

구조 및 의미 검색을 지원하는 비디오 데이터의 모델링

(Video Data Modeling for Supporting Structural and Semantic Retrieval)

복경수[†] 유재수^{**} 조기형^{**}
(Kyoung Soo Bok) (Jae Soo Yoo) (Ki Hyung Cho)

요약 이 논문에서는 비디오 데이터의 논리적 구조와 의미적 내용을 효과적으로 검색하기 위한 비디오 검색 시스템을 제안한다. 제안하는 검색 시스템은 비정형화된 비디오 데이터를 원시 데이터 계층, 내용 계층 그리고 키프레임 계층의 세 계층으로 구성하는 계층화된 모델링을 사용한다. 계층화된 모델링에 존재하는 내용 계층은 비디오 데이터에 대한 논리적인 계층 구조와 의미적 내용을 표현한다. 제안하는 검색 시스템은 모델링에 따라 텍스트 기반의 검색은 물론 시각적인 특징 기반의 유사도 검색을 지원한다. 또한 시공간 관계에 기반한 의미적 내용 검색과 유사도 검색을 지원한다.

키워드 : 비디오 모델링, 구조 검색, 의미 검색

Abstract In this paper, we propose a video retrieval system to search logical structure and semantic contents of video data efficiently. The proposed system employs a layered modelling method that organizes video data in raw data layer, content layer and key frame layer. The layered modelling of the proposed system represents logical structures and semantic contents of video data in content layer. Also, the proposed system supports various types of searches such as text search, visual feature based similarity search, spatio-temporal relationship based similarity search and semantic contents search.

Key words : Video Modeling, Structural Retrieval, Semantic Retrieval

1. 서론

컴퓨터 응용 기술의 발전과 통신 기술의 발달로 일반 사용자들에 의한 멀티미디어 데이터의 사용이 증가되고 있다. 이와 함께 다양한 응용 분야에서 멀티미디어 데이터에 대한 많은 응용 기술들이 개발되고 있다. 멀티미디어 데이터는 이미지, 오디오, 비디오 같은 데이터들이 있으며 일반적으로 텍스트와 숫자 데이터에 비해 대용량이며 비정형화된 데이터 특성을 포함하고 있다.

멀티미디어 데이터의 대표적인 비디오 데이터는 일반적인 멀티미디어의 특징은 물론 이미지 데이터와 같이

시각적인 내용과 오디오와 같은 청각적인 내용을 포함하고 있다. 또한 비디오 데이터 내에는 객체와 사건과 같은 의미적 내용을 포함하고 있다[1, 2]. 비디오 데이터에 포함된 이러한 내용들은 시간의 변화에 따라 그 특성이 변화한다는 특성이 있다.

최근 대용량의 비디오 데이터에 대한 효과적인 저장 및 검색을 수행하기 위한 많은 연구들이 진행되고 있으며 다양한 검색 시스템들이 개발되었다[3]. 비디오 데이터에 대한 검색을 수행하기 위해서는 비디오 분석(video analysis), 비디오 표현(video representation), 비디오 브라우징(video browsing) 그리고 비디오 검색(video retrieval)과 같은 다양한 처리 과정을 수행한다. 비디오 분석은 샷 검출(shot detection), 키프레임 추출(keyframe extraction) 그리고 시각적 특징 추출(visual feature extraction) 등과 같이 비디오 데이터에 대한 처리를 수행한다[4]. 비디오 표현은 비정형화된 비디오 데이터를 논리적 또는 물리적 특징에 따라 비디오의 내

· 본 연구는 한국학술진흥재단 기초학문육성지원사업 인문사회분야지원 일반연구(과제번호:074-DS2501) 연구비 지원에 의하여 수행되었음.

† 비회원 : 충북대학교 정보통신공학과
ksbok@netdb.chungbuk.ac.kr

** 중신회원 : 충북대학교 전기전자컴퓨터공학부 교수
yjs@cbucc.chungbuk.ac.kr
khjoe@cbucc.chungbuk.ac.kr

논문접수 : 2002년 8월 20일
심사완료 : 2003년 3월 10일

용을 표현하는 기법이다. 또한 비디오 브라우징과 검색은 비디오 데이터에 포함된 내용을 쉽게 접근하거나 사용자가 원하는 데이터를 검색할 수 있도록 하는 기능을 수행한다.

비디오 데이터에 대한 효과적인 검색을 수행하기 위해서는 먼저 비정형화된 비디오 데이터에 대한 논리적인 구조 또는 의미적 내용들을 표현하기 위한 모델링 기법이 필요하다. 비디오 데이터에 대한 모델링 기법은 크게 비디오 구조 모델링(modeling the video structure)과 비디오 내용 모델링(modeling the video content)으로 구분할 수 있다[5]. 구조 모델링은 비정형화된 비디오 데이터를 논리적인 계층으로 표현하는 모델링이다. 이에 반해 내용 모델링은 비디오에 포함된 시각적인 특징과 의미적 내용을 표현하는 모델링으로 특징 기반 모델링(feature based modeling)과 의미 모델링(semantic modeling)이 존재한다. 의미 모델링은 비디오 데이터에 포함된 객체와 사건 같은 의미적 내용을 표현하는 모델링이다. 또한 특징 기반 모델링은 비디오에 포함된 이미지에 대한 색상(color), 질감(texture) 그리고 모양(shape)과 같은 시각적 특징을 표현한다. 초기에 비디오 모델링은 대부분은 구조 모델링에 기반하여 의미적 내용을 표현하기 때문에 구조적 내용에 의해 의미적 내용에 대한 의미가 상실되거나 존속 구간을 판별하기 어렵다는 문제점이 있다.

본 논문에서는 비디오 데이터의 구조 검색 및 의미적 내용 검색을 효과적으로 지원하는 VSearcher라는 비디오 검색 시스템을 제안한다. 제안하는 VSearcher는 LVCM(Layered Video Content Modeling)이라는 모델링 기법을 사용하여 비디오 데이터를 표현한다. LVCM은 비디오 데이터에 대한 효과적인 검색을 위해 비디오 데이터를 원시 데이터 계층(raw data layer), 내용 계층(content layer) 그리고 키프레임 계층(keyframe layer)의 세 가지 계층으로 구성한다. 모델링에 따라 VSearcher는 시각적 특징에 기반한 유사도 검색은 물론 텍스트 기반의 구조적 내용 검색과 의미적 내용 검색을 지원한다. 또한 의미적 내용 검색에서는 객체의 공간적인 관계와 객체의 색상 정보에 따른 유사도 검색을 지원한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존에 제안된 비디오 검색 시스템 및 모델링 기법들에 대해 살펴본다. 3장에서는 기존의 검색 시스템 및 모델링 기법의 문제점을 해결하기 위한 모델링 기법에 대해 기술한다. 또한 4장에서는 모델링에 따라 제안하는 검색 시스템에서 제공되는 검색 타입을 기술한다. 5장에서는 제안하는 검색 시스템을 구현한 내용과 기존 시스템과의

비교 분석한 내용을 기술한다. 마지막으로 6장에서는 결론 및 향후 연구 방향에 대해 기술한다.

2. 관련연구

이 장에서는 기존에 비디오 검색을 위해 제안된 여섯 개의 비디오 검색 시스템의 특징을 살펴본다. 또한 비정형화된 비디오 데이터를 표현하기 위해 제안된 비디오 모델링 기법을 구조 모델링과 내용 모델링으로 분류하고 각각에 대한 특징을 살펴본다.

2.1 비디오 검색 시스템

OVID(Object-oriented video Information Database)는 객체지향 모델링에 기반하여 개발된 객체지향 비디오 데이터베이스 시스템이다[6]. OVID에서는 비디오를 프레임의 연속으로 구성되어 있다고 생각하고 각 비디오 프레임들은 하나의 객체로 간주한다. 또한 의미 있는 장면에서 비디오 객체(video object)라는 개념을 이용하여 표현한다. 이러한 비디오 객체는 객체 식별자, 구간 그리고 속성과 속성 값의 집합으로 표현된다.

AVIS(Advanced Video Information System)는 비디오 데이터에 대한 다양한 형태의 질의를 수행하기 위해 메릴랜드(maryland) 대학에서 개발된 원형(prototype) 시스템이다[7,8]. AVIS에서는 비디오 데이터를 객체, 행위 그리고 사건에 의해 표현한다. 사건은 역할과 객체들의 집합에 의해 표현한다.

VideoSTAR(Video SStorage And Retrieval)는 비디오 데이터와 메타 데이터의 재사용과 공유를 지원하기 위해 개발된 비디오 데이터베이스 시스템이다[9,10,11]. VideoSTAR는 비디오 정보 공유, 일반적 모델 개발, 응용 인터페이스 개발, 구조 평가를 위한 실험적 도구 개발을 목표로 개발된 시스템이다. VideoSTAR는 네 개의 서로 다른 저장소로 이루어져 있으며 Video Structure 저장소에는 프레임, 샷, 씬, 시퀀스, 비디오와 같은 비디오 데이터의 구조적 요소들로 구성된다. 또한 일련의 비디오에 대한 특징 정보와 부가적인 정보들은 주석 데이터베이스에 저장한다. 이러한 정보는 사용자 인덱스에 연결되어 사용하며 실세계 객체들과 일반적 비디오 정보 사이에 연결을 수행한다.

JACOB(Just A COntent Based query system for video databases)은 비디오 데이터베이스에서 내용기반 주석 및 질의를 수행하기 위해 개발된 시스템이다[12]. JACOB에서는 색상, 질감, 동작과 같은 일반적인 이미지 특징들을 이용하여 내용기반 검색을 수행한다. JACOB에서는 샷 추출기(shot extractor)에 의해 비디오를 샷들의 열들로 분할한다. 분할된 각 샷은 r-frame이라 불

리는 대표 프레임 추출하고 칼라, 질감, 동작과 같은 특징을 부여한다. 비디오, 샷 그리고 r-frame들은 Shot DB에 저장되며 각 r-frame에서 추출된 특징들은 특징 데이터베이스에 저장한다. 질의 인터페이스에서 입력된 사용자의 질의에 따라 특징 데이터베이스에 저장된 데이터를 분석하여 유사한 r-frame을 탐색하고 가장 유사한 r-frame은 사용자에 의해 선택될 수 있도록 브라우저에 나타낸다.

