

# 계층화된 비디오 내용 모델링 및 브라우징

복 경 수<sup>†</sup>·이 낙 규<sup>†</sup>·허 정 필<sup>†</sup>·유 재 수<sup>††</sup>·조 기 형<sup>†††</sup>·이 병 업<sup>††††</sup>

## 요 약

본 논문에서는 비디오 데이터 대한 구조적 내용과 의미적 내용에 효과적으로 표현하기 위해 비디오 모델링을 제안한다. 또한 사용자가 비디오 데이터에 대한 내용 정보를 쉽게 파악하고 재생할 수 있도록 하는 브라우징 기법을 제안한다. 제안하는 모델링은 원시 데이터 계층, 내용 계층 그리고 키프레임 계층의 세 계층으로 구성되며 비디오 데이터에 대한 논리적인 계층 구조와 의미적 내용은 내용 계층에 표현한다. 비디오 브라우징은 비디오 재생과 내용 정보를 나타내는 두 가지 브라우저를 구현하였다. 비디오 재생 브라우저는 비디오를 재생과 현재 화면 정보를 나타내는 기능을 수행한다. 내용 정보 브라우저는 원시데이터, 구조적 내용 그리고 의미적 내용에 대한 정보를 브라우징하는 기능을 제공한다.

## Layered Video Content Modeling and Browsing

Kyoung Soo Bok<sup>†</sup> · Nak Gyu Lee<sup>†</sup> · Jeong Pil Heo<sup>†</sup>  
Jae Soo Yoo<sup>††</sup> · Ki Hyung Cho<sup>†††</sup> · Byoung Yup Lee<sup>††††</sup>

## ABSTRACT

In this paper, we propose modeling method for video data that represents structural and semantic contents of video data efficiently. Also, a browsing method that helps users easily understand and play the contents of video data is presented. The proposed modeling scheme consists of three layers such as raw data layer, content layer and key frame layer. The content layer represents logical hierarchy and semantic contents of video data. We implement two kinds of browsers for playing video data and providing video contents. The playing browser plays video data and presents the information of currently playing shot. The content browser allows users to browse raw data, structural information and semantic contents of video data.

키워드 : 비디오 모델링(Video Modeling), 브라우징(Browsing), 구조적 내용(Structural Content), 의미적 내용(Semantic Content)

### 1. 서 론

최근 컴퓨터 기술의 발전과 더불어 인터넷의 확산으로 일반 사용자들에 의해 오디오, 이미지, 비디오와 같은 멀티미디어 데이터에 대한 사용이 증가되고 있다. 특히 멀티미디어 데이터의 대표적인 비디오 데이터에 대한 관심이 집중되면서 비디오 데이터에 대한 많은 연구들이 진행되고 있으며 많은 응용 분야에서 사용되고 있다. 일반적으로 비디오 데이터는 텍스트와 숫자 같은 데이터에 비해 대용량이며 비정형화된 데이터 특성을 포함하고 있다. 비디오 데이터는 이미지와 같은 시각적인 특징은 물론 오디오와 같은 청각적인 특성을 포함하고 있으며 시간의 변화에 따라 이러한 특징들이 변화한다는 특성을 갖고 있다. 또한 비디오 데이터에는 특정 객체와 객체들에 의해 발생하는 여러 가지

사건들을 포함하고 있다[1].

비디오 데이터의 사용이 증가되면서 대용량의 비디오 데이터를 효과적으로 저장하고 접근하기 위한 많은 연구들이 진행되었다[2-6]. 비디오 데이터에 대한 접근을 수행하기 위해서는 비디오 분석(video analysis), 비디오 모델링(video modeling), 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. 비디오 모델링은 비디오의 포함된 내용들을 표현하며 비디오 브라우징을 통해 사용자에게 비디오의 내용을 보다 쉽게 이해하고 접근하게 하는 기능을 제공한다. 비디오 검색은 사용자가 관심 있는 비디오의 내용을 검색하는 기능을 수행한다.

비디오 브라우징은 비디오에 포함된 내용들을 쉽게 이해하고 사용자가 원하는 내용에 대한 위치로 직접 이동할 수 있도록 비디오에 포함된 내용들을 나타내는 기능을 수행한다. 비디오 브라우징을 수행하기 위해서는 먼저 비디오를

※ 본 연구는 2003년도 산업자원부 차세대신기술개발 및 ETRI 컴퓨터소프트웨어 연구소 위탁과제에 의해 수행되었음.

† 준 회원 : 충북대학교 대학원 전기전자 컴퓨터공학부

†† 종신회원 : 충북대학교 전기전자 컴퓨터공학부 교수

††† 정 회원 : 충북대학교 전기전자 컴퓨터공학부 교수

†††† 정 회원 : 배재대학교 전자상거래학부 교수  
논문접수 : 2002년 7월 12일, 심사완료 : 2003년 8월 8일

분석하고 분석된 내용을 모델링에 따라 표현하는 과정을 수행해야 한다. 따라서 비디오 모델링은 비정형 특성을 갖는 비디오 데이터에 대한 내용을 브라우징할 수 있도록 비디오 데이터에 대한 논리적 구조 또는 비디오에 포함된 의미적 내용들을 표현한다.

비디오 데이터에 대한 모델링 기법은 비디오의 논리적 구조를 표현하는 구조기반 모델링(structural based modeling)과 비디오에 포함된 객체와 사건과 같은 의미적 내용을 표현하는 의미기반 모델링(semantic based modeling)으로 구분할 수 있다. 초기의 비디오 모델링은 대부분 비디오의 논리적인 구조를 생성하고 이를 통해 비디오의 내용을 표현하기 위한 구조 기반 모델링 기법(structural based modeling)들이 제안되었다[7-9]. 그러나 구조 모델링 기법은 비디오에 포함된 객체와 사건과 같은 의미적 내용(semantic content)을 효과적으로 표현하지 못한다는 문제점이 있다. 따라서 최근 비디오에 포함된 의미적 내용을 효과적으로 표현하기 위한 의미기반 모델링 기법들이 제안되었다[10-15]. 그러나 기존에 제안된 의미 기반 모델링은 구조 기반 모델링 계층 내에 의미적 내용을 표현하기 위한 계층을 추가하거나 기존의 구조적인 계층을 의미적 내용으로 변환하여 사용하고 있다. 따라서 의미적 내용들이 구조적인 계층에 의해 그 의미가 상실되거나 정확한 표현을 하기 어렵다는 문제점이 있었다.

비디오 모델링 기법에 따라 비디오에 포함된 내용들을 표현하면 비디오 브라우징은 모델링에 의해 표현된 내용들을 사용자가 쉽게 이해할 수 있도록 표현하고 접근할 수 있는 기능을 제공해야 한다. 그러나 기존의 브라우징 기법들은 구조 기반 모델링에 의해 비디오 데이터를 표현하고 이를 브라우징하기 때문에 비디오에 포함된 의미적 내용을 효과적으로 표현하지 못한다[17-22].

본 논문에서는 비디오 데이터에 대한 구조적 내용과 의미적 내용을 효과적으로 표현하기 위해 LVCМ(Layered Video Content Modeling)이라는 새로운 비디오 모델링 기법을 제안한다. 또한 제안하는 모델링에 기반하여 비디오 데이터의 내용 정보를 효과적으로 브라우징 할 수 있는 브라우저를 구현한다. 제안하는 모델링에서는 원시 데이터 계층(raw data layer), 내용 계층(content layer) 그리고 키프레임 계층(key-frame layer)의 세 가지 계층으로 구성한다. 비디오 데이터에 포함된 내용을 브라우징하기 위한 방법으로 트리 기반 브라우징(tree based browsing) 기법을 사용한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존에 제안된 비디오 데이터 모델링 기법과 브라우징 기법들을 살펴본다. 3장과 4장에서는 기존의 모델링 기법의 문제점을 해결하기 위한 새롭게 제안한 모델링 기법과 브라우징 기법을 설명한다. 5장에서는 제안하는 모델링 및 브라우징에 대한 우수성을 증명하기 위해 기존에 제안된 기법과의 비교 분석을 수행한다. 마지막으로 6장에서는 결론 및 향후 연구 방향에 대해 기술한다.

## 2. 관련 연구

### 2.1 비디오 모델링

비디오 데이터에 대한 모델링을 수행하기 위해서는 사용자의 요구를 분석하여 사용자가 요구하는 질의 유형을 지원하기 위한 모델을 정립해야 한다. 이러한 비디오 모델링 방법은 크게 구조기반 모델링과 의미기반 모델링으로 분류할 수 있다. 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 의미 기반 모델링이 제안되었다.

