

XML기반 멀티미디어 검색시스템의 설계

윤미희*, 조동욱**

충북과학대학 컴퓨터정보학과

충북과학대학 정보통신학과

e-mail:mihee@ctech.ac.kr

Design of Multimedia Retrieval System based on XML

mihee yoon*, donguk cho**

*Dept of Computer Information Science, Chungbuk Provincial University of Science and Technology

**Dept of Information Communication Engineering, Chungbuk Provincial University of Science and Technology

요 약

컴퓨팅 기술의 발달 및 보편화로 인해 사용자들의 멀티미디어에 대한 요구가 증가하였고, 이러한 요구를 만족시키기 위해서는 단순한 텍스트 형식의 데이터가 아닌 멀티미디어 데이터, 특히 비디오 데이터에 대한 저장, 관리, 검색하는 기능이 필수적이다. 본 논문에서는 비디오데이터에 대한 효율적인 의미검색을 위해 주석기반 검색뿐만 아니라 특징기반 검색을 지원한다. 특히 사용자가 원하는 객체나 장면의 유사성 검색이 가능하며, 장면의 검색 결과로 제시된 장면을 선택한 후 선택된 장면을 기반으로 사용자가 원하는 좀 더 정확한 장면의 검색을 위한 SQBE(scene-query-by-example) 질의가 가능한 XML 기반 멀티미디어 검색시스템을 제안한다.

1. 서론

최근 인터넷 및 초고속통신망의 구축으로 인해 사용자들의 멀티미디어에 정보에 대한 다양한 요구가 증가하였고, 이러한 다양한 요구에 대해 효율적으로 서비스를 제공하기 위해서는 단지 텍스트 데이터뿐만 아니라 이미지, 오디오, 비디오 데이터를 처리할 수 있는 멀티미디어 데이터베이스 시스템을 구축하는 것이 필요하다.[1,2] 다양하고 방대한 비디오데이터를 효율적으로 저장, 검색하는 기능은 비디오 데이터베이스에서 가장 중요한 요소이다. 기존에 개발된 비디오 데이터베이스 시스템으로는 노르웨이 공대의 비디오스타[1], 일본 고베대학의 OVID[3], 미국 메릴랜드 대학의 AVIS[4]등이 있다.

비디오 정보는 방대할 뿐만 아니라, 텍스트나 수식 데이터와 같이 데이터 자체의 속성을 이용하여 처리하기 어렵기 때문에 비디오 데이터베이스에 저장된 객체들의 의미를 기반으로 한 내용기반 검색이 필요하다. 비디오 데이터베이스의 내용기반 검색을 제공하기 위해서는 비디오 데이터의 내용을 구조적으로

체계화한 데이터가 필요하다. 이러한 데이터를 메타 데이터라고 하며, 메타데이터는 내용기반 검색의 가장 필수적인 데이터이다. 그러므로 효율적인 내용기반 검색을 위해서는 메타데이터에 대한 모델링이 필요하다[5,6].

본 논문에서는 유사성 질의와 예제기반 장면질의를 지원하는 멀티미디어 검색시스템은 기존의 선행연구[7]를 바탕으로 원하는 비디오데이터의 검색을 위해 [7]에서 제안한 메타데이터 모델링을 기반으로 비디오 데이터의 내용기반 검색과, 장면과 객체의 유사성 검색이 가능하며, 특히 장면 검색 시 검색결과로 나온 장면을 기준으로 객체의 추가, 삭제를 사용해 더 정확한 질의 결과를 얻을 수 있는 SQBE(scene-query-by-example) 질의가 가능한 멀티미디어 검색시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 기존의 비디오데이터베이스 시스템에 대해 알아보고, 3장에서는 [7]에서 제안한 메타데이터 모델링을 XML 스키마로 변화하는 방법과 질의언어를 설

명하고, 유사성 질의와 SQBE 질의에 대해 설명한다. 4장에서는 장면과 객체에 대한 유사성 질의처리와 SQBE 질의처리에 대해 설명하고, 5장에서는 XML 기반 멀티미디어 검색시스템을 제안하고하고 6장에서는 결론 및 향후 연구과제에 대해 설명한다.

2. 관련연구

본 장에서는 기존의 비디오 데이터베이스 시스템에 대해 설명한다.

AVIS(Advanced Video Information System)[4]는 미국 메릴랜드 대학에서 개발된 비디오데이터베이스 시스템으로 비디오 내용에 나타나는 객체, 사건 행위 유형에 대한 메타데이터를 정의하고, 이들을 비디오 세그먼트와 연계시킴으로써 효율적인 검색 방법을 제안했다. 그러나 AVIS는 상위 레벨의 내용을 중심으로 하는 주석기반의 메타데이터만을 제공하고, 하위레벨을 위한 메타데이터가 없다는 단점이 있다.

Video-Star[1]는 비디오 데이터와 메타데이터의 공유와 재사용을 지원하도록 만든 범용 비디오 데이터베이스 시스템으로, basic context, secondary context 개념을 이용하여 데이터를 공유했다.

