

MPEG-7의 동향

김동욱*, 김진태**

*전주대학교 전자매체공학부

**한서대학교 컴퓨터학과

목 차

- I. MPEG-7의 시작과 목표
- II. MPEG-7의 구성과 요구사항
- III. 비디오 정보 검색
- IV. MPEG-7의 응용분야
- V. MPEG-7의 개발 사례
- VI. MPEG-7 표준화 추진 일정

I. MPEG-7의 시작과 목표

우리 생활에 가까이 다가오고 있는 정보화 사회에서는 사용자가 원하는 다양한 형태의 멀티미디어 정보를 신속하고 효과적으로 찾아내기 위한 노력이 절실히 필요하다. 이를 해결하기 위하여 국제 동영상 부호화 표준화 기구인 ISO/IEC JTC1/SC29/WG11에서는 "Multimedia Content Description Interface"를 위하여 MPEG-7을 1997년에 시작하였다. 즉, MPEG-7에서는 내용 기반 멀티미디어 정보 검색을 효율적으로 지원하기 위한 기술을 개발하고 이를 국제 표준화하고자 하는 것이다. 기존에 표준화되었거나 지금 표준화가 진행되고 있는 MPEG-1, -2 및 -4 등은 오디오/비주얼 데이터의 압축을 그 목표로 하였으나, MPEG-7은 데이터 그 자체가 아닌 데이터의 내용에 대한 표현 방법을 다루는 것이다. 이를 다른 말로 메타데이터(metadata) 또는 bit about bits라고 표현하기도 한다.

MPEG-7에서는 주로 오디오/비주얼 정보(정지화상, 그래픽, 3D 모델, 오디오, 음성, 비디오 등)의 표현을 그 대상으로 하고 있으며, HTML, SGML, 또는 RDF 등이 목표로 하고 있는 텍스트 문서의 표현에 대한 것은 표준화 범위에 포함하지 않는다. 필요한 경우 이러한 문서 포맷에 관한 표준화가 적용될 수 있도록 하는 기능은 갖게 될 것이다. 그림 1은 MPEG-7과 관련된 정보 처리 과정과 MPEG-7이 표준화 하고자 하는 범위를 개략적으로 보여준다.

그림 1에서 특징 추출(feature extraction)방식 및 검색 엔진(search engine)은 표준화의 대상으로 하지 않는다. 왜냐 하면, 이들은 여러 멀티미디어 정보처리 시스템간의 호환성을 획득하는 것과는 상관이 없으며, 업체간의 경쟁이 가능한 부분으로 남겨 놓은 것이 향후 계속적인 기술 발전을 유도하기 위하여 바람직하기 때문이다.

MPEG-7은 멀티미디어 정보의 특징 및 내용 표현 방식에 초점을 맞추어 다음의 사항을 표준화하고자 한다

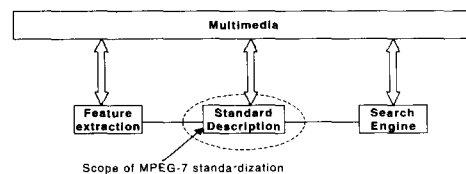


그림 1 MPEG-7 표준화의 범위

- ①기술자(descriptors; D)와 기술 구조(description scheme; DS)
- ②기술 구조를 표현하기 위한 기술정의 언어(description definition language;DDL)
- ③색인, 저장 및 전송을 효율적으로 하기 위하여 사용될 코딩된 기술(coded description)

MPEG-7에서 정의하는 용어는 다음과 같다. 데이터(data)는 표현 형식에 상관없이 오디오/비주얼 정보를 의미하며, 특징(feature)은 오디오/비주얼 정보의 성격이나 속성을 나타낸다.

기술자(descriptor)는 특징을 어떤 값에 연결시켜주는 도구이며, 기술 구조(description scheme)는 데이터를 여러 개의 기술자로 나타내기 위한 그릇이다.

