논문 2008-45CI-5-5

비디오 서버에서 온톨로지를 이용한 의미기반 장면 검색

(Semantic-based Scene Retrieval Using Ontologies for Video Server)

정 민 영*, 박 성 한**

(Min Young Jung and Sung Han Park)

요 약

최근 멀티미디어 정보의 양이 빠른 속도로 증가함에 따라 비디오 자료에 대한 효율적 관리는 매우 중요한 의미를 가지게 되었다. 비디오는 대용량적인 특성과 비정형적인 특성을 가지고 있어 신속하고 효율적으로 비디오 검색을 하기 위해서는 정확한 특징 정보를 추출하여 비디오 색인 구조를 구축해야 한다. 비디오 색인 구조는 전통의 데이터베이스와는 다른 모델링 방법과 검색 방법을 사용한다. 따라서 비디오 색인 구조에서 검색의 속도와 정확도를 향상시키기 위해서는 새로운 비디오 색인 구조가 필요하다. 본 논문에서는 의미적으로 비디오를 장면단위로 검색할 수 있는 비디오 온톨로지 시스템을 제안한다. 비디오 온톨로지 시스템은 장면의 내용에 대한 키워드를 구조화 시킨 장면이름 온톨로지와 장면이 가지는 특징 정보에 대한 정보를 가지는 장면 모델 온톨로지로 구성된다. 장면 이름 온톨로지는 색인된 내용에 대한 의미적 검색이 가능하도록 단어들을 트리 구조로 저장된다. 그리고 장면 모델 온톨로지는 색상, 모양, 재질과 같은 저수준 정보와 객체, 이벤트 같은 고수준 정보의 의미적 차이를 극복해 줌으로써 의미기반 검색이 가능하게 해준다.

Abstract

To ensure access to rapidly growing video collection, video indexing is becoming more and more important. In this paper, video ontology system for retrieving a video data based on a scene unit is proposed. The proposed system creates a semantic scene as a basic unit of video retrieval, and limits a domain of retrieval through a subject of that scene. The content of semantic scene is defined using the relationship between object and event included in the key frame of shots. The semantic gap between the low level feature and the high level feature is solved through the scene ontology to ensure the semantic-based retrieval.

Keywords :: video representation, semantic search, ontologies, video retrieval, semantic

I. 서 론

멀티미디어 콘텐츠의 제작과 전송이 컴퓨터와 통신 관련 기술의 발달에 따라 증가 하고 있다. 이에 따라 최근에 멀티미디어 콘텐츠의 관리가 매우 중요한 문제로 대두 되고 있다. 그 대표적인 연구로써 내용기반 연구를 들 수 있다[1].

내용기반 연구는 검색의 정확도를 높이고, 속도를 빠르게 하기 위하여 사전에 색상과 같은 저수준의 정보를 자동적으로 추출하여 메타데이터로 생성해 놓는 연구 방법이다. 추출된 메타데이터 정보는 멀티미디어 데이터를 관리하고 검색하는데 이용된다. 하지만, 초기의 내용기반 연구들은 메타데이터의 정보표현 및 구조에 대한 표준이 제정되어 있지 않아 일반화된 검색을 지원하는데 어려움이 있었다. 이러한 표준화에 대한 요구와 멀티미디어 검색 분야 지속적인 관심으로 MPEG-7 (Moving Picture Experts Group layer-7) 표준이 제정되면서 검색을 위한 정보 표현 및 저장 방법이 통일 되었다[2].

이처럼 초기의 내용기반 연구는 저수준과 고수준 정보의 관련성을 정의하지 않아 멀티미디어 데이터의 의미적 검색에는 어려움이 있었다. 이러한 문제들을 해결하기 위해서는 저수준 정보와 고수준 특징 정보 사이의

* 학생회원, ** 정회원, 한양대학교 컴퓨터공학부
(Department of Computer Science Eng., Hanyang Univ.)

※ 본 연구는 한국과학재단 목적기초연구 지원(R01-2006-000-10876-0)으로 수행되었음
접수일자: 2008년8월20일, 수정완료일: 2008년9월8일

의미적 차이(semantic gap)를 극복하기 위한 연구가 필요하게 되었다. 이를 위해 온톨로지(ontology)나 관련어 집(thesaurus)에 의해 구조화된 배경 지식(background knowledge)을 이용하는 연구가 시도되고 있다.