VideoQ는 콜롬비아 대학에서 개발된 시스템으로 클라이언트-서버 구조의 비디오 검색 시스템이다[13]. VideoQ는 웹 기반의 비디오 검색 시스템으로 사용자가 썸의 어떤 부분에 대한 동작을 스케치하여 질의를 수행할 수 있도록 구성되어 있다. 클라이언트는 자바 애플릿을 통해 웹 브라우저에서 서로 다른 속성을 가지고 있는 객체들의 집합에 대해 질의를 스케치하게 된다. 속성은 동작, 시간순 순서 및 관계, 모양, 칼라, 질감을 포함하게 된다.

Informedia는 NSF, DARPA, NASA의 후원으로 카네기 멜론 대학에서 개발된 검색 시스템으로 NOD(News-On-Demand) 응용에 적용하도록 설계되었다.[14] Informedia 시스템에서는 비디오와 오디오 데이터를 MPEG-1 압축 형식으로 디지털화하여 저장한다. 이러한 디지털 데이터는 검색을 위해 비디오 자막과 오디오 음성으로부터 오디오의 스토리 경계를 분할한다. 또한 비디오 장면을 분할하고 대표 프레임을 선정하고 분할된 스토리에 대한 정보를 색인으로 구성한다. Informedia에서 질의를 수행하기 위해서 사용자는 음성 자연어로 질의를 수행한다. 사용자에 의한 질의는 Sphinx-II에 의해 음성 질의를 분석하여 디지털 라이브러리를 검색하고 검색 결과를 사용자에게 보여준다.

2.2 비디오 모델링

비디오 데이터에 대한 모델링을 수행하기 위해서는 사용자의 요구를 분석하고 사용자가 요구하는 질의 유형을 지원하기 위한 모델을 정립해야 한다. 이러한 비디오 모델링 방법은 앞에서 언급한 것과 같이 구조 모델링과 내용 모델링으로 구분된다. 초기의 비디오 모델링은 거의 대부분 비디오에 대한 논리적인 구조를 표현하기 위한 구조 기반 모델링에 집중되어 있었다. 그러나 구조기반 모델링은 비디오 내에 포함된 시각적 특징과 객체와 사건과 같은 의미적인 내용을 효과적으로 표현하지 못한다는 문제점이 있었다. 이러한 문제점을 해결하기 위해 내용 모델링 내에 의미적 내용을 표현하는 의미기반 모델링이 제안되었다.

[15]에서는 비디오 데이터를 비디오, 에피소드, 샷의

논리적 계층 구조로 표현하고 샷 내에 장면 구성 정보와 장면 특성 정보를 표현하였다. [15]에서 제안하는 모델링은 비디오의 특징을 규정하는 장면의 특수 효과와 카메라 효과 등의 물리적 정보를 장면 특성 정보에 표현한다. 장면 구성 정보에는 장면의 배경, 움직이는 물체 그리고 상황 등의 정보를 표현하여 비디오에 대한 논리적인 표현을 한다. [9]에서는 비디오 데이터 모델링을 위해 필요한 메타데이터의 공유와 재사용을 지원하기 위해 비디오를 시퀀스, 씬, 샷으로 표현하고 세 가지 주석 타입에 따라 의미적 내용을 표현하였다. [16]에서는 비디오에 대한 구조 정보를 썸과 샷으로 표현하는 객체 지향 기술을 기반으로 복합 객체(complex object) 모델링을 제안하였다. [17]에서는 비디오 데이터에 대한 구조적 내용을 비디오, 씬, 그룹, 샷, 키프레임으로 표현하는 5 계층 모델링을 제안하였다. 또한 샷 영역 검출 알고리즘과 키프레임 추출 기법에 따라 비디오 스트림을 샷으로 분할하고 분할된 샷에서 키프레임을 추출하고 유사한 하위 계층 정보를 이용하여 상위 계층 정보를 생성하는 알고리즘을 제안하였다.

그러나 구조 기반 모델링은 비디오에 포함된 의미적 내용을 표현하지 못하기 때문에 객체와 사건과 같은 의미적 내용을 모델링하기 위한 기법들이 제시되었다. [18]에서는 비디오 데이터에 대한 내용기반 검색을 지원하기 위하여 의미 기반 모델링과 구조 기반 모델링을 통합하는 VIDAM이라는 비디오 데이터 모델링을 제안하였다. VIDAM에서는 오디오와 시각적인 내용을 표현하기 위한 의미 객체(semantic object)와 비디오에서 이용 가능한 공간적-시간적인 정보를 나타내기 위한 구조 객체(structural object)를 표현하였다. [19]에서는 비디오 데이터 모델링을 위해 필요한 메타데이터를 7가지 관점에서 표현하는 GMM(Generic Metadata Model) 모델링을 제안하였다. GMM에서는 비디오 데이터에 대한 논리적 계층 구조를 시퀀스, 씬, 샷으로 표현하고 샷 내에 의미적 내용의 대표적인 객체에 대한 시공간 관계를 표현하였다. [20, 21, 22]에서는 비디오에 포함된 의미적 내용을 표현하기 위해 비디오의 구조적 모델링 계층 내에 객체 또는 사건과 같은 의미적 내용을 표현하는 계층을 포함하는 모델링 기법들을 제안하였다. [20]에서는 구조적 계층 내에 사건과 객체를 표현하는 계층을 포함하고 논리적 계층 구조 또는 씬에 포함된 객체들 사이의 시간 관계성을 표현하는 모델링을 제안하였다. 또한 [21]에서는 비정형 비디오 데이터에 대한 특징 기반 검색과 주석기반 검색을 지원하기 위한 THOMM(Two layered Hybrid Object-oriented Metadata

Model)을 제안하였다. THOMM에서는 비디오 데이터에 대한 논리적 구조 계층 내에 객체를 표현하고 씬과 관련된 객체들에 대한 색상과 시공간 관계를 표현하였다. [22]에서는 비디오, 씬, 샷, 프레임의 4계층 구조 기반 모델링에 객체와 사건 같은 의미적 내용을 표현하기 위해 샷 계층을 시퀀스 계층과 청크 계층으로 세분화하는 5 계층 모델링을 제안하였다. 청크는 새로운 사건의 시작과 종료 또는 샷의 시작과 종료에 의해 분할된다. 실제적인 객체와 사건과 같은 의미적 내용에 대한 표현은 시퀀스, 씬, 비디오 계층에서 이루어진다.

[23, 24]는 비디오에 포함된 의미적 내용을 효과적으로 표현하기 위한 의미 기반 모델링을 제안하였다. [23]에서는 구조적 계층과 독립적으로 비디오에 나타나는 객체와 사건과 같은 의미적 내용을 표현한다. 또한 시간에 따라 객체가 수행하는 동작(action)을 표현하였다. 또한 사건은 객체에 의해 수행하는 동작에 의해 표현된다. [24]에서는 비디오 데이터를 원시 데이터 계층(raw data layer), 특징 계층(feature layer), 객체 계층(object layer) 그리고 사건 계층(event layer)으로 표현하는 모델링 기법을 제안하였다. 객체와 사건과 같은 의미적 내용을 표현하는 계층을 개념 계층(concept layer)이라고 하고 객체 계층의 상위에 사건 계층을 표현한다.

3. 계층화된 비디오 모델링

이 장에서는 비디오 데이터의 효과적인 검색 및 저장을 위한 LVCM(Layered Video Content Modeling)이라는 새로운 비디오 모델링을 제시한다. LVCM는 비디오의 구조적 내용과 의미적 내용을 효과적으로 표현하고 시각적인 특징에 기반한 유사도 검색을 지원하기 위해 비디오를 세 개의 계층으로 표현하는 계층화된 모델링이다. 제안하는 LVCM는 그림 1과 같이 비디오 데이터를 원시데이터 계층(raw data layer), 내용 계층(content-layer), 키프레임 계층(keyframe layer)으로 구성한다.

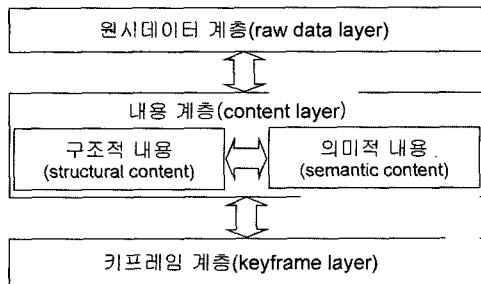


그림 1 계층화된 비디오 모델링

원시데이터 계층은 제안하는 모델링의 최상위 계층으로 비디오 자체가 갖고 있는 가공되지 않은 정보를 나타내는 계층이다. 이러한 원시데이터 계층은 비디오 데이터의 물리적인 특징과 비디오에 제작에 관련된 정보를 표현한다. 원시데이터 계층의 하위에는 비디오의 구조적 내용과 의미적 내용을 표현하는 내용 계층이 존재한다. 내용 계층에 존재하는 구조적 내용 계층은 원시데이터를 비디오, 시퀀스, 씬, 샷과 같은 네 개의 논리적인 계층으로 표현한다. 이에 반해 의미적 내용 계층은 비디오에 포함된 객체, 사건 그리고 배경과 같은 의미적인 내용에 대한 정보를 표현한다. 제안하는 모델링의 최하위 계층에 존재하는 키프레임 계층은 특정 시간 단위 또는 시각적인 특징 변화에 따라 비디오에서 추출한 키프레임에 대한 정보를 나타내는 계층이다.

3.1 원시데이터 계층

제안하는 모델링의 최상위 계층에 존재하는 원시데이터 계층은 물리적인 저장소, 보통 BLOB 영역에 저장된 실제 비디오 데이터 자체에 대한 정보를 표현하는 계층이다. 원시데이터 즉 비디오 데이터 자체는 바이트의 연속으로 텍스트, 오디오 그리고 이미지와 같은 다양한 종류의 비정형 데이터의 혼합체이다. 이러한 원시데이터는 비정형화된 데이터 자체이기 때문에 비디오에 나타나는 어떠한 의미적 표현이나 구조적인 표현을 하지 못한다. 따라서 비디오 자체가 갖고 있는 물리적 특성과 제작에 관련된 특성들을 표현한다. 원시데이터 계층은 (표현 1)과 같이 표현한다. raw_id는 원시 데이터에 부여된 고유 식별자, video_name은 비디오의 제목, blob_id는 비디오 데이터가 저장된 BLOB 공간의 식별자, physical_info는 원시 비디오 데이터의 물리적 특성, production_info는 비디오 제작에 관련된 특성을 표현한다. 비디오 데이터의 물리적 특성을 나타내는 physical_info는 비디오의 상영시간, 포맷 방식, 압축 방식 그리고 초당 프레임 재생률과 같은 물리적 특성을 나타낸다. 또한 비디오 제작에 관련된 특성을 표현하는 production_info는 스태프, 출연진, 감독, 제작자와 같은 비디오 제작에 관련된 특성들을 나타낸다.