II. MPEG-7의 구성과 요구사항

MPEG-7은 문자 이외의 비정형화된 형태의 데이터 내용을 얼마만큼 충실히 분석하고 잘 표현하는가에 대한 부분이 주요 관건이자 핵심부를 이루며, 이는 다시 그래픽 사용자 인터페이스(GUI)를 통해 사용자가 찾고자 하는 데이터를 데이터 내용을 기반으로 하여 유사도가 높은 순서에 따라 빠르고 효과적으로 제시해주는 부분으로 구성된다.

MPEG-7에 관한 요구사항은 기술정의 언어에 대한 요구 사항, 오디오/비주얼 데이터에 공통적으로 적용되는 요구사항, 그리고 오디오 또는 비주얼 데이터별 특성에 연관된 요구사항으로 나누어진다. 우선 기술정의 언어에 대한 요구 사항으로는, ①MPEG-7 응용분야를 지원하는 어떤 기술 구조도 쉽게 생성되어야 하며, ②문법이 모호하지 않고 인터프리터 등에 의해 쉽게 파싱(parsing)될 수 있어야 하며, ③오디오/비주얼 데이터를 포함한 복합적인 멀티미디어 정보에 대해 기술구조를 생성할 수 있어야 하며, ④기술구조 내에 그 코드를 포함시킬 수 있는 기능을 가져야 한다. 이 코드는 표준에서 정의되지 않은 새로운 기술자와 이 기술자에 대한 유사도 측정 기준 등을 사용자가 인식할 수 있도록 하는데 사용된다.

오디오/비주얼 데이터에 공통적으로 적용되는 기술자 및 기술구조에 대한 일반적인 주요 요구 사항으로는, ①데이터에 부가된 주석, 데이터의 시공간적인 구조, 주/객관적 특징, 또는 사건이나 활동 등의 개념까지도 포함하는 모든 특징의 표현을 지원하여야 하고, ②사용자의 요구에 적절히 대응할 수 있는 계층적인 내용압축 및 표현방식을 지원하여야 하며, ③비주얼 질의(query)를 사용해서 오디오 정보를 찾고 오디오 질의를 사용하여 비디오 정보를 찾을 수 있도록 cross-modality를 지원하여야 한다. 또한, ④MPEG-7의 여러가지 기술구조 상호간의 호환성, ⑤특징별 우선순위 지정, ⑥기술자의 스케일러빌리티 지원, ⑦텍스트 기반 기술자의 경우 사용언어의 지정, 모든 언어를 지원할 수 있는 글자 세트, 그리고 언어 사이의 변환을 위한 도구 등이 제공되어야 한다는 것도 주목할 필요가 있다. 기능적인 관점에서의 요구 사항으로는, ①내용기반 검색 및 유사

도 기반 검색 기능의 제공, ②데이터와 동기화되어 제공되거나(streamed) 또는 데이터와 별도로 제공되는(stored) 기술자의 지원, ③대화형 질의의 지원, ④소스 데이터 또는 관련 정보의 위치를 알려줄 수 있는 기능, 그리고 ⑤정보의 개략적인 파악(preview)이 가능하도록 브라우징 기능의 지원 등이 있다. 마지막으로 비주얼 또는 오디오 정보 각각의 특성에 따른 요구 사항으로는 지원되는 특징의 종류, 지원되는 데이터 포맷 및 클래스에 대한 요구 사항이 포함된다.

III. 비디오 정보 검색

1 비디오 인덱싱

(1) 비디오 데이터의 모델링

비디오 검색을 위한 효과적인 데이터베이스의 조직 및 운영을 위해서 고려해야 할 요소들은 다음과 같다. 첫째, 다단계의 비디오 구조를 갖추어야 한다. 이러한 구조는 비디오 정보를 참조하기 용이하고 또한 그 내용을 이해하기 쉬운 형태로 제시한다. 그리고 더 나은 비디오 브라우징을 지원할 수 있으며, 마지막으로 비디오 인덱싱 및 저장 형태를 단순화시킬 수 있기 때문에 이러한 비디오 데이터의 모델링 기법은 매우 중요하다. 둘째, 시간적·공간적 관계가 지원되어야 한다. 비디오 데이터의 주요한 특징 중의 하나가 바로 의미론적으로 볼 때, 시간적·공간적으로 서로 연관되어 있다는 점이다. 셋째, 비디오 데이터 자체 주석을 달아 이를 활용할 수 있어야 한다. 넷째, 비디오 데이터의 비의존성이다. 즉, 동일한 비디오 데이터는 여러가지 서로 다른 경우의 비디오 검색 시에도 참조 대상으로 사용될 수 있어야 한다.