Ⅱ장에서는 이러한 온톨로지나 배경 지식을 이용한 관련연구를 살펴보고, Ⅲ장에서는 장면을 의미적 검색이 가능하도록 색인하기 위한 비디오 온톨로지 시스템을 제안한다. Ⅳ장에서는 본 논문에서 제안하는 방법과 기존의 지식기반 구조를 비교하여 실험하고, Ⅴ장 에서는 결론을 맺는다.

Ⅱ. 관련연구

온톨로지를 이용하여 장면 검색을 하는 연구로는 Hoogs, Stein과 Hollink 등이 제안한 저수준의 특징 정보와 고수준의 특징 정보를 확장 된 WordNet[3]이라는 단어사전을 이용하여 기술 한 연구가 있다[4~7]. 하지만, 이 연구들은 기존의 온톨로지인 WordNet을 이용하여 구조가 너무 방대하고, 복잡하여 실제 동영상 검색을 위한 구조로 적합하지 못하는 문제가 있다. 그리고 색인 구조가 키프레임 영상에 대해서만 기술하기 때문에 비디오 검색이 아니라 마치 이미지 검색과 같아지는 문제가 있다. 본 논문에서는 이러한 문제들을 해결하기 위하여 온톨로지를 이용하여 장면 전체에 대한 의미적 색인 구조를 제안하고자 한다.

Ⅲ. 제안하는 비디오 온톨로지 시스템

본 논문에서 제안하는 의미 기반 검색을 위한 비디오 온톨로지 시스템은 그림 1에서 보이는 것과 같이 크게 네 부분으로 구성된다. 먼저 의미적 장면 생성 과정은

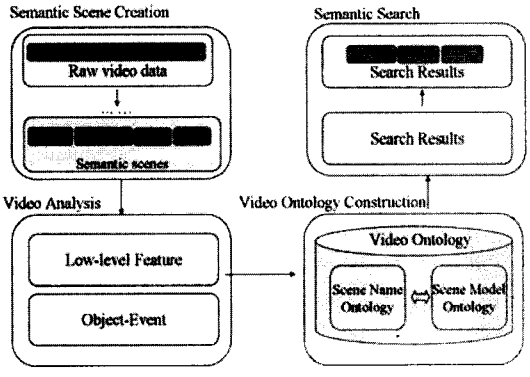


그림 1. 의미기반 검색을 위한 온톨로지 시스템
Fig. 1. Ontologies system for semantic video search.

비디오를 장면단위로 모델링하고 검색하기 위한 전처리 과정이다. 다음으로 비디오 분석과정에서는 MPEG-7표준에서 제공하는 멀티미디어 특징 정보 기술자를 이용하여 동영상의 특징 정보를 추출해 낸다. 비디오 분석 과정을 통해 생성된 정보는 다음단계인 비디오 온톨로지를 구축하는데 이용된다. 그리고 비디오 온톨로지 구축은 두 가지 온톨로지 구성되는데, 의미적 장면이 포함하는 내용의 키워드 트리가 저장되어 있는 장면 이름 온톨로지와 장면이 가지는 멀티미디어 특징정보의 관계성을 저장하는 장면 모델 온톨로지로 이뤄진다. 마지막으로 동영상 검색 부분은 구축된 비디오 온톨로지를 통해 의미적 기반 검색 결과를 보여준다. 각 부분의 자세한 내용은 다음과 같다.

1. 의미적 장면 생성

비디오 검색을 위해 기본이 되는 과정은 비디오 분할이다. 본 논문에서는 그 분할 단위를 '의미적 장면(semantic scene)'이라 정의한다. 의미적 장면 생성과정은 기존의 제안되어진 방법을 순차적으로 실행하여 생성한다. 장르 정보는 온톨로지 도메인을 결정하기 위해 이용되는 정보로써 비디오 전체에 대해 한 번의 정의만 필요하므로 의미적 장면 생성 과정 이전에 수행한다.