[7]에서는 비디오 데이터 모델링을 위해 필요한 메타데이터의 공유와 재사용을 지원하기 위해 비디오 스트림을 시퀀스, 씬, 샷으로 표현하고 세 가지 주식 타입에 따라 의미적 내용을 표현하였다. [8]에서는 비디오에 대한 구조 정보를 씬과 샷으로 표현하는 객체 지향 기술을 기반으로 복합 객체(complex object) 모델링을 제안하였다. [9]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. [10]에서는 비디오 데이터에 대한 내용기반 검색(content based retrieval)을 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 비디오 스트림에서 이용 가능한 공간적-시간적인 정보를 나타내기 위한 구조 객체(structural object)를 표현하였다. [11]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. [12-14]에서는 비디오에 포함된 의미적 내용을 표현하기 위해 비디오의 구조적 모델링 계층 내에 객체 또는 사건과 같은 의미적 내용을 표현하는 계층을 포함하는 모델링 기법들을 제안하였다. [12]에서는 구조적 계층 내에 사건과 객체를 표현하는 계층으로 포함하고 논리적 계층 구조 또는 씬에 포함된 객체들 사이의 시간 관계성을 표현하는 모델링을 제안하였다. 또한 [13]에서는 비정형 비디오 데이터에 대한 특징기반 검색과 주식기반 검색을 지원하기 위한 THOMM(Two lay-

ered 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. [14]에서는 비디오, 씬, 샷, 프레임의 4 계층 구조 기반 모델링에 객체와 사건과 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.

[15, 16]은 비디오에 포함된 의미적 내용을 효과적으로 표현하기 위한 의미 기반 모델링을 제안하였다. [15]에서는 구조적 계층과 독립적으로 비디오에 나타나는 객체와 사건과 같은 의미적 내용을 표현하고 시간에 따라 객체가 수행하는 동작(action)을 표현하였다. 또한 사건은 객체와 객체가 수행하는 동작에 의해 표현하였다. [16]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. [16]에서는 객체와 사건과 같은 의미적 내용을 표현하는 계층을 개념 계층(concept layer)이라 하고 객체 계층의 상위에 사건 계층을 표현한다.

## 2.2 비디오 브라우징

비디오 브라우징은 비디오에 포함된 내용들을 쉽게 이해하고 사용자가 원하는 내용에 대한 위치로 직접 이동할 수 있도록 비디오에 포함된 내용들을 나타내는 기능을 수행한다. [17]에서는 사용자가 원하는 위치로 접근하고 사용자가 원하는 내용인지를 쉽게 파악할 수 있도록 하기 위한 브라우징 기법을 제안하였다. [17]에서 제안하는 브라우징 기법은 비디오의 내용을 나타내기 위한 기본 단위로 샷을 사용하고 각 샷의 내용은 대표 프레임 또는 Rframe을 이용하여 표현한다. 각 샷을 대표하는 Rframe들은 네 개의 동작 추적 영역(motion tracking region)과 샷 길이 지시자(shot length indicator)들로 구성된다. 브라우저는 순서대로 모든 Rframe들을 표시하고 사용자가 Rframe들을 스크롤(scroll)하여 원하는 Rframe을 선택할 수 있다. [18]에서는 [7]에서 제안하는 VideoSTAR 데이터 모델링에 따라 내용 주석 정보와 비디오 구조 정보를 브라우징하는 기능을 제공한다. 사용자가 비디오를 재생하는 동안 비디오 자체가 제공하는 자세한 정보를 얻을 수 있도록 하기 위해 독립적인 브라우징 기법을 이용하여 비디오의 내용 정보와 구조 정보를 출력할 수 있다. 내용 주석 브라우징은 사용자가 데이터베이스로부터 주석 검색을 수행할 수 있도록 하는 기능을 제공한다. 구조 브라우징은 비디오의 계층적인 구조를 제공하여 사용자가 지금 어느 위치에 있는지를 쉽게 판단할 수 있고 사용자가 원하는 위치를 쉽게 이동할 수 있도록 하는 기능을 제공한다. [19]에서는 내용 기반 비디오 브라우징을 지

원하기 위해 비디오의 구조 정보와 시각적인 내용을 표현하기 위한 브라우징 기법을 제안하였다. 비디오에 대한 논리적인 구조를 시퀀스, 씬 그리고 샷의 세 가지 계층 구조로 표현한다. 각 논리적 계층은 고유한 식별자를 갖고 있으며 각 논리적 계층에 대한 정보를 키프레임을 통해 표현한다. 비디오 데이터에 대한 브라우징은 시간 축에 서로 다른 색상으로 비디오의 샷들을 표현하고 브라우저 화면의 하위에 각 샷에서 추출한 키프레임의 리스트와 키프레임에서 추출한 사람의 얼굴을 출력한다. [20]에서는 대용량의 비디오 데이터를 관리하기 위해 ViBE(Video indexing and Browsing Environment)라는 통합 시스템을 제안하였다. ViBE는 비디오를 처리, 표현 그리고 관리하기 위한 다양한 알고리즘과 기술들을 제안하였다. 비디오 데이터는 먼저 샷들로 분할되고 각 샷들을 대표하기 위한 키프레임들을 추출한다. 각 샷들은 하나 이상의 키프레임들을 통해 비디오의 내용을 표현한다. 비디오 데이터에 대한 접근을 효과적으로 수행하기 위해 대용량의 이미지 데이터베이스를 구성하거나 브라우징하기 위해 제안된 유사 피라미드(similarity pyramid) 기법을 이용하여 샷들을 구성하였다. ViBE에서는 사용자의 요구에 따라 동적으로 샷들을 재구성할 수 있다. 유사 피라미드는 대용량의 비디오 데이터를 3차원의 피라미드 구조로 구성하고 유사 피라미드의 각 계층은 2차원 격자 형태로 유사한 샷들을 구성한다. 사용자에게 보여지는 브라우징 환경은 화면의 왼쪽에 유사 피라미드를 출력하고 오른쪽에 관련된 샷들의 집합을 나타낸다.

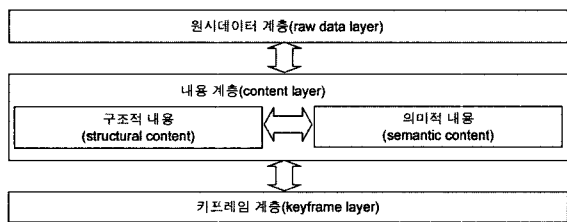
[21]에서는 비디오의 구조 정보를 비디오, 씬, 그룹, 샷 그리고 키프레임의 5 계층으로 표현하고 각 비디오 프레임을 2차원 객체로 생각하고 시간 축을 고려하여 비디오의 흐름을 3차원 공간에 확장한다고 생각하였다. 비디오 데이터의 표현을 위해 3차원 공간을 2차원 화면에 매핑시키기 위한 방법으로 연속적 키 프레임 표현(sequential keyframe representation), 그룹 기반의 표현(group based representation), 씬 기반의 표현(scene based representation) 그리고 비디오 모자이크 표현(video mosaic representation)과 같은 4가지 표현 방법을 고려하였다. 4가지 표현 방법 중 씬 기반의 표현 방법을 가장 효과적인 비디오 브라우징 기법으로 생각하고 비디오의 계층적인 구조를 표현하였다. 씬 기반의 표현은 씬 계층 구조에 대한 주석 정보와 대표 프레임을 나타내고 사용자가 특정 씬을 선택하면 하위의 계층 구조에 대한 주석 정보와 대표 프레임을 나타내도록 하였다. 비디오의 계층 구조를 브라우징하고 사용자가 관심있는 비디오의 특정 위치를 선택하여 재생을 수행할 수 있다. 또한 [22]에서는 비디오 데이터에 대한 재생을 수행하면서 현재 재생되는 화면에 대한 제어와 비디오에서 추출된 샷들에 대한 리스트를 출력하는 브라우저를 제안하였다.

