

비디오의 의미검색과 예제기반 장면검색을 위한 비디오 검색시스템

윤 미 희[†] · 조 동 옥^{††}

요 약

비디오데이터를 효율적으로 처리하기 위해서는 비디오 데이터가 가지고 있는 내용에 대한 정보를 데이터베이스에 저장하고 사용자들의 다양한 질의를 처리할 수 있는 의미검색방법이 요구된다. 본 논문에서는 특징기반 검색과 주석기반 검색을 통합하여 다양한 사용자의 의미검색을 지원하고, 유사성 질의와, SQBE(scene query by example) 질의가 가능한 비디오 검색시스템(Video Retrieval System : VRS)을 제안한다. 사용자는 SQBE 질의를 통해 장면의 검색 결과로 제시된 장면을 기반으로 객체의 추가 삭제를 통해 사용자가 원하는 좀 더 정확한 장면의 검색이 가능하다. 또한 SQBE질의를 위한 질의언어와 이 질의를 처리하기 위한 질의처리알고리즘을 제안하고 장면과 객체의 유사성 검색에 대한 성능평가를 수행했다. 제안된 시스템은 Visual C++과 Oracle을 이용하여 구현되었다.

Video Retrieval System supporting Content-based Retrieval and Scene-Query-By-Example Retrieval

Mi-Hee Yoon[†] · Dong-Uk Cho^{††}

ABSTRACT

In order to process video data effectively, we need to save its content on database and a content-based retrieval method which processes various queries of all users is required. In this paper, we present VRS(Video Retrieval System) which provides similarity query, SQBE(Scene Query By Example) query, and content-based retrieval by combining the feature-based retrieval and the annotation-based retrieval. The SQBE query makes it possible for a user to retrieve scenes more exactly by inserting and deleting objects based on a retrieved scene. We proposed query language and query processing algorithm for SQBE query, and carried out performance evaluation on similarity retrieval. The proposed system is implemented with Visual C++ and Oracle.

키워드 : 멀티미디어(Multimedia), 데이터베이스(Database), 영상처리(Image Processing)

1. 서 론

최근 인터넷 및 초고속통신망의 구축으로 인해 사용자들의 멀티미디어 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 메타데이터라고 하고, 내용기반 검색의 가장 필수적인 데이터이다. 그러므로 효율적인 내용기반 검색을 위해서는 메타데이터에 대한 모델링이 필요하다[5, 6]. 메타데이터의 구성은 컴퓨터가 가지고 있는 내용에 대한 정보를 컴퓨터에 의해 자동적으로 추출한 후 검색에 이용하는 특징기반 검색방법(feature-based retrieval)[7]과 비디오가 가진 의미를 사람이 먼저 파악한 후 자연어를 이용하여 표현하고 검색에 이용하는 주석기반검색방법(annotation-based retrieval)이 있다. 특징 추출에 의해 자동적으로 생성된 데이터에 대한 직접적

[†] 준 회 원 : 도립충북과학대학 컴퓨터정보과학과 교수

^{††} 정 회 원 : 도립충북과학대학 정보통신과학과 교수
논문접수 : 2001년 1월 16일, 심사완료 : 2002년 1월 8일

인 질의를 수행하는 것은 거의 불가능하므로 특징데이터를 사용자가 사용하기 위한 단어로 사상하여 메타데이터를 구성하고, 비디오가 가지고 있는 의미에 대해서는 사용자의 입력에 의해 메타데이터를 구성하여 검색에 활용한다면 효율적인 의미검색이 이루어질 수 있다. 비디오의 정보는 방대할뿐만 아니라, 내포된 정보의 내용들이 다양하여 기존의 데이터베이스 검색 방법으로는 사용자가 원하는 형태와 의미를 정확하게 처리하기 어려우므로, 비디오를 위한 새로운 검색기법이 필요하다. 다양한 사용자의 요구를 처리하기 위해서는 시공간질의, 색상질의, 유사질의와 비디오의 다양한 내용에 대한 질의를 위한 키워드 질의가 필요하다.

