특 집

MPEG-7 표준화 및 내용기반 정보 검색

김 우 생, 김 진 웅*, 임 문 철
광운대학교 전자계산학과, ETRI*

I. 서 론

현재의 세계는 급격한 멀티미디어 정보의 증가와 활용을 경험하고 있다. 이는 다음과 같은 여러 가지 기술 발전에 기인한다. 첫째로 디지털 오디오 비쥬얼(Audio-visual) 데이터 처리 및 압축 기술의 발전과 이와 관련된 국제 표준화의 성공적인 도출이다. MPEG-2 표준은 통신, 방송, 저장 매체에 폭 넓게 적용할 것을 목표로 제정된 디지털 오디오 비디오 데이터의 압축 및 전송에 관한 국제 표준으로서 고선명TV를 포함하는 디지털 TV 방송, 차세대 가전기기의 핵심 장치가 될 DVD (Digital Versatile Disk)의 표준으로 채택되었다. 디지털 방송환경에서는 기존의 아날로그 방송 환경에서보다 훨씬 작은 전송 대역폭에 더 많은 양질의 프로그램을 공급할 수 있게 되었으며, 이미 미국의 DirecTV등에서 150채널 이상의 방송 채널을 갖고 서비스를 제공하고 있다. 둘째로 고성능 개인용 컴퓨터, 대용량 저장 장치의 보편화 및 World Wide Web(WWW)으로 대변되는 컴퓨터 네트워크의 발전에 따라 디지털로 표현된 멀티미디어 정보의 생성, 전송, 가공이 매우 용이해졌다. WWW은 e-mail, newsgroup등을 통한 정보 교환, HomePage를 통한 기업의 홍보, 상품의 광고 및 거래, Web 방송 등 이미 그 활용도는 일상의 주요 활동에 없어서는 안될 수단으로 자리잡고 있다. Yahoo, Alta Vista 등의 Web 검색 엔진이 가장 사람들이 많이 이용하고 있는 Site가 되고 있는 점은 효율적인 정보 검색을 위한 기술과 도구의 필요성이 절실함을 말해주고 있다. 그러나 엄청난 속도로 증가하는 멀티미디어 정보 중에서 사용자가 필요로 하는 내용의 정보를 찾기 위해서는 기존의 키워드 기반의 검색은 한계에 도달한 상황이기 때문에 사용자가 원하는 정보를 내용에 기반하여 검색할 수 있는 방법이 요구되고 있다.

이러한 최근의 기술 발전 추세 및 시장 요구를 바탕으로 하여, 국제 표준화 기구인 ISO와 IEC의 연합기술위원회 산하의 MPEG(공식명칭 : ISO/IEC JTC1 SC29/WG11)에서는 MPEG-7 : Multimedia Content Description Interface 라는 이름으로 멀티미디어 데이터의 내용기반 검색을 위한 내용 표현 방식에 관한 국제 표준화 작업을 시작하였다. 본 고에서는 MPEG-7 표준화의 현재 진행 상황을 살펴보고, 이와 관련된 멀티미디어 내용기반 정보 검색 기법 및 연구 현황에 대하여 설명한다. 2장에서는 MPEG-7 표준화의 목적 및 범위, 응용 분야, 요구사항 및 향후 일정 등을 기술하며, 3장에서는 멀티미디어 정보의 내용 표현 및 검색 연구의 현황과 관련 시스템에 대한 소개를 하고, 마지막으로 4장에서 결론을 맺는다.