의미적 장면 생성의 첫 번째 단계는 동영상을 샷 단위로 분리한다. 분리된 샷들은 내용의 관련성은 있으나 장면 전환 효과로 인해 나뉘진 부분이 다수 존재 한다. 이를 해결하기 위하여 먼저 장면이 전환되는 지점을 검출해 내고, 이를 기준으로 하여 장면 변화에 사용 되지 않은 전 후 프레임을 비교하여 그 프레임들의 차이 값이 적으면 그 샷들은 자동적으로 그룹화 된다. 하지만 이렇게 생성된 비디오 샷들은 아직 의미적으로 장면이 분할되었다고 보기 어렵다. 따라서 사용자의 주관적 판단을 통해 의미적 장면단위로 분할 또는 그룹화해야 한다. 이러한 과정의 최종결과물로써 의미적 장면이 생성된다.

2. 비디오 분석

비디오 분석과정은 비디오 온톨로지 구축에 필요한 정보를 의미적 장면으로부터 추출하여 획득하는 과정이다. 이 과정은 의미적 장면을 구성하는 샷을 기본 단위로 하여 분석한다. 샷은 한 개의 키 프레임으로 정의되며, 키프레임은 하나의 정지 영상 정보이다. 이 키프레임은 장면 모델 온톨로지 구성을 위한 저수준 정보 즉 색상, 모양, 재질 및 모션정보를 포함한다. 그리고 정지

영상은 객체와 이벤트 정보를 포함하고 있으며 이를 본 논문에서는 고수준 정보라 한다. 고수준 정보인 객체는 색상, 재질, 모양의 재질의 저수준 정보를 이용하여 , 이벤트 정보는 이웃한 키프레임의 관계성 정보에 기반을 두어 기술한다. 이벤트는 객체의 움직임에 중점을 두어 기술하기 때문에 모션 정보 중에서도 객체의 방향성과 강도를 이용한다.

3. 장면이름 온톨로지

장면이름 온톨로지는 장면 색인 시 사용되는 객체와 이벤트의 단어들(terms)에 대한 사전이다. 장면이름 온톨로지의 도메인은 장르에 따라 구분되어 저장된다. 단어가 외부적으로 보이는 텍스트는 같더라도 장르에 따라 의미가 다를 수 있기 때문이다. 기본 구조는 단어들을 계층적인 상·하위개념으로 정의한다. 이러한 계층적인 개념 구조는 추상화된 의미도 추론 가능하다. 그림 2에서 보는 바와 같이 장면이름 온톨로지 구축은 단어 트리 구조를 저장하는 과정이 선행되어야 한다. 주제들은 특정 장르에 포함되고, 각각 계층적인 구조를 가진다. 주제들의 트리 관계는 자식 노드는 부모노드의 하위 개념이다. 각 주제들이 가지는 객체와 이벤트 단어 트리는 일정한 규칙을 가지고 저장된다. 주제는 주 객체를 루트 노드로 하여 정의한다.

그림 2의 (A)는 주 객체의 이벤트 관계를 나타내며, 왼쪽 자식(left child)이 주 객체의 이벤트가 된다. 그리고 오른쪽 형제(right sibling)는 주 객체와 관계를 갖는 객체들의 노드들이다. 주 객체가 다른 객체(O_b)와의 관계는 그림 2의 (B)와 같이 왼쪽 자식 노드가 부모노드와의 관계를 표현하는 이벤트가 된다. 그리고 오른쪽 형제 노드들은 O_b가 가지는 이벤트나 O_b와의 관계를

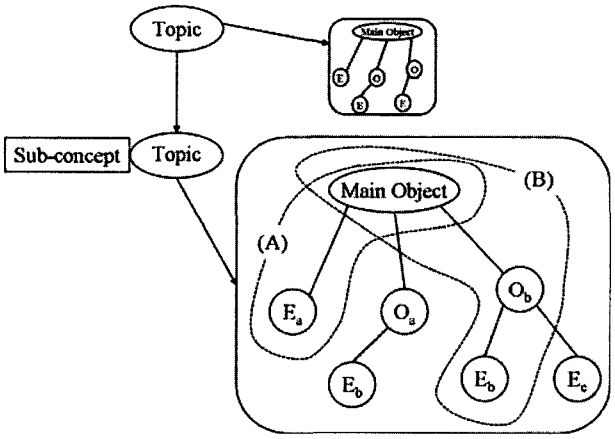


그림 2. 장면이름 온톨로지 단어 트리 구조
Fig. 2. The tree structure of scene name ontology.