본 논문에서는 기존의 연구[8]를 확장하여, 특징기반 검색과 주석기반 검색을 통합하여 다양한 사용자의 의미검색을 지원하고 유사성 질의와, 결과로 제시된 장면을 기반으로 사용자는 더 상세한 질의를 위해 이 장면의 객체를 기준으로 관심 있는 객체의 추가 삭제들 통해 사용자가 원하는 더 정확한 장면의 검색이 가능한 비디오 검색시스템(Video Retrieval System : VRS)을 제안한다. 본 논문에서 제안하는 비디오 검색시스템은 세 개의 계층을 가진 통합 객체지향 메타데이터 모델(Three layered Hybrid Object-oriented Metadata Model)은 원시데이터 계층, 비디오의 의미검색을 위해 특징과 주석을 통합하는 메타데이터계층과 질의 재형성을 위한 의미계층으로 구성되며 모델링을 위해 통합 모델링 언어(UML)를 사용하였다. THOMM을 기반으로 원하는 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 예제기반 장면질의(Scene Query By Example : SQBE)가 가능하다. 본 논문에서는 유사성 질의를 위한 알고리즘과, SQBE 질의를 위한 질의처리 알고리즘을 제안한다. 또한 유사성 질의로 나온 결과를 가지고 장면과 객체에 대한 성능평가를 수행한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 기존의 비디오데이터베이스 시스템에 대해 알아보고, 3장에서는 세 개의 계층을 가진 통합 객체지향 메타데이터 모델에 대해 설명하고 이 메타데이터 모델을 기반으로 비디오를 검색하기 위한 질의언어를 설명한다. 4장에서는 장면과 객체에 대한 유사성 질의와 SQBE 질의를 처리하기 위한 질의 처리 알고리즘을 제시하고, 5장에서는 VRS의 구조를 설명하고 6장에서는 유사성 질의의 성능평가를 설명하고 7장에서는 구현 예를 설명하고 마지막으로 8장에서는 결론을 맺는다.

2. 관련 연구

VIDEO-STAR[9]는 비디오와 메타데이터의 공유와 재사용을 지원하기 위해 관계형 데이터베이스 모델을 사용한 데이터베이스 시스템이다. 메타데이터의 구성은 분할기법을 통해 비디

오데이터를 구조화하고, 구조화된 비디오데이터에 대해 people, location, event의 메타데이터를 구성하고, 이를 다시 기본, 일차, 이차 컨텍스트로 구분하여 재사용과 공유를 용이하게 하며, 고정된 속성을 통해 사용자가 쉽게 질의를 구성할 수 있다. 그러나 메타데이터가 people, location, event의 내용만으로 한정되어 있어 다양한 의미검색이 불가능하고, 질의를 처리하기 위해서는 비디오데이터베이스를 검색할 때 'Decompose', 'Map to composition'등의 대응연산을 수행하여 해당 비디오데이터와 메타데이터를 관련시켜야 하므로 질의처리가 복잡하다.

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

Venus[8]는 특징기반 검색을 위해 개발된 비디오 데이터베이스 시스템이다. 이미지 처리 기술을 사용하여, 비디오의 각 프레임에 나타나는 객체들의 시간관계, 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.

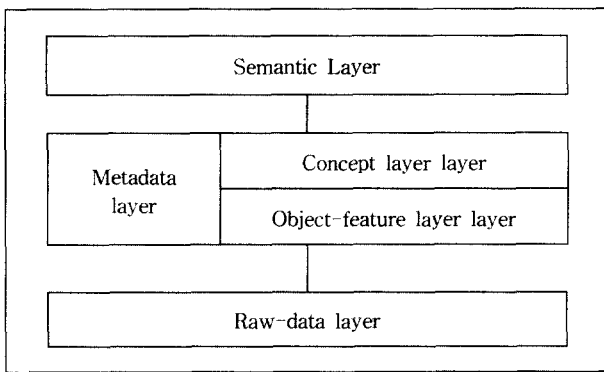
OVID[3]는 비디오데이터를 모델링하고 검색하기 위해 객체지향접근 방법을 사용한 객체지향 비디오데이터베이스 시스템이다. OVID는 의미 있는 장면을 비디오 객체라는 개념을 도입하고 각 비디오 객체마다 포함된 내용을 설명하는 속성과 속성값을 부여하여 메타데이터를 구성하였다. OVID는 비디오객체를 검색하기 위해 Video SQL이라는 사용자 인터페이스를 제공한다. OVID는 비디오 객체모델이 스키마를 사용하지 않았기 때문에 사용자는 각 비디오객체마다 완전한 속성과 속성값을 지정해야 하고 비디오데이터의 구조가 일반화되어 있지 않아 질의형식을 지정할 수 없다.

3. 메타데이터 모델과 비디오 질의언어

본 논문에서는 비디오정보를 다루기 위한 비디오 분할과정과 추출된 특징들을 색인하는 과정은 포함하지 않고 추출된 특징들과 비디오 내용에 대한 데이터의 모델링을 수행한다. 본 논문에서 제안하는 세 개의 계층을 가진 통합 객체지향 메타데이터 모델(Three layered Hybrid Object-oriented Metadata Model : THOMM)은 원시데이터 계층, 비디오의 의미검색을 위해 특징과 주석을 통합하는 메타데이터계층과 질의 재형성을 위한 의미계층으로 구성되며 모델링을 위해 통합 모델링 언어(UML)를 사용하였다.