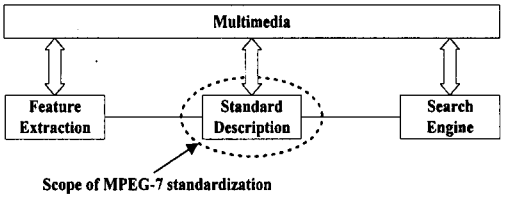
II. MPEG-7 표준화 동향

1. 목적 및 범위[1]

'한 장의 그림은 천 마디의 말로도 표현이 어렵다.(A picture is worth a thousand words.)'든지, '보기(듣기) 전에는 믿기 어렵다.(Seeing (Hearing) is believing.)' 등의 말은 멀티미디어 정보의 속성을 단적으로 드러내고 있다. 일반적으

로 비디오 Archive로부터 'Terminator II 영화에서의 오토바이가 질주하는 장면'을 찾으려고 하거나, 또는 '오늘 내가 즐겨 시청한 것과 유사한 종류의 프로그램'을 방송하는 TV 채널을 찾으려고 하는 경우, 이를 키워드 기반의 표현 및 검색 기술로 구현하기란 대단히 어렵다. 이러한 문제를 해결하기 위해 내용 기반 멀티미디어 정보 검색을 효율적으로 지원하기 위한 기술을 개발하고 이를 국제 표준화하고자 하는 것이 MPEG-7이다. 즉, 기존에 표준화되었거나 지금 표준화가 진행되고 있는 MPEG-1/2/4 등은 오디오 비쥬얼 데이터의 데이터 압축을 그 목표로 하였으나, MPEG-7은 데이터 그 자체가 아닌 데이터의 내용에 대한 표현 방법을 다루는 것이다. 이를 다른 말로 '메타데이터(Metadata)', 또는 'Bits about bits'라고 표현하기도 한다.

MPEG-7에서는 주로 오디오비쥬얼 정보(정지화상, 픽쳐, 그래픽, 3D 모델, 오디오, 스피치, 비디오)의 표현을 그 대상으로 하고 있으나, HTML, SGML, 또는 RDF등이 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, 이 중 MPEG-7이 표준화 하고자 하는 범위를 개략적으로 보여준다.



〈그림 1〉 MPEG-7 표준화의 범위

이 중 특징 추출(Feature Extraction) 방식 및 검색 엔진(Search Engine)은 표준화의 대상으로 하지 않는다. 왜냐 하면, 이들은 여러 멀티미디어 정보 처리 시스템간의 호환성을 획득하는 것과는 상관이 없으며, 업체간의 경쟁이 가능한 부분으로 남겨 놓는 것이 향후 계속적인 기술 발전을 유도하기 위하여 바람직하기 때문이다. MPEG-7은 멀티미디어 정보의 특징 및 내용 표현 방식에 초점을 맞추어 다음의 사항을 표준화하고자 한다.

- 기술자(Descriptors : D)와 기술 구조(Descri-ption Scheme : DS)
- 기술 구조를 표현하기 위한 기술정의언어(Description Definition Language : DDL)
- 색인, 저장 및 전송을 효율적으로 하기 위하여 사용될 코딩된 기술(Coded Description)

MPEG-7에서 정의하는 용어는 다음과 같다. 데이터(Data)는 표현 형식에 상관없이 오디오비쥬얼 정보를 의미하며, 특징(Feature)은 오디오비쥬얼 정보의 성격이나 속성을 나타낸다. 기술자(Descriptor)는 특징을 어떤 값에 연결시켜주는 도구이며, 기술 구조(Description Scheme)는 데이터를 여러 개의 기술자로 나타내기 위한 그릇이다. 표 1에 이 용어들의 개념을 명확히 하기 위하여 대표적인 특징 형태, 특징 및 기술자의 예를 보였다.

2. 응용 분야[3]

MPEG-7은 매우 다양한 여러 분야에 사용될 수 있으며, 주요 응용 분야는 다음과 같다.