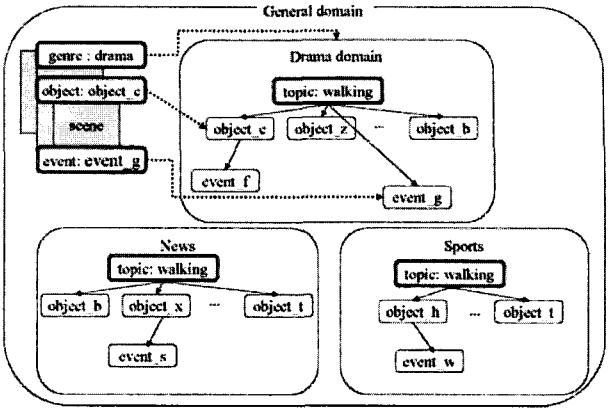


그림 3. 장면이름 온톨로지 내 단어 속성 그래프
Fig. 3. The graph of the word properties in scene name ontology.

갖는 객체 노드들이 된다. 이러한 원리로 장면이름 온톨로지의 단어 트리를 완성하게 된다.

구축 되어진 장면 이름 온톨로지를 이용하여 실제 의미적 장면이 색인되는 과정은 그림 3과 같다. 의미적 장면 생성 과정에서 얻어진 장르정보인 'drama'는 'Drama'도메인을 결정하기 위한 정보로 사용되고, 객체와 이벤트 정보인 'object_c'와 'event_g'는 'walking'이라는 주제로 색인하기 위한 정보를 사용된다.

4. 장면모델 온톨로지

장면 모델 온톨로지는 그 주제들이 가질 수 있는 저수준 정보와 객체와 이벤트 같은 고수준 정보의 관계성을 정의하여 저장하는 것을 목적으로 한다. 즉, 장면 모델은 장면내의 이벤트와 객체들의 관계 속에서 의미적으로 정의 내려진다. 특히 이때 이벤트는 단순히 하나

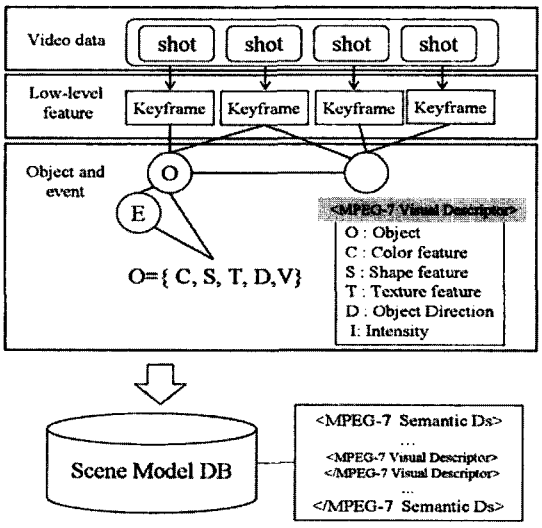


그림 4. 장면 모델 온톨로지 구성
Fig. 4. Scene model ontology construction.

의 키프레임인 정지 영상만으로 정의되는 것이 아니라, 각 키 프레임 영상에 존재하는 객체들의 간의 관계성에 의해 정의된다. 즉, 이벤트의 방향성은 샷 간의 동일 객체의 움직임으로 정의 되어 지고, 속도는 동일 객체가 샷과 샷 사이의 시간 간격을 기준으로 정의한다. 장면 모델 온톨로지는 장면이 가지는 저수준 정보와 고수준 정보를 모델링 하여 하나의 장면으로 정의하고 있는 데이터베이스이다. 그 생성 과정을 살펴보면 그림 4와 같다.