한편의 비디오다큐먼트는 필름이론(film theory)에 의해 시퀀스(sequence), 장면(scene), 화면(shot), 프레임(frame)으로 구성된다. TV뉴스를 예로 들면 "1998년 10월26일 KBS 9시 뉴스"가 하나의 완전한 비디오다큐먼트가 되고, 하나의 기사가 시퀀스가 되며, 시퀀스는 앵커맨이 나오는 장면에서 다음 기사가 나오기 전까지의 장면의 집합으로 이루어진다. 본 논문에서는 동일시간, 동일장소에서 일어나는 화면의 집합인 장면을 처리의 기본단위로 한다.

비디오를 위한 세 개의 계층을 가진 객체지향 통합 메타데이터 모델(TOHHM)은 (그림 7)과 같다. 각 계층에 대한 정의는 다음과 같다.



(그림 1) 세 개의 계층을 가진 통합 객체지향 메타데이터 모델

(정의 1) 원시데이터계층(Raw-data layer)
물리적인 비디오스트림을 표현하는 계층

(정의 2) 메타데이터계층(metadata layer)
비디오에 대한 정보와 비디오스트림에 포함된 의미를 기술하기 위한 계층으로 개념계층(concept layer)과 객체특징계층(object-feature layer)의 부계층(sub-layer)으로 구성된다.

(정의 2-1) 개념 계층(concept layer)
비디오 정보와 비디오스트림에 나타나는 추상적인 의미를 표현하기 위한 계층

(정의 2-2) 객체특징계층(object-feature layer)
장면에 나타나는 객체들의 특징을 기술하기 위한 계층

(정의 3) 의미 계층(semantic layer)
메타데이터계층에 나타나는 의미들 사이의 연관관계를 표현하는 계층으로 IF-THEN규칙의 형태로 표현된다. 본 논문에서는 IF-THEN규칙을 지식이라 칭하고 지식베이스를 구성한다.

메타데이터계층의 개념계층은 주석기반질의를 지원하고 객체특징계층(object-feature layer)은 특징기반질의를 지원한다. 주석기반질의는 한편의 비디오에 대한 질의나 기사의 질

의, 인물이나 사건과 같은 질의가 가능하고, 한 장면에 나타나는 관심 있는 객체의 특징질의는 객체특징계층을 이용하여 객체의 시간관계질의, 공간관계질의, 한 장면에 나타나는 객체들간의 시공간상관관계의 질의가 가능하다. 또한 메타데이터계층을 이용하여 유사성질의를 지원할 수 있다. 의미계층은 사용자의 주관적인 관점으로 쓰여진 질의를 처리하기 위한 계층으로 질의의 결과가 존재하지 않을 경우 이 계층의 지식을 사용하여 질의를 재형성 할 수 있다. 원시데이터계층은 원시비디오객체로 구성되고 메타데이터계층은 비디오, 시퀀스, 장면, Object, 객체특징 객체로 구성된다.

메타데이터 모델인 THOMM을 기반으로 본 논문에서 제안하는 비디오 검색시스템에서 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. 특히 장면에 대한 질의 시 주어진 장면과 유사한 장면을 검색하기 위한 유사성 구문과, 주어진 장면을 기반으로 객체의 추가, 삭제를 통해 검색의 범위를 축소하고 확대하면서 사용자가 원하는 좀 더 정확한 검색을 지원하는 예제기반 장면 질의(SQBE)가 가능하다. SQBE질의에 대한 BNF는 다음과 같다[1].

```

    <SQBE> ::= <scene_id> Insert <object_exp> | <scene_id> Delete
    <object_exp>
    
```

4. 유사질의

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 객체특징 계층을 이용한 통합유사성 질의로 구분된다. 통합유사성 질의는 첫번째로 개념 계층을 사용하여 주석기반 검색을 수행한 후 나온 결과에 객체 특징 계층을 이용하여 특징기반 검색을 수행한

1) 비디오다큐먼트, 시퀀스, 장면에 대한 BNF는 [8]에 설명되어 있다.