[23]에서는 키프레임을 통해 비디오의 내용을 쉽게 접근할 수 있도록 하기 위한 브라우징 기법을 제안하였다. [23]에서 제안하는 브라우저는 비디오에서 추출한 키프레임을 통해 비

디오의 내용을 브라우징하고 키프레임을 통해 직접 재생할 수 있는 기능을 통합하였다. [24]에서는 인터넷을 통해 비디오의 내용에 대한 시각적인 요약 정보를 제공하기 위한 브라우징 기법을 제안하였다. [24]에서는 비디오의 구조적인 내용을 썸, 샷 그리고 키프레임으로 구성하고 하나의 샷에서 하나 이상의 키프레임을 추출한다. 제안된 브라우저의 헤더 부분에는 비디오의 이름, 길이, 샷과 썸의 개수를 출력한다. 또한 썸, 샷 그리고 샷에서 추출한 키프레임의 순서로 구조적 내용에 대한 시각적인 요약 정보를 제공한다.

### 3. 비디오 데이터 모델링

이 장에서는 비디오 데이터의 구조적 내용과 의미적 내용을 효과적으로 표현하기 위한 LVCМ이라는 새로운 비디오 모델링 기법을 제시한다. 제안하는 모델링 구조는 (그림 1)과 같이 원시데이터 계층, 내용 계층, 키프레임 계층의 세 계층으로 구성된다.



(그림 1) 계층화된 비디오 모델링

원시데이터 계층은 비디오 자체가 갖고 있는 가공되지 않은 정보를 나타내는 계층으로 비디오 데이터의 물리적인 특징과 비디오 제작에 관련된 정보를 표현한다. 원시데이터 계층의 하위에는 비디오의 구조적 내용과 의미적 내용을 표현하는 내용 계층이 존재한다. 내용 계층에 존재한 구조적 내용 계층은 원시데이터를 비디오, 시퀀스, 썸, 샷과 같은 4개의 논리적인 계층구조로 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(Raw Data Layer)

원시데이터 계층은 제안하는 모델링의 최상위 계층으로 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 비디오 자체가 되기 때문에 비디오에 나타나는 어떠한 의미적 표현이나 구조적인 표현을 하지 못하고 단지 비디오 자체가 갖고 있는

상영시간, 포맷 방식, 압축 방식과 같은 물리적 특성과 스타프, 출연진, 감독, 제작자와 같은 비디오 제작에 관련된 특성들을 기술한다. 원시데이터 계층은 <표 1>과 같이 표현한다. *video\_id*는 원시 데이터에 부여된 고유 식별자, *video\_name*은 비디오의 제목, *blob\_id*는 비디오 데이터가 저장된 BLOB 공간의 식별자, *physical\_info*는 원시 비디오 데이터의 물리적 특성, *production\_info*는 비디오 제작에 관련된 특성을 표현한다.

<표 1> 원시 데이터의 표현

속 성	설 명
<i>video_id</i>	비디오 식별자
<i>video_name</i>	비디오 제목
<i>blob_id</i>	물리적 저장 공간에 대한 식별자
<i>physical_info</i>	원시 데이터에 대한 물리적 특성
<i>production_info</i>	원시 데이터를 제작에 관련된 특징

#### 3.2 내용 계층(Content Layer)

원시데이터 계층의 하위에는 비디오의 논리적 내용과 의미적 내용을 표현하는 내용 계층이 존재한다. 논리적 내용 계층은 비디오의 논리적 구조를 나타내는 계층으로 비디오, 시퀀스, 썸, 샷의 네 가지 계층 구조로 표현한다. 또한 의미적 내용 계층은 비디오에 포함된 객체, 사건 그리고 배경을 표현하는 계층이다.

##### 3.2.1 구조적 내용 계층(Structural Content Layer)

내용 계층에 존재하는 구조적 내용 계층은 비정형화된 비디오의 논리적인 구조를 표현하기 위한 계층이다. 구조적 내용 계층은 비디오의 논리적 구조를 표현하기 위해 비디오, 시퀀스, 썸, 샷으로 구성된다. 원시 데이터에 대한 구조 모델의 각 계층을 생성하는 과정은 다음과 같다.

첫째, 원시 데이터에서 샷 영역 검출(shot boundary detection) 기법을 이용하여 자동적으로 샷들을 생성한다.

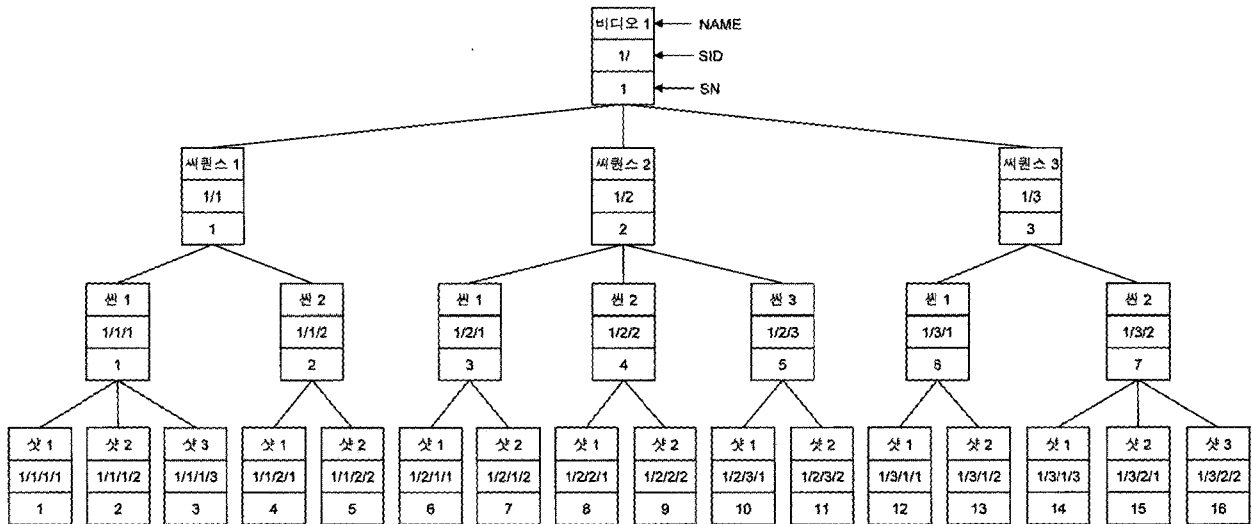
둘째, 샷 영역 검출에 의해 샷들이 추출되면 각 샷에 대해 하나 이상의 키프레임을 추출하고 샷에 대한 대표 키프레임을 기록한다.

셋째, 각 샷에서 추출된 키프레임 중 샷을 대표할 수 있는 대표 키프레임을 저장하고 샷에 순서 정보 SN(Sequential Number)를 부여한다.

넷째, 샷 계층이 생성되면 연속된 샷을 이용하여 썸 계층을 생성하고 각 썸을 대표하는 대표 키프레임과 SN을 부여한다.

다섯째, 썸 계층을 이용하여 시퀀스 계층을 위와 동일한 방법으로 구성하고 비디오 계층을 생성한다.

여섯째, 샷 레벨부터 비디오 레벨까지의 구조적 계층을 생성하고 SN이 부여되면 각 구조 계층에 고유한 식별자 SID (Structure ID)를 할당한다.



(그림 2) 구조적 내용 계층의 식별자 부여 방법

구조적 내용 계층에서는 각 구조 계층에 대한 시각적인 요약 정보를 표현한다. 이를 위해 샷 레벨에서 추출된 키프레임을 이용하여 각 계층에서는 구조 계층을 대표하기 위한 키프레임을 선택하여 각 구조 계층 내에 표현한다. 또한 각 구조적 내용 계층에서는 각 논리적 구조 계층에 대한 계층 정보와 순서 정보를 표현하기 위한 고유한 식별자를 부여한다.

각 논리적 계층에 부여된 계층 정보 SID는 상위 계층과 하위 계층의 논리적 계층 구조를 표현하기 위해 사용하고 순서 정보 SN은 동일한 계층에서 존재하는 구조 정보들 사이의 순서를 표현하기 위해 사용된다. (그림 2)는 각 구조 계층에 대한 내용이 생성되면 각 구조적 내용 계층에 고유한 식별자 SID와 SN을 부여하는 과정을 보여준다. 먼저 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 구조와 독립적으로 동일한 계층에서 순서 정보를 표현하기 위해 사용된다. 예를 들어, “시퀀스 2”에 “씬 1”은 씬 계층에서 3번째 존재하기 때문에 SN을 3으로 부여한다.