의미적 장면은 내부에 샷 정보를 가진다. 각 샷들이 가지고 있는 키프레임을 추출한다. 각 샷들의 키프레임 정보에 기반 하여 저수준 정보를 추출하고, 객체와 이벤트의 관계성도 분석한다. 객체와 이벤트의 관계성은 저수준 정보와 함께 기술 기술되어 장면 모델 온톨로지에 저장된다. 객체는 색상, 모양, 재질과 같은 저수준 정보에 의해 기술되어지며, 이벤트는 객체의 방향성과 움직임 강도에 의해 기술된다. 의미적 장면이 하나의 모델로 정의 되어, 저장 되는 내부 메타데이터 구조는 MPEG-7 Semantic Ds를 이용하여 객체와 이벤트의 관계성을 기술하고, MPEG-7 비주얼 디스크립터를 이용하여 저수준 정보를 추출하여 저장하는 구조이다. 장면 모델 온톨로지 내에 저장될 때 사용되는 언어는 이러한 관계성을 정의 할 수 있는 메타데이터 구조인 RDFS (Resource Description Framework Schema)를 이용하여 저장된다. RDF는 다양한 메타데이터 사이의 연결을 위하여 Semantic, Structure, Syntax에 대한 표준화를 제공한다. RDF는 세 가지의 특성을 표현하고 있으며, 이는 각각 객체, 속성 그리고 값이다. 그리고 RDF Schema는 자원들의 특성을 기술하며, 자원과 자원사이의 관계와 제약 조건을 기술하고 정의하는 기능을 가진다. 장면 모델 온톨로지의 기본 메타데이터 구조는 그림 5에서 바와 같다. 그림 5에 있는 장면 모델 기본구조가 파생되어 나가면서 하나의 의미적 장면 나아가서는 동영상 전체에 대해서 정의하여 저장된다.

```
<?xml version="1.0" encoding="UTF-8"?>
<video_ontology>
 <cotent title="title_name" genre="genre_id">
  <scene id="scene_id" topic="topic_name" relation="rel_id">
    <object id="object_id" name="object_name">
      <color>  value   </color>
      <texture> value   </texture>
      <shape>  value   </shape>
    </object>
    <event id="event_id">
      <direction> value </direction>
      <intensity> value </intensity>
    </event>
  </scene>
 </cotent>
</video_ontology>
```

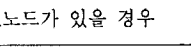
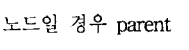
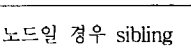
그림 5. XML을 이용한 장면모델 저장 구조
Fig. 5. Scene model framework using XML.

5. 의미적 검색

본 논문은 사용자의 질의의 결과가 정확하게 매칭 되는 부분이 없다면 의미적 질의문을 생성하는 과정을 제안한다. 즉 온톨로지 계층 구조 속에서 추론을 통해 사용자의 질의가 확장되어 의미적 질의 그룹 생성 한다. 그 과정을 살펴보면 사용자의 질의 결과가 직접적으로 매칭되면 그 결과를 보여준다. 하지만 매칭되는 결과가 없으면, 사용자의 질의문이 장면이름 온톨로지 계층구조상에 어디에 위치 하냐에 따라 각각 다른 시멘틱 쿼리를 생성한다. 표 1에서 보는 바와 같이 자식 노드가 있으면 자식노드를 검색할 수 있는 쿼리를 생성하고, 단말노드인 경우는 부모 노드와 형제 노드를 검색할 수 있는 쿼리를 생성한다.

의미적 검색은 사용자의 질의가 입력되면 장면을 정의하는 주제 키워드를 장면이름 온톨로지 내에서 트리 구조를 검색하여 결과를 보여준다. 매칭 되는 결과는 주제 키워드뿐만 아니라, 그 주제 키워드의 하위 개념도 검색에 포함된다. 그리고 객체와 이벤트의 단어가 질의로 입력되는 경우에는 장면이름 온톨로지 내에서 객체나 이벤트 정보를 이용하여 주제를 결정하여, 동일한 주제 키워드를 가지는 의미적 장면들의 결과를 보여준다. 장면이름 온톨로지는 장르에 따라 도메인이 구분되어져 있기 때문에 검색의 범위가 넓어지는 문제의 보완이 가능하다. 따라서 장르에 따라 도메인이 구분되어 있으므로 장르에 따른 검색 결과도 보여주는 것이 가능하다.

표 1. 시멘틱 질의
Table 1. Semantic query search.

	검색 노드 위치
child노드가 있을 경우	
leaf 노드일 경우 parent	
leaf 노드일 경우 sibling	

IV. 성능 평가

본 논문에서 제안하는 의미기반 시스템과 기존의

표 2. 실험 데이터
Table 2. An experimental data.