다. 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. 통합유사성 질의는 객체의 통합유사성 질의와 장면에 통합유사성 질의로 구성되며, 질의 처리시 주석기반 검색을 먼저 수행한 후 나온 결과에 대해 특징기반 검색을 수행한다. 장면에 유사성 검색시 사용자는 주어진 장면에 배경에 더 많은 관심이 있을 수도 있고 또는 사람이나 동물과 같은 객체에 더 많은 관심이 있을 수도 있다. 이런 사용자의 관점을 고려하기 위해 각각의 객체가 유사성에 미치는 영향을 가중치로 표현하고, 장면에 유사성 정도를 계산하는데 사용한다. 장면에 있는 i 번째 객체의 총 유사성의 정도는 다음과 같다.

$$similarity(O_{qi}, O_{si}) = \frac{similarity_{color} * W_{ci} + similarity_{spatial} * W_{si}}{W_{ci} + W_{si}} * W_{bi}$$

(단 O_{qi} 는 장면에의 한 객체, O_{si} 는 검색대상 장면에의 한 객체, W_{bi} 는 배경이나 개체에 대한 가중치, W_{ci} 는 색상의 가중치, W_{si} 는 공간의 가중치)

[8]에서 제안한 유사성 식을 기반으로 객체와 장면에의 유사성 검색을 위한 알고리즘은 다음과 같다. Search_similarity 알고리즘은 주석기반과 특징기반을 혼합한 유사성 질의 처리와 특징기반의 유사성 처리가 가능한 유사성 알고리즘이다.

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 장면에의 총 유사성 정도를 계산한다.

유사질의는 특징기반 검색만을 수행하는 유사질의와 주석기반과 특징기반 검색을 혼합하여 사용하는 통합 유사성 질의로 나뉜다. 통합유사성 질의는 개념 계층을 이용함으로써 검색영역과 검색시간을 줄일 수 있다. 또한 객체특징 계층을 통하여 유사성 정도를 계산하고 사용자의 관심도를 고려함으로써 사용자가 원하는 가장 유사한 장면, 객체의 검색이 가능하다.

```

Algorithm Search_similarity
if (객체의 유사성 질의)
  if (주석기반 유사성 질의)
    유사 객체 검색
  else // 특징기반 질의
    색상과 공간에 의한 유사 객체
else if (장면에의 유사성 질의)
  result_list = ∅
  if (주석기반 유사성 질의)
    for (주석기반에 의해 선택된 장면 Sq)
      가중치를 이용한 장면에의 유사성 검색
  else // 특징기반
    for (모든 장면 Ss)
      가중치를 이용한 장면에의 유사성 검색
    
```

5. SQBE 질의

SQBE 질의란 주석기반 질의의 결과로 생성된 장면을 기반으로 사용자가 원하는 좀 더 정확한 장면에의 검색을 지원하는 검색기법이다. 즉 사용자가 장면에의 속성과 속성값을 가지고 주석기반의 질의를 수행하고, 그 질의 결과로 얻어진 하나의 장면을 선택하면 그 장면이 SQBE의 질의장면이 된다. 선택된 질의장면을 기반으로 사용자는 객체의 추가와 삭제들 통해 사용자가 원하는 더 정확한 장면에의 검색이 가능하다. 객체의 추가 삭제시 사용자는 객체가 가지는 속성, 객체의 이름, 역할, 색상 위치의 속성값을 부여할 수 있다. 객체의 추가가 선택되면 기존의 장면이 가지는 모든 객체와 현재 추가된 객체를 가지고 있는 장면을 검색한다. 만일 검색 결과가 존재한다면 검색장면을 띄워 아이콘의 형태로 디스플레이하고, 디스플레이된 장면 중 하나의 장면을 선택하면 그 장면이 상영된다. 만일 객체의 삭제가 선택되면 현재의 장면에의 객체 중 삭제가 될 객체를 제외한 나머지 객체들만을 가지는 장면을 검색한다.

객체의 추가, 삭제시 선택된 객체의 평가가 필요하다. 첫째로, 객체의 이름이나 역할이 선택되면 객체클래스를 검색하고 해당되는 객체를 가진 장면과 질의장면과 같은 객체를 가진 장면에의 교집합을 구한다. 만일 교집합의 결과가 공집합이 아니라면 결과를 디스플레이한다. 둘째로 객체의 색상이나 위치가 선택되면 객체클래스와 객체특징클래스를 조인(join)하고 색상과 위치의 평가가 시작된다. 색상의 평가는 색상의 유사성 식을 기반으로 질의 객체와 유사한 색상을 가진 객체를 산출한 후 객체의 이름을 검색하고 위의 첫 번째 과정을 수행한다. 객체의 위치평가는 절대 좌표를 기반으로 화면의 중심좌표와 객체의 좌표를 비교하여 현재 주어진 위치와 일치하는 객체를 검색한 후 첫 번째 과정을 수행한다. 사용자는 SQBE를 사용하여 장면에의 질의결과로 나온 한 장면을 기반으로 객체의 추가 삭제를 통해 사용자가 원하는 더 정확한 장면에의 검색이 가능하다.