비디오는 프레임(frame), 샷(shot), 장면(scene)의 세 가지 구성요소로 이루어진다. 프레임이란 비디오를 구성하는 최소 단위로 필름 한 장에 해당하는 하나의 개별 영상을 나타낸다. 샷은 하나의 카메라로 촬영한 영상을 나타내며, 샷 내에서는 필름이 끊기지 않고 연속적으로 연결되어 있다. 장면은 연속된 일련의 샷으로 구성되며, 주인공이나 특정한 장소와 같이 하나의 대상을 연속하여 촬영한 영상을 나타낸다. 비디오 분할기술은 장면전환 효과에 의해 발생하는 샷 사이의 경계를 검출하는 방법으로, 샷은 비디오 분할의 기본단위로 사용된다.

(2) 비디오 데이터 인덱싱

비디오 데이터는 대용량으로 인하여 비디오 데이터를 액세스하고 검색하는데 많은 시간이 소비된다. 따라서, 이를 용이하게 하기 위해 비디오 인덱싱 기법이 필요하게 된다. 전통적인 텍스트 기반의 데이터베이스 시스템에 비해 비디오 인덱싱은 훨씬 어렵고 복잡하다. 텍스트 기반은 단지 몇 개의 키워드를 이용하여 데이터베이스를 구성할 수 있는 반면, 비디오 인덱싱의 경우에는 무엇을 특징정보로 사용할 것인지에 대한 구분부터 불명확하다.

현재 연구되고 있는 비디오 인덱싱 기법으로는 특징 정보를 어떻게 유도하는가에 따라 주석기반 인덱싱(annotation-based indexing), 특징기반 인덱싱(feature-based indexing), 그리고 영역국한 인덱싱(domain-specific indexing) 등으로 분류할 수 있다.

첫째, 주석 기반 인덱싱인 경우, 비용 효율 면에서 떨어지고 시간 소비가 많다. 또한 응용 분야에 의존적이기 때문에 서로 다른 분야에 있어서는 동일하게 적용하기가 곤란하다. 주석 작업에 있어서도 사용자의 주관적 관점이 개입될 소지가 많기 때문에 데이터의 내용을 충실히 표현하는데 제한이 따른다는 단점을 갖고 있다. 따라서 이러한 방법은 적은 양의 비디오 데이터를 데이터베이스로 구축하는데 적합하다.

둘째, 특징 기반 인덱싱인 경우 특징 정보에 대한 인덱싱 처리를 완전히 자동적으로 수행하는 것을 목표로 하는데, 이는 비디오 분할 기법, 텍스트 기반에서의 키워드에 해당하는 대표 프레임의 설정 및 특징 정보 추출 기법 등과 같은 영상 처리 알고리즘에 의해 수행 가능하다. 특징 정보로는 색상, 텍스처, 객체의 움직임 정보, 위치 정보 등을 들 수 있다.

셋째, 영역 국한 인덱싱인 경우 응용 대상이 특정 범위로 제한된 형태의 비디오 데이터에 대해 논리적인 구조를 사용하여 상부에서 점차 하부 구조적인 형태로 나누고 이를 분석하는 방법이다. 예를 들어 뉴스 프로그램에서 진행자가 주요 뉴스를 소개한 후 이를 다시 기자가 상세 보도하는 형식으로 구성되는데, 각각의 주요 뉴스들을 상부 구조로 설정하고 그에 해당하는 기사 내용을 하부 구조로 나누어 이를 의미론적으로 결합시켜 해석하는 방식을 취한다. 이러한 방식은 적용 대상이 극히 한정돼 있고 무엇보다 분할된 비디오 데이터로부터 의미를 부여하는 자체가 난해하고 제한되어 있다는 단점을 갖는다.

