

# MPEG-7 기반 의미적 메타데이터 모델을 이용한 멀티미디어 주석 및 검색 시스템의 개발

안 형 근<sup>†</sup> · 고 재 진<sup>††</sup>

## 요 약

최근 멀티미디어 정보의 양이 매우 빠른 속도로 증가함에 따라 멀티미디어 데이터에 대한 다양한 검색은 매우 중요한 이슈가 되고 있다. 멀티미디어 데이터를 효율적으로 처리하기 위해서는 멀티미디어 데이터가 가지고 있는 의미 내용을 추출할 수 있는 의미 기반 검색 기법이 필요하다. 기존 연구되어온 멀티미디어 데이터의 검색은 주석 기반 검색, 특징 기반 검색, 주석과 특징 기반 검색의 통합 검색시스템이 있다. 이러한 시스템들은 검색 데이터의 생성을 위해 주석자의 많은 노력과 시간을 요구하고 특징 추출을 위한 복잡한 계산을 요구하며, 생성된 데이터는 변화되지 않는 정적인 검색을 수행하는 단점이 있다. 또한, 인간에게 좀 더 친숙하고 의미적인 형태의 검색 방법을 제공하지 못하고 있는 실정이다.

본 논문에서는 MPEG-7을 이용하여 멀티미디어 데이터를 구조적으로 표현하고 효율적으로 추출하기 위한 의미적 메타데이터 기반의 멀티미디어 주석 및 검색시스템(S-MARS)을 구현 제안한다. 본 시스템은 멀티미디어 데이터에 대한 주석이나 검색, 브라우징을 위한 그래픽 인터페이스를 제공하며 멀티미디어 정보를 표현하기 위해 의미적 메타데이터 모델을 기반으로 구현하였다. 멀티미디어 데이터에 대한 의미적 메타데이터 모델은 MPEG-7 표준에 정의되어 있는 멀티미디어 묘사 스키마를 기반으로 XML 스키마를 이용하여 작성하였다. 결론적으로, 제시한 멀티미디어 데이터에 대한 의미적 메타데이터를 XML 형태로 표현하고, XML을 지원하는 데이터베이스 시스템을 이용하여 표준적인 데이터의 상호 교환이 용이하게 이루어질 수 있으며, 의미적 메타데이터를 활용하여 삽입 기반 검색 알고리즘 방법을 제공함으로써 검색에 대한 정확성과 사용자의 검색 만족도를 극대화시킬 수 있다.

키워드 : MPEG-7 묘사, 의미 주석, 메타데이터, S-MARS, 삽입 검색

## Development of Multimedia Annotation and Retrieval System using MPEG-7 based Semantic Metadata Model

An Hyoung Geun<sup>†</sup> · Koh Jae Jin<sup>††</sup>

### ABSTRACT

As multimedia information recently increases fast, various types of retrieval of multimedia data are becoming issues of great importance. For the efficient multimedia data processing, semantics-based retrieval techniques are required that can extract the meaning contents of multimedia data. Existing retrieval methods of multimedia data are annotation-based retrieval, feature-based retrieval and annotation and feature integration based retrieval. These systems take annotator a lot of efforts and time and we should perform complicated calculation for feature extraction. In addition, created data have shortcomings that we should go through static search that do not change. Also, user-friendly and semantic searching techniques are not supported.

This paper proposes to develop S-MARS(Semantic Metadata based Multimedia Annotation and Retrieval System) which can represent and extract multimedia data efficiently using MPEG-7. The system provides a graphical user interface for annotating, searching, and browsing multimedia data. It is implemented on the basis of the semantic metadata model to represent multimedia information. The semantic metadata about multimedia data is organized on the basis of multimedia description schema using XML schema that basically comply with the MPEG-7 standard. In conclusion, the proposed scheme can be easily implemented on any multimedia platforms supporting XML technology. It can be utilized to enable efficient semantic metadata sharing between systems, and it will contribute to improving the retrieval correctness and the user's satisfaction on embedding based multimedia retrieval algorithm method.

Key Words : MPEG-7 Description, Semantic Annotation, Metadata, S-MARS, Embedding Retrieval

### 1. 서 론

<sup>†</sup> 정 회 원 : 울산대학교 컴퓨터정보통신공학부 외래교수

<sup>††</sup> 정 회 원 : 울산대학교 컴퓨터정보통신공학부 교수  
논문접수 : 2007년 3월 24일, 심사완료 : 2007년 8월 23일

최근 비약적인 멀티미디어 기술의 발전과 함께 수많은 정보들이 디지털화 되었으며, 다양한 디지털 입력장치와 저장

장치의 발달로 인해 오디오, 이미지, 비디오와 같은 많은 양의 멀티미디어 데이터가 만들어지고 있다[1]. 또한 인터넷의 확산은 멀티미디어 데이터의 양적 팽창을 돕는데 큰 역할을 하고 있다. 따라서 찾고자 하는 데이터를 인터넷 상이나 데이터베이스에서 효율적으로 검색할 수 있는 시스템의 필요성이 증대되고 있으며, 멀티미디어 데이터가 포함하고 있는 고급수준의 의미적(semantic) 정보까지 자유롭게 표현할 수 있는 방법도 대두되고 있다. 멀티미디어 데이터를 사용하는 사용자의 요구를 충족하기 위해서는 방대한 정보를 가지고 있는 멀티미디어 데이터를 데이터베이스화하여 효율적인 유지와 관리가 필요하다[2]. 이를 위하여 멀티미디어의 제작 및 지식 정보뿐만 아니라 멀티미디어 데이터 자체를 다양한 방법으로 분석한 후 정보들을 추출하여 체계적으로 인덱싱하고 분류 및 표현하는 것은 무엇보다 중요하다[3, 4].

기존의 검색에는 사용되는 방법에 따라 내용 기반 검색(content-based retrieval)과 주석 기반 검색(annotation-based retrieval)으로 나누어진다. 전자는 멀티미디어 데이터가 가지는 다양한 시각적인 특징 정보를 데이터로 추출하여 저장하고 이를 바탕으로 사용자로부터 질의를 처리하는 방법이다[5]. 이 방법은 멀티미디어 데이터가 가지는 의미 정보의 표현이 어려울 뿐만 아니라 의미 검색의 결과 또한 부정확하다. 반면에 후자는 멀티미디어 데이터에 대한 의미 정보를 사용자가 직접적으로 분석하여 비정형적 텍스트나 정의된 형식의 메타데이터를 주석으로 작성하여 저장한 후에 질의 키워드와 주석되어진 텍스트를 비교하여 연관된 데이터를 찾는 방법이다[6]. 이 방법은 사용자가 표현할 수 있는 의미 정보를 텍스트의 주석으로 처리한다. 하지만 주석함에 있어 많은 시간과 노력을 요구하고 사용자의 주관적인 관점의 주석 및 질의 키워드로 인하여 의미적 차이(semantic gap)가 발생하여 검색 결과의 부정확성과 검색의 한계가 존재한다[7].

상기에서 제시된 방법에서 멀티미디어 데이터의 표현을 보다 자유롭게 할 수 있고 정확한 의미 검색이 가능하기 위하여 MPEG-7 표준을 채택하였다. MPEG-7은 상호호환성의 문제를 해결하기 위한 메타데이터의 표현과 다양한 응용 및 확장을 위하여 제정된 표준으로 오디오 · 비주얼 데이터를 표현하는데 유연하고 확장 가능한 프레임워크를 제공한다. 따라서 본 논문에서는 사용자의 편의성과 멀티미디어 데이터의 자유로운 표현을 위한 MPEG-7 기반 의미적 메타데이터를 이용한 멀티미디어 주석 및 검색시스템(Semantic-based Metadata Annotation & Retrieval System, S-MARS)을 구현 제안한다. 본 시스템에서 멀티미디어 데이터의 표현을 위한 주석은 구조적 형태의 멀티미디어 데이터가 포함하고 있는 객체를 작은 단위의 의미 엔티티인 노드(node)와 관계 속성을 링크(link)로 표현하는 작업을 선행연구[21]에서 정의하였으며, [21]에서 구현된 GUI 기반의 SMAT(Semantic Metadata Annotation Tool)은 노드와 링크를 도구에서 사각형과 선분의 도형으로 표현하고 해당 도형을 작업 화면에 드로잉하고 트리 또는 그래프로 디자인하여 멀티미디어 데이터를 표현하는 방법을 채택하고 있으며, 결과는 MPEG-7

기반 XML 문서로 저장되었다. 멀티미디어 데이터의 검색은 질의 키워드를 어떻게 선택하느냐가 검색 결과의 정확성을 결정한다. 기존의 주석 기반 검색은 텍스트 매칭 방법이기 때문에 사용자마다 일관되지 않는 질의 키워드를 이용하여 의미적 차이로 인해 부정확한 검색 결과를 보여준다. 본 구현시스템의 검색에서는 [21]의 분류 데이터베이스에 저장된 주석을 위한 의미 엔티티(event, time, person, place, object, etc)들을 검색의 질의 키워드로 재사용(reusability)을 하기 때문에 사용자마다 검색을 위한 질의 키워드 사용에 일관성을 가지며 검색의 정확성을 좀 더 높일 수 있다. 또한 사용자 편의성을 위하여 검색을 위한 사용자 인터페이스에서도 [21]의 드로잉 방법을 채택하여 질의 키워드를 생성하는 질의 트리를 이용하여 기술한다.