본 논문에서 제안하는 SQBE 알고리즘은 다음과 같다. SQBE 알고리즘은 장면에의 질의로 나온 결과 중 원하는 장면을 선택한 후 그 장면을 기반으로 원하는 객체를 추가 삭제할 경우 먼저, 제시된 객체를 평가하고 질의장면이 가지는 객체와 현재 추가 혹은 삭제될 객체를 이용하여 원하는 장면에의 검색하기 위해 Search_scene 알고리즘을 호출한다.

```

Algorithm SQBE
// 선택된 객체의 평가를 통해 객체의 이름 검색
if (object.name = "" or object.role = "")
// 객체의 이름과 역할이 선택되지 않았다면
if not (object.color = "" or object.position = "")
  evaluate color and spatial of object
  select object.name
  from meta-db
    
```

```

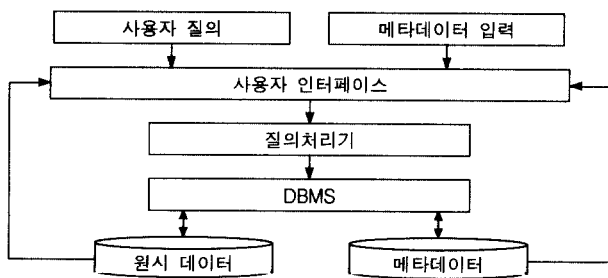
where object.id = object.feature.id and object.color = "selected
RGB" and
      object.position = "selected position"
endif
else if not (object.role = '') // 객체의 역할이 선택되었다면 객체의
이름 검색
      select object.name
      from meta-db
      where object.role = "selected role"
endif
// 선택된 객체 이름과 예제 장면을 기반으로 원하는 장면 검색
call Search_scene(object.name)
call Play_result
    
```

```

Algorithm Search_scene(object.name)
// 선택된 예제 장면인 scene과 객체 이름을 기반으로 원하는 장면 검색
if (insert)
  select scenel.who, scenel.what
  from meta-db
  where scenel.who = (scene.who union object.name)
else if (delete)
  select scenel.who, scenel.what
  from meta-db
  where scenel.who = (scene.who difference object.name)
endif
    
```

6. 비디오 검색 시스템의 구조

비디오 검색 시스템의 구조는 (그림 2)와 같다. VRS는 사용자 인터페이스, 질의처리기, 상용의 DBMS, 메타데이터베이스, 원시데이터, 지식베이스로 구성된다. 사용자는 사용자 인터페이스를 이용해 데이터의 입력과 질의를 수행한다. 메타데이터의 입력은 동영상 플레이어에 의해 플레이되는 비디오를 보면서 해당 속성에 속성값을 부여한다. 특히 색상과 위치는 해당 객체를 사각형의 영역으로 선택함에 의해 절대 위치가 계산되어 입력되고, 가장 유사한 색상을 색상테이블에서 선택함에 의해 그 색상의 RGB값이 입력된다. 원시데이터는 프레임으로 구성된 디지털 비디오이고, DBMS는 상용의 데이터베이스관리자로서 메타데이터베이스를 검색한다.



(그림 2) VRS의 구조

질의처리기는 사용자로부터 질의를 입력받아 질의를 분석한 후 상용의 DBMS에서 사용되는 SQL문을 생성한다. 상용의 DBMS는 해당하는 질의의 결과를 찾기 위해 메타데이터

베이스를 탐색하고 탐색결과를 플레이한다. 질의처리기는 질의분석기 질의어 생성기로 구성되어 있다. 질의어 생성기는 사용자의 질의를 입력으로 받아 상용의 DBMS에서 사용되는 SQL문을 생성하는 전처리기이다.

7. 유사성 질의의 성능평가

본 장에서는 본 논문에서 제시하는 장면의 유사성 질의와 객체의 유사성 질의에 대한 성능 평가의 결과를 제시한다.

7.1 객체의 유사성 성능평가

객체의 유사성 질의에 대한 성능평가를 위해 사용자의 관점에서 사용자의 평가기준으로 관련 있는 객체를 사용자가 미리 선택한 후, 유사성 질의에 의해 나온 결과들을 기반으로 recall과 precision을 계산하여 성능 평가를 수행했다. 질의 결과를 기준으로, recall과 precision은 사용자가 평가기준을 두고 미리 정의해 놓은 관련 있는 객체와 검색의 결과로써 산출된 객체의 집합으로 나뉜다. recall과 precision은 다음과 같이 정의된다.

(정의 1) Recall(객체)

사용자에 의해 미리 정의된 관련 있는 객체의 수와 검색된 객체의 수의 비율을 의미한다. 100% recall의 의미는 관련 있는 모든 객체를 검색했음을 의미한다.

Recall_{객체} =

$$\frac{\text{사용자에 의해 정의된 관련 있는 객체} \cap \text{검색된 관련 있는 객체}}{\text{사용자에 의해 정의된 관련 있는 객체}}$$

(정의 2) Precision(객체)

검색된 객체의 총 개수에 대한 검색된 관련 있는 객체의 수를 의미한다. 100% precision의 의미는 모든 검색된 객체가 관련이 있음을 의미한다.