위에 방법에 의해 논리적 계층 정보와 순서 정보가 부여되면 각 계층에 대한 상세한 정보를 표현해야 한다. 먼저 샷 (shot) 계층은 구조적 내용 계층의 최하위 계층으로 하나의 카메라 움직임에 관한 기본 단위로 1대의 카메라로 연속되게 기록되는 프레임들의 연속이다. 이러한 샷은 장면의 전환이 없는 임의의 개수의 프레임으로 구성된다. 비디오에서 샷의 생성은 하나의 카메라가 연속된 움직임에 의해 기록되는 컷(cut)을 검출하기 위한 샷 영역 검출(shot boundary detection)에 수행된다. 샷 계층의 표현은 <표 2>와 같다. *shot\_id*는 (그림 2)와 같이 샷에 대한 식별자를 부여한 것

으로 “*videoID/sequenceID/sceneID/shotID*”의 조합으로 비디오 계층부터 샷 계층까지의 순서 정보를 조합하여 생성한다. *shot\_sn*은 샷 계층에서 현재 샷의 순서 정보, *start\_frameID*과 *end\_frameID*은 샷에 대한 존속 구간을 나타내기 위한 시작 프레임과 종료 프레임을 나타내며 *frame\_num*은 하나의 샷에 나타난 프레임 수를 나타낸다. 또한 *camera\_action*과 *special\_effect*는 샷에 나타난 카메라 효과와 특수 효과를 나타내며 *annotation*은 사용자 또는 편집자에 의해 부여된 주석 정보를 나타낸다. 마지막으로 *ref\_image*은 하나의 샷을 대표하는 키프레임 이미지를 나타낸다.

<표 2> 샷 계층의 표현

속 성	설 명
<i>shot_id</i>	샷 식별자
<i>shot_name</i>	샷 이름
<i>shot_sn</i>	샷 순서 정보
<i>start_frameID</i>	샷의 시작 프레임 식별자
<i>end_frameID</i>	샷의 종료 프레임 식별자
<i>frame_num</i>	샷의 프레임 수
<i>camera_action</i>	샷에서 발생한 카메라 효과
<i>special_effect</i>	샷에서 발생한 특수 효과
<i>annotation</i>	샷에 대한 주석 정보
<i>ref_image</i>	샷에 대한 대표 이미지

샷 계층의 상위에는 연속된 샷들을 연결한 씬 계층이 존재한다. 씬은 동일 시간, 동일한 장소에서 발생한 내용으로 하나 이상의 샷으로 구성된다. 씬 계층의 표현은 <표 3>과 같다. *scene\_id*는 씬에 대한 식별자로 “*videoID/sequenceID/sceneID*”의 조합으로 비디오 계층부터 씬 계층까지의 순서 정보를 조합하여 생성한다. *scene\_sn*은 씬 계층에서 현재 씬의 순서 정보, *shot\_num*은 현재 씬에 포함된 샷의 개수를 나타낸다.

〈표 3〉 씬 계층의 표현

속 성	설 명
<i>scene_id</i>	씬의 식별자
<i>scene_name</i>	씬의 이름
<i>scene_sn</i>	씬의 순서 정보
<i>start_frameID</i>	씬의 시작 프레임 식별자
<i>end_frameID</i>	씬의 종료 프레임 식별자
<i>shot_num</i>	씬에 포함된 샷의 개수
<i>annotation</i>	씬에 대한 주석 정보
<i>ref_image</i>	씬에 대한 대표 이미지

씬 계층의 상위에는 특정 조건에 따라 연속된 씬들을 연결한 시퀀스 계층이 존재한다. 시퀀스는 연결 가능한 이야기로 분리할 수 있는 샷들로 구성되며 하나 이상의 씬으로 구성된다. 시퀀스 레벨의 표현은 <표 4>와 같다. *sequence\_id*는 시퀀스에 부여된 식별자로 "videoID/sequenceID"의 조합으로 표현된다. 또한 *scene\_num*은 현재 시퀀스에 포함된 씬의 개수를 나타낸다.

〈표 4〉 시퀀스 계층의 표현

속 성	설 명
<i>sequence_id</i>	시퀀스의 식별자
<i>sequence_name</i>	시퀀스의 이름
<i>sequence_sn</i>	시퀀스의 순서 정보
<i>start_frameID</i>	시퀀스의 시작 프레임 식별자
<i>end_frameID</i>	시퀀스의 종료 프레임 식별자
<i>scene_num</i>	시퀀스에 포함된 씬의 개수
<i>annotation</i>	시퀀스에 대한 주석 정보
<i>ref_image</i>	시퀀스에 대한 대표 이미지

구조 기반 모델의 최상위에는 비디오 계층이 존재한다. 하나의 비디오는 하나 이상의 시퀀스로 구성되며 원시 비디오 데이터 전체의 구조적인 정보를 표현한다. 비디오 계층의 표현은 <표 5>와 같다. *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. 비디오 계층에

〈표 5〉 비디오 계층의 표현

속 성	설 명
<i>video_id</i>	비디오의 식별자
<i>video_name</i>	비디오의 이름
<i>sequence_num</i>	비디오에 포함된 시퀀스의 개수
<i>scene_num</i>	비디오에 포함된 씬의 개수
<i>shot_num</i>	비디오에 포함된 샷의 개수
<i>frame_num</i>	비디오에 프레임의 개수
<i>annotation</i>	비디오에 대한 주석 정보
<i>ref_image</i>	비디오에 대한 대표 이미지

서 표현되는 *sequence\_num*, *scene\_num*, *shot\_num*은 논리적 구조 계층에 표현된 전체적인 하위 구조 계층의 개수를 나타낸다.

### 3.2.2 의미적 내용 계층(Semantic Content Layer)

구조적 내용 계층은 비디오의 논리적 구조를 표현하는 계층이다. 이에 반해 의미적 내용 계층은 비디오에 포함된 객체, 사건 그리고 배경과 같은 비디오에 포함된 의미적인 내용을 표현하는 계층이다. 의미적 내용 계층은 구조적인 내용과 독립적으로 구성된다. 의미적 내용 계층을 구성하기 위해서는 다음과 같은 과정을 수행한다.

첫째, 의미적인 내용을 표현하기 위해 먼저 원시 데이터에서 각각의 의미적 내용을 추출하고 의미적 내용에 대한 시각적 요약 정보를 나타낼 키프레임을 추출한다.

둘째, 원시 데이터에서 추출된 각각의 의미적 내용들은 먼저 의미적 내용 테이블(semantic content table)에 등록하고 식별자를 부여한다.

셋째, 의미적 내용 테이블에 등록된 각각의 의미적 내용에 대한 구체적인 표현을 기술한다.

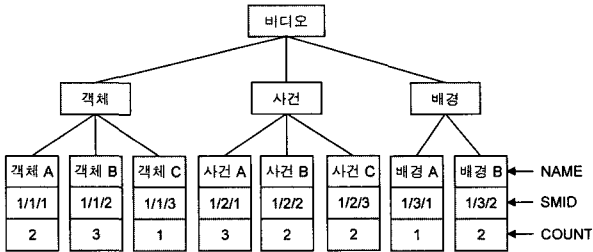
의미적 내용 테이블은 원시 데이터와 실제적인 객체, 사건 그리고 배경과의 관계를 매핑시키기 위한 테이블로 원시 데이터에서 추출한 의미적 내용에 대한 요약 정보를 표현한다. 의미적 내용 테이블은 다음 <표 6>과 같이 구성된다. *semantic\_type*은 표현하려는 의미적 내용의 타입을 설정하는 부분으로 표현하는 의미적 내용이 객체, 사건 그리고 배경 중 어떤 것인지를 설정한다. *semantic\_name*은 의미적 내용에 부여된 이름, *semantic\_id*는 의미적 내용에 부여된 식별자로 "semantic\_id/semantic\_type/id"의 조합으로 구성된다. 이때 *semantic\_id*는 원시데이터에 대한 의미적 내용 테이블에 부여한 식별자로 원시 데이터의 식별자에 대한 참고 기능을 수행한다. 또한 *semantic\_type*은 의미적 내용의 타입으로 객체는 1, 사건 2 그리고 배경 3으로 할당하고 *id*는 의미적 내용에 부여된 식별자 번호를 나타낸다. *semantic\_num*은 해당 의미적 내용이 비디오 내에 나타나는 횟수를 나타낸다.

