾는 방법이다. 이 방법은 사람이 자동화된 방법을 알아내기 힘든 동영상 자료나 복잡한 이미지 자료의 의미를 쉽게 모델링 할 수 있지만, 대용량의 데이터에 대하여 사람이 일일이 주석을 작성해야 하며 데이터에 대한 일관성을 잃기 쉽다는 단점이 있다.

따라서 이 두 가지 방법이 가지는 장점을 함께 수용하고 단점을 상호 보완하여 사용한다면 효과적인 검색을 할 수 있을 것이다. 본 논문에서는 MPEG-7을 기반으로 한 통합 멀티미디어 검색시스템을 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 통하여 기존의 멀티미디어 검색 및 주석 시스템에 대해 간단히 기술하였으며 3장에서는 주석 기반과 내용 기반 검색 기법과 이를 보완하기 위한 새로운 데이터 검색 기법들을 통합하는 메타 데이터 모델링 방법과 주석 처리기법을 설명한다. 4장에서는 스키마를 이용한 통합 검색 및 주석 시스템의 구현과 실험에 대해서 기술한다. 마지막으로 5장에서는 결론과 향후 계획에 대해서 논의한다.

### 2. 관련연구

멀티미디어 검색 시스템은 검색하고자 하는 데이터의 종류에 따라 크게 이미지와 비디오 검색 시스템으로 나눌 수 있다. 이미지 검색 시스템으로는 UC Berkeley에서 개발한 데이터베이스인 POSTGRES를 기반으로 하여 개발한 이미지 검색 시스템인 Chabot[1]이 있으며, 각 이미지가 가질 수 있는 속성들과 개체들에 대한 설명을 이용한 텍스트 기반의 검색과 사용자에

의해 다양한 형태로 정의될 수 있는 concept query와 색상 정보를 이용한 내용 기반 검색을 지원한다. QBIC[2]은 IBM에서 개발한 정지영상 및 동영상 검색 엔진이며 QBE(Query by Example)를 통한 유사도 질의와 색상이나 질감 등을 이용한 사용자의 스케치를 통해서 질의를 할 수 있게 해주는 상용화된 영상 검색 시스템이다. SMOOTH Video DB[3]는 비디오 인덱싱 모델인 VIDEX를 기반으로 하여 인덱스 데이터베이스를 구축하였고 이벤트, 개체, 사람, 위치와 같은 high-level의 의미 정보들과 히스토그램과 같은 물리적인 low-level 정보들을 저장하여 텍스트 기반의 질의를 지원한다. Vane은 Tcl/Tk를 이용한 유연한 질의 및 주석 인터페이스와 SGML DTD를 이용한 데이터 모델을 구현하였다.

### 3. 메타 데이터 모델링

#### 3.1 이미지 메타데이터

이미지에서의 메타데이터 기술구조는 이미지의 의미적인 표현이 가능한 의미기술구조(Semantic DS)와 정형적인 데이터를 표현할 수 있는 메타 정보 기술구조(Meta Information DS)로 크게 나눌 수 있다. 의미 기술구조는 이미지가 표현하고 있는 객체들에 대한 상세한 정보와 움직임이나 상태 등에 대한 구체적인 정보를 표현하기 위해서 5WH(When, Where, Why, What, Who, How)와 Event 등에 대해서 기술하고 있다. 이에 반해서 메타정보 DS는 이미지의 정형적인 데이터를 기술하는 기술구조로서 의미 기술구조에 비해 객관적이며 이미지의 저작권, 제작자의 정보와 이미지에 대한 간단한 설명정보를 기술할 수 있다.

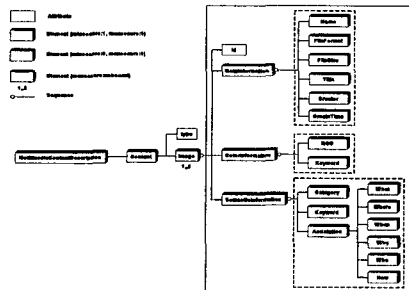


그림 1 이미지 메타데이터 XML Schema Diagram

그림 1은 이미지 메타데이터를 위한 XML Schema Diagram이다. 각 이미지들은 위에서 언급했던 이미지 기술구조(Image DS)와 일반적인 분류 체계에 따라 사람(Human), 동물(Animal), 식물(Plant), 예술(Art), 스포츠(Sport), 건물(Architecture), 풍경(Landscape)으로 7가지로 나누었다. 각 이미지는 물리적인(Physical) 정보들과 메타(Meta) 정보, 색상(Color) 정보, 내용(Semantic) 정보로 구성되어 있다.