이 후 본 논문의 구성은 다음과 같이 전개된다. 1장 서론에 이어 2장에서는 관련연구로 기존 주석 도구와 통합 멀티미디어 시스템의 구현 내용에 대하여 간략하게 소개한다. 3장에서는 멀티미디어 데이터의 의미 정보를 새로운 의미적 메타데이터 모델을 통하여 표현하는 방법과 그에 따른 구성요소인 MPEG-7 묘사들에 대하여 살펴본다. 4장에서는 구현된 주석 및 검색 시스템과 그에 적용된 사용자 인터페이스 작업처리, 삽입검색 알고리즘에 의한 검색처리 방법에 대하여 기술한다. 5장에서는 실험을 통해 본 시스템에 대하여 평가를 하고, 마지막으로 6장에서는 결론을 통한 향후 연구방향을 살펴보고 끝을 맺는다.

## 2. 관련연구

멀티미디어 검색 시스템들은 다양한 연구들에 의해 설계 및 구현되어 왔다. 멀티미디어 데이터 검색은 다양한 정보를 이용하여 검색 수행하기 때문에 차별화 된 사용자 인터페이스의 구성도 검색 시스템의 매우 중요한 요소가 되기 때문에 현재까지 구현된 시스템의 사용자 인터페이스에 대한 관련기술을 본 장에서 설명한다.

멀티미디어 데이터의 주석만을 위한 대표적 도구로 IBM에서 개발된 VideoAnnEx[8,9]는 입력되는 멀티미디어 데이터를 연속적으로 선택하여 자유롭게 분해를 할 수 있고, 각 해당 세그먼트는 인터페이스를 통하여 주석 항목을 선택하고 비정형 텍스트(FreeTextAnnotation) 형태로 저장한다. 하지만 검색에 대한 기능이 없는 것이 단점이다. Ricoh Movie Tool[10]은 MPEG-7 표준을 따르는 비디오 내용 기술들을 대화적으로 생성하기 위한 도구이다. 사용자에게 편의성을 제공하는 빌트인 인터페이스를 통하여 구조적 기반의 MPEG-7 묘사들을 대화적으로 생성하지만, 멀티미디어 데이터의 특징 기반이기 때문에 추상의 고급수준 메타데이터를 가진 의미 묘사의 표현은 가능하지 않다는 단점이 있다. MPEG-7 Annotation Tool[11]은 EPFL 연구소에서 개발되었으며, 각 세그먼트에 사용자가 주석 할 수 있는 계층적인 시간분류를 제공하고 있으며, 비정형적 텍스트의 형태로 메타데이터 값이 저장된다.

통합적 멀티미디어 검색시스템에 대한 이전 연구의 대부분은 특징의 유사성 또는 키워드에 의한 텍스트 매칭 검색이다. 먼저 QBIC(Query by Image and Video Content)[12] 시스템은 IBM에서 개발한 정지영상 및 동영상 검색 엔진으로써 예제를 통한 유사도 질의 및 스케치에 의한 질의, 색상, 질감 등을 지원한다. 점차 기능이 향상되고 있으며 장면 전환 검색 기능과 스토리보드에 의한 비디오 요약 출력 기능도 첨가되어 있다. Chabot[13] 시스템은 POSTGRES 데이터베이스를 기반으로 멀티미디어 데이터의 단순 속성, 객체에 대한 설명 정보를 이용한 텍스트 질의와 색상 분석을 통한 내용질의를 지원한다. 또한 보다 정확한 멀티미디어 표현 방법을 위한 내용 기반 검색의 Virage[14] 시스템도 있다. 이 시스템은 텍스트, 정지 영상, 동영상 등을 검색할 수 있는 API를 갖추고 있으며 특정 멀티미디어 이미지를 위한 추가적인 모듈에 대한 확장이 용이하도록 제작되었다. VisualSEEK[15]는 인터넷에서 영상정보를 검색할 수 있는 사용자 인터페이스를 제공하고 있는 검색 엔진이다. 초기 화면에는 저장하고 있는 데이터들을 분류하기 위한 메뉴 화면이 뜨고 이를 통해 주제별로 검색을 할 수 있도록 해 놓았다. 색상과 공간 질의를 지원하며 이와 함께 질의 키워드를 입력하여 검색할 수 있는 SAFE라는 엔진을 포함하고 있는 시스템이며, 주로 색상의 히스토그램과 영역, 크기, 공간적 위치들을 추가적으로 사용하는 방법이다. MARS(Multimedia Analysis and Retrieval System)[16] 시스템은 멀티미디어 검색과 데이터베이스 관리를 통합하기 위하여 설계 및 개발되었으며, 다음과 같은 4개의 부 영역 카테고리인 멀티미디어 내용 표현, 멀티미디어 정보 검색, 멀티미디어 특징 인덱싱, 멀티미디어 데이터베이스 관리 등이 있다. 본 시스템은 내용 기반의 저장과 검색을 위해 고급수준의 상위객체 멀티미디어 정보를 지원하고, 특징에 따른 가중치 값을 조정하여 각기 다른 방법의 유사도 측정도 가능하다.

최근 연구에는 정확한 의미 검색을 위하여 메타데이터를 이용한다. 메타데이터를 구성하기 위해서 멀티미디어 데이터가 가지는 의미 정보를 객체로 표현하여 객체들 간의 관계를 이용함으로써 보다 정확한 검색이 이루어지는 연구도 진행되었다. 멀티미디어 데이터가 가지는 의미 정보의 사건, 객체 등에 대한 메타데이터를 정의하고 그러한 사건들 간의 관계를 형성하고 세그먼트 간의 연계가 이루어질 수 있도록 제안된 AVIS(Advanced Video Information System)[2], 메타데이터의 공유와 재사용의 용이함을 가지고 있어 질의를 쉽게 할 수 있는 VideoSTAR(Video Storage and Retrieval)[17], 객체들에 속성 값을 부여하여 메타데이터를 구성한 OVID(Object-oriented Video Information Database)[18] 등의 시스템이 있다. 또한 CREAM(Creating Relational Annotation-based Metadata)[19]은 전형적인 주석 프레임워크로 클래스와 관계 인스턴스를 포함하는 관계적 메타데이터를 구성한다.

상기에서 살펴본 대부분의 관련 시스템들은 멀티미디어 데이터를 표현하는데 특징 정보나 텍스트로 멀티미디어 데이터의 의미 정보를 표현하고 질의 키워드로 검색을 수행하

였다. 일부 시스템들은 멀티미디어 데이터를 효과적으로 표현하기 위하여 사건, 객체와 속성의 메타데이터를 정의하고 해당 값들은 텍스트로 표현하였다. 하지만 사건과 객체 서로 간의 관계를 이해하는데 어려움이 있으며, 또한 기존의 텍스트 입력방법은 표준의 텍스트 키워드가 없기 때문에 동일한 멀티미디어 데이터에 대한 사용자의 표현이 상이하여 일관되지 못하고 주관성에 의하여 많은 텍스트 키워드가 생성된다는 단점이 있다. 따라서 본 논문은 기존 시스템에서 멀티미디어 데이터의 사건과 객체, 관계 속성을 텍스트로 표현하는 방법보다 효율적으로 표현할 수 있는 시스템을 구현하는데 그 목적이 있다. 이러한 목적을 위하여 구현 시스템은 사용자들에게 익숙한 드로잉을 이용한 GUI 환경을 기본으로 하며, 또한 멀티미디어 표현 결과를 이해하기 쉽고, 작업에 사용하는 구성요소들을 확장 및 재사용할 수 있어야 한다. 그리고 모든 작업의 기본은 공통적인 공유를 위하여 표준성을 고려하여야 한다는 것이다.

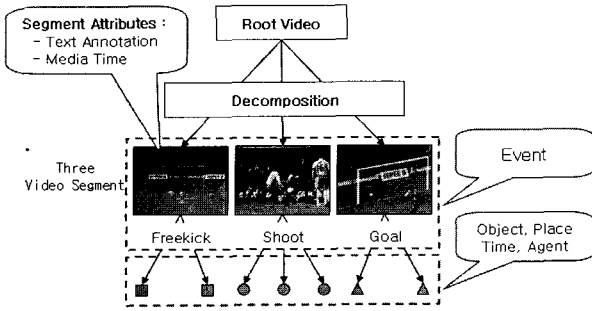
### 3. 멀티미디어 데이터의 의미표현

의미화는 멀티미디어 데이터의 논리적인 표현 방법으로 기존의 비정형 텍스트와 구조적 텍스트 메타데이터로부터 통계 및 마이닝에 의해서 새로운 의미 혹은 지식의 정보를 추출하고 표현한 것을 의미한다. 그리고 멀티미디어 데이터가 가지는 의미 정보를 효과적이고 체계적으로 구성한 것을 의미적 메타데이터라고 정의한다. 의미적 메타데이터를 위해서는 멀티미디어 데이터의 의미 정보를 일반화할 수 있는 추상의 개념이 필요하며 이를 통한 구조적 정보를 의미화하는 과정이 필요하다.

#### 3.1 멀티미디어 표현과 MPEG-7 묘사

멀티미디어 데이터의 의미적 메타데이터 표현으로 널리 사용될 수 있는 표준으로 MPEG-7을 들 수 있다. MPEG-7은 다양한 응용과 확장을 위하여 제정된 표준으로 오디오비주얼 데이터를 기술하는데 유연하고 확장 가능한 프레임워크를 제공한다. MPEG-7은 기술자(D, Descriptor), 묘사구조(DS, Description Scheme), 묘사정의언어(DDL, Description Definition Language) 및 기술을 인코딩하기 위한 방법을 표준화하며, 또한 MPEG-7 멀티미디어 묘사구조(Multimedia Description Scheme)에서 제공하는 메타데이터를 카테고리별로 총 8개의 파트로 구성하고 있다. 따라서 MPEG-7은 멀티미디어 데이터와 관련된 모든 메타데이터를 구조적으로 기술할 수 있으며, 또한 의미 정보를 기술하기 위한 규약과 형식을 제공하고 있다.