〈표 6〉 의미적 내용 테이블

필 드	설 명
<i>raw_id</i>	원시데이터에 대한 식별자
<i>semantic_type</i>	의미적 내용의 타입(객체, 사건, 배경)
<i>semantic_name</i>	의미적 내용의 이름(NAME)
<i>semantic_id</i>	의미적 내용에 부여된 식별자(SMID)
<i>semantic_num</i>	의미적 내용이 나타나는 횟수(COUNT)

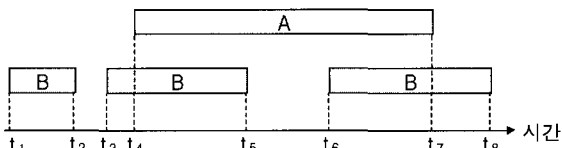
제안하는 모델링에서는 구조적 내용 계층에서와 동일하게 각각의 의미적 내용에 대한 식별자를 (그림 3)과 같이 부여한다. 의미적 내용 계층에서 식별자는 구조적 내용 계층과 달리 계층적인 상하 관계 또는 순서 정보가 필요 없

다. 따라서, 의미적 내용 테이블에 존재하는 *semantic\_id*는 단지 의미적 내용에 따른 분류를 수행할 수 있도록 한다.



(그림 3) 의미적 내용 식별자 부여 방법

의미적 내용 테이블에 비디오에 존재하는 객체, 사건 그리고 배경에 대한 정보를 등록하면 실제적인 의미적 내용 각각에 대한 정보는 객체, 사건, 배경 테이블에 표현한다. 하나의 의미적 내용은 비디오에서 여러 번 반복하여 나타날 수 있다. 의미적 내용 계층에서는 의미적 내용에 대한 존속 구간을 명확히 표현하기 위해 동일한 이름을 갖는 의미적 내용이 불연속적으로 나타날 경우 서로 다른 식별자를 부여한다. 예를 들어, (그림 4)와 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 식별자를 부여하기 위해서 먼저 의미

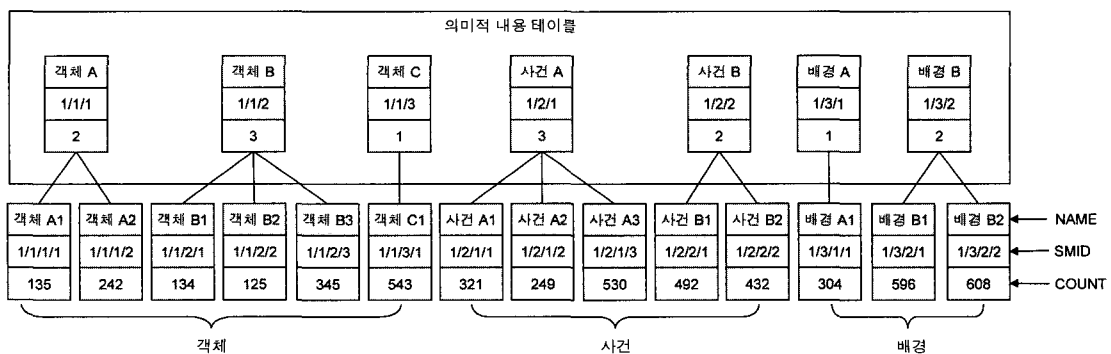
적 내용 테이블에 등록된 각각의 의미적 내용에 대한 식별자 SMID와 의미적 내용이 나타나는 횟수 COUNT 값이 필요하다. SMID와 COUNT를 이용하여 실제적인 객체, 사건 그리고 배경에 대한 식별자를 부여한다. (그림 5)는 각 의미적 내용에 대한 식별자를 부여하는 방법을 나타낸 것이다. 객체, 사건 그리고 배경에 대한 식별자는 “SMID/NUMBER” 형태로 의미적 내용 테이블에 기록된 SMID를 기준으로 “/”로 구분된 문자 뒤에 COUNT 수만큼의 식별자를 부여하여 동일한 이름을 갖는 객체가 나타나는 순서 정보를 표현한다. 이때,  $1 \leq \text{NUMBER} \leq \text{COUNT}$ 이다.

객체는 하나의 비디오에서 의미적인 정보를 표현하기 위해 연속된 프레임에 나타난 특정 사물이나 사람으로 의미적 내용 테이블에 등록된 객체에 대한 실제적인 정보를 표현한 것이다. 의미적 내용 테이블에 등록된 객체에는 각 객체의 이름과 객체가 나타나는 수가 표현되어 있다. 이를 통해 객체는 연속된 프레임에 나타난 객체에 대한 정보를 표현하고 (그림 5)에서 설명한 것과 같이 고유한 식별자 SMID를 부여한다.

객체는 시간에 따라 다양한 동작을 수행하며 여러 사건과 관련되어 나타난다. 따라서 객체가 시간에 따라 수행하는 동작과 시간의 변화에 따라 객체에 관련된 사건에 대한 정보를 표현해야 한다. 또한 객체의 존속 구간을 명확히 판단할 수 있도록 객체가 나타나는 시간을 표현해야 한다. 의미적 내용 계층에 존재하는 객체의 표현은 <표 7>과 같다.

<표 7> 객체 표현

속 성	설 명
<i>object_id</i>	객체 식별자(SMID)
<i>object_name</i>	객체의 이름(NAME)
<i>start_frameID</i>	객체의 시작 프레임 식별자
<i>end_frameID</i>	객체의 종료 프레임 식별자
<i>frame_num</i>	객체가 나타나는 프레임 수(COUNT)
<i>actionlist_pointer</i>	객체가 시간의 변화에 따라 수행한 동작
<i>eventlist_pointer</i>	객체가 시간의 변화에 따라 참여한 사건
<i>annotation</i>	객체에 대한 주석 정보
<i>ref_image</i>	객체에 대한 대표 이미지



(그림 5) 의미적 내용 식별자 부여

*object\_id*는 연속된 프레임에 나타난 객체에 부여된 객체 식별자 *SMID*이고 *object\_name*은 객체의 이름으로 의미적 내용 테이블에 등록된 객체의 이름을 나타낸다. 또한 *start\_frameID*와 *end\_frameID*는 객체가 나타나는 시작 프레임과 종료 프레임으로 객체가 나타내는 존속 구간을 판별하기 위한 것이다. *actionlist\_pointer*와 *eventlist\_pointer*는 객체가 시간에 따라 수행하는 동작과 객체에 관련된 사건에 대한 정보를 나타낸다. *ref\_image*는 객체의 시각적인 요약 정보를 나타내는 대표 이미지를 나타낸다.

사건은 특정 객체에 발생하는 의미적인 내용으로 시간의 변화에 따라 여러 객체들에 의해 발생될 수 있다. 따라서 하나의 사건은 특정 객체와 연관되어 있다. 객체는 시간에 변화에 따라 다른 객체 또는 사건들과 관련되어 있기 때문에 시공간적인 표현이 요구된다. 그러나 사건은 시간의 연속된 흐름에 따라 전개되기 때문에 공간적인 관계가 존재하지 않고 단지 시간적인 관계의 표현이 필요하다. 의미적 내용 계층에 나타나는 사건의 표현은 <표 8>과 같다. *event\_id*는 연속된 프레임에 나타난 사건에 부여된 고유한 식별자 *SMID*이고 *event\_name*은 사건의 이름으로 의미적 내용 테이블에 등록된 사건의 이름을 나타낸다. 객체와 동일하게 *start\_frameID*와 *end\_frameID*는 사건이 나타나는 시작 프레임과 종료 프레임으로 사건의 존속 구간을 나타내기 위해 사용된다. *objectlist\_pointer*는 사건에 관련된 객체에 대한 정보를 나타내며 *ref\_image*는 사건에 대한 시각적인 요약 정보를 나타내는 대표 이미지이다.

<표 8> 사건의 표현

속 성	설 명
<i>event_id</i>	사건 식별자(SMID)
<i>event_name</i>	사건의 이름(NAME)
<i>start_frameID</i>	사건의 시작 프레임 식별자
<i>end_frameID</i>	사건의 종료 프레임 식별자
<i>frame_num</i>	사건이 나타나는 프레임 수(COUNT)
<i>objectlist_pointer</i>	사건 관련된 객체
<i>annotation</i>	사건에 대한 주석 정보
<i>ref_image</i>	사건에 대한 대표 이미지

배경은 객체와 사건이 나타내는 배경 정보를 나타내는 것으로 시간의 변화에 따라 특정 객체 또는 사건이 발생하는 배경에 대한 정보를 나타낸다. 또한 객체와 사건과 같이 불연속된 프레임에 나타난 배경에 유일한 식별자를 부여한다. 배경의 표현은 <표 9>와 같다. *bg\_id*는 연속된 프레임에 나타난 배경에 부여된 식별자 *SMID*이고 *bg\_name*은 배경의 이름으로 의미적 내용 테이블에 등록된 배경의 이름을 나타낸다. *start\_frameID*과 *end\_frameID*은 배경이 나타나는 시작 프레임과 종료 프레임을 나타내며 배경의 존속 구간을 판별하기 위해 사용된다. *ref\_image*는 배경에 대한 시각적인 요약 정보를 나타내기 위한 대표 이미지이다.