<raw_id, video_name, blob_id, physical_info, production_info> (표현 1)

3.2 내용 계층

원시데이터 계층의 하위에는 실제적으로 비디오에 대한 구조적 내용과 의미적 내용을 표현하는 내용 계층이 존재한다. 구조적 내용 계층은 비디오 데이터를 논리적 구조에 의해 표현하는 계층이다. 구조적 내용 계층은 원시 데이터를 비디오, 시퀀스, 씬 그리고 샷의 네 개의

계층 구조에 의해 표현한다. 또한 의미적 내용 계층은 원시 데이터에 포함된 객체, 사건 그리고 배경과 같은 의미적 내용을 표현하는 계층이다. 내용 계층에는 그림 2와 같이 구조적 내용과 의미적 내용 그리고 구조적 내용과 논리적 내용 사이의 관계를 매핑하기 위한 구조-의미 매핑 테이블(structure-semantic mapping table)이 존재한다. 제안하는 모델링은 의미적 내용에 대한 존속 구간과 의미적 내용들 사이의 연관 관계를 효과적으로 표현하기 위해 구조적 내용과 의미적 내용을 독립적으로 표현하고 있다. 따라서 검색 과정에서 구조적 내용에 포함된 의미적 내용을 검색하는 혼합 검색을 지원하기 위해 구조-의미 매핑 테이블이 필요하다. 구조-의미 매핑 테이블은 의미적 내용이 포함된 샷에 대한 정보를 표현한다. 구조적 내용 계층에서는 각 구조적 내용들에 대한 논리적 구조와 순서 정보를 표현하기 위한 구조 정보 식별자를 부여한다. 이러한 구조 정보 식별자는 상위 계층과 하위 계층에 대한 비디오의 논리적인 계층 구조를 표현하기 때문에 샷에 대한 구조 정보 식별자를 통해 상위 계층에 대한 식별자를 판별할 수 있다. 따라서 의미적 내용이 포함된 샷에 대한 식별자만으로도 상위 계층에 대한 검색이 가능하다.

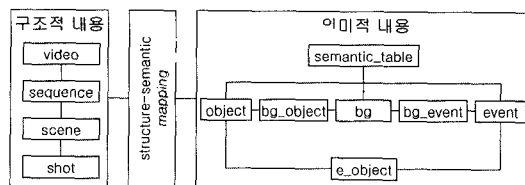


그림 2 내용 계층과 매핑 테이블

의미적 내용 계층은 비디오에 나타난 객체, 사건 그리고 배경에 대한 내용을 표현한다. 각각의 의미적 내용 계층을 표현하기 위해서는 먼저 *semantic_table*에 원시 데이터에서 추출한 의미적 내용에 대한 총괄적인 정보를 표현한다. 실제적인 객체, 사건 그리고 배경에 대한 내용들은 *object*, *event* 그리고 *bg* 테이블에 표현한다. 의미적 내용에 나타나는 각 내용들은 시간의 변화에 따라 서로 연관되어 존재한다. 예를 들어 하나의 사건은 여러 객체들의 연관된 동작에 의해 표현된다. 또한 하나의 배경에는 여러 사건과 객체들이 함께 존재할 수 있다. 이러한 의미적 내용들 사이의 연관 관계를 표현하기 위해 *e_object*, *bg_object* 그리고 *bg_event* 테이블이 존재한다. *e_object* 테이블은 하나의 사건에 연관된 객체들에 대한 내용을 표현하기 위한 테이블이다. 또한

*bg_object*과 *bg_event* 테이블은 배경에 연관된 객체와 사건에 대한 내용을 표현하기 위한 테이블이다.

3.2.1 구조적 내용 계층

내용 계층에 존재하는 구조적 내용 계층은 비정형화된 비디오를 논리적인 구조에 의해 데이터를 표현하는 계층이다. 구조적 내용 계층은 비디오의 논리적 구조를 표현하기 위해 비디오, 시퀀스, 씬, 샷으로 구성된다. 구조적 내용 계층의 구성하기 위해서는 먼저 원시 데이터를 샷 영역 검출에 의해 샷으로 분할하고 분할된 각 샷에서 하나 이상의 키프레임을 추출한다. 샷 계층이 구성되면 연관된 샷들을 이용하여 씬 계층을 구성하고 씬을 대표하는 대표 프레임 추출한다. 이와 유사한 방법으로 시퀀스 계층과 비디오 계층을 구성한다.

각 구조적 내용 계층은 구조 정보 식별자, 순서 정보, 하위 계층의 수, 시간 구간 그리고 각 구조 계층에 대한 대표 프레임에 의해 표현한다. 구조적 내용 계층에 부여되는 식별자는 논리적 구조 계층에 대한 계층 정보와 순서 정보를 표현하기 위한 식별자를 부여한다. 구조적 내용에 부여된 시간 구간은 각 구조적 내용에 대한 존속 구간을 나타내기 위해 사용되며 대표 프레임은 시각적인 요약 정보를 나타내기 위해 사용된다.

각 구조적 내용 계층에 부여된 구조 정보 식별자 SID(Structure Identifier), 순서 정보 SN(Sequential Number) 그리고 하위 계층의 수 SC(Structure Count)는 논리적 계층 구조를 표현하기 위해 사용되며 상위 계층과 하위 계층 간의 관계를 표현한다. 구조적 내용에 대한 SID와 SN 그리고 SC를 표현하기 위해 [그림 3]과 같이 샷 계층에서부터 비디오 계층까지의 계층 구조를 트리 형태로 표현한다. 샷 계층으로부터 시퀀스 계층까지의 논리적인 구조가 생성되면 먼저 비디오 계층은 원시 데이터에 대한 구조적 내용을 식별하기 위해 비디오 계층에 대한 SID, SN 그리고 SC를 부여한다. 비디오 계층에 대한 정보를 생성하면 이를 통해 각 계층에 대한 SID, SN 그리고 SC를 구조적 계층 순서에 따라 생성한다.

각 논리적 계층에 부여된 구조 정보 식별자 SID는 상위 계층과 하위 계층의 논리적 계층 구조를 표현하기 위해 사용된다. SID는 상위 계층과 하위 계층의 구분은 "/"로 표현하고 동일한 상위 계층에 포함된 하위 계층은 "/" 이후에 순서 정보를 부여한다. 예를 들어, "비디오 1"의 "시퀀스 2"는 "1/2"로 부여하고 "시퀀스 2"의 "씬 1"은 "1/2/1"로 부여한다. 이에 반해 SN은 상위 계층과의 계층 구조를 표현하는 것과 별개로 동일한 계층에서 순서 정보를 표현하기 위해 사용된다. 예를 들어, "시퀀스

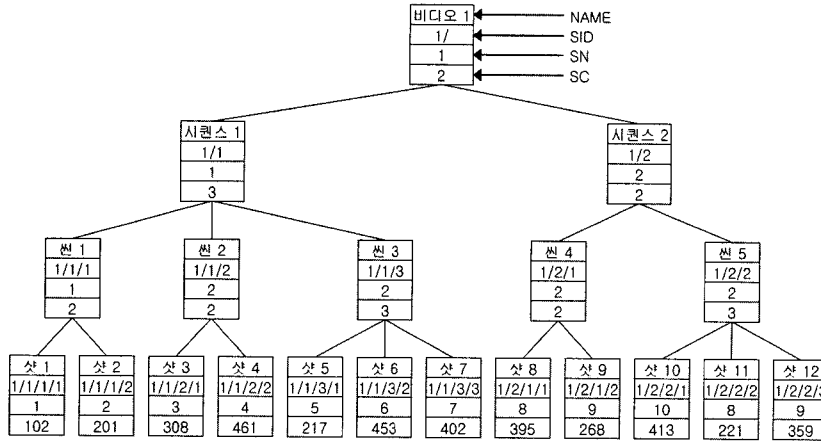


그림 3 구조적 내용 식별자 부여

스 2"에 "씬 1"은 "시퀀스 2"에서는 첫 번째 씬이지만 씬 계층에서 3번째 존재하기 때문에 SN을 3으로 부여한다. SC는 하위 계층에 존재하는 구조적 내용의 수를 나타내는 것으로 비디오 계층에는 하위 계층에 존재하는 시퀀스의 수를 나타내고 시퀀스 계층에는 하위에 존재한 씬의 수를 나타낸다. 그러나 샷 계층에는 샷에 포함된 프레임 수를 나타낸다.

샷(shot) 계층은 구조적 내용 계층의 최하위 계층으로 하나의 카메라 움직임에 관한 기본 단위로 하나의 카메라에 의해 연속되게 기록되는 프레임들의 연속이다. 이러한 샷은 장면의 전환이 없는 임의의 개수의 유사한 프레임으로 구성된다. 비디오에서 샷의 생성은 하나의 카메라가 연속된 움직임에 의해 기록되는 컷(cut)을 검출하기 위한 샷 영역 검출에 의해 이루어진다. 이와 같이 검출된 각 샷에서 하나 이상의 키프레임을 추출한다. 추출된 키프레임은 키프레임 계층에 저장한다. 샷 계층의 표현은 (표현 2)와 같다. 샷 계층에 표현되는 shot_id는 그림 3에서 같이 각 샷에 부여된 구조 정보 식별자 SID로 비디오 계층부터 샷 계층까지의 논리적인 계층 구조 정보를 나타낸다. shot_name은 각 샷에 부여된 이름, shot_sn은 샷 계층에서 샷의 순서 정보 SN이다. frame_num은 샷을 구성하는 프레임 수를 나타내는 SC이다. duration은 샷이 나타나는 존속 구간을 나타내는 것으로 샷의 시작 프레임 start_frame과 종료 프레임 end_frame에 의해 나타낸다. 또한 camera_action과 special_effect는 샷에 나타난 카메라 효과와 특수 효과를 나타낸다. annotation은 사용자 또는 편집자에 의해 부여된 주석 정보를 나타낸다. 마지막으로 ref_image는 하나의 샷을 대표하는 대표 프레임을 나타낸다.