멀티미디어 데이터는 형식에 따라 그 특징이 상이하여 다양한 메타데이터를 구축하지만 일반적으로 멀티미디어 데이터의 구조적 표현은 계층적 트리 형태로 표현되며, 각 노드에는 이미지 카테고리 정보 또는 비디오 샷과 장면에 대한 정보를 포함시킨다. 이러한 정보에는 매체의 물리적 정보와 의미적 정보가 포함되며, 해상도, 컬러, 세그먼트의 시작과



(그림 1) 비디오 데이터의 구조적 트리

```
<Mpeg7>
<Description xsi:type="ContentEntityType">
<MultimediaContent xsi:type="VideoType">
<Video id="root video">
<!-- MediaTime description -->
<MediaTime>
<MediaTimePoint>T00:00:00</MediaTimePoint>
<MediaDuration>PT0M30S</MediaDuration>
</MediaTime>
<!-- Temporal Decomposition description -->
<TemporalDecomposition gap="false" overlap="false">
<VideoSegment id="Freekick">
<!-- TextAnnotation description -->
<TextAnnotation>
<FreeTextAnnotation>Freekick</FreeTextAnnotation>
</TextAnnotation>
<MediaTime>
<MediaTimePoint>T00:00:00</MediaTimePoint>
<MediaDuration>PT0M15S</MediaDuration>
</MediaTime>
</VideoSegment>
<VideoSegment id="Shoot">
<!-- TextAnnotation description -->
<!-- MediaTime description -->
</VideoSegment>
<VideoSegment id="Goal">
<!-- TextAnnotation description -->
<!-- MediaTime description -->
</VideoSegment>
</TemporalDecomposition>
</Video>
</MultimediaContent>
</Description>
</Mpeg7>
```

(그림 2) MPEG-7 묘사의 XML 문서

끝 시간, 샷에 대한 대표 프레임이 속하며, 의미적 정보에는 멀티미디어 데이터에서 얻을 수 있는 내용이 포함된다. 본 논문에서 다루게 될 내용은 MPEG-7 기반으로 서술된 주석 정보를 말하며, MPEG-7 표준 문서에서 멀티미디어 데이터에 대한 기술은 사건(event), 객체(object), 행위자(agent), 장소(place), 시간(time) 등으로 정의하고 이들의 조합에 의하여 나타낼 수가 있다. 멀티미디어 데이터에 대한 표현의 예시로서 (그림 1)은 비디오에 대한 세그먼트 트리를 보여주고 있으며, (그림 2)는 예시에 대한 기본적인 MPEG-7 묘사의 XML 문서를 보여주고 있다.

다양한 종류의 메타데이터는 주어진 멀티미디어 데이터에 관련시키기 위하여 MPEG-7 묘사의 많은 형식들을 사용한다. 묘사들 중에 구조 묘사, 의미 묘사, 분류 묘사 등이 본 논문에서 적용된 MPEG-7의 표준 묘사라고 할 수 있다.

- 구조 묘사(Structural Description)
 

주어진 멀티미디어 데이터의 구조적 계층 정보를 말한다. 예를 들어 멀티미디어 데이터와 관련된 특징(feature)을 가지는 샷(shot)의 시간적인 분해나 객체들의 분해를 말한다.
- 분류 묘사(Classification Description)
 

멀티미디어 데이터를 적절하게 기술하기 위한 특정 도

메인 영역에 있는 의미적 엔티티들의 집합을 말한다. ClassificationScheme(CS)으로 정의된다.

- 의미 묘사(Semantic Description)
 

의미적 엔티티들 서로 간에 연관된 관계집합으로 구성된 것을 말한다.

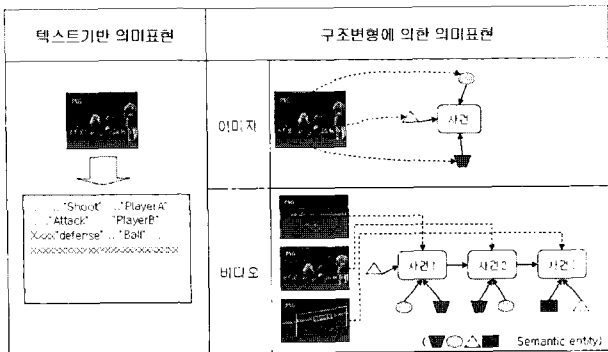
MPEG-7 구조 묘사 및 분류 묘사는 의미 묘사를 생성하기 위하여 조합되어지며, 후에 의미적 메타데이터로 구성되고 주석되어진다. 또한 의미 묘사 생성을 위한 작업은 사용자 인터페이스를 통하여 트리 또는 그래프 형태의 드로잉 작업으로 가능해진다.

### 3.2 멀티미디어 데이터의 추상과 변형

멀티미디어 데이터의 구조적 분해에 따른 의미 묘사를 위해서는 추상의 기술을 허용해야 한다. 추상은 멀티미디어 데이터에 적당한 값을 하나 또는 그 이상의 일정한 표현으로 적용되는 것을 의미한다. 본 논문에서는 크게 미디어 추상(media abstraction)과 형식적 추상(formal abstraction) 등으로 나누어 설명할 수 있다.

미디어 추상은 멀티미디어 데이터가 가지는 의미 정보를 직관적으로 표현하는 것을 의미한다. 예를 들어 “축구선수가 공을 차는 장면”, “프리킥 상황에서 심판이 호각을 부는 장면” 등과 같이 사용자에게 의해서 표현된 시간적인 의미 그 자체의 비정형 텍스트 값을 말한다. 형식적 추상은 표현하고자 하는 공통적 패턴에 도메인으로부터 여러 값들을 적용하여 표현하는 것을 의미하며 이러한 멀티미디어 표현 방법을 시스템에 적용하기 위한 기술이기도 하다. 공통적 패턴에는 값을 대입하기 위한 프레이스 홀더(placeholder) 또는 변수(variable)를 가지고 있으며, 해당 변수 부분에 주어진 도메인으로부터 값들을 적용하여 다양한 멀티미디어 표현이 이루어진다. 예를 들어 “어떤 사람이 공을 차는 장면”에서 “어떤 사람”은 바로 변수가 되는 것이며, 이 “어떤 사람”에 “축구 선수”, “어린 아이” 등과 같이 도메인에서 선택된 값으로 멀티미디어 표현이 이루어진다. “축구 선수가 공을 차는 장면”, “어린 아이가 공을 차는 장면” 등과 같이 “어떤 사람”에 적용되는 도메인 값에 의해서 다양하게 표현되어진다.

위 내용으로 기존의 멀티미디어 데이터의 의미적 표현은 사용자가 멀티미디어 데이터를 직접 분석함으로써 적절한 내용을 텍스트 형태로 입력하는 미디어 추상이 대부분이다. 미디어 추상은 객체들 간의 의미 관계를 표현함에 있어 너무 제한적이며 사용자마다 주관적이기 때문에 주석을 위한 키워드가 공통적인 표준을 가질 수가 없다. 이와 같은 방법은 사용자에게 많은 시간과 노력을 요구하며 동일한 멀티미디어 데이터에 대한 주석 및 질의 키워드가 일관되지 않고 표현이 부정확할 수가 있다. 반면에 형식적 추상은 객체들 간의 관계를 일관성 있게 해석할 수 있는 표준적인 모델을 정립하여 의미 정보를 자유롭고 정확하게 표현할 수 있다. 이러한 추상을 위해서는 주어진 멀티미디어 데이터를 작은 단위로 분리하고 각 단위는 의미 사건 엔티티와 의미 엔티



(그림 3) 멀티미디어 데이터의 의미 표현

티로의 의미 개념화할 필요가 있다. 실례로 이미지는 자체가 하나의 멀티미디어 데이터이기 때문에 하나의 사건으로, 비디오의 경우 여러 개의 세그먼트로 표현되어 다수 개의 사건들로 대응되도록 의미 개념화를 한다. 즉, 멀티미디어 데이터를 사건의 집합체라고 할 수 있으며, 시스템 측면으로 이러한 사건을 의미 사건 엔티티로 표현을 한다. 멀티미디어 데이터 표현은 각 의미 사건 엔티티들에 해당 의미 엔티티들과 관계 속성 등으로 즉, {의미 사건 엔티티}={의미 엔티티 : 객체, 행위자, 시간, 장소, 속성, 기타}와 같이 구성한다. 의미 엔티티는 다수 개를 사용하여 의미 사건 엔티티에 구성할 수 있다.

(그림 3)는 멀티미디어 데이터에 대한 의미 정보를 표현함에 있어 기존의 비정형적 텍스트 표현 방법과 본 논문의 시스템에서 제시하는 구조 변형에 의한 의미 사건 표현 방법을 보여주고 있다.

(그림 3)의 구조 변형에 의한 의미표현의 예시로 “프리킥 상황에서 선수A가 공을 차는 장면”의 이미지 데이터가 있으면 의미 사건 엔티티는 이미지 데이터가 가지는 중심 내용으로 정의할 수 있다. 본 예시에서는 “프리킥”을 의미 사건 엔티티로 정의하며 다른 이름으로 정의하여도 된다. 따라서, 물리적인 이미지 데이터는 논리적인 “프리킥”과 동일한 의미를 가지게 된다. 의미 사건 엔티티는 다른 의미 엔티티를 구성하는데 이러한 요소로 “선수A”, “공”, “차다” 등이 될 수 있다. 추가적인 의미 엔티티로 “심판”, “관중”, “상대 선수”, “잔디 색상” 등 다양하게 기술할 수 있다. 의미 표현의 결과는 {“프리킥”}={“선수A”, “공”, “차다”, “심판”,

“관중”, “상대 선수”)로 기술할 수 있으며 표현에 따라 의미 엔티티를 다양하게 구성할 수 있다.