2 비디오 검색

(1) 비디오 데이터 질의 및 검색

비디오 데이터베이스의 목적은 비디오 데이터의 내용물을 사용자에게 효과적이고 쉽게 제공할 수 있도록 하는데 있다. 비디오 데이터 검색 처리는 전형적으로 다음과 같은 처리 단계를 고려할 수 있다. 첫째, 사용자는 그래픽 사용자 인터페이스를 통해 질의를 행한다. 둘째, 이 질의에 대한 처리 및 평가를 통해 얻어진 특징이나 값을 이용하여 데이터베이스에 저장된 비디오 데이터와 정합시키고 검색한다. 마지막으로 비디오 데이터 질의 및 검색 결과를 적당한 형태의 그래픽 사용자 인터페이스 상에 디스플레이시킨다. 여기서 고려되어야 할 점은 질의 내용으로부터 얻어질 수 있는 특징 정보의 추출이다. 특징 정보는 비디오 데이터베이스 구성시 적용되는 특징 추출 요소들이 그대로 적용될 것이다. 또한 공간적 정보인지 시간적 정보인지에 대한 구분도 필요하다. 질의에 대한 응답의 경우 매칭 정확도에 따라 응답의 내용이 달라질 것이다.

표 1 내용 기반 검색의 응용 사례

개발 기관	시스템 명칭	웹사이트 도메인 주소
IBM	QBIC	http://www.qbic.almaden.ibm.com
Virage	Virage	http://www.virage.com
Informix Software	Visual Intelligence	http://illustra.com
Interpix Software	Image Surfer	http://isurf.interpix.com
Columbia 대학	ADVENT	http://www.ctr.columbia.edu/advent
Chicago 대학	WebSeer	http://infolab.cs.uchicago.edu/webseer

현재 정지 영상을 대상으로 하는 내용 기반 영상 검색 시스템이 인터넷을 통하여 시범적으로 서비스가 제공되고 있으며, 표 1은 대표적인 내용 기반 영상 검색 시스템들이다. 이들 각각의 시스템들은 사용자가 원하는 형태의 질의 영상에 대한 것이 아니라 질의 자체를 데이터베이스 내의 영상으로 국한시키고 이로부터 각기 시스템 내에 저장된 데이터베이스 영상들을 서로간의 특징 정보를 비교하여 유사도가 높은 순서대로 결과를 제공하도록 되어 있다. 시간적 특성을 고려한 비디오 데이터의 경우에는 서비스 내용이 배제되어 있는 상황이다.

(2) 정보 검색 기술

멀티미디어 정보의 검색에 있어 사용되는 방법으로 다음과 같은 텍스트 기반 검색과 내용 기반 검색의 두 가지 방식으로 크게 분류할 수 있다. 첫 번째 방식으로 검색의 대상이 되

는 모든 멀티미디어 데이터에 사람이 직접 색인을 첨가하고, 사용자 또한 주제를 이용하여 원하는 정보를 검색하는 텍스트 기반 방식이다. 이 방식은 비정형적인 멀티미디어 데이터에 사람이 직접 의미 정보를 기술하기 때문에 제한된 범위내에서는 효율적인 검색이 가능하다는 장점을 갖고 있으나 대용량의 데이터에 사람이 일일이 색인을 첨가하므로 시간과 비용이 많이 소요되며, 색인을 첨가하는 사람과 검색하는 사용자의 관점이 불일치할 경우 검색 효율이 크게 떨어진다. 또한 멀티미디어 데이터가 가지는 복잡한 속성을 텍스트만으로는 충분히 표현할 수 없다는 단점을 갖고 있다.