#### 3.2 비디오 메타데이터

비디오에서의 메타데이터 기술구조 역시 이미지와 비슷하게 메타 정보와 의미 정보 그리고 물리적인 정보로 크게 나눌 수 있으며, 이러한 정보들은 비디오

데이터를 장면(Scene), 샷(Shot), 프레임(Frame)의 계층적 구조를 가지는 트리 형태로 표현한다고 볼 때 하나의 노드들로 표현될 수 있다.

그림 2는 비디오 메타데이터를 위한 XML Schema Diagram이다. 이미지와 유사하게 각 비디오는 메타정보와 샷 계층 검색 알고리즘을 통하여 검색된 각각의 샷 들에 대한 프레임정보들과 이벤트들에 대해 기술하고 이들을 다시 상위 엘리먼트인 장면에 대한 프레임과 이벤트 정보들과 묶어 의미적인 정보를 표현한다. 카테고리(Category)는 비디오 파일이 어느 범주에 속해 있는지를 알 수 있게 해주며 크게 뉴스(News), 스포츠(Sports), 영화(Movie)와 교육(Education)으로 나누었다. 각 범주는 하위범주(Sub Category)를 가지며 이는 동적으로 생성되는 인터페이스를 통해서 다양하게 주석을 입력할 수 있다.

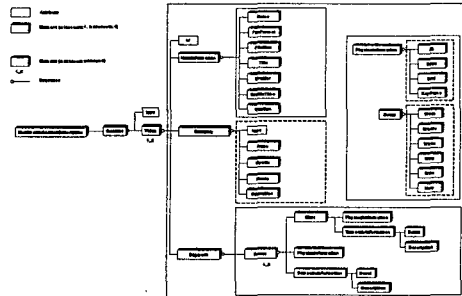


그림 2 비디오 메타데이터 XML Schema Diagram

### 4. 통합 멀티미디어 검색 시스템

#### 4.1 시스템 구성도

그림 3은 본 논문에서 구현한 시스템의 전체 구성도를 나타낸다.

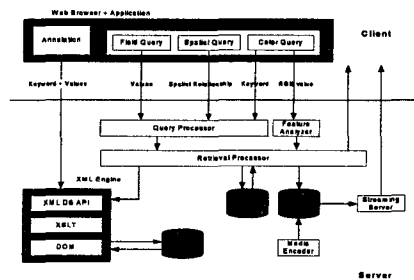


그림 3 전체 시스템 구성도

클라이언트에서는 주로 웹 브라우저나 어플리케이션에서의 인터페이스를 이용하여 주석을 입력하거나 키워드, 필드 값의 선택이나 공간질의를 이용한 주석 및 질의 처리가 이루어진다. 서버에서는 오프라인에서 검색된 샷 정보와 색상 히스토그램 및 공간 관계 분석을 이용하여 이미지 및 동영상에 대한 특징들을 분석 및 주석 처리하고, 온라인에서는 분석 및 주석 데이터들의 저장 및 추출을 위한 XML DB API, XSLT, DOM 등을 포함하는 XML Engine을 통하여 XML 데이터베이스와 연동한다.

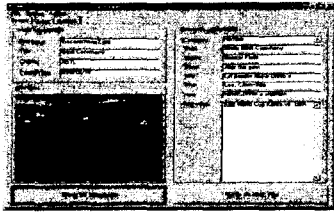


그림 4 이미지를 위한 주식 처리 인터페이스

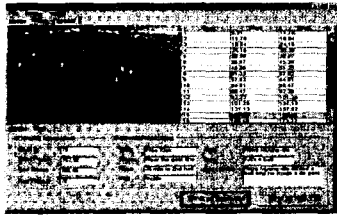


그림 5 비디오를 위한 주식 처리 인터페이스

그림 4와 그림 5는 이미지 데이터와 비디오 데이터에 대해 주식 정보를 기술할 수 있는 인터페이스를 보여주고 그림 6은 검색 결과를 브라우저하는 인터페이스를 나타낸다.