### 3.3 MPEG-7 기반 의미적 메타데이터 모델

기존 시스템에서 자유롭게 표현하지 못한 멀티미디어 데이터의 의미적 표현을 기술하기 위해서 메타데이터 표준 구조를 변경하여 의미적 메타데이터 모델로 구성한다. 의미적 메타데이터 모델은 MPEG-7 기반이며 정확하고 빠른 결과를 얻기 위하여 구조적인 XML 문서로 저장 처리한다.

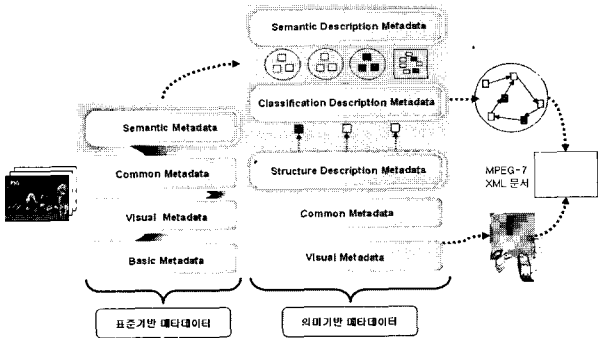
멀티미디어 데이터에 대한 메타데이터 표준 구조는 대부분이 크기와 비율, 타입 등과 같이 저급수준의 기본적인 특징 데이터의 기본 메타데이터. 전체적인 평균 색상이나, 질감, 일부 추출되어진 영역, 모양과 같은 특정 추출의 비주얼 메타데이터. 주제, 제목, 저자, 생성일 등과 같이 DC(Dublin Core) 표준 안을 사용하여 기술한 공통 메타데이터. 개념 혹은 의미 정보나 지식 정보 등과 같은 비정형 텍스트 기반으로 기술된 의미 메타데이터와 같이 4개의 메타데이터 영역으로 구성된다. 메타데이터 표준 구조에서의 의미 표현은 오직 비정형 텍스트로만 표현 가능하여 사용자들의 주관적인 표현에 의해 상이한 의미를 부여할 뿐만 아니라 표현이 자유롭지 못하며 검색에서 부정확한 결과를 얻을 수가 있다. 이러한 문제점을 위하여 모든 사용자들이 멀티미디어 데이터의 의미 표현을 하기위해 사용할 수 있는 구성 요소들을 공유할 수 있어야 하며, 그러한 요소들이 표준을 따르고 추가적인 확장을 자유로이 할 수 있는 새로운 구조 변형이 필요하다. 따라서 메타데이터 표준 구조를 MPEG-7 기반 의미적 메타데이터 모델로 변형하여 시스템에 적용을 한다. 해당 모델에 적용하기 위해서는 멀티미디어 데이터의 의미 정보를 일반화할 수 있는 추상의 개념과 구조적 변형이 필요하며 형식적 추상을 적용하게 된다.

본 논문에서 이용할 의미적 메타데이터 모델의 구성을 간략하게 살펴보면, 메타데이터 표준 구조에서 기본 메타데이터와 비주얼 메타데이터는 비주얼 특징 메타데이터로 통합을 하였으며, 의미 정보 표현의 정확성과 확장을 위하여 의미 메타데이터를 구조 묘사 메타데이터, 분류 묘사 메타데이터, 의미 묘사 메타데이터 등으로 세분화 하였으며, 각 메타데이터 영역은 <표 1>과 같이 기술할 수 있다.

의미적 메타데이터 모델은 고급수준의 의미 정보에 핵심

<표 1> 의미적 메타데이터 모델의 구조

영역별 메타데이터	기술 내용
비주얼 메타데이터	크기, 비율, 타입, 해상도, 컬러 등과 같이 MPEG-7 MDS(Part 5) 비주얼 기술자 요소들을 사용하여 기술하며, 추가적으로 디지털 장비의 EXIF 정보를 포함한다. 표준의 기본과 비주얼의 통합이다.
공통 메타데이터	표준 기반 메타데이터 구조의 공통 메타데이터 값에서 추가적으로 메타데이터에 대한 메타데이터를 제시하여 메타데이터 식별을 가능하게 하고 있다.
구조 묘사 메타데이터	비디오 세그먼트, 이미지 객체 등 멀티미디어 데이터의 형식에 따른 구조적 분해를 하고 그에 따른 사건의 의미를 부여한다. 또한 논리적으로 6W1H의 정형적 형식에 의한 기본 개념적 의미를 부여한다.
분류 묘사 메타데이터	의미 엔티티(6W1H에 의한 사건, 장소, 시간 등)와 의미 관계 속성 등과 같은 멀티미디어 데이터의 의미 정보를 위한 특정 키워드 집합을 말하며, 분류 데이터베이스로 저장할 하며 재사용 가능하다.
의미 묘사 메타데이터	멀티미디어 데이터가 가지는 의미 정보를 사건과 사건, 사건과 장소, 장소와 시간 등과 같이 추상의 수준에서 의미 엔티티들 서로 간의 의미 관계를 MPEG-7 표준안을 사용하여 기술한다.



(그림 4) MPEG-7 기반 의미적 메타데이터 모델

을 두고 있기 때문에 표준 기반 구조의 의미 메타데이터 영역을 확장하고 재사용하기 위하여 보다 구조적으로 세분화하였으며; (그림 4)의 음영 부분은 MPEG-7 표준의 프레임 워크를 이용한 의미 구조의 확장 부분을 보여주고 있다.

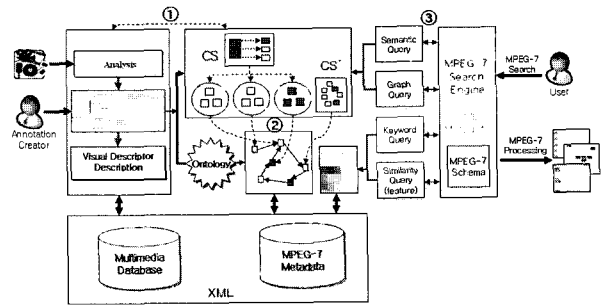
#### 4. 멀티미디어 주석 및 검색 시스템의 구현(S-MARS)

##### 4.1 전체적인 시스템 구조

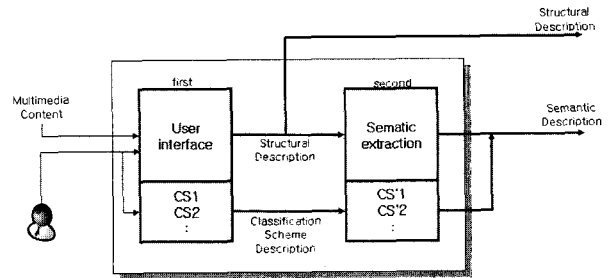
본 장에서는 MPEG-7 기반의 의미적 메타데이터 모델을 이용한 통합 멀티미디어 주석 및 검색 시스템(S-MARS)을 구현 소개한다. 기존의 인기 있는 시스템 및 도구에서 사용한 텍스트 표현의 애로점을 개선한 드로잉 기반의 대화형 그래픽 사용자 인터페이스를 채택하고 있으며 주석과 검색에 필요한 모든 의미 정보를 구조화하여 작은 단위의 의미 엔티티로 변형하여 데이터베이스화하였기 때문에 표준적인 작업이 이루어질 뿐만 아니라 재사용성과 검색을 위한 사용자에게도 정확한 결과 및 작업의 편의성을 제공하고 있다. (그림 5)는 의미적 멀티미디어 주석 및 검색 시스템의 전체적인 구조를 도식화하여 보여주고 있다.

MPEG-7 기반으로 표현된 의미적 멀티미디어 데이터를 검색하기 위해서 시스템의 전체적인 구조를 크게 의미 엔티티 생성부, 의미 모델 생성부, 의미 데이터 검색부 등과 같이 세 부분으로 나누어 설명할 수 있다.

- 의미 엔티티 생성부(그림 5-①) : 주어진 멀티미디어 데이터들은 MPEG-7 기반의 정형적인 메타데이터 형식에 의하여 자동 및 반자동 처리 표현되고, 또한 구조적 분해에 의한 의미 엔티티들은 온톨로지에 의해서 구분하고 재사용을 위해서 분류 데이터베이스에 저장한다. 자동 추출에 관련된 메타데이터는 저급수준의 컬러, 크기, 질감 등이며, 반자동에 의한 추출로는 의미 엔티티, 의미 속성과 같은 고급수준의 의미적 정보에 대한 처리이다. 또한 비디오 데이터인 경우는 분해를 통하여 세그먼트의 단위로 분해를 하고 추상화를 통하여 표현한다.
- 의미 모델 생성부(그림 5-②) : 전처리 과정에서 생성 저장된 의미 엔티티들을 이용하여 키워드가 아닌 트



(그림 5) S-MARS 시스템의 전체적인 구조



(그림 6) 의미 엔티티 추출을 위한 프레임워크

리 및 그래프 표현으로 주석 및 질의 모델을 생성한다. 이를 위한 인터페이스는 대화식의 사용이 간편한 그래픽 사용자 환경이며, 이렇게 구성된 최종 결과는 MPEG-7 묘사 기반이며 XML 문서로 저장된다.

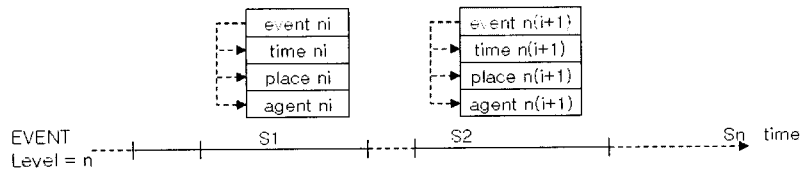
- 의미 데이터 검색부(그림 5-③) : XML 문서로 표현된 멀티미디어 데이터를 검색하기 위해 사용자 질의를 질의 트리 모델로 표현하고 구조적인 문서의 효율적인 검색을 위하여 삽입 검색 알고리즘을 적용 멀티미디어 데이터의 의미 검색이 가능하도록 한다. 질의 키워드의 의미적인 추론이 가능하도록 온톨로지를 구축하여 지원하고 있으며, 전처리 단계의 의미 모델 생성부에서와 같이 질의 트리 모델을 위한 사용자 인터페이스를 제공하고 있다.