<shot_id, shot_name, shot_sn, frame_num, duration, camera_action, special_effect, annotation, ref_image>

(표현 2)

샷 계층의 상위에는 연속된 샷들로 구성된 씬 계층이 존재한다. 씬은 동일한 장소에서 동일한 객체가 나타내는 하나 이상의 연속된 샷으로 구성된다. 씬 계층의 표현은 (표현 3)과 같이 scene_id, scene_name, scene_sn, duration, shot_num, annotation 그리고 ref_image에 의해 나타낸다. 이때, scene_id는 씬에 부여된 구조 정보 식별자 SID로 비디오 계층에서부터 씬 계층까지의 논리적 계층 정보를 나타낸다. 또한 scene_sn은 씬 계층에서 현재 씬의 순서 정보 SN을 나타내며 shot_num은 현재 씬에 포함된 샷의 개수 SC를 나타낸다. duration과 annotation은 샷 계층과 동일하며 ref_image은 씬에 대한 대표 키프레임을 나타낸다.

<scene_id, scene_name, scene_sn, duration, shot_num, annotation, ref_image>

(표현 3)

씬 계층의 상위에는 연속된 씬들로 구성된 시퀀스 계층이 존재한다. 시퀀스는 하나의 연결 가능한 이야기를 표현할 수 있는 내용으로 하나 이상의 씬으로 구성된다. 시퀀스 레벨의 표현은 (표현 4)과 같이 sequence_id, sequence_name, sequence_sn, duration, scene_num, annotation 그리고 ref_image에 의해 표현된다. 이때, sequence_id는 시퀀스에 부여된 구조 정보 식별자 SID로 비디오 계층에서 시퀀스 계층까지의 논리적인 계층 정보를 나타낸다. 또한 sequence_sn과 시퀀스 계층에서 현재 시퀀스의 순서 정보 SN을 나타내며 shot_num은 현재 시퀀스에 포함된 씬의 개수 SC를 나타낸다.

<sequence_id, sequence_name, sequence_sn,

duration, scene_num, annotation, ref_image>

(표현 4)

구조적 내용 계층의 최상위 계층에는 비디오 계층이 존재한다. 하나의 비디오는 하나 이상의 시퀀스로 구성되며 원시 비디오 데이터 전체의 구조적인 정보를 표현한다. 비디오 계층의 표현은 (표현 5)과 같이 원시 데이터에서 논리적 계층에 대한 전체적인 정보를 표현한다. *video_id*는 구조적인 계층의 최상위인 비디오 계층에 부여된 식별자이다. 하나의 원시 데이터에 대해 사용자 또는 편집자에 따라 여러 구조적인 내용 정보를 표현할 수 있다. 따라서 하나의 원시 데이터에 대한 *video_id*는 하나 이상 존재할 수 있다. *raw_id*는 원시 데이터에 식별자로 구조적 내용을 통해 원시 데이터를 접근할 수 있다. 비디오 계층에서 표현되는 *sequence_num*, *scene_num*, *shot_num*은 논리적 구조 계층에 표현된 전체적인 하위 구조 계층의 개수를 나타낸다.

<*video_id, raw_id, sequence_sn, scene_num, shot_num, annotation, ref_image*> (표현 5)

3.2.2 의미적 내용 계층

의미적 내용 계층은 비디오에 포함된 객체, 사건 그리고 배경과 같은 의미적인 내용을 표현하는 계층이다. 의미적 내용 계층은 구조적인 내용과 독립적으로 구성되며 의미적 내용과 구조적 내용 사이의 관계는 [그림 2]와 같이 구조-내용 매핑 테이블을 통해 이루어진다. 구조적 내용 계층에서는 샷을 분할하고 분할된 각 샷에서 하나 이상의 키프레임을 추출하여 구조적 내용 계층에 대한 대표 프레임을 표현하였다. 이와 유사하게 의미적 내용 계층에서는 각 의미적 내용을 추출하면 추출된 의미적 내용을 대표하기 위한 하나 이상의 키프레임을 추출한다. 추출된 키프레임 중 의미적 내용을 대표하기 위한 대표 프레임을 의미적 내용 계층에 표현한다.

원시 데이터에서 추출된 의미적 내용은 실제적인 객체, 사건 그리고 배경과 같은 정보에 대한 식별자를 부여하기 위해 구조적 내용 계층에서 부여된 SID, SN 그리고 SC와 유사하게 SMCID, FC를 부여한다. 구조적 내용은 상위 계층과 하위 계층 사이의 논리적 계층 관계와 동일 계층에서 현재의 구조적 내용에 대한 순서 정보 SN이 중요하다. 그러나 의미적 내용은 서로 다른 의미적 내용들 사이의 순서 정보가 필요하지 않다. 따라서 의미적 내용에서는 구조적 내용과 달리 SN과 같은 순서 정보를 표현하지 않고 단지 의미적 내용 식별자 SMCID(SeMantic Count Identifier)와 의미적 내용에 포함된 프레임 수 FC(Frame Count)만을 표현한다.

의미적 내용 계층에 존재하는 객체, 사건 그리고 배경

에 대한 정보를 객체, 사건, 배경 테이블에 표현한다. 그러나 하나의 의미적 내용은 비디오에서 한번 이상 반복하여 나타날 수 있다. 즉, 동일한 의미적 내용이 불연속적인 시간에 다른 의미적 내용과 함께 반복하여 나타날 수 있다. 따라서 동일한 의미적 내용이 불연속적인 시간에 나타나는 경우 존속 구간을 명확히 표현하기 위해서 다른 식별자를 부여한다. 예를 들어, 그림 4와 같이 의미적 내용 A와 B가 비디오 내에 나타나는 경우를 생각해 보자. 의미적 내용 A는 $[t_4, t_7]$ 동안 비디오 내에서 한번만 나타나지만 의미적 내용 B는 $[t_1, t_8]$ 동안 세 번 나타나고 있다. 만약 의미적 내용 B에 하나의 식별자를 부여한다면 실제로 의미적 내용이 나타나지 않은 시간 동안에도 의미적 내용이 존재하는 것으로 인식될 수 있다. 따라서 의미적 내용에 대한 존속 구간을 명확히 판단하고 동일한 의미적 내용에 대한 순서 정보를 표현하기 위한 식별자 부여방법이 필요하다.

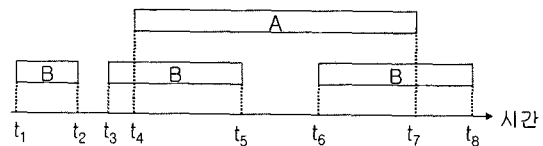


그림 4 의미적 내용에 대한 존속 구간

의미적 내용에 대한 의미적 내용 식별자 SMCID를 표현하기 위해 먼저 비디오에서 추출된 의미적 내용을 객체, 사건 그리고 배경에 따라 각각 저장하고 의미적 내용에 대한 이름 NAME을 *semantic_table*에 등록한다. *semantic_table*에 의미적 내용에 대한 이름이 등록되면 그림 5와 같이 의미적 내용을 트리 형태로 구성한다. *semantic_table*에는 먼저 의미적 내용의 이름 NAME, 의미적 내용에 대한 식별자 SMID(SeMantic Identifier) 그리고 의미적 내용이 연속적인 시간에 나타나는 수 SC(Semantic Count)를 기록한다. SMID는 *semantic_table*에 등록된 NAME에 따라 부여된 유일한 식별자로 의미적 내용에 따라 부여한다. SMID는 원시 데이터에서 의미적 내용 테이블에 부여된 *semantic_id*, 의미적 내용이 타입을 나타내는 *semantic_type* 그리고 동일한 의미적 내용 타입 내에서 부여된 *semantic_number*를 "/"에 의해 조합하여 부여한다. 즉, SMID는 "sem_id/sem_type/sem_num"로 표현되며 *sem_type*은 객체는 1, 사건 2 그리고 배경은 3의 값이 부여된다. 예를 들어 객체 B의 SMID가 "1/1/2"일 때 의미적 내용에 부여된 식별자 *sem_id*가 1, *sem_type*은 객체이며 객체

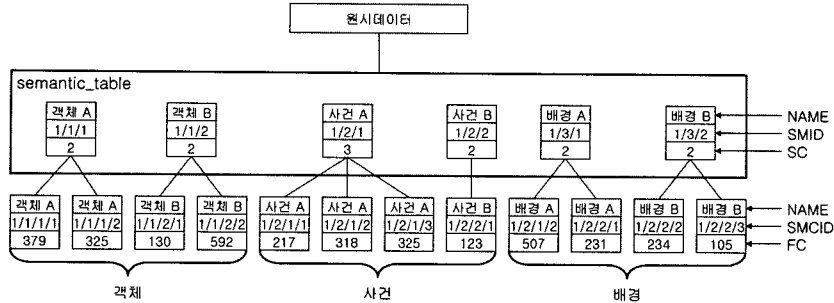


그림 5 의미적 내용 식별자 부여

B에 부여된 값이 2이라는 것을 의미한다. *semantic_table*에 의미적 내용에 대한 SMID와 SC가 할당되면 실제로 객체, 사건 그리고 배경 테이블에 저장된 각각의 의미적 내용에 SMCID와 FC를 할당한다. SMCID는 SMID와 SC를 "/"으로 조합하여 생성한다. 즉 동일한 이름을 갖는 의미적 내용에 대한 SMID를 불연속된 시간에 존재하는 각각에 "SMID/COUNT"에 의해 새로운 식별자를 부여한다. 단, $1 \leq \text{COUNT} \leq \text{SC}$ 이다. 의미적 내용 계층에서 부여된 SMCID는 구조적 내용 계층과 달리 계층적인 상하 관계가 필요 없다. 따라서, 의미적 내용 테이블에 존재하는 SMID와 의미적 내용이 연속된 시간에 나타나는 순서 정보를 이용하여 SMCID를 부여한다.