Precision_{객체} =

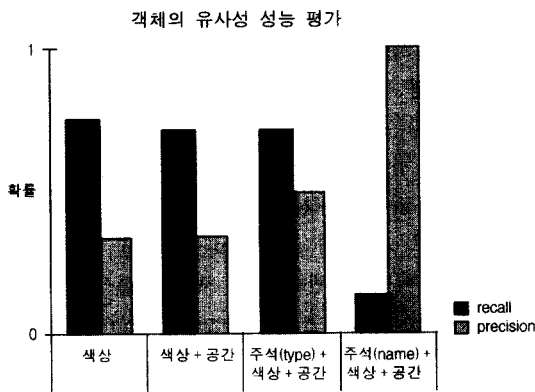
$$\frac{\text{사용자에 의해 정의된 관련 있는 객체} \cap \text{검색된 관련 있는 객체}}{\text{사용자에 의해 검색된 관련 있는 객체}}$$

객체를 기반으로 첫 번째 색상, 두 번째 색상과 공간, 세 번째 주석기반(type)과 색상, 공간, 네 번째 주석기반(name)과 색상, 공간을 가지고 검색한 결과를 나타내는 것으로 예를 들어 "검정 옷을 입고 왼쪽에 있는 사람을 찾아라"의 결과로 나온 객체 중 원하는 한 객체를 선택한 후 그 객체의 색상, 공간을 기준으로 유사질의를 수행한 결과는 <표 1>과 같다. 150개의 다른 특징을 가진 객체 중 사용자가 관련 있다고 미리 정의한 객체는 24개이다. 시스템에 의해 검색된 객체는 임계값인 색상의 유사성 확률이 0.6보다 크고, 공간의 유사성 확률이 0.6보다 큰 객체이다. <표 1>의 성능평가에 대한 차트는 (그림 3)과 같다. 즉 주석기반과 색상 공간을 가지고 유

사성 검색을 한 경우가 가장 정확도(precision)가 높다.

<표 1> 객체의 유사성 질의에 대한 성능평가

번호	주석기반	색상	공간	Recall	Precision
1	x	o	x	75%	33%
2	x	o	o	71%	34%
3	사람	o	o	71%	49%
4	김종국	o	o	13%	100%



(그림 3) 객체의 유사성 평가결과

7.2 장면의 유사성 성능평가

장면의 유사성 질의에 대한 성능평가를 위해 사용자의 관점에서 사용자의 평가기준으로 관련 있는 장면을 사용자가 미리 선택한 후, 유사성 질의에 의해 나온 결과들을 기반으로 recall과 precision을 계산하여 성능 평가를 수행했다. 장면의 유사성 질의에 대한 성능평가는 색상, 색상과 공간관계, 통합 유사성관계, 배경과 객체의 가중치 값을 달리 했을 경우를 나누어 성능을 평가한다. 질의 결과를 기준으로, recall과 precision은 사용자가 평가기준을 두고 미리 정의해 놓은 관련 있는 장면과 검색의 결과로써 산출된 장면의 집합으로 나뉜다. recall과 precision은 다음과 같이 정의된다.

(정의 1) Recall(장면)

사용자에 의해 미리 정의된 관련 있는 장면의 수와 검색된 장면의 수의 비율을 의미한다. 100% recall의 의미는 관련 있는 모든 장면을 검색했음을 의미한다.

$$Recall_{\text{장면}} = \frac{\text{사용자에 의해 정의된 관련 있는 장면} \cap \text{검색된 관련 있는 장면}}{\text{사용자에 의해 정의된 관련 있는 장면}}$$

(정의 2) Precision(장면)

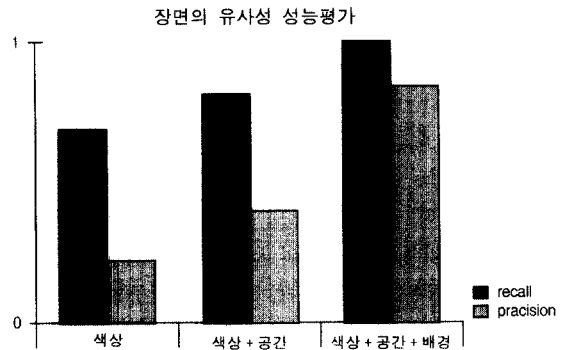
검색된 장면의 총 개수에 대한 검색된 관련 있는 장면의 수를 의미한다. 100% precision의 의미는 모든 검색된 장면이 관련이 있음을 의미한다.