##### 4.2 MPEG-7 기반 의미 엔티티 생성과 주석처리

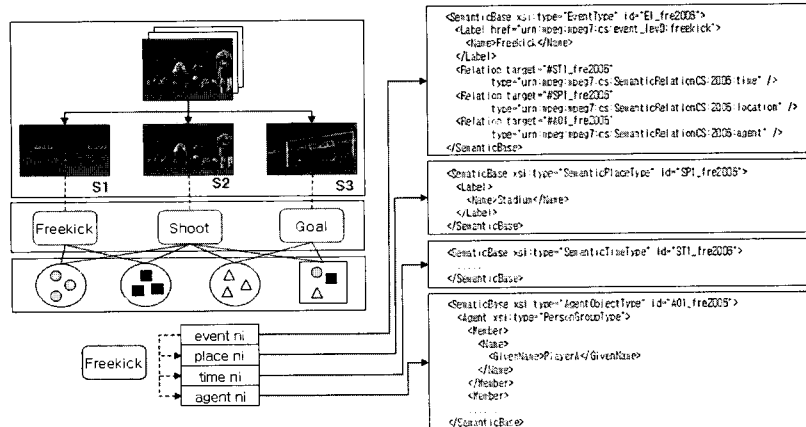
###### 4.2.1 의미 엔티티 생성처리

의미적 메타데이터는 앞 장에서 정의한 의미적 메타데이터의 구조 묘사, 분류 묘사를 이용한 의미 묘사의 관계 연결에 의해서 기술된다. 이러한 조합에 의한 의미 정보의 생성을 위해서는 구조적 분해와 추상화를 위하여 사용자 인터페이스 처리를 하게 되며 위의 (그림 6)에서 보여주고 있다.

구조 묘사 단계에서는 주어진 멀티미디어 데이터의 비주얼 기술자 추출, 세그먼트 단위의 분해, 물리적인 세그먼트의 의미 사건 추상화, 의미 사건 구성을 위한 의미 엔티티들의 추출, 추가적인 비정형 텍스트 메타데이터 작성 등의 처리를 한다. 본 단계에서는 주어진 멀티미디어 데이터를 시간적 분해에 의해서 n개의 서로 다른 세그먼트로 분해를 한다. 분해된 세그먼트는 물리적이며 논리적인 의미 정보를



(그림 7) 구조적 분해 후의 사건 의미화



(그림 8) 사건 의미화에 따른 의미 엔티티들의 XML 문서 표현

가지고 있지 못하다. 분해된 세그먼트는 의미 사건으로 추상화하고 실제 세그먼트들이 가지는 고유의 의미 정보를 적당한 함축적인 의미 사건 엔티티들로 일대일 대응 관계를 성립시킨다. 또한 의미 사건은 시간(<SemanticTimeType>), 장소(<SemanticPlaceType>), 행위자(<AgentObjectType>) 등과 같은 의미 엔티티들과 조합을 하여 멀티미디어 데이터를 표현하게 된다. 멀티미디어 데이터의 의미 정보를 표현하는 의미 사건 엔티티와 의미 엔티티들은 MPEG-7 기반 분류 묘사의 CS(Classification Scheme)로 정의하고 사용할 수 있으며 CS 데이터베이스를 구성하여 저장하고 온톨리지를 이용하여 분류할 수 있다. (그림 7)은 멀티미디어 데이터의 구조적 분해 후 의미 사건 엔티티와 의미 엔티티를 가지는 두 개의 세그먼트 구조(S1, S2)를 보여주고 있다. 각 세그먼트(S1, S2)는 의미 사건 엔티티이며 다른 의미 엔티티를 가지게 된다.

예로 세 개의 세그먼트(S1, S2, S3)로 분해가 되는 “축구 경기의 프리킥 상황에서 슛하여 골인하는 장면”의 멀티미디어 데이터가 주어졌을 때, 각 세그먼트 S1, S2, S3은 물리적이며 추상화를 통하여 세그먼트의 S1은 “Freekick”, S2는 “Shoot”, S3은 “Goal”로 일대일 대응하여 의미 사건 엔티티를 정의한다. 이러한 의미 사건들에는 다양한 의미 정보를 가진 의미 엔티티들을 추가적으로 구성하여 의미 사건과 관련시킬 수 있다. 의미 엔티티로는 “Referee”, “PlayerA”, “PlayerB”, “Ball”, “Whistle”, “Blow” 등과 같이 기본적으로 부여할 수 있고, 요구조건에 따라 시스템에서 새로이 추가적으로 생성하여 CS를 구성할 수 있다. 세그먼트 중의 의미 사건 “Freekick”의 멀티미디어의 구성요소는 {“Referee”, “PlayerA”, “Ball”},

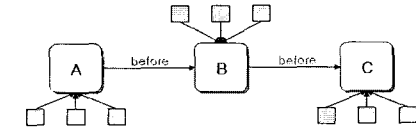
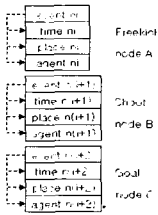
{“Referee”, “PlayerA”, “PlayerB”, “Ball”} 등 의미 엔티티를 어떻게 구성하느냐에 따라 다양하게 의미 표현을 할 수 있다. 그러나 의미 사건을 구성하는 의미 엔티티들의 값은 항상 동일하다는 것을 알 수가 있다. (그림 8)은 위 “Freekick” 의미 사건에 따른 의미 엔티티의 구성과 XML 문서를 보여 주고 있다.

#### 4.2.2 의미 묘사 표현과 XML 주석

의미 모델 생성 단계는 전처리에서 생성된 의미 사건 엔티티 및 의미 엔티티를 사용자 인터페이스를 이용하여 트리 또는 그래프로 디자인하여 주석과 질의를 표현할 수 있는 드로잉 단계이다. 생성된 분류 묘사의 의미 엔티티들은 노드로 표현하며, 의미 엔티티 서로 간의 관계 연결은 방향성을 가지는 링크로 정의한다. 모든 멀티미디어 데이터의 의미 표현은 노드와 링크의 상호 관계적인 연결을 통하여 생성되며 결과 문서는 MPEG-7 기반의 XML 문서로 저장된다.

예시로 상기에서 제시된 “축구(Soccer)” 멀티미디어 데이터에서 “Freekick(node A)”, “Shoot(node B)”, “Goal(node C)” 등의 세 개의 세그먼트를 의미 사건 엔티티들로 추상 표현하였으며, 의미 사건들은 의미 엔티티들로 구성되었다. 최종적으로 인터페이스에서는 관계(링크)를 통하여 의미 묘사를 표현하며 관계로는 의미 사건들 서로 간의 관계와 의미 사건 엔티티와 의미 엔티티들 간의 관계 등이 있다. 서로 간의 관계와 구성은 다음과 같이 전개 할 수 있다.

- “Freekick(node A)”는 “Shoot(node B)” 보다 이전에 발생되어야 한다(*before*).



(그림 9) 의미 묘사 표현

```
<Semantics>
  <Graph>
    <Node id="Freekick" href="#E1_fre2006"/>
    <Node id="Shoot" href="#E1_sho2006"/>
    <Node id="Goal" href="#E1_goa2006"/>
    <Relation source="#Freekick" target="#Shoot"
      Type="urn:mpeg:mpeg7:cs:TemporalRelationCS:2006:before" />
    <Relation source="#Shoot" target="#Goal"
      Type="urn:mpeg:mpeg7:cs:TemporalRelationCS:2006:before" />
  </Graph>
</Semantics>
```

(그림 10) 의미 사건들 서로 간의 의미기술 XML 문서

- “Shoot(node B)”는 “Goal(node C)” 보다 이전에 발생 되어야 한다(before).
- 각 의미 사건의 노드들은 의미 표현을 위하여 의미 엔티티들을 연결한다.
  - “Freekick(node A)” = {“PlayerA”, “PlayerB”, “Ball”}
  - “Shoot(node B)” = {“Referee”, “PlayerA”, “Ball”}
  - “Goal(node C)” = {“Referee”, “Whistle”, “Blow”}

위와 같이 전개된 사건들의 노드 표현과 사건과의 관계를 표현하면 아래 (그림 9)와 같이 표현할 수 있다(회색의 사각형은 의미 엔티티를 의미한다.)

MPEG-7 기반 위 표현을 위해서는 (그림 10)과 같이 묘사 구조의 <Graph> 태그를 사용한다(<Graph> 요소는 <Semantics> 요소의 자식 요소들 중 하나이다). 의미 사건 엔티티들은 Node 태그를 이용하며, “before”와 같은 관계 속성은 <Relation> 태그를 이용한다. 시간적 관계의 경우는 CS의 <TemporalRelationCS>를 사용하여 표현한다.