객체는 하나의 비디오에서 의미적인 정보를 표현하기 위해 연속된 프레임에 나타난 특정 사물이나 사람으로 *semantic_table*에 등록된 객체에 대한 실제적인 정보를 표현한 것이다. *semantic_table*에 등록된 객체에는 객체의 이름 NAME, 의미적 내용에 대한 식별자 SMID (SeMantic Identifier) 그리고 동일한 이름을 갖는 객체가 나타나는 수 SC(Semantic Count)를 기록한다. 이를 통해 그림 5와 같이 객체에 대한 고유한 식별자 SMCID를 부여한다. 객체는 시간에 따라 다양한 동작을 수행하며 비디오 내에 나타난다. 따라서 객체는 시간의 변화에 따라 객체가 수행하는 동작에 대한 정보를 표현해야 한다. 또한 객체의 존속 구간을 명확히 판단할 수 있도록 객체가 나타나는 시간을 표현해야 한다. 의미적 내용 계층에 존재하는 객체는 (표현 6)과 같이 *object_id*, *object_name*, *duration*, *action_list*, *annotation* 그리고 *ref_image*에 의해 표현된다. *object_id*는 연속된 프레임에 나타난 객체에 부여된 객체 식별자 SMCID이고 *object_name*은 객체의 이름으로 *semantic_table*에 등록된 객체의 이름을 나타낸다. *duration*은 객체의 존속 구간을 나타내는 것으로 샷의 시작 프레임 *start_frame*

과 종료 프레임 *end_frame*에 의해 나타낸다. 또한 *annotation*은 객체에 부여된 주석 정보이며 *ref_image*는 하나의 객체를 대표하는 대표 프레임을 나타낸다. *action_list*는 객체가 시간의 변화에 따라 수행하는 동작에 대한 정보를 나타내는 것으로 (표현 7)과 같이 하나의 객체가 n번의 동작을 수행한다고 할 때 객체가 수행하는 동작 *action_i*와 동작이 수행되는 시간 구간 *I_i*에 의해 표현된다. 단 $1 \leq i \leq n$ 이고 *I_i*는 동작이 시작되는 *start_frame*과 종료되는 *end_frame*에 의해 나타낸다.

<object_id, object_name, duration, action_list, annotation, ref_image> (표현 6)

<(action₁, I₁), (action₂, I₂), ..., (action_n, I_n)>

(표현 7)

사건은 특정 객체들의 상호 작용에 의해 발생하는 의미적인 내용으로 여러 객체들에 의해 발생할 수 있다. 따라서 하나의 사건은 특정 객체와 연관되어 존재한다. 따라서 시간의 변화에 따라 사건에 관련된 객체들에 대한 시간 관계를 표현해야 한다. 의미적 내용 계층에 나타나는 사건은 (표현 8)과 같이 *event_id*, *event_name*, *duration*, *annotation*, *ref_image*에 의해 표현한다. *event_id*는 연속된 프레임에 나타난 사건에 부여된 고유한 식별자 SMCID이고 *event_name*은 사건의 이름으로 *semantic_table*에 있는 사건의 이름을 나타낸다. *duration*, *annotation* 그리고 *ref_image*는 객체의 표현과 유사하다. 또한 사건에서는 시간의 변화에 따라 사건에 관련된 객체를 표현하기 위해 그림 2에서 보는 것과 같이 *e_object* 테이블을 사용한다. *e_object*는 (표현 9)과 같이 사건에 관련된 객체의 식별자 *object_id_i*와 객체가 사건에 관련된 시간 구간 *I_i*을 기록한다. 단 $1 \leq i \leq m$ 이고 *I_i*는 동작이 시작되는 *start_frame*과 종료되는 *end_frame*에 의해 나타낸다.

<event_id, event_name, duration, annotation, ref_image> (표현 8)

$\langle (object_id_1, I_1), (object_id_2, I_2), \dots, (object_id_n, I_n) \rangle$ (표현 9)

배경은 객체와 사건이 나타내는 배경 정보를 나타내는 것으로 시간의 변화에 따라 특정 객체 또는 사건이 발생하는 배경에 대한 정보를 나타낸다. 따라서 배경에서는 배경에 포함된 객체 또는 사건이 발생하는 배경에 대한 정보를 표현해야 한다. 배경의 표현은 (표현 10)과 같이 *bg_id*, *bg_name*, *duration*, *annotation* 그리고 *ref_image*에 의해 표현한다. *bg_id*는 연속된 프레임에 나타난 배경에 부여된 식별자 SMCID이고 *bg_name*는 배경의 이름으로 *semantic_table*에 등록된 배경의 이름을 나타낸다. *duration*, *annotation* 그리고 *ref_image*은 객체와 동일하다. 사건과 유사하게 배경에서는 배경에 나타난 객체 또는 사건이 발생하는 배경에 대한 정보를 나타내기 위해 그림 2에서 보는 것과 같이 *bg_object*와 *bg_event* 테이블을 사용한다. *bg_object*는 (표현 11)과 같이 배경에 나타난 객체가 *k*개 일 때 객체에 대한 식별자 *object_id*와 객체가 나타난 시간 구간 I_i 에 의해 나타낸다. 또한 *bg_event*는 (표현 12)에서 보는 것과 같이 배경에서 발생한 사건에 대한 식별자 *event_id*와 객체가 나타난 시간 구간 I_i 에 의해 나타낸다.

$\langle bg_id, bg_name, duration, annotation, ref_image \rangle$ (표현 10)

$\langle (object_id_1, I_1), (object_id_2, I_2), \dots, (object_id_n, I_n) \rangle$ (표현 11)

$\langle (event_id_1, I_1), (event_id_2, I_2), \dots, (event_id_n, I_n) \rangle$ (표현 12)

3.3 키프레임 계층

키프레임 계층은 제안하는 모델링의 최하위에 존재하는 계층으로 특정 시간 단위 또는 시각적인 특징 변화에 따라 구조적 내용과 의미적 내용 계층에서 추출한 키프레임에 대한 정보를 표현하는 계층이다. 구조적 내용 계층에 존재하는 샷에서는 하나의 샷을 대표하는 하나 이상의 키프레임을 추출한다. 또한 의미적 내용 계층에서는 객체, 사건 그리고 배경을 대표하기 위한 하나 이상의 키프레임을 추출한다. 구조적 내용 계층과 의미적 내용 계층에서 추출한 키프레임들은 각각의 내용 정보들에 대한 시각적인 요약 정보는 물론 검색 과정에서 시각적인 특징 기반의 검색을 지원할 수 있는 기능을 제공한다.

키프레임 계층에서는 검색 과정에서 시각적 특징 기반의 검색을 수행할 수 있도록 프레임에 대한 시각적 특징 즉, 색상(color), 질감(texture) 그리고 모양(shape)과 같은 시각적 특징 정보를 표현해야 한다. 또한 하나

의 키프레임은 하나 이상의 객체를 포함하고 있다. 따라서 키프레임 계층에는 키프레임에 포함된 객체에 대한 식별자와 객체가 갖고 있는 시각적인 특징 정보와 공간적인 위치 정보를 표현해야 한다. 키프레임에 대한 시각적인 특징은 이미지 처리(image processing) 과정에서 자동적으로 추출된 특징 벡터를 기록한다. 이미지 처리에 의해 추출된 특징 벡터는 검색 과정에서 이미지에 대한 특징 벡터를 이용하여 유사도 검색을 수행할 수 있다. 키프레임 계층의 표현은 (표현 13)과 같이 *keyframe_id*, *visual_featurelist*, *kf_objectlist*, *event_id*, *bg_id* 그리고 *shot_id*에 의해 표현한다. *keyframe_id*는 키프레임 식별자로 원시 데이터에서 키프레임이 몇 번째 프레임인지를 판별하기 위해 사용된다. *visual_featurelist*는 키프레임에 대한 시각적인 특징 정보를 나타낸 것으로 이미지 처리 과정에서 자동적으로 추출된 특징 정보를 표현한다. *event_id*와 *bg_id*는 키프레임이 관련된 사건과 배경에 대한 식별자를 나타낸다. 또한 *shot_id*는 키프레임이 속한 샷에 대한 식별자이다. 키프레임 계층에 키프레임이 속한 *shot_id*로 샷 계층 상위에 존재하는 썸, 시퀀스 그리고 비디오 계층에 대한 식별자를 판단할 수 있기 때문에 키프레임이 속한 모든 구조적 내용 계층에 대한 식별자를 기록하지 않는다. *kf_objectlist*는 키프레임에 포함된 객체에 대한 정보를 나타낸 것으로 객체의 식별자와 객체의 시각적인 특징 정보를 기록한다. (표현 14)는 *kf_objectlist*를 나타낸 것이다. 키프레임 내에 *n*개의 객체가 포함되어 있을 경우 *object_id*는 키프레임에 포함된 객체의 식별자이고 *color_i*은 키프레임 내에서 객체가 갖는 시각적인 특징 정보로 RGB 값으로 나타낸다. 단, $1 \leq i \leq n$ 이다. 또한 *mbr_i*는 키프레임 내의 공간적인 위치를 나타낸 것으로 MBR(Minimum Boundary Region)로 표현한다.

$\langle keyframe_id, visual_featurelist, kf_objectlist, event_id, bg_id, shot_id \rangle$ (표현 13)

$\langle (object_id_1, color_1, mbr_1), (object_id_2, color_2, mbr_2), \dots, (object_id_n, color_n, mbr_n) \rangle$ (표현 14)

객체의 공간적인 위치를 나타내기 위한 MBR은 특정 시간에 나타난 객체의 공간적 위치를 각 축의 시작 위치와 끝 위치를 이용하여 표현한다. 그림 6은 객체에 대한 공간적 위치를 MBR로 표현하는 과정을 나타낸 것이다. 그림 6에서 보는 것과 같이 객체의 공간적인 위치는 (x_1, x_2, y_1, y_2) 으로 표현한다. 이때, x_1 과 x_2 는 객체가 포함하고 있는 X축의 시작 위치와 끝 위치를 나타낸다. 또한 y_1 과 y_2 는 Y축의 시작 위치와 끝 위치를 나타낸다.

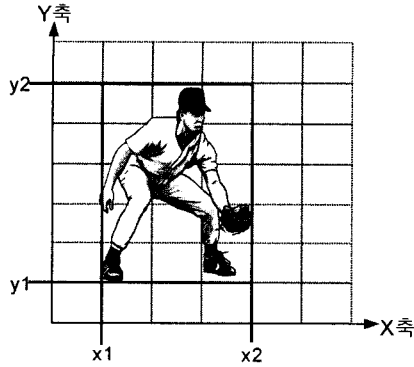


그림 6 객체의 공간적인 위치 표현

4. 비디오 검색

본 장에서는 계층화된 비디오 모델링에 따라 제안하는 비디오 검색 시스템은 다양한 검색을 지원한다. 제안하는 VSearcher에서는 텍스트 기반의 검색은 물론 시각적인 특징 기반의 유사도 검색과 시공간 관계에 기반한 의미적 내용 검색과 유사도 검색을 지원한다. 또한 구조적 내용과 의미적 내용에 대한 혼합 검색을 지원한다.