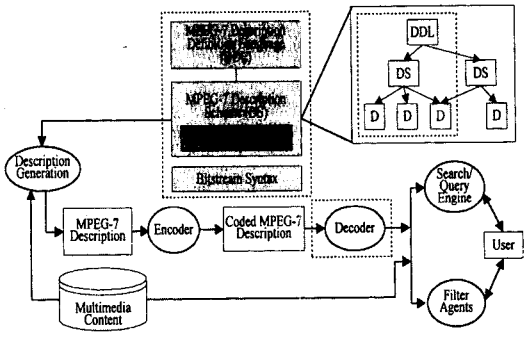
- 교육에서의 멀티미디어 자료 이용
- 언론 분야(이름, 음성 또는 얼굴 정보를 사용하여 어떤 정치가의 연설 장면을 찾을 경우)
- 여행 정보 안내, 지리 정보 서비스
- 역사 박물관, 미술품 전시 등의 문화적인 서비스
- 게임 검색, 가라오케에서의 노래 검색 등 오락 분야
- 사람의 특징 인식 등 범죄 수사
- 지도 제작, 생태 조사, 자연 자원 관리 등의 원격 탐사
- 원격 구매
- 건축, 부동산, 내부 장식 관련 정보 검색
- 영화, 비디오, 라디오 정보의 저장 및 검색

한 종류의 기술자나 기술구조가 위에 열거한 모든 응용 분야를 다 만족시키는 것은 어려운 일이

〈표 1〉 오디오비쥬얼 데이터의 대표적인 특징 및 기술자의 예[2]

특징 형태(Feature Type)	특징(Features)	기술자(Descriptors)
N-dimensional spatio-temporal structure	duration of a music segment	time code, etc.
	trajectory of objects	chain code, etc.
Statistical Information	color	color histogram, etc.
	audio frequency content	average of frequency components, etc.
Objective features	color of an object	color histogram, etc.
	shape of an object	a set of polygon vertices, a set of moments, etc.
Subject features	emotion(happiness, angry, sadness, etc.)	a set of eigenface parameters, text, etc.
	style	text, etc.
Concepts	event	text, etc.
	activity	text, etc.

며, 효율적이지도 않다. 따라서, MPEG-7에서는 기술자나 기술구조의 제안 사항을 주요 응용 분야별로 나누어 표준화 할 방침을 갖고 있다. 응용 분야와 관련하여 논의가 진행되고 있는 것 중의 하나는 'pull model'과 'push model'의 구분이다. 전자는 국지적으로나 네트웍 상에 있는 데이터베이스로부터 사용자의 질의에 가장 가까운 데이터를 찾아주는 응용 분야이고, 후자는 방송에서와 같이 정보 제공자로부터 일방적으로 데이터가 공급될 때, 이로부터 원하는 정보를 필터링하여 주는 응용 분야이다. 두 응용 분야의 주요 특징상 차이점은 후자의 경우 정보 검색시 실시간성이 강하게 요구된다는 것이다. 그림 2에는 MPEG-7의 표준화 요소와 응용 시스템 기능 구성 간의 관계를 보였다.



〈그림 2〉 MPEG-7의 표준화 요소와 응용 시스템 기능 구성

3. 요구 사항[4]

MPEG-7에 관한 요구 사항은 기술정의 언어에 대한 요구 사항, 오디오비쥬얼 데이터에 공통적으로 적용되는 요구 사항, 그리고 오디오 또는 비쥬얼 데이터별 특성에 연관된 요구 사항으로 나누어진다. 우선 기술정의 언어에 대한 요구 사항으로는, (1)MPEG-7 응용 분야를 지원하는 어떤 기술구조도 쉽게 생성되어야 하며, (2)문법이 모호하지 않고 인터프리터 등에 의해 쉽게 파싱(parsing)될 수 있어야 하며, (3)오디오비쥬얼 데이터를 포함한 복합적인 멀티미디어 정보에 대한 기술구조를 생성할 수 있어야 하며, (4)기술구조 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.

오디오비쥬얼 데이터에 공통적으로 적용되는 기술자 및 기술구조에 대한 일반적인 주요 요구 사항으로는, (1)데이터에 부가된 주석, 데이터의 시공간적인 구조, 주/객관적 특징, 또는 사건이나 활동 등의 개념까지도 포함하는 모든 특징의 표현을