두 번째 방식은 멀티미디어 데이터의 내용을 대표할 수 있는 특징을 추출하여 이를 기반으로 색인과 검색을 수행하는 내용 기반 검색 방식이다. 이 방식은 멀티미디어 데이터로부터 특징 정보를 자동으로 추출하여 색인 과정에 사용하므로 시간 및 인력의 소모를 줄일 수 있는 장점을 갖지만 멀티미디어 데이터로부터 정확한 특징 내용을 추출하기가 어렵다는 단점을 갖고 있다. 하지만 최근에는 컴퓨터 비전이나 영상 처리, 컴퓨터 그래픽스 등과 같은 여러 분야의 연구 결과를 복합적으로 이용하여 멀티미디어 데이터로부터 보다 정확한 특징 정보를 추출하려는 연구가 진행 중이다.

위의 두 가지 방식 중 현재 주로 연구되고 있는 내용 기반 검색 기술은 대상에 따라 정지 영상 검색과 비디오 검색의 두 가지로 나누어진다. 정지 영상 검색에 사용되는 특징 정보로는 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 관계 등이 있다. 한편 내용 기반 비디오 검색 기술은 대상이 갖고 있는 속성이 일반 텍스트나 정지 영상이 지닌 특성 외에 시간적인 성분과 비정형적인 구조를 내포하고 있으므로 훨씬 더 복잡한 처리 기술을 요구하는데 이를 다시 분류해 보면 다음과 같다. 즉, 비디오 데이터가 압축된 데이터나 아니냐에 의존하여 처리 기법이 달라지며, 대부분의 경우 비디오 데이터는 압축된 상태로 저장 및 전송되며 이를 대상으로 처리하는 것이 일반적이다. 각각의 경우에 대해 살펴보면 다음과 같다. 먼저 비압축 상태에서의 비디오 분할은 사용하는 특징의 크기에 따라 화소 단위의 분할 방법, 부분 영역 단위의 분할 방법, 프레임 단위의 분할 방법으로 나누어진다.

화소 단위의 분할 방법은 연속된 두 프레임에서 상응하는 두 화소 사이의 특징차를 구하여 그 차이가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 검출하는 방법으로 주로 화소의 밝기 값이나 예지 등을 특징으로

사용한다. 화소 단위의 분할 방법은 처리 시간이 빠른 반면 잡음이나 영상내 물체의 이동에 민감하다는 단점을 가지고 있다.

부분 영역 단위의 분할 방법은 한 장의 프레임을 다수개의 부분 영역으로 나누고 연속된 두 프레임에서 상응하는 부분의 영역간의 특징차가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 검출하는 방법이다. 부분 영역 단위의 분할 방법에서는 부분 영역별 평균 밝기 값이나 밝기 히스토그램, 색상 히스토그램, 움직임 벡터 등을 특징값으로 처리하며 히스토그램 기법이 가장 많이 사용된다. 가장 일반적으로 사용되는 프레임 단위의 분할 방법에서는 한 장의 프레임에 대하여 대표적인 특징값을 구하고, 각 프레임간 특징값의 차이가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 검출한다. 프레임 단위의 분할 방법에서는 밝기 히스토그램, 색상 히스토그램, 차영상의 히스토그램 등을 특징값으로 사용한다.

한편, 일반적으로 비디오 데이터는 용량이 매우 크기 때문에 데이터베이스의 효율적인 관리를 위해 압축된 형태로 저장, 관리된다. 압축된 비디오 데이터의 경우 별도의 복호화 처리 과정 없이 직접적으로 장면의 경계 프레임을 검출하는 비디오 분할 기법들이 여러가지 제시되고 있다. DCT 계수를 이용하는 방법, 움직임 벡터를 이용하는 방법, 이 두 가지를 혼용하는 방법 등이 있다.

DCT 계수를 사용하는 방법은 인접한 DCT 계수의 차가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 추출하는 방법으로, MPEG 형식으로 압축된 비디오에 적용할 경우 DCT 계수 추출이 용이하고 처리 시간이 빠르다는 장점을 갖는 반면, 페이드(fade), 디졸브(dissolve), 패닝(panning) 등의 카메라 움직임에 따른 점진적 장면 전환은 검출하지 못한다는 단점이 있다. 움직임 벡터를 이용하는 방법은 카메라의 이동에 의해 발생하는 점진적 장면의 변화도 분석할 수 있어 DCT 계수 적용법의 단점을 보완할 수 있는 반면, 잡음이나 물체의 움직임에 민감하고 또 처리 속도가 일반 경계 검출 방식에 비해 느리다는 단점을 갖고 있다. 이러한 관계를 이용하여 두가지를 혼용한 시스템이 적용되기도 한다.

