

MPEG-7 기반의 Dynamic Description Scheme 설계 및 구현

이용남^o 고재진 최기호

광운대학교 컴퓨터공학과, KETI, 광운대학교 컴퓨터공학과

yonalee@korea.com, jaejini@keti.re.kr, khchoi@gwu.kwangwoon.ac.kr

A Design and Implementation of Description Scheme based on MPEG-7

Yong-Nam Lee^o Jae-Jin Go Ki-Ho Choi

Dept. of Computer Engineering, Kwangwoon Univ., KETI, Dept. of Computer Engineering Kwangwoon Univ

요 약

본 논문은 MPEG-7을 기반으로 내용기반 검색을 위한 자동화 시스템을 구현하고자 한다. 하위레벨 특징(Low-level feature) 추출에서 DDL(Description Definition Language) 작성까지 자동화 시스템을 설계 및 구현하고, 프로듀서의 입장에서 고려된 고정적인 DS(Fixed Description Scheme)에 대응하는 유동적인 DS(Dynamic Description Scheme)를 이용한 사용자 중심의 개인적인 비디오 검색 시스템 구현을 목적으로 한다.

1. 서 론

디지털 오디오 비주얼(Audio-Visual) 데이터 처리 및 압축 기술의 발전과 관련 국제 표준화의 성공적 도출, 고성능 PC, 대용량 저장 장치의 보편화 및 네트워크의 발전에 따라서 디지털 멀티미디어 정보의 생성, 전송, 가공이 매우 용이해져 다양한 멀티미디어 데이터의 증가 및 활용 욕구가 강해졌다.[1] 현재 WWW에서는 주로 문서 정보(textual information)에 대한 검색만이 가능한 상태이고 멀티미디어 데이터베이스에서 영상의 색상(color), 질감(texture), 물체의 모양(shape)에 대한 정보 등의 하위레벨 특징(Low-Level) 들을 이용한 검색이 가능하다. 또한 현재 대부분의 검색 프로그램은 HTML을 이용한 문서 기반의 정보를 검색하고 있기 때문에 영상, 동영상 등을 참조할 수는 있으나 영상이나 동영상 내에 설명을 덧붙일 수는 없다.[3]

MPEG위원회는 현재 내용기반으로 하는 검색한계를 확장해서 내용기반 멀티미디어 정보검색을 효율적으로 지원하기 위한 MPEG-7을 만들었다. [2][3][4] 멀티미디어 데이터가 갖는 풍부한 정보에 비추어 보아, MPEG7에는 앞서 언급한 하위레벨 내용과 의미 정보가 포함된 내용이 있는 상위레벨 두가지 방향의 접근 및 해법이 제시될 것으로 예상된다. 따라서 현재 MPEG7의 유용성 및 표준으로서의 성공은 이 측면을 어떻게 적절히 수용하는가에 하는데 크게 좌우될 것이다.

본 논문은 2장에서 MPEG-7과 내용기반 검색에 대한 내용을 소개하고 3장에서는 제안된 시스템을 소개하였으며, 4장에서는 소개한 시스템을 구현하고, 5장에서는 결론을 기술하였다.

2. 내용기반검색과 MPEG-7

내용기반 검색의 의미는 이미지, 비디오, 오디오 등 멀티미디어 자료를 검색할 때 자료에 추가된 주석을 이용하여 검색하는 텍스트 기반 검색과는 달리, 멀티미디어 자료를 분석한 특징 값을 이용함으로써 검색하고자 하는 내용에 유사한 멀티미디어 자료를 검색하는 기술을 말한다.[1] 현재 텍스트 검색엔진 시장에는 6700여개의 서버를 가진 구글(google)이라는 성능 좋은 엔진이 등장했지만, 멀티미디어 검색을 하고자 할 때 텍스트 설명을 만드는 사람과 검색하는 사람과의 관점 차이로 검색 효율이 떨어질 수 있기 때문에 사용자가 원하는 정보의 내용에 기반한 "멀티미디어 내용기반검색"이 요구된다. 동영상내 객체의 움직임을 찾아내는 VideoQ 시스템에서는 사용자가 움직이는 스케치(animated sketch)질의를 한다. 하지만 이 시스템에서 사용자는 이미 비디오 검색 시스템 제공자에 의해 결정되어진 시퀀스(sequence) 결과만을 보게 된다.

MPEG-7은 기존의 MPEG 표준들과는 다르게 특별한 검색 요구에 만족하는 오디오 비디오 정보를 표현하는 표준으로써 "멀티미디어 내용기술 관계 (Multi-media Content Description Interface)"로 제정됐다. 따라서 MPEG-7은 데이터 그 자체가 아닌 데이터의 내용에 대한 특징 표현 방법으로 검색하는 표준으로 디스크립션(Description)은 콘텐츠 자체와 연결되어 사용자가 관심 있는 멀티미디어 자료를 빠르고 효율적으로 찾을 수 있게 한다.[3][4] MPEG-7은 다양한 형태의 멀티미디어 정보를 기술하기 위해 사용되는 표준 기술집합(Descriptor set)을 결정하는 것이다.[1] MPEG-7의 전문용어(Terminology)로 D, DS, DDL등을 소개한다.