<표 9> 배경의 표현

속 성	설 명
<i>bg_id</i>	배경 식별자(SMID)
<i>bg_name</i>	배경의 이름(NAME)
<i>start_frameID</i>	배경의 시작 프레임 식별자
<i>end_frameID</i>	배경의 종료 프레임 식별자
<i>frame_num</i>	배경이 나타나는 프레임 수(COUNT)
<i>annotation</i>	배경에 대한 주석 정보
<i>ref_image</i>	배경에 대한 대표 이미지

3.3 키프레임 계층(Keyframe Layer)

키프레임 계층은 제안하는 모델링의 최하위에 존재하는 계층으로 특정 시간 단위 또는 시각적인 특징 변화에 따라 원시 데이터에서 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 배경에 대한 시각적인 요약 정보를 표현하기 위한 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 계층에서 표현된 각 키프레임은 구조적 내용과 의미적 내용에 대한 시각적인 요약 정보를 표현하기 위해 사용된다. 따라서 키프레임 계층에서 표현되는 각 키프레임은 관련된 구조적 내용 계층과 의미적 내용 계층에 대한 정보를 표현해야 한다. 이를 통해 특정 내용에 대한 시각적인 요약 정보를 표현할 때 사용될 키프레임을 판별할 수 있다.

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. 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 <표 10>과 같다. *keyframe\_id*는 키프레임 식별자로 원시 데이터에서 몇 번째 프레임에서 추출되었는지를 판별하기 위해 사용된다. *visual\_featurelist*는 키프레임에 대한 시각적인 특징 정보를 나타낸 것으로 이미지 처리 과정에서 자동적으로 추출된 특징 정보를 표현한다. *objectlist\_pointer*는 키프레임에 포함된 객체에 대한 식별자를 나타낸 것으로 하나의 키프레임에는 여러 객체가 동시에 나타날 수 있으므로 이에 대한 정보를 표현한 테이블에 대한 포인터이다. *event\_id*와 *bg\_id*는 키프레임에 관련된 사건과 배경에 대한 식별자를 나타낸다. 또한 *shot\_id*는 키프레임에 속한 샷에 대한 식별자로 샷에 대한 시각적인 요약 정보를 표현



하기 위해 사용된다.

〈표 10〉 키프레임 계층의 표현

속 성	설 명
keyframe_id	키프레임 식별자
visual_featurelist	키프레임에 대한 시각적 특징
objectlist_pointer	키프레임에 속한 객체에 정보 리스트
event_id	키프레임이 속한 이벤트 식별자
bg_id	키프레임이 속한 배경 식별자
shot_id	키프레임이 속한 샷 식별자

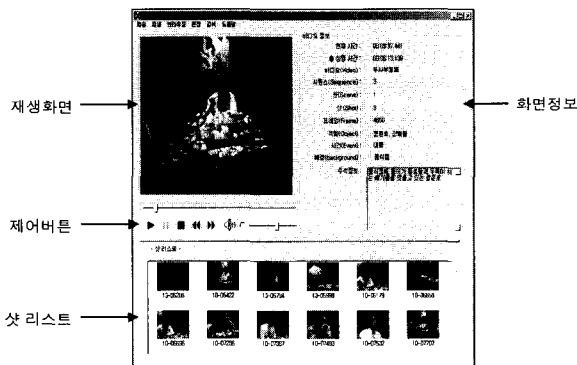
#### 4. 브라우징 인터페이스

본 논문에서는 제안하는 모델링에 기반하여 크게 [비디오 재생]과 [내용 정보]의 두 가지 브라우저를 구현하였다. 먼저 [비디오 재생]은 비디오를 재생과 제어에 관련된 기능과 현재 재생되는 화면에 대한 정보를 나타내는 기능을 수행한다. [내용 정보]는 원시데이터, 구조적 내용 그리고 의미적 내용에 대한 정보를 브라우징하는 기능을 제공한다.

본 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.

##### 4.1 비디오 재생

비디오 데이터 재생을 위한 인터페이스는 비디오의 재생을 위한 제어 기능은 물론 현재 재생되고 있는 화면에 대한 정보와 다음에 재생될 샷 정보를 나타낸다. (그림 6)은 비디오 재생 인터페이스를 나타낸 것이다. 비디오 데이터 재생 인터페이스는 [비디오 재생 및 제어 기능], [현재 화면 정보] 그리고 [샷 리스트]의 세 부분으로 구성되어 있다.



(그림 6) 재생 및 현재 화면 정보

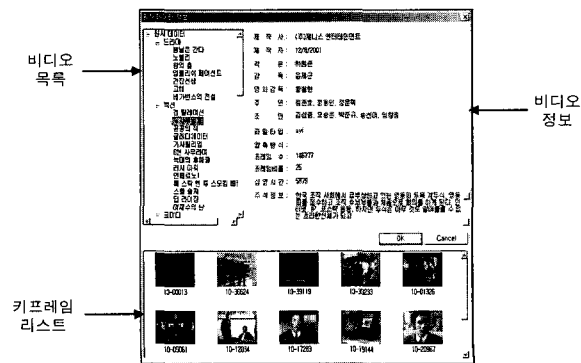
상위 왼쪽 부분은 [비디오 재생 및 제어 기능] 부분으로 비디오 데이터의 재생 화면과 재생 및 볼륨 제어 기능으로 구성되어 있다. 화면 재생에 관련된 기능으로 재생, 일시정지, 정지, 되감기, 빨리감기 기능과 볼륨 제어 기능으로 음

소거 기능과 소리크기 조절 기능 등이 있다. 상위 오른쪽 부분은 [현재 화면 정보] 부분으로 현재 재생되는 화면에 대한 정보를 나타내는 부분이다. [현재 화면 정보] 부분에 나타나는 부분은 원시 데이터 계층에 존재하는 비디오의 일반적인 내용과 현재 재생되고 있는 샷에 대한 정보를 표현한다. 또한 현재 재생되는 화면에 나타나는 객체, 사건 그리고 배경과 같은 의미적 내용을 함께 나타낸다. 하위 오른쪽 부분에 존재하는 [샷 리스트]는 다음에 재생될 샷에 대한 정보를 출력하는 부분이다. [샷 리스트]에 출력되는 이미지는 현재 재생되고 있는 샷 이후에 재생될 샷에 대한 대표 이미지이다. 구조적 내용 계층에 존재하는 샷 계층에는 각 샷을 대표하는 대표이미지를 포함하고 있다. 이러한 대표이미지는 키프레임 계층에 표현되어 있다. 사용자는 [샷 리스트]에 존재하는 샷의 대표 이미지를 통해 다음에 재생될 내용을 파악할 수 있다. 사용자가 특정 샷을 재생하고 싶을 때는 [샷 리스트]에 존재하는 샷의 대표 이미지를 선택하면 사용자가 선택한 이미지를 포함하는 샷을 재생시킬 수 있다.

##### 4.2 내용 정보 브라우징

내용 정보 브라우징 인터페이스는 원시 데이터 자체에 대한 정보뿐만 아니라 비디오에 포함된 구조적 내용과 의미적 내용에 대한 정보를 브라우징하는 기능을 제공한다. 내용 정보 브라우징은 [원시 데이터 브라우저], [구조적 내용 브라우저] 그리고 [의미적 내용 브라우저]로 구성되어 있으며 텍스트 기반의 요약 정보와 시각적인 요약 정보를 함께 제공한다.