4.2.3 의미 주석을 위한 사용자 인터페이스 구현

주석을 위한 사용자 인터페이스는 주어진 멀티미디어 데

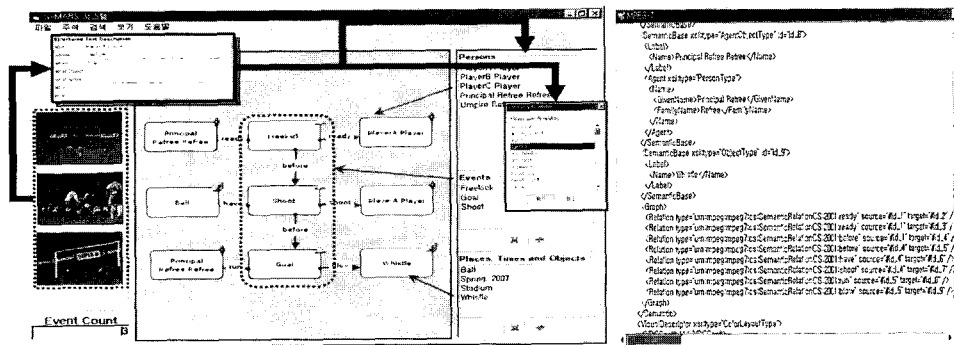
이터를 일차적으로 분석한 후 MPEG-7 기반의 비주얼 기술자 및 그 외 멀티미디어 데이터의 물리적 정보를 자동으로 추출하고, 사용자에게 의해 표현되는 의미 정보들은 메타데이터 입력기를 통하여 입력이 된다. 비디오의 경우는 세그먼트 단위의 분해 작업이 있으며, 이 후 메타데이터 입력기를 통하여 입력된 기본 정보들은 의미 엔티티들의 키워드들로 추출되어 분류 데이터베이스에 저장하게 되고, 사용자 인터페이스의 구성 요소인 의미 엔티티를 구성하게 된다. 사용자 인터페이스에서 의미 사건 엔티티와 의미 엔티티들은 사각형의 노드로서 표현되며 서로 관련된 노드들의 표현은 선분의 링크로 연결한다. 또한 노드들의 관계를 보다 자세히 표현하기 위하여 링크에는 속성 값을 부여하고 있다. 의미 엔티티들은 사용자 인터페이스를 통하여 삭제, 추가, 수정이 모두 가능하다. 이러한 의미 엔티티들은 멀티미디어 데이터의 의미 주석을 위하여 사용자 인터페이스에서 드로잉 기반으로 의미 주석을 생성한다. (그림 11)은 본 장에서 제시한 “Soccer” 멀티미디어 데이터의 의미 주석 작업 화면을 보여 주고 있다. 사용자 인터페이스 우측에는 분류 데이터베이스에 저장된 의미 사건 및 의미 엔티티의 목록이 있으며, 해당 목록의 요소를 드로잉하여 트리 또는 그래프로 표현하여 의미 주석을 생성하게 된다. 주석 결과는 XML문서로서 저장되면 (그림 11)에서 보여주고 있다.

4.3 의미적 멀티미디어 데이터의 검색

4.3.1 의미 검색을 위한 삽입 검색 알고리즘

MPEG-7 묘사는 멀티미디어 묘사구조의 인스턴스이며 XML 문서로 표현한다. 다양한 모델들이 멀티미디어 데이터를 기술하기 위하여 XML 문서들로 제안되었다. W3C에서는 DOM(Document Object Model)이라는 모델을 제시하였으며, 이 모델은 XML 문서를 순서 레이블 트리(ordered label tree)로 표현하였다.

따라서 본 논문에서는 트리 구조의 XML 문서를 효과적으로 검색할 수 있는 삽입 검색 알고리즘(Embedding Retrieval Algorithm)[20]을 멀티미디어 데이터의 의미 검색에 적용하였다. 삽입검색 알고리즘에서는 기본적인 두 가지의 조건을 제시하고 있다. 첫째는 전처리 과정으로 모든 트리의 노드 순행으로 후위순행(postorder)하여 노드를 정리하고, 상향식



(그림 11) 사용자 인터페이스를 통한 의미 주석 작업화면 및 주석 결과 XML 문서



```

Input : Q - a list of query nodes sorted in postorder
        D - a list of data nodes
Output : Boolean to indicate if a match is found
1: Match := True
2: i := 1 // i는 인덱스 포인터
3: while (Match = True and i <= |Q|) do { // |Q|는 질의 노드 수 }
4:   M(qi) := {}
5:   if(children(qi) = {} ) then { //질의트리에서 말노드 처리 }
6:     for all d ∈ leaves(D) such that (label(d)=label(qi) or label(d) ⊆ label(qi))
7:       do { // label(d) ≠ label(qi)이면 포함해야 한다. }
8:         M(qi) := M(qi) ∪ {d}
9:     end for
10:  else { //질의트리에서 루트노드 또는 내부노드 처리 }
11:    for all d ∈ D such that (label(d)=label(qi)) do
12:      S := {}
13:      for all qj ∈ children(qi) do
14:        S' := {}
15:        for all dm ∈ M(qj) such that (d is an ancestor of dm) do
16:          S' := S' ∪ {dm}
17:        end for
18:        S := S ∪ S'
19:      end for
20:      if S ≠ {} then
21:        M(qi) := M(qi) ∪ {S}
22:      end if
23:    end for
24:  end if
25:  if M(qi) = {} then
26:    Match := False // 검색처리 종료
27:  else
28:    i := i + 1
29:  end if
30: end while
31: return Match:
    
```

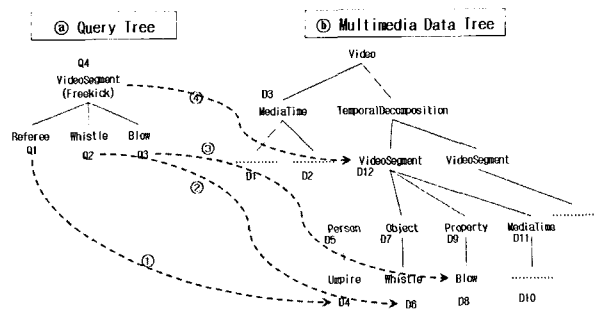
(그림 12) 의미 검색을 위한 삽입 검색 알고리즘

(bottom-up) 방법으로 처리를 해야 한다는 것이며, 둘째로 검색 처리에 있어 노드들 간의 관계(예, 조상노드-자식노드)와 개념을 항상 보존해야 한다는 것이다. 추가적으로 본 논문의 멀티미디어 데이터 검색을 위한 질의 키워드는 질의 트리로 표현하고 검색 처리 과정에서 노드 서로 간의 의미가 동등하거나 포함 관계의 규칙을 위하여 온톨로지를 이용한다.

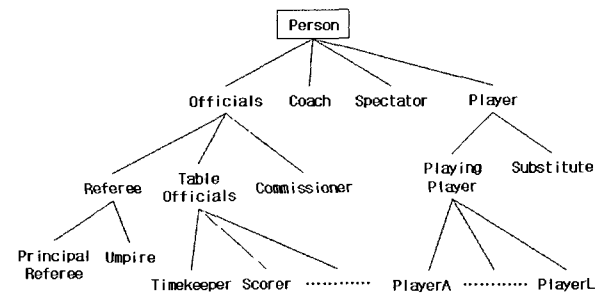
삽입 검색 알고리즘은 사용자 질의 트리에서 단말 노드(leaf node), 루트 노드(root node) 또는 내부 노드(internal node) 등과 같이 해당 질의 노드 위치에 따라 서로 다른 처리를 하게 된다. (그림 12)는 멀티미디어 데이터 검색을 위한 삽입 검색 알고리즘이다.

삽입 검색 알고리즘에서 Q는 사용자 질의 노드이고 D는 데이터베이스에 저장된 데이터노드를 의미하며 후위순회 처리하여 정렬되어 있다. 각각의 질의 단말 노드(qi)에 대해 알고리즘은 M(qi)라는 매칭 집합을 구성한다(line 5-9). 질의 단말 노드와 데이터의 단말 노드가 같은 값을 가지거나 데이터 단말 노드가 질의 단말 노드에 포함되는 값일 경우 매칭 집합을 구성한다(line 6). 이러한 포함 처리는 개념 값들 사이의 관계를 체크하는데 사용한다. 관련된 다른 노드들(내부노드와 루트노드)은 알고리즘에서 현재 질의 노드의 삽입 과정을 하나씩 구성한다. qi를 현재 내부 질의 노드라고 하고, 알고리즘은 qi와 같은 값을 가지는 모든 데이터 노드들을 추출하게 된다(line 11). 이후 qi를 루트로 하는 질의 서브트리를 d를 루트로 하는 모든 데이터 서브트리로 삽입 처리한다(line 13-19). 동일한 질의 노드에 대해서는 데이터 트리에 몇 번의 삽입으로 얻을 수가 있다.