IV. MPEG-7의 응용분야

MPEG-7은 매우 다양한 분야에 응용될 수 있다. 몇 가지 응용 예를 보면 다음과 같다.

- 디지털 도서관 (영상 목록, 음악 사전 등)
- 멀티미디어 사전 (직업별 전화번호부 (yellow pages) 등)
- 방송 매체 선택 (라디오 채널, TV 채널 등)
- 멀티미디어 편집 (개인 전자 우편 서비스, 미디어 제작(media authoring) 등)

그 외에도 다음과 같은 분야에 응용할 수 있다.

- 교육에서의 멀티미디어 자료 이용
- 언론분야 : 특정한 정치가의 이름, 말, 또는 얼굴로 그의 연설을 검색
- 여행 정보 안내, 지리 정보 서비스
- 역사 박물관, 미술품 전시 등의 문화적인 서비스
- 오락 : 게임 검색, 노래방의 노래 검색
- 조사 업무 : 얼굴 특징 인식, 법정 조사 (forensics)
- 의료 응용 분야
- 지도 제작, 생태 조사, 자연 자원 관리 등의 원격 탐사
- 쇼핑 : 자기가 원하는 물건 검색, 원격 구매
- 건축, 부동산, 인테리어 디자인
- 사회 생활 : 중매 서비스
- 영화, 비디오, 라디오 정보의 저장 및 검색

MPEG-7에서 사용자의 질의(query)에 대한 응답 방법은 표준화 범위 밖에 있다. 따라서 표준화 기술과 연계가 가능한 어떤 유형의 시청각 정보도 어떤 유형의 질의 자료에 의해서 검색될 수 있다. 예를 들어 비디오, 음악, 연설 등의 질의를 통해서 우리가 원하는 비디오 자료를 검색할 수 있다. 위와 같은 다양한 응용 분야들을 한 종류의 기술자나 기술구조가 만족시키는 것은 어려운 일이며, 효율적이지도 않다. 따라서, MPEG-7에서는 기술자나 기술구조의 제안 사항을 주요 응용 분야별로 나누어 표준화 할 방침을 가지고 있다. 응용 분야와 관련하여 논의가 진행되고 있는 것 중의 하나는 'pull model'과 'push model'의 구분이다. 전자국지적으로나 네트워크 상에 있는 데이터베이스로부터 사용자의 질의에 가장 가까운 데이터를 찾아주는 응용분야 이고, 후자는 방송에서와 같이 정보 제공자로부터 일방적으로 데이터가 공급될 때, 이로부터 원하는 정보를 필터링 하여 주는 응용 분야이다. 두 응용 분야의 주요 특징상 차이점은 후자의 경우 정보 검색시 실시간이 강하게 요구된다는 것이다.

V. MPEG-7의 개발 사례

- 문서, 영상 검색과의 연결 (linking text and image retrieval)
현재 대부분의 검색 프로그램은 HTML을 이용한 문서 기반의 정보를 검색한다. 따라서 영상, 동영상 등을 참조할 수는 있으나 영상이나 동영상 내에 설명을 덧붙일 수는 없다. 이러한 문제점을 해결하기 위해 여러 가지 도구들이 개발되고 있다. XML(eXtensible Markup Language)은 사용자에게 자신이 작성한 설명들을 정의할 수 있도록 하고 있으나 그 설명들이 각각의 영상 객체에 직접 연관되지는 않는다. SGML은 현재 존재하는 인터넷의 검색 도구를 이용하여 적당한 영상을 검색할 수 있도록 영상 내에 설명을 추가하는 방식을 포함하고 있다. HyTime은 동영상과 오디오를 포함하는 관련된 문서와 영상을 위치에 대한 설명과 함께 위치시키는 기술을 갖고 있다.