4.1 원시 데이터 검색

제안하는 모델링의 최상위 계층에 존재하는 원시 데이터 계층에는 비디오의 이름, 제작에 관련된 특성 그리고 물리적인 특성을 포함한다. 원시 데이터 검색은 비디오의 이름 또는 제작자, 출연진, 감독 등과 같은 제작에 관련된 특성들에 의한 검색을 수행한다. 또한 비디오 데이터의 물리적인 특성 즉, 사항들에 의한 파일 포맷, 압축 방식 등과 같은 특성을 이용한 검색을 수행한다. 일반적으로 원시 데이터 검색은 대부분 텍스트 기반의 검색을 수행한다.

4.2 유사도 검색

키프레임 계층에는 구조적 내용과 의미적 내용에서 추출된 키프레임에 대한 정보를 나타낸다. 이러한 키프레임에는 키프레임에 대한 시각적인 특징 정보와 프레임에 나타난 객체에 대한 공간적인 위치와 색상 정보를 표현한다. 유사도 검색은 키프레임 계층에 표현된 프레임 자체의 시각적인 특성과 프레임에 나타난 객체의 색상 정보를 이용하여 검색을 수행한다. 프레임 계층에 표현된 프레임 자체의 시각적인 특징은 색상, 모양 그리고 질감을 표현하고 객체의 색상 정보는 RGB 값으로 표현한다.

프레임의 시각적 특징에 대한 정보 즉, 색상, 모양 그리고 질감이 $F(C_F, S_F, T_F)$ 일 때 사용자의 질의

$Q(C_Q, S_Q, T_Q)$ 에 대한 유사도 검색 $SF(F, Q)$ 는 (식 1)과 같다. 이때, F_1, F_2, F_3 는 색상, 모양 그리고 질감에 부여된 가중치로 가중치의 합은 1이다. 즉, $F_1 + F_2 + F_3 = 1$ 이 된다. 또한 $MaxF$ 는 사용자의 질의와 프레임의 시각적 특징의 최대 차이이다. 따라서 $0 \leq SF(F, Q) \leq 1$ 이 된다.

$$SF(K, Q) = 1 - \frac{F_1|C_Q - C_F| + F_2|S_Q - S_F| + F_3|T_Q - T_F|}{MaxF} \quad (\text{식 1})$$

프레임 계층에 표현하는 객체의 색상 정보는 RGB 값으로 나타내며 프레임에 나타난 객체의 색상이 $O(R_O, G_O, B_O)$ 일 때 사용자의 질의 $Q(R_Q, G_Q, B_Q)$ 에 대한 유사도 검색 $SC(O, Q)$ 는 (식 2)과 같다. 이때, C_1, C_2, C_3 는 R, G, B 각각에 부여된 가중치로 가중치의 합은 1이다. 즉, $C_1 + C_2 + C_3 = 1$ 이 된다. 또한 $MaxC$ 는 사용자의 질의와 객체의 색상의 최대 차이이다. 따라서 $0 < SC(O, Q) \leq 1$ 이 된다.

$$SC(O, Q) = 1 - \frac{C_1|R_Q - R_O| + C_2|G_Q - G_O| + C_3|B_Q - B_O|}{MaxC} \quad (\text{식 2})$$

프레임의 시각적 특징이나 객체의 색상 정보를 이용한 유사도 검색은 구조 검색 또는 의미 검색을 위해 텍스트 기반 검색과 함께 사용한다. 구조적 내용과 의미적 내용을 생성하는 과정에서 각 내용에 대한 시각적인 특징 기반의 검색을 수행하기 위해 추출한 키프레임에 대한 정보를 이용하여 유사도 검색을 수행한다.

4.3 구조 검색

제안하는 모델링에 존재하는 내용 계층에는 비디오를 논리적인 계층구조로 표현하는 구조적 내용 계층이 존재한다. 구조적 내용 계층에는 논리적 계층 구조에 대한 식별자, 존속 구간 그리고 사용자 또는 편집자에 의해 부여된 주석 정보를 표현하고 있다. 또한 샷 계층에서는 샷 내에 표현된 특수 효과나 편집 효과 등을 표현하고 있다. 구조 검색은 구조적 내용 계층에 표현된 주석 정보를 기반으로 텍스트 기반 검색을 수행할 수 있다. 또한 샷에 부여된 특수 효과나 편집 효과를 통한 검색도 가능하다.

각 구조적 내용에는 구조적 내용이 나타나는 시작 프레임과 종료 프레임으로 존속 구간을 나타내고 있다. 이러한 존속 구간은 [25]에서 언급한 시간 관계성에 기반한 검색을 지원할 수 있다. 구조 검색을 수행하기 위해서는 검색 계층과 조건 계층을 선택해야 한다. 검색 계층은 실제적으로 검색을 수행하여 결과 값으로 나타낼

구조적 계층을 나타내는 것이다. 이에 반해 조건 계층은 검색 계층을 검색하기 위해 조건을 검사하기 위한 계층을 나타낸다. 예를 들어, “주어진 텍스트를 만족하는 샷을 포함하는 시퀀스를 찾아라”라는 검색을 수행할 때, 검색 계층은 시퀀스이고 조건 계층은 샷이 된다. 이러한 경우 먼저 주어진 텍스트를 포함하는 샷을 찾는다. 검색 조건을 만족하는 샷을 검색하면 샷의 식별자를 이용하여 현재 샷 상위에 존재하는 시퀀스를 검색한다. 이러한 검색을 수행하기 위해서는 각 논리적 계층 구조에 대한 식별자와 순서 정보를 이용하여 논리적 계층 구조에 대한 검색을 수행한다.

구조적 내용 계층에 대한 유사도 검색은 키프레임 계층에 대한 키프레임에 대한 시각적인 특징 정보를 이용하여 수행한다. 조건 계층 S 에 포함된 키프레임의 수가 n 이라고 할 때, 질의 Q 에 대한 구조적 내용에 대한 유사도 $SS(S, Q)$ 는 (식 3)과 같다. 만약 조건 계층과 검색 계층이 다르다면 조건 계층의 구조적 내용 식별자 SID를 이용하여 검색 계층에 대한 구조적 내용 식별자를 계산하여 결과 값을 반환한다.

$$SS(S, Q) = \text{Max}\{SF(F_i, Q)\}, 1 \leq i \leq n \quad (\text{식 3})$$

4.4 의미 검색

의미 검색은 비디오에 포함된 객체, 사건 그리고 배경과 같은 의미적 내용을 검색하는 것으로 제한하는 모델링의 의미적 내용 계층에 표현된 정보를 이용하여 검색을 수행한다. 의미적 내용에 대한 검색은 의미적 내용에 포함된 의미적 내용의 이름과 존속 구간 그리고 주석 정보를 이용하여 검색을 수행할 수 있다. 의미적 내용의 이름과 주석 정보를 이용한 검색은 텍스트 기반의 검색을 수행하며 존속 구간에 대한 검색은 구조 검색과 동일하게 시간 관계성에 기반한 검색을 수행할 수 있다. 또한 의미 검색은 서로 다른 의미적 내용과 연관된 내용에 대한 검색을 수행할 수 있다. 이렇게 서로 다른 의미적 내용에 대한 연관 관계를 통한 검색을 수행하기 위해서는 의미적 내용들 사이의 관계를 매핑하는 e_object , bg_object 그리고 bg_event 테이블을 이용하여 수행한다. e_object 는 사건과 관련된 객체에 대한 검색을 수행할 수 있으며 bg_object 는 bg_event 배경에 관련된 객체나 사건에 대한 검색을 수행할 수 있다.

구조적 내용 계층에 대한 유사도 검색과 유사하게 시각적 특징에 기반한 의미적 내용에 대한 유사도 검색을 수행한다. 의미적 내용 SM 이 n 개의 키프레임을 갖고 있다고 할 때, 질의 Q 에 대한 의미적 내용에 대한 유사도 $SMS(SM, Q)$ 은 (식 4)과 같이 계산한다.

$$SMS(SM, Q) = \text{Max}\{SF(F_i, Q)\}, 1 \leq i \leq n \quad (\text{식 4})$$

의미적 내용에 존재하는 객체는 위에서 언급한 일반적인 의미 검색은 물론 시간의 변화에 따라 객체가 수행하는 동작에 대한 내용을 이용한 검색도 가능하다. 또한 객체는 하나의 키프레임 내에 객체가 갖고 있는 공간적인 위치를 표현하고 있다. 일반적인 공간 관계성에 기반한 검색은 [26]에서 제안한 관계성을 기반으로 수행되며 본 논문에서 객체들 사이의 거리에 기반한 유사도 검색을 새롭게 제안한다. 두 객체들 사이의 거리 유사성을 정의하기 위해 먼저 특정 키프레임 t 에 나타난 두 객체 O_i 와 O_j 에 대한 거리를 (식 5)과 같이 정의한다. 특정 프레임 t 에서 객체 O_i , O_j 의 MBR을 $MBR_i^t = (X_i^t, Y_i^t)$, $MBR_j^t = (X_j^t, Y_j^t)$ 라 할 때 두 객체 O_i , O_j 의 거리 $Dist(O_i, O_j, t)$ 은 (식 5)과 같이 계산한다. 단, $X_i^t = [X_{si}^t, X_{fi}^t]$, $Y_i^t = [Y_{si}^t, Y_{fi}^t]$, $X_j^t = [X_{sj}^t, X_{fj}^t]$, $Y_j^t = [Y_{sj}^t, Y_{fj}^t]$ 이다. $Dist(O_i, O_j, t)$ 는 두 객체에 대한 공간적인 위치를 MBR로 표현했을 때 객체에 대한 MBR 사이의 최소 거리로 정의한다. 이때, X 와 Y 는 각 축에서 MBR들 사이의 거리이며 $MaxDist$ 는 MBR들 사이의 최대 거리를 나타낸다.