알고리즘을 위한 예시로 "Soccer" 멀티미디어 데이터에서 "프리킥 상황에서 심판이 호각을 불다"라는 장면을 검색을 위해서는 다음과 같이 설명할 수 있다. 기존의 주석 기반 검색은 질의 키워드로 "프리킥 상황에서 심판이 호각을 불다"라는 텍스트를 입력하여 매칭 검색을 하였다. 질의 키워드에 따라 "프리킥 상황에서 부심이 호각을 불다"라는 텍스트를 입력할 경우 심판과 부심의 텍스트가 매칭이 되지 않



(그림 13) 삽입 검색 알고리즘 적용 예시



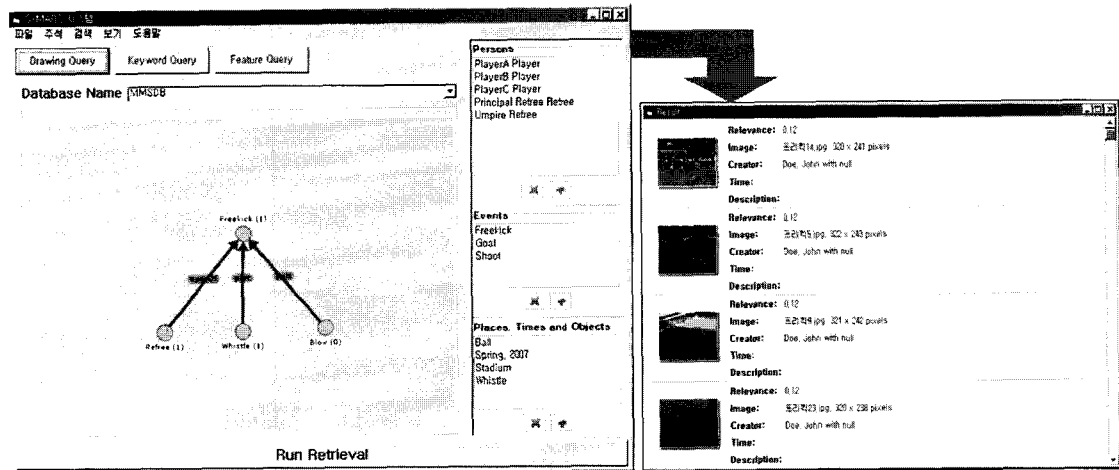
(그림 14) "Soccer"의 "Person" 역할 온톨로지

기 때문에 부정확한 검색 결과가 제시될 수 있다. 제안하는 시스템에서는 질의 키워드를 {Freekick : Referee, Whistle, Blow}와 같이 구성할 수 있으며, 해당 질의 노드(Q)는 (그림 13: ②)와 같이 질의 트리로 표현할 수 있으며, (그림 13: ⑥)는 MPEG-7 데이터베이스에 저장된 MPEG-7 묘사의 데이터노드(D)들이다. Q1, Q2와 Q3은 후위순회 정렬된 질의 노드의 순서번호이며, D1, D2에서 D12까지도 후위순회 정렬된 데이터노드의 순서번호이다. ①~④까지의 점선 화살표는 질의 노드와 데이터 노드 사이의 검색을 위한 매칭 처리를 의미한다. 이러한 매칭 처리에서 노드들은 조상-자식 관계 및 개념적인 의미 또한 보존을 해야 한다.

(그림 13)에서 ①과 같은 검색의 경우 질의 노드(Q1:Referee)와 데이터노드(D1:Umpire)는 서로 단어는 상이하지만 개념적인 의미는 동일하며 서로 포함 관계로 성립하게 된다. 이러한 의미적인 추론 개념을 검증하기 위하여 본 논문에서는 축구 경기에 대한 "Person" 역할 온톨로지를 (그림 14)와 같이 정의하고 있다. 이 결과 Q1의 매칭으로 M(Q1)은 {D4} 집합을 구성하게 된다. Q2, Q3, Q4의 경우 D6, D8, D12와 일치하기 때문에 D12를 루트로 하는 서브트리를 구성하고 매칭 집합을 구성하게 된다. 예시의 Q4에 대한 질의 처리 결과는 M(Q4)={D4, D6, D8, D12}와 같으며, 그 결과 멀티미디어 데이터를 검색의 결과 데이터로 한다.

#### 4.3.2 의미 검색을 위한 사용자 인터페이스 구현

S-MARS 시스템은 기존 검색 방법을 통합하기 위하여 주석 기반의 Keyword Query, 특징 기반의 Feature Query, 질의 모델 기반의 Drawing Query 등을 지원하고 있다. 본 논문에서 제안하는 검색은 Drawing Query를 통한 멀티미디어



(그림 15) 사용자 인터페이스를 통한 질의 모델 생성 작업화면 및 검색 결과

어 검색을 의미한다. 기존의 검색 방법인 Keyword Query 및 Feature Query에 대해서는 생략하며 Drawing Query에 대한 시스템 구현만 기술한다.

앞 장까지 “Soccer” 멀티미디어 데이터를 분석하여 사용자 인터페이스를 통하여 드로잉하고 주석 처리하였다. 검색을 위한 사용자 인터페이스에서도 마찬가지로 주석 처리에서 생성된 의미 엔티티를 질의 모델의 생성에 재사용한다. 질의 모델은 사용자 인터페이스를 통하여 노드와 링크로 표현되는 질의 트리를 의미하며 질의 키워드가 되는 것이다. 질의 모델 생성을 위한 사용자 인터페이스에서도 의미 엔티티는 추가, 수정, 삭제가 가능하다. (그림 15)는 “Soccer” 관련 멀티미디어 데이터를 저장하고 있는 멀티미디어 데이터베이스에서 “프리킥 상황에서 심판이 호각을 부는 장면”을 검색하기 위하여 사용자 인터페이스에서 의미 엔티티 {Freekick : Referee, Whistle, Blow}을 이용하여 질의 트리를 구성하는 작업 화면과 그 검색결과를 보여주고 있다. 주어진 질의에 대한 질의 모델은 루트노드는 사건의 “Freekick”으로 하며 단말 노드는 “Referee”, “Blow”, “Whistle” 등으로 구성되며 이러한 노드들은 사용자 인터페이스 우측에 있는 의미 엔티티를 재사용하여 드로잉 작업으로 할 수 있다. 의미 엔티티들은 주석 작업 과정에서 분류 데이터베이스에 저장되어 있다.

### 5. 실험 및 평가

본 장에서는 구현시스템에 적용된 삽입 검색 알고리즘의 드로잉 기반 검색이 기존의 텍스트 매칭을 이용한 주석 기반 검색에 비해 의미적 질의 키워드를 적용할 경우 얼마나 검색에서 정확성을 보이는지 실험을 통하여 평가한다. 본 실험은 펜티엄-IV 2.4Ghz CPU와 1GB 메인 메모리를 탑재한 윈도우 XP 환경에서 진행을 하였다. 실험 대상의 멀티미디어 데이터는 데이터의 손실과 전송률을 향상시키기 위해서 RAID 저장시스템에 저장하였다. 실험에 사용되는 멀티미디어 데이터는 [22]의 각종 스포츠에 관련된 비디오 클립

과 이미지로 구성되어 있으며, 데이터의 수는 이미지 500개와 비디오 클립 14편 정도를 대상으로 하였다. 비디오는 대략 8,720개의 키 프레임을 검출하였으며 또한 샷의 이미지 정보도 저장하고 있다. 실험의 모든 멀티미디어 데이터는 MPEG-7 기반의 구조적 XML 문서로 주석 저장되어 있다.

본 실험의 검색 개념은 “Soccer”에 관련된 키워드 중에 심판과 관련된 “Referee”, “Umpire”, “Principal Referee” 등이 내포하고 있는 동일한 개념적 관계를 파악하여 검색결과를 보여주는가 하는 것이다. (그림 14)에서 “Referee”는 “Umpire”와 “Principal Referee”의 부모 노드이며 포괄적 의미를 가지고 있으며 “Umpire”와 “Principal Referee”는 단말 노드이며 독립적이고 부모 노드에 포함되는 관계를 가지고 있다. 실험을 위하여 이미지와 비디오 데이터에 대해 상에서 기술하는 개념적인 관계를 포함하는 세 가지의 질의 키워드를 제시하고 기존의 주석 기반 검색 방법과 본 논문에서 제안하는 드로잉 기반 검색 방법에 적용하여 의미적 검색 결과를 비교 평가한다.

실험을 위한 세 가지 질의 키워드(Q)는 다음과 같다. “프리킥 상황에서 호각을 부는 장면”의 Q1:{Freekick, Whistle, Blow}, “프리킥 상황에서 주심이 호각을 부는 장면”의 Q2:{Freekick, Principal Referee, Whistle, Blow}, “프리킥 상황에서 심판이 호각을 부는 장면”의 Q3:{Freekick, Referee, Whistle, Blow} 등이다. 실험의 검색 결과에 따른 정확한 적합성을 평가하기 위하여 재현률(Recall)과 정확률(Precision)에 의한 성능평가 척도를 이용하여 실험하였다. 여기서 정확률은 시스템이 찾은 결과 중 정확하게 찾은 비율을 나타내고, 재현률은 전체 찾아야 할 결과 중 시스템이 정확하게 찾은 비율로 계산 된다. <표 2>, <표 3>은 질의에 대한 검색 후의 정확률과 재현률의 결과를 보여주고 있다.

실험 결과의 표에서 알 수 있듯이 대부분의 질의에서 기존의 주석 기반 검색보다는 제안 검색이 더 정확하다는 것을 볼 수 있다. 질의 Q1에서 주석 검색은 “Freekick”의 의미 “Soccer”와는 관계없이 “Whistle, Blow”와 텍스트 매칭된 모든 스포츠 멀티미디어 데이터를 검색하는 의미적 차이

〈표 2〉 이미지에 대한 질의 결과

질의	검색 방법	retrieved	relevant	relevant in DB	precision	recall
Q1	주석검색	71	39	42	54.93%	92.86%
	제안검색	43	41	42	95.35%	97.62%
Q2	주석검색	41	33	37	80.49%	89.19%
	제안검색	38	36	37	94.74%	97.30
Q3	주석검색	71	20	47	28.17%	42.55%
	제안검색	47	47	47	100.00%	100.00%

〈표 3〉 비디오에 대한 질의 결과

질의	검색 방법	retrieved	relevant	relevant in DB	precision	recall
Q1	주석검색	47	21	24	44.68%	87.50%
	제안검색	25	22	24	88.00%	91.67%
Q2	주석검색	21	9	13	42.86%	69.23%
	제안검색	12	12	13	100.00%	92.31%
Q3	주석검색	53	7	21	13.21%	33.33%
	제안검색	21	21	21	100.00%	100.00%

의 결과를 보였다. 반면에 제안 시스템의 검색은 드로잉을 이용하여 “Freekick”을 루트로 “Whistle”과 “Blow”를 자식 노드로 하는 질의 트리를 구성하여 검색한 경우 정확한 결과를 얻을 수가 있었다. 또한, 의미 개념적인 실험으로 Q2에서 주석검색의 경우 질의 키워드의 “Principal Referee”에 대해서만 텍스트 매칭이 존재하는 경우에 결과를 제시하지만, 제안 시스템의 검색은 “Principal Referee”와 동일한 의미를 가지거나 포괄관계의 “Referee”까지 의미 검색이 이루어졌다. Q3에서 제안 시스템의 검색은 주석 검색 방법과는 달리 질의 키워드인 “Referee” 뿐만 아니라 개념적으로 동일한 의미를 가지는 자식인 단말 노드의 “Principal Referee”, “Umpire”까지 의미 검색이 이루어졌다. 결론적으로 질의 Q2, Q3에서 “Referee”와 같은 포괄적인 의미 검색 질의를 할 경우가 그렇지 않은 경우보다 더 높은 재현율과 정확률을 보이고 있음을 알 수 있었다.