[원시 데이터 브라우저]는 비디오 데이터의 가공되지 않는 원시 데이터 자체에 대한 정보를 제공한다. [원시 데이터 브라우저]는 (그림 7)과 같이 세 부분으로 구성되어 있다. 상위 왼쪽 부분은 데이터베이스에 저장된 비디오의 목록을 나타낸다. 비디오의 목록은 장르별로 구분하여 비디오 데이터에 목록을 나타낸다. 상위 오른쪽 부분은 비디오에 대한 정보를 나타내는 부분으로 원시데이터 계층에 표현된 비디오 자체가 갖고 있는 가공되지 않는 정보를 표현한다. 원시 데이터 브라우저의 하위 부분은 원시 데이터에 대한 시각적인 요약 정보를 나타낸다. 원시 데이터의 시각적인

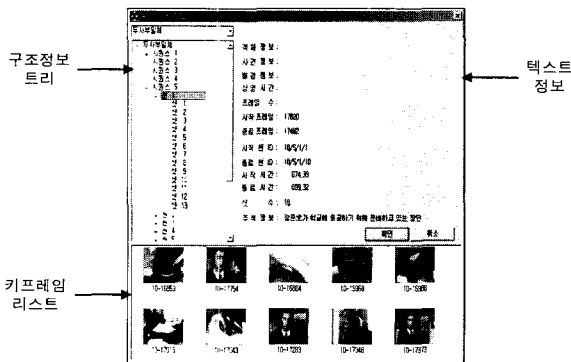


(그림 7) 원시 데이터 브라우저

요약 정보를 통해 사용자가 비디오에 대한 요약 정보를 볼 수 있다.

[원시 데이터 브라우저]는 비디오 자체가 갖고 있는 정보를 텍스트로 표현하고 비디오에 대한 시각적인 요약 정보를 나타낸다. 이에 반해 [구조적 내용 브라우저]과 [의미적 내용 브라우저]는 비디오에 대한 논리적인 계층 구조와 객체, 사건 그리고 배경과 같은 의미적 내용에 대한 정보를 나타낸다. [구조적 내용 브라우저]는 (그림 8)과 같이 비디오의 논리적인 계층 구조를 비디오, 시퀀스, 씬, 샷의 4 계층으로 표현되며 제안하는 모델링의 구조적 내용 계층에 표현된 정보를 나타낸다.

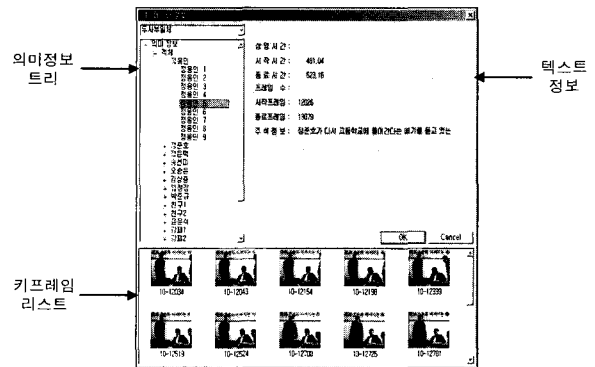
[구조적 내용 브라우저]는 [원시데이터 브라우저]와 유사하게 텍스트 기반의 요약 정보와 시각적인 요약 정보를 나타낸다. [구조적 내용 브라우저]는 세 부분으로 구성되어 있다. 상위 왼쪽 부분은 비디오의 논리적인 계층 구조를 트리 형태로 표현하는 부분이다. 상위 오른쪽 부분은 논리적 계층 구조에 대한 정보를 나타내는 부분으로 트리 형태로 표현된 논리적 계층의 이름을 선택하면 이에 대한 텍스트 기반의 정보를 나타낸다. 논리적 계층 구조에 대한 텍스트 기반의 요약 정보는 논리적 계층 구조 내에 표현된 구조적 내용과 논리적 계층 구조 내에 포함된 의미적 내용을 나타낸다. 논리적 계층 구조에 포함된 의미적 내용은 제안하는 모델링의 내용 계층에 존재하는 구조적 내용 계층과 의미적 내용 계층에 대한 매핑 테이블(mapping table)을 이용하여 구조적 내용에 포함된 의미적 내용을 표현한다. [구조적 내용 브라우저]의 하위 부분은 논리적 계층 구조에 대한 시각적인 요약 정보를 나타낸다. 각 논리적 계층에 대한 시각적인 요약 정보는 현재 선택된 계층 구조 하위에 존재하는 계층 구조에 대한 대표 이미지를 출력한다. 예를 들어 시퀀스에 대한 시각적인 요약 정보는 선택된 시퀀스 하위에 존재하는 씬들에 대한 대표 이미지를 출력한다.



(그림 8) 구조적 내용 브라우저

[의미적 내용 브라우저]는 비디오에 포함된 객체, 사건 그리고 배경과 같은 의미적 내용 정보를 나타낸다. (그림 9)는 [의미적 내용 브라우저] 화면을 나타낸다. [의미적 내용 브라우저]는 [구조적 내용 브라우저]와 유사하다. 상위 왼쪽 부분은 비디오에 나타난 의미적 내용을 객체, 사건,

배경에 의해 분류하고 각 의미적 내용에 대한 리스트를 트리 형태로 표현하는 부분이다. 상위 오른쪽 부분은 의미적 계층 구조에 대한 정보를 나타내는 부분이다. [의미적 내용 브라우저]의 하위 부분은 의미적 내용 계층에 대한 시각적인 요약 정보를 나타낸다. 의미적 내용에 대한 시각적인 요약 정보는 모델링의 키프레임 계층에 표현되어 있으며 사용자가 특정 객체의 이름을 선택하면 객체의 이름에 해당하는 식별자를 통해 키프레임 계층에 존재하는 키프레임 리스트를 출력한다.



(그림 9) 의미적 내용 브라우저

5. 비교 분석

본 논문에서는 비디오 데이터를 효과적으로 표현하기 위한 새로운 모델링과 모델링에 따라 표현된 정보를 효과적으로 브라우징할 수 있는 다양한 브라우저를 구현하였다. 이 장에서는 제안된 비디오 브라우징 기법과 모델링에 대한 비교 분석을 통해 제안하는 모델링 및 비디오 브라우징 기법의 우수성을 증명한다. 제안하는 비디오 모델링 및 브라우징 기법을 비교를 수행하기 위해 [17-21]에서 제안된 기법과의 비교를 수행한다. 편의상 [17-21]을 CBVS(Content-based Browsing of Video Sequence), VideoSTAR, SCVB(Structured and Content-based Video Browsing), VIBE, UFVB(Unified Framework of Video Browsing)라 한다. 또한 본 논문에서 제안하는 기법을 LCVMB(Layerd Content Video Modeling and Browsing)이라 한다. 기존의 모델링 및 브라우징 기법에 대한 비교 분석은 <표 11>과 같다.

제안하는 모델링은 계층화된 비디오 모델링으로 원시 데이터 계층 내에 비디오 자체가 갖고 있는 정보를 표현한다. 또한 내용 계층 내에 구조적 내용은 물론 의미적 내용을 표현한다. 따라서 기존의 브라우징 기법에 비해 원시 데이터에 대한 정보는 물론 의미적 내용에 대한 정보를 브라우징할 수 있는 기능을 제공한다. [17]에서 제안된 CBVS과 제안하는 LCVMB는 비디오 데이터 재생 및 제어 기능을 제공한다. 그러나 CBVS에서는 비디오 데이터의 재생 및 제어 기능을 물론 재생되고 있는 현재 화면에 대한 정보와 다음에 재생될 샷 리스트를 제공한다.

〈표 11〉 비교 분석

	모델링		브 라 우 저			
	구조적 내 용	의미적 내 용	재 생	원 시 데이 터	구조적 내 용	의미적 내 용
CBVS	○	×	○	×	○	×
VideoSTAR	○	×	×	×	○	×
SCVB	○	×	×	×	○	×
ViBE	○	×	×	×	○	×
UFVB	○	×	×	×	○	×
LCVMB	○	○	○	○	○	○

LCVMB는 기존의 모든 브라우징 기법들과 같이 구조적 내용에 대한 정보를 브라우징하는 기능을 제공한다. 그러나 기존의 브라우징 기법은 샷에서 추출된 키프레임 리스트를 통해 시각적인 요약 정보만을 제공한다. 이에 반해 제안하는 LCVMB는 트리 기반으로 비디오 데이터의 구조적 내용을 표현하는 것은 물론 각 구조적 내용에 대한 텍스트 요약 정보와 키프레임 리스트를 통한 시각적인 요약 정보를 제공한다.