$$Precision_{\text{장면}} = \frac{\text{사용자에 의해 정의된 관련 있는 장면} \cap \text{검색된 관련 있는 장면}}{\text{검색된 관련 있는 장면}}$$

총 107개의 장면 중 사용자가 관련 있다고 정의해 놓은 16개의 장면을 기반으로 첫 번째 색상, 두 번째 색상과 공간 주석기반 색상 공간 개체와 배경을 가중치를 가지고 검색한 결과를 나타내는 것으로 질의 “초록의 운동장을 배경으로 파란색 옷을 입은 포수가 임의의 위치에 나오는 장면을 검색하라”의 결과 중 하나의 장면을 선택한 후 유사성 검색을 한 결과는 <표 2>와 같다. (그림 4)에서 보이는 것처럼 색상 공간 배경을 기반으로 유사성 검색을 했을 경우 가장 좋은 성능을 나타낸다.

<표 2> 장면의 유사성 질의에 대한 성능평가

번호	주석기반	색상	공간	배경	Recall	Precision
1	o	o	x	x	69%	23%
2	o	o	o	x	81%	40%
3	o	o	o	o	100%	84%



(그림 4) 장면의 유사성 평가결과

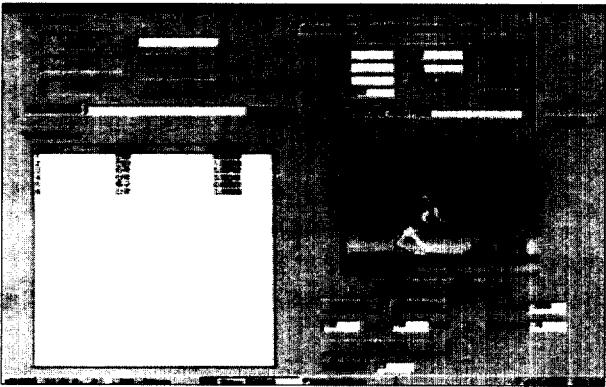
8. 구현 예

본 논문에서 제안한 비디오 검색시스템을 스포츠 뉴스로 도메인으로 구현했다. 장면 질의 “후구도예고스케가 중앙에 출현하고, 그 후에 미후라가 포수의 왼쪽에 위치하는 장면을 찾아라”의 검색 결과 중 한 장면인 (그림 5-1)의 장면을 선택한 후, 이 장면과 유사한 장면을 찾아라의 검색 결과 장면은 (그림 5-2)와 같다.

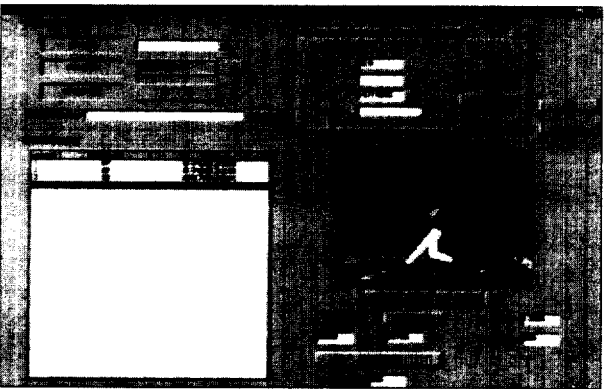
SQBE 질의를 위한 질의 장면은 (그림 6-1)과 같다. (그림 6-1)의 질의 장면은 SQBE 기준 장면을 위해 먼저 유사성 장면을 검색하기 위해 “후구도예고스케”가 나오는 장면을 질의 한 것이다. 사용자는 장면의 질의결과로 나온 한 장면을 기준으로 객체의 추가 삭제를 이용해 좀 더 사용자가 원하는 유사한 장면의 검색이 가능하다. 유사 장면 검색결과 한 장면을 선택한 후, 그 장면을 기반으로 “파랑색 옷을 입은 포수를 추가하라”는 SQBE 질의는 다음과 같다.

```
Find Scene
From meta-db
Where Scene-id Insert Object.role = "포수", Object.color = "파랑"
```

기준 장면을 기반으로 위의 SQBE 질의를 이용한 SQBE 검색 결과는 (그림 6-2)이다. 사용자는 SQBE 질의를 통해, 장면의 질의로 나온 결과를 기반으로 객체의 추가, 삭제를 이용해 원하는 더 정확한 장면의 검색이 가능하다. 본 비디오 검색시스템은 Windows NT 4.0에서 Visual C++ 과 Oracle 7을 이용하여 구현되었다.