또 다른 의미 검색으로 “드리블하여 쏘는 장면”의 멀티미디어 데이터를 검색할 경우, “드리블”, “쏘”와 같은 키워드는 “Soccer”의 타 스포츠 종목에도 적용되는 키워드이다. 이러한 키워드를 질의 키워드 (“Dribble”, “Shoot”, “Ball”)로 하여 멀티미디어 데이터를 검색할 경우 “Dribble”과 “Shoot”의 의미를 포함하고 있는 농구, 핸드볼 등과 같은 타 스포츠 멀티미디어 데이터를 검색하게 된다. 이와 같이 스포츠 종목에서 사용하는 연관된 어휘의 키워드로 구성된 온톨로지에 의하여 공통된 키워드의 의미에 따른 멀티미디어 데이터에 대한 검색이 이루어질 수가 있다.

## 6. 결 론

본 논문에서는 다양한 의미 정보를 가지고 있는 멀티미디어 데이터를 자유롭게 표현하여 주석하거나 정확한 검색의 결과를 얻을 수 있는 시스템(S-MARS)을 구현 제안하고 성

능을 평가하였다. 본 시스템은 멀티미디어 데이터의 의미 주석을 위하여 MPEG-7 기반의 의미적 메타데이터를 이용하였으며, 해당 구조에 적용하기 위하여 주어진 멀티미디어 데이터를 세그먼트의 단위로 분해하고 해당 세그먼트를 논리적인 의미 사건 엔티티로 추상화하였다. 의미 사건 엔티티는 시간, 장소, 행위자, 객체 등의 의미 엔티티들과 서로 관련지어 멀티미디어 데이터의 의미적 메타데이터를 생성하였다. 이러한 일련의 작업은 대화식 사용자 인터페이스를 통하여 드로잉 작업으로 가능하며, 노드와 링크로 이루어진 트리와 그래프로 표현하였다. 결과 의미적 메타데이터는 MPEG-7 기반의 구조적 XML 문서로 저장하였다. 또한, 의미 주석 과정에서 생성된 의미적 엔티티들은 분류 데이터베이스에 저장되어 주석을 위하여 재사용이 가능하고 작업 상황에 따라 추가, 삭제, 수정이 가능하다. 멀티미디어 검색에서는 주석과정에서 생성된 의미 엔티티를 재사용함으로써 검색을 위한 사용자들의 일관된 질의가 이루어지므로 기존 시스템에서 상이한 질의 키워드의 사용으로 인한 부정확한 검색결과를 해결할 수가 있었다. 질의를 위한 질의 키워드는 사용자 인터페이스를 통하여 질의 모델을 생성하였으며, 질의 모델은 사용자 인터페이스를 통하여 의미 엔티티를 드로잉하여 질의 트리로 표현하였다. 검색방법으로 MPEG-7 기반으로 표현된 멀티미디어 데이터 기술에 대한 추론을 위하여 도메인 온톨로지 확장된 삼입 검색 알고리즘을 이용하였다. 해당 알고리즘은 모든 노드를 후위순행으로 하였으며 상향식의 방법으로 처리하였다. 질의 트리의 단말 노드와 데이터 노드 사이의 매칭을 우선으로 하였으며, 내부 노드를 처리를 하여 매칭 값을 결과로 하는 멀티미디어 데이터의 검색이 이루어졌다. 실험 결과에서 알 수 있듯이, 검색 키워드를 포괄적인 의미의 질의를 할 경우가 그렇지 않은 경우보다 정확률과 재현율이 더 높아짐을 알 수 있었다. 또한 구현 시스템은 주석과 검색을 위한 작업 환경이 대화식 사용자 인터페이스의 드로잉 기반 작업으로 가능하므로 사

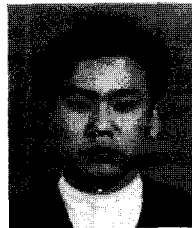
용자에게 편의성을 제공해 주며, 이러한 과정에서 의미적 엔티티를 데이터베이스로부터 재사용할 수 있으며 상황에 따라 의미 엔티티를 표준에 의한 확장을 할 수 있다는 특징을 가지고 있다.

향후 연구로 멀티미디어 질의 결과를 평가하기 위한 의미적인 유사성 측정 방법을 연구 계획하고 있으며, 이러한 유사성의 결과 값을 기반으로 전체적인 멀티미디어 데이터의 유사도 맵을 시각적인 결과로 제시할 수 있는 인터페이스를 연구하여 구현할 계획에 있다.

### 참 고 문 헌

- [1] Dongge Li, Ishwar K. Sethi, Nevenka, Thomas McGee, "Classification of general audio data for content-based retrieval," Pattern Recognition Letters, Vol.22, No.5, pp.533-544, 2001.
- [2] Sibel Adali, Kasim S. Candan, Su-Shing Chen, Kutluhan Erol, and VS Subrahmanian, "The Advanced Video Information System : data structure and query processing," Multimedia System, pp.172-186. 1996.
- [3] Baeza-Yates and Ricardo A., "Modern Information Retrieval," Addison-Wesley, 1999.
- [4] Nevenka Dimitrova, Avidesh Zakhor, and Thomas. Huang, "Applications of video-content analysis and retrieval," IEEE Multimedia, Vol.9 No.3 pp.42-55, 2002.
- [5] M. Flickner et al., "Query by Image and Video Content : The QBIC System," Computer, Vol.28, No.9, pp.23-32, 1995.
- [6] B. Y. Ricardo and R. N. Berthier, Modern Information Retrieval, ACM press, 1999.
- [7] Tsuhan Chen. Low-Level Features to High-Level Semantics : Are We Bridging the Gap? In EWIMT, London, UK, November 2004.
- [8] <http://www.research.ibm.com/VideoAnnEx>
- [9] M. Naphade, C.-Y. Lin, J. R. Smith, B. L. Tseng, and S. Basu. Learning to annotation video database. In SPIE Electronic Imaging 2002 - Storage and Retrieval for Media Database, San Jose, CA, USA, January 2002.
- [10] <http://www.ricoh.co.jp/src/multimedia/MovieTool/index.html>
- [11] <http://itswww.epfl.ch/~newuma/>
- [12] Flickner, Myron, et. al., "Query by Image and Video Content : The QBIC System," IEEE Computer, Vol.28, No.9, September, 1995.
- [13] 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.
- [14] N. Kosugi, Y. Nishihara and T. Stakata, "A Practical Query -By-Humming System for a Large Music Database," Proc of ACM Multimedia 2000 Conference, November, 2000.
- [15] J. R. Smith and S. F. Chang. "VisualSEEK : a fully automated content-based image query system," ACM Multimedia, Boston, May, 1996.
- [16] T. S. Huang, S. Mehrotra, and K. Ramchandran, "Multimedia Analysis and Retrieval System(MARS) project," Proc. of the 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval, University of Illinois at Urbana-Champaign, March, 1996.
- [17] R. Hjelmsvold, "VideoSTAR-A Database for Video Information Sharing," Ph. D. Thesis, Norwegian Institute of Technology, 1995.
- [18] Eitetsu Oomoto, Katsumi Tanaka, "OVID : Design and Implementation of a Video Object Database System," IEEE TKDE, Vol.5, No.4, pp.629-643, 1993.
- [19] S. Handschuh, S. Staab. A. Maedche, "CREAM-Creating relational metadata with a component-based, ontology-driven annotation framework," K-CAP'01, pp.76-88, Victoria, Canada, Oct. 2001.
- [20] T. Schlieder and F. Nanumann, "Approximate tree embedding for querying XML data," In proceedings of the ACM SIGIR Workshop on XML and Information Retrieval, July 2000.
- [21] 안형근, 고재진, "의미적 멀티미디어 메타데이터 생성을 위한 MPEG-7 기술 기반 주석도구의 개발," 정보처리학회 논문지, 제14-D권, 제1호, pp.35-44, 2007.
- [22] <http://ghs.pasco.k12.fl.us/> - Pictures

### 안 형 근



e-mail : hkahn@mail.ulsan.ac.kr

2000년 한국방송통신대학교 컴퓨터학과 (이학사)

2003년 울산대학교 정보통신대학원 정보통신공학과(공학석사)

2006년 울산대학교 대학원

컴퓨터정보통신공학부(박사과정 수료)

1997년~2004년 현대오토시스템 기술지원부

2004년~2006년 (주)CFIC 기업부설연구소 연구소장

관심분야: 멀티미디어DB, DB설계/분석, ERP, BPM, Workflow

### 고 재 진



e-mail : jjkoh@mail.ulsan.ac.kr

1972년 서울대학교 응용수학과(공학사)

1981년 서울대학교 대학원 계산통계학과 (이학석사)

1990년 서울대학교 대학원 컴퓨터공학과 (공학박사)

1975년~1979년 한국후지쓰(주) 기술개발부 사원

1979년~현 재 울산대학교 컴퓨터정보통신공학부 교수

관심분야: DB시스템, 전문가 시스템, DB설계, ERP