- 정보 공급자 관점에서의 멀티미디어 색인 (a content provider's view of multimedia indexing)

프랑스 파리의 INA(Institut National de l'Audiovisuel)는 모든 프랑스 방송에 대한 공식적인 자료를 관리하고 있다. INA는 지난 10년 동안 8만여개의 TV 방송과 3만 5천여개의 라디오 방송을 위하여 컴퓨터 데이터베이스를 이용하여 프로그램을 검색하고 있다. 그들은 현재 1년에 약 3만 5천여개의 프로그램을 검색하고 있다. 검색은 현재 색인 키워드(thesaurus keywords)와 이차적 기술(secondary descriptions)의 조합을 이용하여 수동으로 이루어지고 있다. 그러나 INA의 검색에서는 동사(verb)를 지원하지 않아 행동이나 사건 사이의 관계에 대하여 정확하게 분류하기가 어렵다.

INA는 이 거대한 자료들을 멀티미디어 지식창고(multimedia knowledge base)로 만들려고 한다. 이를 위해서 그들은 위에서 나타난 문제점들을 해결하고 자료들의 기술에 대한 표준화된 어휘(vocabulary)들을 개발하고 있다. 이 어휘들은 저장된 행동의 요약(summary of actions) 뿐만 아니라 촬영 유형(shot type), 기간(duration), 카메라 움직임(camera motion), 촬영 장소(location of shot), 배우, 시선(gaze direction), 상의 위치(screen position), 그리고 기타 관심 있는 사항들에 대한 것들을 포함하고 있다. 따라서 이 어휘들이 완성되면 영상과 음성을 검색할 수 있는 강력한 도구가 될 것이다.

● 비디오 메일 검색 프로젝트
(the video mail retrieval project)
URL의 Martin Brown은 VMR(video mail retrieval) 프로젝트를 맡고있다. VMR 프로젝트팀은 비디오와 음성이 합쳐진 메일(mail)을 분석하고, 메일 내용을 찾을 수 있는 방법에 대해서 조사하였다. 현재 알려진 사용자의 제한된 어휘를 이용한 음성 인식은 가능하지만 훈련되지 않은(untrained) 사용자의 음성 인식률은 50% 이하에 그치고 있다.

이러한 문제점들을 해결하기 위하여 VMR 팀은 말을 하는 데 사용되는 음소(phonemes)를 모아서 각각의 단음절(phones)을 분석하였다. 그리고 이를 조합함으로써 단어에 대한 인식률이 높아지게 되었다. 이 방법에 학습(learning) 알고리즘을 덧붙여서 비디오 메일을 검색할 수 있는 적당한 방법이 기대되고 있다.

● 콜럼비아 대학의 S. F. Chang은 WWW상에 저장된 영상을 검색하는 도구를 개발하였다. 이 도구는 오프라인(off-line)으로 영상을 검색하거나 온라인(on-line)으로 브라우징(browsing) 하는 것이 모두 가능하다. 여기서 데이터 검색을 용이하게 할 수 있도록 데이터 베이스를 설계하기 위해서 어떤 형태로 질문을 받아가 하는 것은 중요한 문제이다. 콜럼비아 대학의 WebSeek이란 도구는 각각의 영상을 고정된 항목들에 따라 계층적으로 분류하였다. 현재 WebSeek에서는 백만 개 이상의 GIF와 JPEG 영상 그리고 만 개 이상의 비디오 클립에 대해서 분석하는 데 사용되고 있다.

콜럼비아 대학에서는 WWW 상에서 MPEG 동영상을 브라우징하고 편집할 수 있는 WebClip 도구와 객체 기반 비디오 질의 시스템인 VideoQ를 개발하였다. 이러한 도구들은 입력되는 압축된 데이터 스트림을 분석함으로써 장면 변환을 90%이상 검출할 수 있는 장면 변환 검출기(scene change detector) 들을 가지고 있다.