$$Dist(O_i, O_j, t) = \begin{cases} \sqrt{X^2 + Y^2}, & t \in [t_s^i, t_e^i], [t_s^j, t_e^j] \\ MaxDist, & \text{기타} \end{cases} \quad (\text{식 5})$$

$$\text{이때, } X = \begin{cases} X_{sj}^t - X_{fi}^t, & X_{sj}^t > X_{fi}^t \\ X_{si}^t - X_{fj}^t, & X_{si}^t > X_{fj}^t \\ 0, & \text{기타} \end{cases}, \quad Y = \begin{cases} Y_{sj}^t - Y_{fi}^t, & Y_{sj}^t > Y_{fi}^t \\ Y_{si}^t - Y_{fj}^t, & Y_{si}^t > Y_{fj}^t \\ 0, & \text{기타} \end{cases} \text{이다.}$$

객체들 사이의 거리가 정의되면 객체들 사이의 특정 프레임 t 에 나타난 객체들 사이의 거리 유사도 $SD(O_i, O_j, t)$ 은 (식 6)과 같이 계산한다. 특정 프레임 t 에서 객체의 거리 유사도는 0~1로 계산되며, 두 객체의 상대적 거리의 차이에 따라 가까울수록 1로 그렇지 않으면 0이 된다.

$$SD(O_i, O_j, t) = 1 - \frac{Dist(O_i, O_j, t)}{MaxDist} \quad (\text{식 6})$$

객체 프레임 t 에 나타난 객체들 사이의 거리를 통해 두 객체들 사이의 거리 유사도 $SD(O_i, O_j)$ 는 (식 7)과 같이 계산한다. 두 객체 O_i 와 O_j 의 존속 구간이 $[t_s^i, t_e^i]$, $[t_s^j, t_e^j]$ 라 할 때, 객체들 사이의 거리 유사도 $SD(O_i, O_j)$ 는 두 객체가 나타나는 존속 구간에서 상대적 거리의 유사도를 구한 것이다. 객체의 거리 유사도는 0~1로 계산되며 두 객체의 상대적 거리의 차이에 따라 가까울수록 1로 그렇지 않으면 0이 된다. 이때, $t_s = \text{MIN}(t_s^i, t_s^j)$,

$t_j = \text{MAX}(t_j^i, t_j^j)$ 이다.

$$SD(O_i, O_j) = \frac{\sum_{t=t_s}^{t_e} \text{Dist}(O_i, O_j, t)}{t_j - t_s} \quad (\text{식 7})$$

키프레임 계층에는 객체의 공간적인 위치 뿐만 아니라 객체의 색상 정보를 RGB 형태로 표현하고 있다. 이러한 객체의 색상 정보를 통해 질의 Q에 대한 색상 유사도 검색을 수행할 수 있다. 키프레임에 나타난 객체 O가 n개의 프레임에 포함되어 있을 때 질의 Q에 대한 색상 유사도 $SCS(O, Q)$ 는 (식 8)과 같다.

$$SCS(O, Q) = \text{Max}\{SC(O_i, Q)\}, 1 \leq i \leq n \quad (\text{식 8})$$

4.5 혼합 검색

제안하는 검색 시스템은 텍스트 기반의 원시 데이터 검색, 구조 검색 그리고 의미 검색을 지원한다. 또한 시각적인 특징 및 객체의 색상 정보와 공간적인 관계에 기반한 유사도 검색을 지원한다. 이러한 검색은 독립적인 검색으로 수행되기도 하지만 텍스트 기반 검색과 유사도 검색을 동시에 수행하는 혼합 검색도 지원한다. 제안하는 검색 시스템은 이러한 혼합 검색은 물론 구조-의미 매핑 테이블을 통한 구조 검색과 의미 검색에 대한 혼합 검색도 지원한다.

5. 시스템 구현 및 비교 분석

5.1 시스템 구현

이 장에서는 비디오 데이터에 대한 구조 검색과 의미 검색을 수행하기 위한 비디오 검색 시스템 Vsearcher에 대해 기술한다. 이 논문에서 구현하는 브라우저는 Windows 2000 Server 운영체제 환경에서 구현되었으며 모델링에서 제안된 정보들을 표현하기 위해 Microsoft SQL Server 2000 데이터베이스를 사용하였다. 구현 언어로는 Microsoft Visual C++ 6.0과 Microsoft Directshow 8.1를 사용하였다.

그림 7은 원시 데이터 검색과 구조 검색을 위한 사용자 인터페이스 화면이다. 원시 데이터 및 구조 검색을

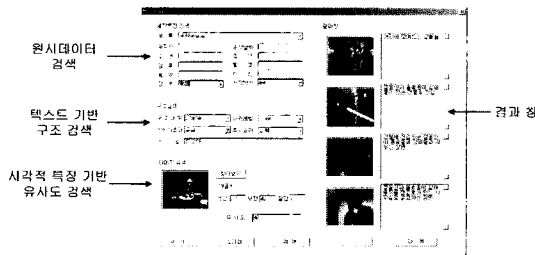


그림 7 원시 데이터 검색 및 구조 검색

수행하는 인터페이스 화면은 검색 조건을 입력하는 부분과 검색 결과를 나타내는 두 부분으로 이루어져 있다. 일반적으로 사용자는 비디오의 이름과 같이 원시 데이터 자체가 갖고 있는 특성을 이용하여 비디오를 검색하고 이러한 비디오 데이터에 대한 구조 검색을 수행할 수 있다. 이러한 검색을 가능하게 하기 위해 원시 데이터 검색과 구조 검색 인터페이스는 왼쪽 상단 부분에 존재하는 원시 데이터 검색 인터페이스를 통해 비디오의 이름, 제작에 관련된 특성 그리고 물리적인 특징에 기반하여 검색을 가능하게 한다. 또한 원시 데이터 검색 인터페이스의 하단 부분에는 원시 데이터에 대한 구조 검색을 수행할 수 있는 인터페이스 화면을 구성하였다.

구조 검색은 텍스트 기반의 검색과 시각적 특징 기반의 유사도 검색을 모두 가능하도록 하였으며 텍스트 기반의 검색과 유사도 검색의 혼합 검색도 가능하다. 시각적 특징 기반의 유사도 검색은 예제 기반의 검색(query by example)과 스케치 기반 검색(query by sketch) 방법이 존재한다. 본 논문에서는 예제 기반의 검색을 이용한 유사도 검색을 사용하였다. 시각적 특징 기반의 유사도 검색 화면에 [찾아보기]를 선택하면 검색을 위해 사용될 수 있는 예제 이미지 리스트가 존재하고 예제 이미지에서 하나를 선택하여 검색을 수행할 수 있다. 또한 유사도 검색을 위한 가중치와 유사도를 지정할 수 있다. 검색 인터페이스 화면의 오른쪽은 검색 결과를 나타내는 부분이다. 검색 결과는 사용자에게 시각적인 요약 정보와 텍스트 정보를 함께 제공한다. 검색 결과에 나타난 시각적인 요약 정보는 검색 계층에 나타난 대표 프레임을 나타낸다.

구조 검색과 함께 비디오에 포함된 의미적 내용을 검색하는 의미 검색 화면은 그림 8과 같다. 의미 검색 인터페이스는 객체, 사건, 배경을 검색하기 위한 조건을 입력하는 부분과 검색 결과를 나타내는 두 부분으로 이루어져 있다. 각 의미적 내용에 대한 검색 조건은 시간

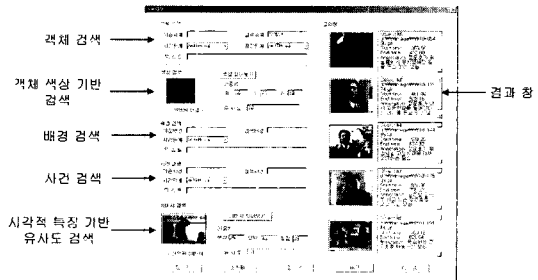


그림 8 의미 검색

관계에 기반한 검색을 수행하기 위한 기준을 입력하는 부분과 텍스트 기반의 검색을 수행하기 위한 텍스트 입력 부분을 포함하고 있다. 또한 객체에 대한 검색은 키 프레임에 나타난 객체의 색상 정보를 이용한 유사도 검색을 수행하기 위해 색상 정보를 입력하는 부분과 공간 관계를 입력하는 부분이 추가적으로 구성되어 있다. 또한 객체의 색상 정보를 이용한 유사도 검색에 대한 가중치와 유사도를 입력할 수 있다.

구조 검색과 동일하게 의미적 내용에 대한 시각적 특징 기반의 유사도 검색을 수행할 수 있도록 예제 이미지를 선택하기 위한 부분과 가중치 및 유사도를 입력하는 부분을 포함하고 있다. 의미 검색 인터페이스의 오른쪽 부분은 검색 결과를 나타내는 부분으로 구조 검색과 동일하게 시각적인 요약 정보와 텍스트 정보를 함께 나타낸다. 시각적인 요약 정보는 의미적 내용 계층에 표현된 각각의 의미적 내용에 대한 대표 프레임으로 검색 결과에 나타난 이미지를 선택하면 검색된 결과를 재생할 수 있다.

그림 9는 구조 검색 및 의미 검색을 수행한 결과에 대한 대표 이미지를 선택하여 검색 결과를 재생하는 화

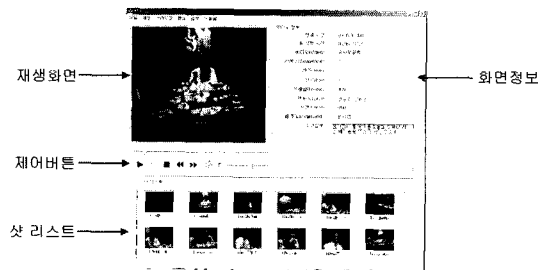


그림 9 재생 화면

면이다. 재생 인터페이스는 [재생 화면], [화면 정보], [제어 버튼] 그리고 [샷 리스트]로 구성되어 있다. [재생 화면]은 현재 재생되고 있는 화면을 나타낸 것으로 검색된 결과에 대한 시작 프레임부터 종료 프레임까지의 시간 정보를 이용하여 재생 시간이 결정된다. [화면 정보] 부분은 현재 재생되고 있는 화면에 대한 구조적 내용과 의미적 내용을 나타내는 부분이다. [제어 버튼] 부분은 재생되는 데이터에 대한 제어를 수행하기 위한 버튼으로 이루어져 있다. 마지막으로 [샷 리스트]는 다음에 재생될 샷에 대한 대표 이미지를 나타낸 것이다.