동영상 이름	동영상 길이	샷의 수	의미적 장면의 수
coffee prince 1	61 min.	135 개	22 개
coffee prince 2	60 min.	142 개	15 개
coffee prince 3	59 min.	125 개	26 개
coffee prince 4	58 min.	167 개	26 개
coffee prince 5	60 min.	116 개	30 개

Stein등이 제안한 지식 기반 검색 시스템[6]을 비교하여 분석한다. 지식 기반 검색 시스템은 저수준의 특징 정보와 고수준의 특징 정보를 확장 된 WordNet이라는 단어사전을 이용하여 기술한 연구로써 키프레임 이미지 대해서 색인 작업을 수행한다. 의미기반 시스템의 온톨로지 데이터와 지식기반 시스템이 가지는 지식 데이터를 각각 구축한다. 이 구축된 데이터의 양을 늘려가며 성능평가를 수행한다. 이러한 구축된 각각의 메타데이터는 검색을 쉽게 하기 위해 파싱 하여 구조화된 표로 데이터베이스에 저장 한다. 실험 데이터는 표 2와 같고, 의미적 장면은 기존의 제안되어진 방법을 순차적으로 실행하여 생성 뒤 실험을 수행한다.

비디오 검색 시스템의 효율성 평가는 식(1)과 식(2)에 표현된 검색 엔진 성능 평가에 이용되는 정확도(precision)와 재현율(recall)을 사용한다.

$$\text{precision}=\frac{|\{\text{relevant scenes}\}\cap\{\text{retrieved scenes}\}|}{|\{\text{retrieved scenes}\}|}\times 100 \quad (1)$$

$$\text{recall}=\frac{|\{\text{relevant scenes}\}\cap\{\text{retrieved scenes}\}|}{|\{\text{relevant scenes}\}|}\times 100 \quad (2)$$

정확도는 식(1)에서 보는 바와 같이 검색 장면의 수중에서 실제 관련 있는 장면의 비율이다. 그림 6은 이러한 장면 검색의 정확도를 보여준다. 정확도는 초기에 구축된 온톨로지의 크기가 작을 때는 기존의 연구방법과 큰 차이가 없다. 초기에 정확도가 떨어지는 이유는 매칭되는 결과가 없을 경우 시멘틱 검색을 통해 유사한 장면의 결과를 나타낼 수 있는데, 구축된 온톨로지 양이 적어 정확한 검색 결과를 포함하기 어렵다. 중간 부분의 정확도가 떨어지는 이유도 온톨로지 데이터가 쌓여 유사 검색의 결과가 더 많아지기 때문이다. 하지만 데이터가 많이 쌓일수록 제안하는 방법의 정확도가 높아진다. 제안하는

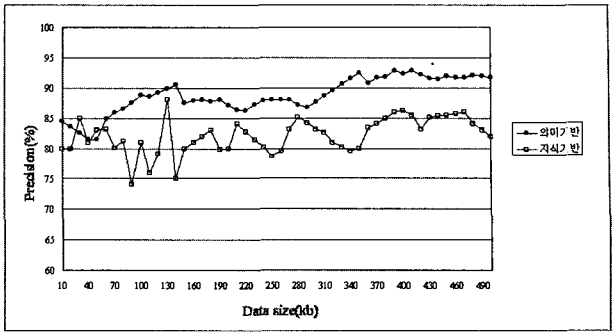


그림 6. 동영상 검색의 정확도
Fig. 6. Precision of video retrieval.

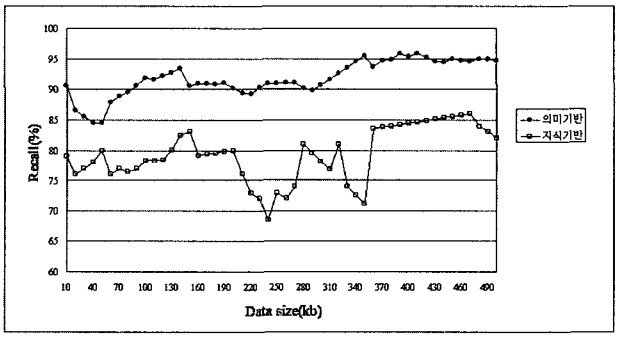


그림 7. 동영상 검색의 재현율
Fig. 7. Recall of video retrieval.