2.1 D(Descriptor)

어떤 표현 값을 하나의 특징에 대응시키는 역할을 하며[1] 복합적일 수도 있다. 특징을 정의하기 위한 선택

스(syntax)과 시맨틱(semantics)을 정의하며, 해당 특징을 디스크립터 값(Descriptor Value)으로써 나타낼 수 있다. [2][3]

2.2 DS(Description Scheme)

D의 구조적 조합으로, 이 구조는 도큐먼트의 주석을 다는(annotate) 사용될 수 있고 또 몇몇 특징의 조합을 통해 보다 상위 레벨의 개념을 잘 표현할 수 있게 해준다.[1][3] DS는 오디오 비주얼 데이터의 일반적인 특성과 함께 응용영역 세부적인 특성을 함께 수용할 수 있도록 설계되어 있으며, 일반적인 DS는 멀티미디어 데이터의 변하지 않는 메타데이터를 기술하는데, 이들 메타데이터는 제작, 생성, 사용과 관리 등과 관련된 정보들이다. 오디오-비주얼 콘텐츠의 특정 DS는 이들 데이터의 신호적 특성이나 특징, 모델 및 시맨틱을 표현하는 방법을 제공한다. 다른 DS들로는 이들 정보를 이용하여 효과적으로 네비게이션(navigation)하고 접근할 수 있는 방법을 제공한다.

2.3 DDL(Description Definition Language)

DDL은 사용자가 생성한 DS나 D에 대한 멀티미디어 표현의 기초를 제공한다. DDL은 DS나 D를 합칠 수 있는 구문 규칙을 제공하며, 새로운 DS나 D를 만들 수 있는 언어이며 기존의 DS를 확장 변형 할 수 있게 한다.[2][3] DDL은 UML과 같은 모델링 언어가 아니라 DS, D등으로 오디오 비디오 데이터를 모델링한 결과를 표현하는 스키마(Schema) 언어이다.[3][5] DS들의 엘리먼트(element)들과 DS들간의 공간적(spatial), 시간적(temporal), 구조적(structural), 개념적(conceptual) 관계를 표현할 수 있어야 한다. [5]

3. 제안된 시스템

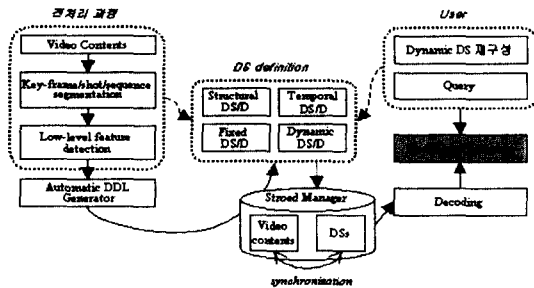


그림 1 전체 시스템 구성도

3.1 자동화된 DDL

기존의 내용기반 검색에서 하위레벨 특징의 추출은 자동화되어 있다. 그러나 표현방법이 통일된 MPEG7에서는 이러한 하위레벨 뿐 아니라 상위레벨 내용까지 기술할 수 있으므로 본 논문에서는 추출된 특징 및 특징 값

(D), 그리고 다음에 소개한 네 개의 DS를 연결하여 DDL의 자동화시스템을 구현한다.

3.2 제안된 DS

본 논문에서는 기존의 컬러(color), 텍스처(texture), 모션(motion) 등의 하위레벨 특징에 초점을 두었던 내용기반 검색을[6][8] 개선해 인간의 개념적인(perceptual) 기술에 의한 검색 시스템을 구현한다. 본 논문에서 제안한 네 개의 DS는 다음과 같다.

3.2.1 Structural DS

일반적으로 장면분할 작업을 통하여 동영상상을 구조화하는데[7] 특히, 장면전환이 이루어진 프레임에 대표 프레임으로 삼아 키프레임(key-frame)으로 지정한다.[1] 장면 전환에는 칼라에 대한 분포율을 정보로 사용하고, 계산이 빠르며, 카메라의 위치 변화, 조명, 크기 변화에 민감하지 않은 방법으로 칼라히스토그램을 이용한다.[7] 칼라히스토그램은 (1)과 같이 나타내며, 이들의 차이를 구하는 식은 (2)를 주로 사용하며, 칼라공간 별로 따로 계산하기 위해 (3)을 사용했다.

$$H_i = \sum_{j=1}^n h_j \tag{1}$$

$$d(I_i, I_{i+1}) = \sum_{j=1}^n |H_i(j) - H_{i+1}(j)| \tag{2}$$

$$d(I_i, I_{i+1}) = \sum_{j=1}^n |H_i(j) - H_{i+1}(j)| + |H_i^r(j) - H_{i+1}^r(j)| + |H_i^g(j) - H_{i+1}^g(j)| \tag{3}$$

H_i 는 화상 안에서 칼라 i 의 전체 칼라 히스토그램을 나타내며, n 은 칼라 버킷의 수이다. 장면전환이 검출되는 임계값은 반복실험을 통해 48%이상의 오차를 장면전환으로 검출하며, 키프레임을 선택한다. 또한 비디오 편집자 또는 프로듀서에 의해, 시간과 줄거리 중심의 시퀀스(sequence)로 나누고, 그 안에서 다시 이벤트 중심으로 샷(shot)을 구별해 낸다. 각각은 고유한 인덱스로 정렬한다.

3.2.2 Temporal DS

일반적으로 멀티미디어 데이터는 그 자체가 시간의 순서로써 표현되어진다. 따라서 동기화는 사건의 시간적 순서를 확실히 하여 사건과 스토리를 전개시킨다. 그림 2에서처럼 시간의 흐름에 의한 비디오 데이터의 특성을 고려하여 Structural DS에서 결정된 키프레임의 현재시간(current time)과 프로듀서 및 사용자에게 의해서 선택된 샷과 시퀀스의 기간(duration)으로 동기화 시켜 검색시 디스플레이한다.

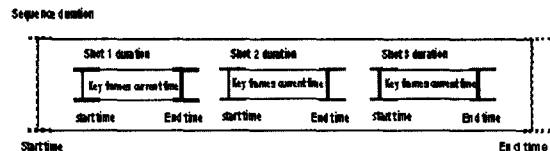


그림 2 Temporal DS의 구조