5.2 비교 분석

제안하는 비디오 모델링 및 검색 시스템의 우수성을 보이기 위해 기존에 제안된 비디오 검색 시스템과의 비교 분석을 수행한다. 제안하는 시스템은 모델링 관점과 검색 관점에서 관련 연구에서 언급한 OVID[6], AVIS[8], VideoSTAR[11], JACOB[12], VideoQ[13]과의 비교 분석을 수행한다.

표 1은 기존의 검색 시스템과의 비교 분석한 결과이다. 표 1에서 보는 것과 같이 기존의 검색 시스템은 구조적 내용과 의미적 내용을 모두 표현하지 못하기 때문에 다양한 형태의 검색 타입을 지원하지 않는다. 또한 AVIS와 VideoSTAR은 의미적 내용 검색을 지원하지만 의미적 내용을 구조적 계층 내에 표현하고 있다. 또한 객체, 사건, 배경과 같은 의미적 내용을 모두 표현하지 못하고 있다. 따라서 의미적 내용의 존속 구간을 명확하게 표현하지 못하고 의미적 내용들 사이의 연관 관계성을 통한 검색을 수행하지 못한다.

제안하는 Vsearcher는 구조적 내용과 의미적 내용을 모두 표현하는 것은 물론 내용들 사이의 관계성과 시각적 특징을 모두 표현하고 있다. 따라서 구조적 계층 사

표 1 시스템 비교 분석

비교항목 시스템	모 델 링			검색 타입	
	구조적 내용	의미적 내용	시각적 특징	구조 검색	의미 검색
OVID	일부 지원	일부 지원	지원하지 않음	텍스트 기반 검색	텍스트 기반 검색
AVIS	지원하지 않음	지원	지원하지 않음	지원하지 않음	텍스트 기반 검색 연관 관계 검색
VideoSTAR	지원	지원	지원하지 않음	텍스트 기반 검색	텍스트 기반 검색
JACOB	일부 지원	지원하지 않음	지원	시각적 특징 검색	제공하지 않음
VideoQ	일부 지원	일부 지원	지원	지원하지 않음	시각적 특징 검색
Informedia	지원하지 않음	일부 지원	지원하지 않음	지원하지 않음	텍스트 기반 검색 음성 인식 검색
VSearcher	지원	지원	지원	텍스트 기반 검색 구조적 계층 검색 시각적 특징 검색	텍스트 기반 검색 의미적 내용 검색 시각적 특징 검색 연관 관계 검색

이의 관계 또는 의미적 내용들 사이의 관계성에 대한 검색에 기반한 검색+을 지원한다. 또한 내용 계층의 하위 레벨에 프레임 계층을 표현하기 때문에 시각적 특징 기반의 검색을 모두 지원한다.

6. 결론 및 향후 연구

본 논문에서는 비디오 데이터에 대한 구조 검색과 의미 검색을 지원하는 비디오 검색 시스템을 제안한다. 제안하는 검색 시스템은 계층화된 모델링에 따라 데이터를 표현한다. 또한 구조적 내용과 의미적 내용에 대한 표현을 내용 계층에 표현하고 이를 통한 다양한 형태의 검색을 지원한다. 제안하는 모델링의 최하위에 존재하는 키프레임 계층은 구조적 내용과 의미적 내용에 대한 시각적인 특징 기반의 유사도 검색을 지원할 수 있도록 한다. 향후 연구로 보다 다양한 형태의 검색을 지원하기 위한 공간간 관계와 시간의 변화에 따라 공간적인 위치가 변하는 움직임 객체에 대한 연구를 계속 수행할 예정이다.

참 고 논 문

- [1] 황규영, 박종목, 노웅기, 박병권, "멀티미디어 데이터베이스", *한국정보과학회 논문지*, 제 12권 제 7호, pp.59-69, 1994.
- [2] W. I. Grosky, "Multimedia Information Systems", *IEEE Multimedia*, Vol. 1, No. 1, pp.12-24, 1994.
- [3] 박유미, 최영일, 진성일, 허대영, 최동훈, "비디오 데이터베이스의 내용 기반검색을 위한 메타데이터의 분류와 모델링", *한국전자통신연구원 주간기술동향*, pp12-26, 1997.
- [4] Philippe Aigrain, HongJiang Zhang and Dragutin Petkovic, "Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review", *Multimedia Tools and Applications*, Vol. 3, No. 3, pp.179-202, 1996.
- [5] M. Petkovic and W. Jonker, "An Overview of Data Models and Query Languages for Content-based Video Retrieval", *Pro. International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, l'Aquila, Italy, July 2000.
- [6] E. Oomoto and K. Tanaka, "OVID: Design and Implementation of a Video-Object Database System", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, No. 4, pp.629-643, 1993.
- [7] S. Adali, K. S. Candan, S. S. Chen, Kutluhan Erol and V. S. Subrahmanian, "The Advanced Video Information System: Data Structures and Query Processing", *Multimedia Systems*, Vol. 4, No. 4, pp.172-186, 1996.
- [8] S. Adali, K.S. Candan, K. Erol and V.S. Subrahmanian, "AVIS: An Advanced Video Information System", *The Institute for Systems Research*, University of Maryland, Technical Report, TR 97-44, pp.369-373, 1997.
- [9] Rune Hjelsvold and Roger Midtstraum, "Modelling and Querying Video Data", *Pro. the 20th International Conference on Very Large Databases*, Santiago, Chile, September 1994.
- [10] Rune Hjelsvold, Roger Midtstraum, and Olav Sandsta, "Searching and Browsing a Shared Video Database", *Pro. the International Workshop on Multi-Media Database Management Systems*, Blue Mountain Lake, New York, August 28-30 1995.
- [11] Rune Hjelsvold, "VideoSTAR-A Database for Video Information Sharing", *Dr.Ing. Thesis*, Norwegian Institute of Technology, November 1995.
- [12] M. La Cascia, and E. Ardizzone, "JACOB: Just a content-based query system for video databases", *Pro. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, 1996.
- [13] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong, "VideoQ: An Automatic Content-Based Video Search System Using Visual Cues", *Pro. ACM Multimedia '97 Conference*, Seattle, WA, November 1997.
- [14] Hauptmann, A., Witbrock, M., "Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval.", *Intelligent Multimedia Information Retrieval*, Mark T. Maybury, Ed., AAAI Press, pp. 213-239, 1997
- [15] 권성복, 조영우, 김영모, "구조화된 논리적 비디오를 이용한 비디오 브라우징 및 검색 시스템", *한국 정보처리학회 논문지* 제4권 6호 pp.1443-1452, 1997.
- [16] 류은숙, 이규철, "구조화된 비디오 문서의 데이터 모델 및 질의어와 색인 기법", *한국멀티미디어학회 논문지*, 제 1권 1호, pp.1-17, 1998.
- [17] Yong Rui and Thomas Huang, "Chapter 9.2:A Unified Framework for Video Browsing and Retrieval", in *The Image and Video Processing Handbook edited by Alan Bovik*, Academic Press, pp.705-715, 2000.
- [18] Srinivasan, U. and Riessen, G, "A Video Data Model for Content-Based Search", *Pro. the Eighth International Workshop on Database and Expert Systems (DEXA'97)*, Toulouse, September 1997.
- [19] Y. M. Park, Y. K. Kim, S. I. Jin and W. Choi, "Hierarchical Structure-based Metadata Model for Video Database Applications", *ISCA COMPUTERS AND THEIR APPLICATIONS*, pp.242-245, March. 1998.

- [20] 최지희, 용환승, "시간 관계성을 기반으로 한 비디오 데이터 모델의 설계 및 구현", *한국멀티미디어학회 논문지*, 제2권 3호, pp.252-264, 1999.
- [21] 윤미희, 윤용익, 김교정, "비디오의 의미검색과 유사성 검색을 위한 통합비디오정보시스템", *한국정보처리학회 논문지*, 제 6권 8호, pp.2031-2041, 1999.
- [22] Al Safadi, L.A.E. and Getta, J.R., "Semantic Modeling for Video Content-based Retrieval Systems", *Computer Science Conference, 2000. ACSC 2000. 23rd Australasian*, pp.2-9, 2000.
- [23] Jia-Ling Koh, Chin-Sung Lee and Chen, A.L.P., "Semantic video model for content-based retrieval", *Pro. IEEE International Conference on Multimedia Computing and Systems*, Vol. 2, pp.472-478, 1999.
- [24] M. Petkovic and W. Jonker, "A Framework for Video Modeling", *Pro. Eighteenth IASTED International Conference Applied Informatics*, Innsbruck, Austria, February 2000.
- [25] J. F. Allen, "Maintaining knowledge about temporal intervals", *Communications of ACM*, Vol.26, No.11, pp.832-843, 1983.
- [26] J. Z. Li, M. T. Ozsu, D. Szafron, "Modeling of Video Spatial Relationships in an Objectbase Management System", *Pro. International Workshop on Multimedia DBMS*, Blue Mountain Lake, NY, pp. 124-133, 1996.



조 기 형

1966.2 인하대학교 전기공학과 공학사.
1984.8 청주대학교 산업공학과 공학석사.
1992.2 경희대학교 전자공학과 공학박사.
1981~1988 충주공업전문대학 조교수.
1996~1999 전기전자컴퓨터공학부 교
재 충북대학교 전기전자컴퓨터공학부 교
수. 관심분야는 데이터베이스시스템, 화상처리 및 통신, 통
신 프로토콜, 분산 객체 컴퓨팅 등



복 경 수

1998 충북대학교 수학과 이학사. 2000
충북대학교 정보통신공학과 공학석사.
2000~현재 충북대학교 정보통신공학과
박사과정. 관심분야는 자료 저장 시스템,
멀티미디어 데이터베이스, 고차원 색인
구조, 시공간 색인구조 등



유 재 수

1989 전북대학교 컴퓨터공학과 공학사.
1991 한국과학기술원 전산학과 공학석사.
1995 한국과학기술원 전산학과 공학박사.
1995~1996.8 목포대학교 전산통계학과
전임강사. 1996.8~현재 충북대학교 전기
전자컴퓨터공학부 부교수. 관심분야는 데
이타베이스 시스템, 정보검색, 멀티미디어 데이터베이스, 분
산 객체 컴퓨팅 등