6. 결론 및 향후 연구

본 논문에서는 비디오 데이터에 대한 내용 정보를 효과적으로 브라우징하기 위한 계층화된 내용 모델링과 브라우징 기법을 제안한다. 제안하는 모델링은 비디오 데이터를 원시 데이터 계층, 내용 계층 그리고 키프레임 계층으로 표현한다. 내용 계층은 구조적 내용과 의미적 내용을 표현하고 내용 계층에 존재하는 구조적 내용과 의미적 내용 사이의 관계는 매핑 테이블을 통해 표현한다. 키프레임 계층은 구조적 계층과 의미적 내용에서 추출한 키프레임을 통해 시각적인 요약 정보를 제공한다. 본 논문에서는 제안하는 모델링에 따라 비디오의 재생과 관련된 브라우저와 원시 데이터 계층, 내용 계층에 존재하는 각 내용 정보를 나타낼 수 있는 브라우저를 구현하였다. 향후 연구 방향으로 제안하는 모델링에 따라 비디오 데이터를 저장하고 검색을 수행할 수 있는 비디오 검색 시스템을 개발할 예정이다.

참 고 문 헌

[1] 박유미, 김용걸, 진성일, 김선주, "비디오 데이터베이스 구축을 위한 메타데이터 모델링 기술", 정보과학회지, 제15권 제11호, pp.473-482, 1997.  
 [2] E. Oomoto and K. Tanaka, "OVID : Design and Implementation of a Video-Object Database System," IEEE Transaction on Knowledge and Data Engineering, Vol.5, No.4, pp.629-643, 1993.  
 [3] R. Hjelsvold, "VideoSTAR - A Database for Video Information Sharing," Ph.D. Thesis, Norwegian Institute of

Technology, 1995.  
 [4] M. La Cascia and E. Ardizzone, "JACOB : Just a content-based query system for video databases," ICASSP-96, pp.7-10, 1996.  
 [5] 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, 1997.  
 [6] Wactlar, H., "Informedia - Search and Summarization in the Video Medium," Pro. Imagina 2000 Conference, Monaco, 2000.  
 [7] Rune Hjelsvold and Roger Midtstraum, "Modelling and Querying Video Dat," VLDB '94, pp.686-694, 1994.  
 [8] 류은숙, 이규철, "구조화된 비디오 문서의 데이터 모델 및 질의어와 색인 기법", 한국멀티미디어학회논문지, 제1권 제1호, pp.1-17, 1998.  
 [9] Yong Rui, Thomas, S. Huang and Sharad Mehrotra, "Constructing Table-of-Content for Videos," ACM Multimedia Systems Journal, Special Issue Multimedia Systems on Video Libraries, Vol.7, No.5, pp.359-368, 1999.  
 [10] Uma Srinivasan and Gerrit Riessen, "A Video Data Model for Content-Based Search," DEXA Workshop 1997, pp. 178-182, 1997.  
 [11] Y. M. Park, Y. K. Kim, S. I and Jin, W. Choi, "Hierarchical Structure-based Metadata Model for Video Database Applications," ISCA COMPUTERS AND THEIR APPLICATIONS, pp.242-245, 1998.  
 [12] 최지희, 용환승, "시간 관계성을 기반으로 한 비디오 데이터 모델의 설계 및 구현", 한국멀티미디어학회논문지, 제2권 제3호, pp.252-264, 1999.  
 [13] 윤미희, 윤용익, 김교정, "비디오의 의미검색과 유사성검색을 위한 통합비디오정보시스템", 정보처리논문지, 제6권 제8호, pp.2031-2041, 1999.  
 [14] Al Safadi, L. A. E and Getta, J. R., "Semantic modeling for video content-based retrieval systems," Computer Science Conference, ACSC 2000, 23rd Australasian, pp.2-9, 2000.  
 [15] Jia-Ling Koh, Chin-Sung Lee and Chen, A. L. P., "Semantic video model for content-based retrieval," Multimedia Computing and Systems, IEEE International Conference on, Vol.2, pp.472-478, 1999.  
 [16] M. Petkovic and W. Jonker, "A Framework for Video Modeling," Eighteenth IASTED International Conference Applied Informatics, Innsbruck, Austria, 2000.  
 [17] F. Arman, R. Depommier, A. Hsu and M-Y. Chiu, "ContentBased Browsing of Video Sequences," Pro. ACM International Conference on Multimedia, pp.97-103, 1994.  
 [18] Rune Hjelsvold, Roger Midtstraum and Olav Sandsta, "Searching and Browsing a Shared Video Database," Pro. 1995 International Workshop on Multi-Media Database

Management Systems, pp.90-98, 1995.

- [19] Jong Jiang Zhang and Wei-Ying Ma, "Structured and content-based video browsing," Pro. Signals, Systems & Computers, Conference Record of the Thirty-Second Asilomar Conference on, Vol.1, pp.910-914, 1998.
- [20] J. Chen, C. M. Taskiran, A. Albiol, C. A. Bouman and E. J. Delp, "ViBE : A Video Indexing and Browsing Environment," Pro. SPIE Conference on Multimedia Storage and Archiving Systems IV, Vol.3846, pp.148-164. 1999.
- [21] Yong Rui and Thomas Huang, "Chapter 9.2 : A Unified Framework for Video Browsing and Retrieval," The Image and Video Processing Handbook edited by Alan Bovik, Academic Press, pp.705-715, 2000.
- [22] Francis Li, Anoop Gupta, Elizabeth Sanocki, Li-wei He and Yong Rui, "Browsing Digital Video," In CHI Letters : Human Factors in Computing Systems, CHI 2000, pp.169-176, 2000.
- [23] A. Divakaran and R. Cabasson, "Content-based browsing system for personal video recorders," Pro. International Conference on Consumer Electronics, pp.18-20, 2002.
- [24] M. Guillemot, P. Wellner, D. Gatica-Perez and J-M. Odo-bez, "A Hierarchical Keyframe User Interface for Browsing Video over the Internet," Pro. the 9th International Conference on Human-Computer Interaction, 2003.



**복 경 수**

e-mail : ksbok@netdb.chungbuk.ac.kr  
 1998년 충북대학교 수학과 이학사  
 2000년 충북대학교 정보통신공학과 공학석사  
 2000년~현재 충북대학교 전기전자 컴퓨터공학부 박사과정

관심분야 : 자료 저장 시스템, 멀티미디어 데이터베이스, 이동객체 데이터베이스, 고차원 색인 구조, 시공간 색인구조 등



**이 낙 규**

e-mail : true01@netdb.chungbuk.ac.kr  
 2001년 충북대학교 정보통신공학과 공학사  
 2003년 충북대학교 전기전자 컴퓨터공학부 공학석사  
 2003년~현재 충북대학교 전기전자 컴퓨터공학부 박사과정

관심분야 : XML, 멀티 미디어 검색, 자료 저장 시스템, 시공간 색인 구조



**허 정 필**

e-mail : hjungpil@netdb.chungbuk.ac.kr  
 2002년 충북대학교 정보통신공학과 공학사  
 2002년~현재 충북대학교 전기전자 컴퓨터공학부 석사과정  
 관심분야 : 멀티 미디어 검색, 자료 저장 시스템, 고차원 색인 구조



**유 재 수**

e-mail : yjs@cbucc.chungbuk.ac.kr  
 1989년 전북대학교 컴퓨터공학과 공학사  
 1991년 한국과학기술원 전산학과 공학석사  
 1995년 한국과학기술원 전산학과 공학박사  
 1995년~1996년 목포대학교 전산통계학과 전임강사

1996년~현재 충북대학교 전기전자 컴퓨터공학부 부교수 및 컴퓨터정보통신연구소  
 관심분야 : 데이터베이스 시스템, 정보검색, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅 등



**조 기 형**

e-mail : khjoe@cbucc.chungbuk.ac.kr  
 1966년 인하대학교 전기공학과 공학사  
 1984년 제주대학교 산업공학과 공학석사  
 1992년 경희대학교 전자공학과 공학박사  
 1981년~1988년 충주공업전문대학 조교수  
 1996년~1999년 전기전자공학부장

1988년~현재 충북대학교 전기전자 컴퓨터공학부 교수 및 컴퓨터정보통신연구소  
 관심분야 : 데이터베이스시스템, 화상처리 및 통신, 통신 프로토콜, 분산 객체 컴퓨팅 등



**이 병 엽**

e-mail : bylee@pcu.ac.kr  
 1991년 한국과학기술원 전산학과 공학사  
 1993년 한국과학기술원 전산학과 공학석사  
 1997년 한국과학기술원 경영정보공학 공학박사  
 1997년~2003년 대우정보시스템 eBiz 사업본부

2003년~현재 배재대학교 전자상거래학부 전임강사  
 관심분야 : 데이터마이닝, XML, 인공지능, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅 등