연구 방법은 키프레임뿐만 아니라, 장면에 포함된 샷들 사이의 관계성을 정의해 놓음으로써 기존의 연구방법에 비해 정확한 검색을 지원한다.

식 (2)에 표현된 재현율은 전체의 관련 있는 장면 중에서 실제 검색된 비율을 보여준다. 기존의 방법은 키프레임에 대해서만 색인하는 구조이므로 관련 있는 장면을 모두 검색해내는 데는 문제가 있다. 그러므로 그림 7에서 보는 바와 같이 제한하는 방법에 비해 좋은 성능을 보이지 못한다. 반면 제안하는 방법은 장면의 내용을 포함 하는 온톨로지로 구축함으로써 관련 있는 장면을 대부분 검색해 낸다. 그리고 시멘틱 검색를 지원함으로써 실제로 텍스트가 매칭 되는 것뿐만 아니라, 장면이름 온톨로지 구조를 통해 도메인내의 포함된 하위 개념도 함께 검색해 냄으로써 관련 있는 장면을 검색해 낸다. 하지만 기존의 연구는 이러한 의미적 검색을 지원하지 않아 낮은 재현율을 보인다.

V. 결 론

의미 기반 검색은 고수준 정보와 저수준 정보 사이의 의미적 간극을 줄이는 과정이 필요하다. 본 논문에서는 저수준 정보와 고수준 정보 사이의 관계성을 비디오 온

톨로지의 구축을 통하여 정의한다. 이러한 비디오 온톨로지 구축은 키워드 검색처럼 단순히 텍스트 매칭을 벗어나 장면을 정의하는 단어의 의미를 분석함으로써, 단어의 모양은 다르지만 의미적으로 근접해 있는 장면들을 모두 검색해낸다. 그리고 제안하는 방법은 장면내의 샷의 관계성을 기술하고, 의미적 검색을 지원하여 지식기반 시스템에 비해 더 나은 성능을 보인다. 추후에 비디오 온톨로지 데이터가 다량으로 구축되게 되면 사용자의 개입 없이 자동적인 동영상 분석도 가능하리라고 본다.

참 고 문 헌

[1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Huang Qian, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, "Query by image and video content: The QBIC system," IEEE Computer, vol. 28, no. 9, pp.23-32, September 1995.

[2] J. M. Martinez, Overview of the mpeg-7 standard. technical report 5.0, ISO/IEC, Singapore, 2001.

[3] C. Fellbaum, WordNet: An Electronic Lexical Database, Bradford Book, 1998.

[4] L. Hollink, A.T. Schreiber, J. Wielemaker, and B. Wielinga, "Semantic annotation of image collections," In Proc. of Knowledge Markup and Semantic Annotation Workshop, USA, 2003.

[5] A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer., "Video content annotation using visual analysis and a large semantic knowledgebase," In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp.379, 2003.

[6] G.C. Stein, J. Rittscher, and A. Hoogs. Enabling "Video annotation using a semantic database extended with visual knowledge." In Proc. of ICME, vol. 1, pp.161-164, 2003.

[7] L. Hollink, M. Worring, A.T. Schreiber, "Building a visual ontology for video retrieval", In Proc. of the ACM Multimedia, pp.479-482, 2005.

저 자 소 개

정 민 영(학생회원)

2002년 한남대학교 멀티미디어과 학사

2004년 한양대학교 컴퓨터공학과 석사

2004년~현재 한양대학교 컴퓨터공학과 박사 과정

<주관심분야 : 영상 처리, 비디오 검색, 지식 기반 시스템>

박 성 한(정회원)

1970년 한양대학교 전자공학과 학사

1973년 서울대학교 전자공학과 석사

1984년 미국 텍사스 주립대 전기 및 컴퓨터공학과 박사

2003년 대한전자공학회 회장

2005년~2007년 WFEO 정보통신의장

1986년~현재 한양대학교 컴퓨터공학과 교수

<주관심분야: 영상처리, 컴퓨터 네트워크 및 이동 센서네트워크 >