Venus[8]는 특징기반 검색을 위해 개발된 비디오 데이터베이스 시스템이다. 이미지 처리 기술을 사용하여, 비디오의 각 프레임에 나타나는 객체들의 시간 관계, 공간관계를 메타데이터로 구축하였다. Venus는 특징기반 질의언어인 CVQL(Content-based Video Query Language)을 제공한다. 그러나 비디오 내용에 대한 상위 레벨의 의미를 위한 메타데이터가 없으므로 다양한 질의를 구성할 수 없다.

3. 메타데이터 모델과 비디오 질의언어

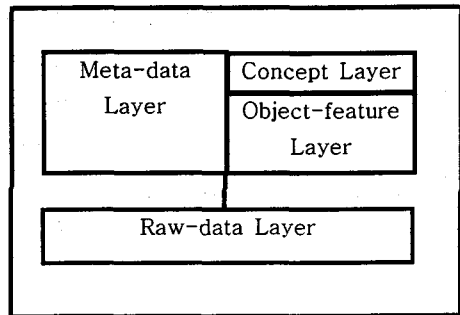
[7]에서 제안한 두 개의 레이어를 가진 통합 객체 지향 데이터 모델(Two layered Hybrid Object-oriented Metadata Model : THOMM)은 원시데이터 계층(raw-data layer)과 비디오의 의미검색을 위해 특징과 주석을 통합하는 메타데이터 계층(meta-data layer)으로 구성되며 모델링을 위해 객체지향 모델링 기법을 사용하였다.

한편의 비디오는 필름이론에 의해 시퀀스, 장면, 화면, 프레임으로 구성된다. 본 논문에서는 동일시간, 동일 장소에서 발생하는 화면의 집합인 장면을 처리의 기본 단위로 한다.

주석기반 질의와 시공간 질의 유사성질의, 장면을 기반으로 사용자가 원하는 좀 더 정확한 장면을 검

색하기 위한 SQBE 질의를 지원하기 위한 THOMM은[7]은 (그림 1)과 같다.

비디오의 물리적인 비디오스트림을 위한 원시데이터계층(raw-data layer)과 원시 데이터의 내용을 기술하기 위한 메타데이터계층(meta-data layer)으로 구성되며 메타데이터 계층은 다시 주석 기반 질의를 위한 개념계층(concept layer)과 특징기반질의를 위한 객체특징계층(object-feature layer)으로 구성된다.



(그림 1) 통합 객체지향 메타데이터 모델 (THOMM)

XML 스키마가 의 표기법은 제한한 업체마다 조금씩 다르기는 하지만 데이터의 형식 및 반복 횟수의 지정 등 그 자체가 XML 구조로 되어 있다는 것은 공통점이다[8].

객체들 사이의 관계는 4가지(복합, 집합, 상속, 연관)의 관계성을 가지고 있다. 이것은 XML 스키마에 의해 표현될 수 있다[9]. THOMM을 UML로 모델링한 후 [9]에서 기술한 XML 스키마 표현 방법에 의해 XML로 변경될 수 있다.

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 사용하여 관련된 항목들에 관계서는 key/keyref를 pair로 생성

해서 처리한다.

본 논문에서 제안하는 멀티미디어 검색 시스템에서 사용되는 비디오 질의언어의 구문은 다음과 같다 [7].

```
FIND <result>
FROM <meta-db>
WHERE <predicate_clause>
```

FIND절은 사용자가 검색하고자 하는 대상으로 비디오다큐먼트, 시퀀스, 장면, 객체가 될 수 있다. FROM절은 질의의 탐색영역을 정의하며 WHERE절은 검색조건을 정의한다.

유사 질의는 주어진 장면과 가장 유사한 장면이나 주어진 객체와 가장 유사한 객체를 검색하기 위해 메타데이터 계층을 사용한다. 유사질의는 객체특징 계층만을 이용한 색상이나, 공간의 유사 질의, 개념 계층과 객체특징 계층을 이용한 통합유사성 질의로 구분된다.

유사 질의는 주어진 장면과 가장 유사한 장면이나 주어진 객체와 가장 유사한 객체를 검색하기 위해 메타데이터 계층을 사용한다. 유사질의는 객체특징 계층만을 이용한 색상이나 공간의 유사성 질의, 개념 계층과 객체 특징 계층을 이용한 통합유사성 질의로 구분된다. 통합유사성 질의는 첫번째로 개념 계층을 사용하여 주석기반 검색을 수행한 후 나온 결과에 객체 특징 계층을 이용하여 특징기반 검색을 수행한다.

유사객체 검색은 주석기반에 의해 선택된 객체에 공간, 색상의 유사성 정도를 계산하여 임계치 이상의 유사성 정도를 갖는 객체를 검색한다. 색상과 공간에 의한 유사객체 검색은 주석기반이 아닌 특징기반의 유사성 검색을 위해 객체특징계층의 객체들에 대해 유사성의 정도를 검색한다. 가중치를 이용한 장면 검색은 객체의 타입에 따라 유사성의 정도를 계산하고 장면의 총 유사성 정도를 계산한다.