이 시스템은 프레임들이 트리 구조(tree structure)를 통해 계층적으로 분석된 프레임들 사이의 관계를 검색 하는 데 많은 이점이 있다. 앞으로 해결해야 할 과제는 여러 검색 알고리즘에 대한 비교 분석 도구가 아직 없다는 것이다.

이 밖에도 대표적인 비디오 검색 시스템으로는 다음과 같은 것이 있다.

- 미국 IBM사의 QBC
- 싱가포르 ISS의 SWIM

● CMU(Canegie Mellon Univ.)의 Infomedia

VI. MPEG-7 표준화 추진 일정

현재 MPEG-7 표준화는 그 요구 사항을 정의하는 단계에 있으며, 표준 시험용 데이터(standard test material)를 수집하고 제안서의 평가 기준 및 방법도 아울러 만들어가고 있다. MPEG-7에서 취하고 있는 표준화 과정을 살펴보면, MPEG-1, -2와 같이 우선 제안요구서(call for proposal; CFP)가 공고되고 나면 일정한 기간동안 기술 제안서를 받아서 그 중 요구사항이 잘 반영되고 객관적으로 성능이 우수한 제안서들을 바탕으로 통합 시험 모델(experimentation model; XM)을 만든다. XM을 바탕으로 각 세부 항목별 공동 실험(core experiment)을 통합 성능 및 기능 보완, 표준안으로서의 통합 기능 검증을 거쳐 WD(working draft), CD(committee draft)등을 만들어 간다. 이 단계를 협력 단계(collaboration phase)라고 한다. CD에서 실질적인 기술적 사항들이 모두 확정되고, 그 이후에는 편집 상의 보완 및 국가별 투표를 거쳐(DIS) 국제 표준(IS)으로 확정된다. MPEG-7의 표준화 작업 일정은 표 2와 같다.

표 2 MPEG-7의 workplan

work	time
Call for Proposals (CFP)	1998년 10월
Experimental Model (XM)	1999년 3월
Working Draft (WD)	1999년 12월
Committee Draft (CD)	2000년 10월
Final CD	2001년 2월
Draft IS	2001년 7월
International Standard (IS)	2001년 11월

참고문헌

- [1] ISO/IEC JTC1/SC29/WG11, "MPEG-7: Context and Objective," MPEG98/N2207 Mar. 1998.
- [2] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Evaluation Ad Hoc Group: Paris Meeting Report," MPEG98/M3785, Jul. 1998.
- [3] MPEG Requirement Group, "MPEG-7: Proposal Package Description (PPD) - Draft v1.0," ISO/IEC JTC1/SC29/WG11

- N1923, Oct. 1997.
- [4] 김진태, 김동욱, "내용 기반 비디오 인덱싱 및 검색 기술", 한국멀티미디어학회지, 제2권 제1호, 1998. 6.
 - [5] 김우생, 김진웅, 임문철, "MPEG-7 표준화 및 내용기반 정보 검색", 전자공학회지, 제25권 제8호, pp. 26-39, 1998. 8.
 - [6] 이상미, "MPEG-7 발전 방향", 한국전자통신연구원 주간기술동향, 제838호, pp. 1-11, 1998. 3.
 - [7] 김해광, "MPEG-7 표준 : 멀티미디어 데이터 기술 인터페이스를 위한 부호화 기술", 한국통신학회지, 제15권 제12호, pp. 74-84, 1998. 12.

저자소개



김 동 욱

- 1987년 2월 성균관대학교 전자공학과 공학사
- 1992년 2월 중앙대학교 대학원 전자공학과 공학석사
- 1996년 8월 중앙대학교 대학원 전자공학과 공학박사
- 1997년 3월~1998년 2월 청운대학교 전임강사
- 1998년 3월~현재 전주대학교 전자매체공학부 조교수



김 진 태

- 1987년 중앙대학교 전자공학과 졸업(공학사)
- 1989년 중앙대학교 대학원 전자공학과 졸업(공학석사)
- 1993년 중앙대학교 대학원 전자공학과 졸업(공학박사)
- 1995년~현재 한서대학교 컴퓨터과학과 조교수
- 관심분야: 영상압축, 비디오 인덱싱 및 검색, 워터마킹