(그림 5-1) 기준 장면



(그림 5-2) 유사성 장면



(그림 6-1) SQBE 질의장면



(그림 6-2) SQBE의 객체 추가

9. 향후 연구과제

본 논문에서는 주석기반과 특징기반을 통합한 효율적인 의미검색을 지원하고, 장면과 객체의 유사검색과 예제기반 장면 검색을 지원하는 비디오 검색시스템을 제안했다. VRS는 한편의 비디오를 필름이론에 의해 비디오다큐먼트, 시퀀스, 장면, 객체로 나누고 주석기반 검색과 특징기반 검색을 지원하기 위해 THOMM을 사용한다. 또한 유사한 장면과 유사한 객체를 검색하는 유사검색이 가능하다. 유사검색은 객체특징 계층만을 이용하는 특징기반 검색과 THOMM의 메타데이터의 두 부계층을 이용한 검색이 가능하다. 즉 개념 계층에 의한 주석기반 검색 후 객체 특징 계층에 의한 특징기반 검색을 수행하는 통합유사성 질의로 구별된다. 통합유사성 질의는 검색의 영역을 축소함으로써 검색시간을 줄일 수 있고 더 정확한 유사 질의 처리가 가능하다.

사용자에 의해 선택된 장면을 기준으로 좀 더 정확한 검색을 위한 SQBE 질의는 사용자가 기준 장면을 선택한 후 사용자가 원하는 객체의 추가 삭제를 통해 검색의 영역을 축소 삭제하면서 더 정확한 질의결과를 얻을 수 있다. 본 논문에서는 SQBE 질의를 위한 질의언어와 질의처리 알고리즘을 제안했다. 또한 객체와 장면의 유사성 결과의 성능 평가를 위해 recall과 precision을 사용하였다. 그 결과 주석과 특징을 통합한 유사성 검색이 더 좋은 성능을 가지고 있음을 알 수 있다.

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. 기존의 비디오 데이터베이스 시스템과의 비교는 <표 3>과 같다.

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 질

의결과를 얻을 수 있는 방법에 대한 연구가 수행되어야 한다.

〈표 3〉 시스템들간의 비교

시스템 질의유형	특정기반 질의		주석기반 질의		제안된 시스템	
	VENUS	KMeD	AVIS	OVID	Video- Star	VRS
객체기반 질의	가능	가능	가능	가능	가능	가능
사건기반 질의	불가능	불가능	가능	가능	가능	가능
시공간질의	가능	가능	불가능	가능	시간 질의만 가능	가능
색상질의	불가능	가능	불가능	불가능	불가능	가능
유사질의	불가능	가능	가능	불가능	불가능	가능
SQBE 질의	불가능	불가능	불가능	불가능	불가능	가능
분 석	특정기반 의료분야 에만 적용 검색	주석기반 의료분야 에만 적용 검색	질의 구성 과 처리가 어려움	공간 질의 불 가능 내용 한정	주석기반, 특정기반 질의, 유사 질의 처리 효율 높음 일 수 있음	

참 고 문 헌

[1] R. Hjelmsvold, "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] Tony C. T. Kuo and Arbee L. P. Chen, "A Content-Based Query Language For Video Databases," In Proc. of IEEE Multimedia '96, 1996.
 [8] 윤미희외, "비디오의 의미검색과 유사성 검색을 위한 통합비디오정보시스템", 정보처리학회논문지, 제6권 제8호, Aug. 1999.
 [9] R. Hjelmsvold, "VIDEO-STAR - A Database for Video Information Sharing," Ph.D. Thesis, Norwegian Institute of Technology, Nov. 1995.

윤 미 희

e-mail : mihee@ctech.ac.kr

1990년 숙명여자대학교 전자계산학과 졸업 (학사)

1992년 숙명여자대학교대학원 전자계산학과 (석사)

1999년 숙명여자대학교대학원 전자계산학과 (박사)

2000년~현재 독립충북과학기술대학교 컴퓨터정보과학과 교수

관심분야 : 멀티미디어 데이터베이스, 멀티미디어 시스템, 지식베이스, 인터넷 방송

조 동 욱

e-mail : ducho@ctech.ac.kr

1983년 한양대학교 공대 전자공학과(공학사)

1982년~1983년 (주)신도리코 기술연구소 장학생 겸 연구원

1983년~1985년 한양대학교 대학원 전자공학과(공학석사)

1985년~1989년 한양대학교 대학원 전자통신공학과(공학박사)

1991년~2000년 서원대학교 정보통신공학과 부교수

2000년~현재 독립충북과학기술대학교 정보통신과학과 교수

1996년 한국통신학회, 한국통신학회 충북지부 학술상 수상

1997년 한국통신학회 공로상 수상

1999년 한국통신학회, 한국통신학회 충북지부 학술상 수상

2001년 한국정보처리학회 우수논문상 수상

관심분야 : 영상처리, 영상 통신, 멀티미디어 시스템, 생체인증, 생명정보