또한 주어진 장면을 기반으로 객체의 추가, 삭제들 통해 검색의 범위를 축소하고 확대하면서 사용자가 원하는 좀 더 정확한 검색을 지원하는 예제기반 장면 질의 (SQBE)가 가능하다.

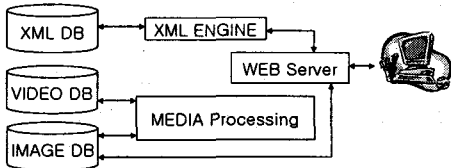
4. 질의처리

유사 질의는 주어진 장면과 가장 유사한 장면이나 주어진 객체와 가장 유사한 객체를 검색하기 위해 메타데이터 계층을 사용한다. 유사질의는 객체특징 계층만을 이용한 색상이나 공간의 유사성 질의, 개념

계층과 객체 특징 계층을 이용한 통합유사성 질의로 구분된다. 통합유사성 질의는 첫번째로 개념 계층을 사용하여 주석기반 검색을 수행한 후 나온 결과에 객체 특징 계층을 이용하여 특징기반 검색을 수행한다. 즉 개념 계층의 주석기반 검색에 의해 찾아진 장면에 대해 그 장면에 포함된 객체와 주어진 장면에 포함된 같은 객체 사이의 공간 관계, 색상 관계의 특징을 비교하여 유사성 검색을 수행한다.

5. XML 기반 멀티미디어 검색시스템

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 객성이나 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 웹 브라우저를 통해 브라우저된다. 비디오의 경우 계층적인 구조로 구성되었기 때문에 시퀀스, 장면, 샷을 기반으로 검색하고 추출할 수 있다. 멀티미디어 검색 시스템을 위한 구조도는 (그림 2)과 같다. XML DB는 컴퓨터에 의해 자동 추출된 이미지와 비디오의 특징과 사람에 의해 기술된 주석에 의해 전처리 과정에서 생성된 데이터베이스이다. 사용자 질의가 입력되면 XML 엔진이 XML 데이터베이스와 연동하여 해당하는 이미지나 비디오를 추출하고 추출된 결과를 사용자의 웹브라우저나 동영상 플레이어를 이용하여 보여주게된다. 비디오 데이터베이스와 이미지 데이터베이스는 원시 데이터가 저장된 곳으로 추출된 결과에 따라 해당 이미지가 있는 파일이나, 비디오의 해당 프레임을 보여주기 위해 액세스하고 그 결과를 미디어 처리과정을 거쳐 웹서버를 통해 사용자의 컴퓨터에 보여주게된다.



(그림 2) XML기반 멀티미디어 검색시스템의 구조

6. 결론

본 논문에서는 주석기반과 특징기반을 통합한 효율적인 의미검색을 지원하고, 장면과 객체의 유사검색과 예제기반 장면 검색을 지원하는 XML기반 멀티미디어검색시스템을 제안했다.

본 논문에서 제안한 비디오 검색시스템은 주석기반과 특징기반을 통합한 의미검색과 특징기반 검색이 가능하고 장면과 객체의 유사성 검색이 가능하다. 특히 장면 검색 시 검색결과로 나온 장면을 기준으로 객체의 추가, 삭제를 사용해 사용자가 원하는 더 정확한 질의 결과를 얻을 수 있는 SQBE 질의가 가능하다.

향후 연구과제로는 효율적인 검색을 위한 인덱싱 기법과 음성 인식 및 자연어 이해 기술을 통합하여 메타데이터를 생성하고 질의할 수 있는 기법의 연구가 필요하다. 또한 이렇게 생성된 메타데이터를 기반으로 웹상에서 실시간으로 원하는 질의결과를 얻을 수 있는 방법에 대한 연구가 수행되어야 한다.

참고문헌

- [1] R. Hjelsvold, "VideoSTAR - A database for Video Information Sharing," Ph. D. Thesis, Norwegian Institute of Technology, Nov. 1995.
- [2] John Z. Li, M. Tamer Ozsu, and Duane Szafron, "Querying Languages in Multimedia Database Systems," TR95-12, The university of Alberta, Canada, 1995.
- [3] E. Oomoto and K. Tanaka, "OVID: Design and Implementation of a Video-Object Database System," IEEE Trans. on Knowledge and Data Engineering, Vol.5, No.4, 1993.
- [4] Sibel Adali, et. al., "The Advanced Video Information System : data structures and query processing," Multimedia Systems, 1996.
- [5] S. Smoliar and H. Zhang, "Content-Based video Indexing and Retrieval," IEEE Multimedia, 1994.

[6] W. Klas and A. Sheth, "Metadata for Digital Media : Introduction to the special issue," SIGMOD Record, Vol.23, No.4, Dec. 1994.

[7] 윤미희외, "비디오의 의미검색과 유사성 검색을 위한 통합비디오정보시스템", 한국정보처리학회 논문지 제 6권 8호, Aug. 1999.

[8] 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.

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