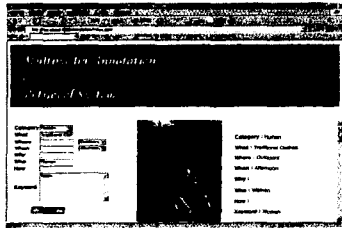


그림 6 검색 및 브라우징 인터페이스

#### 4.2 구현 및 테스트

본 시스템은 운영체제는 윈도우 2000 서버, 구현도구는 Java 2D와 JAI(Java Advanced Imaging)를 사용하여 이미지 공간 분석 및 질의 인터페이스를 구현하였고, JMF(Java Media Framework)와 ActiveX Control을 사용하여 비디오 검색 및 주석을 가능하게 하였다. 스트리밍 서비스를 위하여 윈도우 미디어 서버를 사용하였으며, 웹 서버는 톰캣(Tomcat)을 사용하였고 스크립트 언어는 JSP를 사용하였다. 확장성과 편의성을 고려하여 어플리케이션과 웹을 모두 지원하며 XML 전용 데이터베이스인 eXcelon을 사용하였다.

본 시스템의 성능분석을 위해 1,000개의 코렐(Corel) 이미지에 대해서 주식 입력 및 오브젝트간의 공간 관계에 대한 처리과정을 거쳐 eXcelon 데이터베이스에 저장하였고, 비디오의 경우는 영화, 스포츠, 뉴스 등의 데이터를 사용하였다. 그림 7은 기본 검색과 고급 검색을 사용하여 비디오 데이터에 대한 질의를 처리하는데 소요되는 시간을 나타내며 질의 처리시에 기본 검색이 고급 검색보다 더 많은 시간이 소요되었는데, 이것은 사용자의 입력 값과 주석의 엘리먼트 값을 XPath와 XQuery를 이용하여 비교하는 고급 검색

의 경우 XML 문서에서 특정 노드의 값만 비교하는데 반해, 기본 검색은 한정된 부분이 아닌 트리 형태의 주식 문서에 존재하는 모든 노드의 값을 비교하기 때문이다.

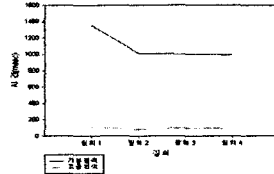


그림 7 검색별 질의 처리시간

표 1은 “빨간색 차 옆에 서 있는 사람의 이미지를 찾아라” 라는 질의에 따라 수행된 결과를 나타낸다.

Query	Relevant	Relevant	Precision
Red / Person	53	23	43.4
Red / Car	366	23	6.2
Red / Person & Car	35	23	65.7
Left or Right / Red / Person & Car	24	23	95.8

표 1 질의 결과

#### 5. 결론 및 향후 계획

본 논문에서는 사용자가 이미지와 비디오 데이터를 가지고 효율적인 검색을 하는 방안으로 멀티미디어 데이터에 대한 메타 데이터를 구축하고 좀 더 쉽고 편리한 인터페이스를 통한 XML 기반의 통합된 멀티미디어 주식 및 검색 시스템을 구현하였다.

향후 연구는 좀 더 자동화된 멀티미디어 데이터 분석 기법과 오디오를 포함한 대용량 멀티미디어 데이터의 효율적인 인덱싱 기법 등에 대한 연구와 XML 기반의 오디오 검색 시스템을 포함한 통합 검색 시스템을 구현하는 것이다.

#### 6. 참고 문헌

- [1] V. Ogle and M. Stonebraker, "Chabot: Retrieval from a Relational Database of Images," IEEE Computer, vol. 28, no. 9, pp. 40-48, 1995.
- [2] M. Flickner, et al., "Query by Image and Video Content: The QBIC System," Computer, vol. 28, no. 9, pp. 23-32, 1995.
- [3] H. Kosch, et al., "The SMOOTH Video DB - Demonstration of an integrated generic indexing approach," ACM Multimedia Conference, pp. 495-496, Los Angeles, USA, October-November 2000.
- [4] E. Hwang and S. Lee, "Spatial Similarity and Annotation-Based Image Retrieval System," IEEE 4th International Symposium on Multimedia Software Engineering, California, Dec. 2002.
- [5] E. Hwang and Y. Nam, "XVIDEO: XML Based Video Annotation and Retrieval," Proc. of Int'l Conf. on Internet Computing, Las Vegas, June 2002.