지원하여야 하고, (2)사용자의 요구에 적절히 대응할 수 있는 계층적인 내용 압축 및 표현 방식을 지원하여야 하며, (3)비쥬얼 질의를 사용해서 오디오 정보를 찾고 오디오 질의를 사용하여 비디오 정보를 찾을 수 있도록 Cross-modality를 지원하여야 한다. 또한, (4)MPEG-7의 여러 가지 기술 구조 상호간의 호환성, (5)특징별 우선 순위 지정, (6)기술자의 Scalability 지원, (7) 텍스트 기반 기술자의 경우 사용 언어의 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, (1)내용 기반 검색 및 유사도 기반 검색 기능의 제공, (2)데이터와 동기화 되어 제공되거나 (Streamed) 또는 데이터와 별도로 제공되는 (Stored) 기술자의 지원, (4)대화형 질의의 지원, (5)소스 데이터 또는 관련 정보의 위치를 알려줄 수 있는 기능, 그리고 (6)정보의 개략적인 파악 (preview)이 가능하도록 브라우징(Browsing) 기능의 지원 등이 있다. 마지막으로 비쥬얼 또는 오디오 정보 각각의 특성에 따른 요구 사항으로는 지원되는 특징의 종류, 지원되는 데이터 포맷 및 클래스에 대한 요구 사항이 포함된다.

4. 향후 일정 및 전망

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[5]. MPEG에서 취하고 있는 표준화 과정을 살펴보면, 우선 제안 요구서(Call for Proposal : CFP)가 공고되고 나면 일정한 기간동안 기술 제안서를 받아서 그 중 요구 사항이 잘 반영되고 객관적으로 성능이 우수한 제안서 들을 바탕으로 통합 시험 모델(Experimentation Model : XM)을 만든다. XM은 MPEG-2에서의 TM(Test Model), MPEG-4의 VM(Verification Model)과 동일한 역할을 한다. 이 단계까지를 경쟁 단계(Competitive Phase)라고 하며, 이 후에는 XM을 바탕으로 각 세부 항목별 공동 실험(Core Experiment)을 통한 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. 표 2에 MPEG-7의 향후 일정을 보였다.

〈표 2〉 MPEG-7 표준화 일정

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

III. 내용기반 정보 검색

멀티미디어 데이터는 다양한 형태의 데이터들로 구성되어 있을뿐더러 데이터의 크기가 방대하기 때문에 MPEG-7에서는 효과적이고(우리가 원하는 데이터를 얼마나 정확하게 찾을 수 있는가 하는 관점)도 효율적인(얼마나 빠르게 원하는 데이터를 찾을 수 있는가 하는 관점) 멀티미디어 데이터 내용 검색을 위한 표현 방법을 표준화 하고자 한다.

멀티미디어 데이터가 갖는 풍부한 정보에 비추어 보아, MPEG-7에는 두 가지 방향의 접근 및 해법이 제시될 것으로 예상된다. 그 하나는 상위 레벨 내용에 기반한 검색과 다른 하나는 그보다 하위 레벨 내용에 기반한 검색 방법이다. 하위 레벨의 내용에는 키워드, 형태, 색, 크기, 위치, 방향성, 멜로디 등이 포함될 수 있고, 상위 레벨의 내용에는 의미 정보들이 포함될 수 있다. 하위 레벨의 내용에 의한 검색은 사용자 질의가 부자연스러운 점은 있으나 시스템이 자동으로 검색할 수

있다는 장점이 있는 반면, 상위 레벨의 내용에 의한 검색은 좀더 자연스런 질의를 할 수 있으나 현재의 컴퓨터 기술로는 사용자의 개입을 필요로 한다는 단점이 있다. MPEG-7의 유용성 및 표준으로서의 성공은 이러한 두 가지 측면을 어떻게 적절히 수용하는가 하는데 크게 좌우될 것이다.

본 장에서는 멀티미디어 데이터를 대상으로 현재까지 연구되거나 개발된 내용기반 정보 검색 기법들에 관해서 설명하고자 한다. 멀티미디어 데이터는 크게 음성이나 음향정보를 포함한 오디오 데이터, 사진이나 그래픽들을 포함한 정지 영상 데이터, 비디오를 포함하는 동영상 데이터로 구분 할 수 있다. 따라서 내용 기반 검색 기법을 이러한 세 가지 데이터와 연관시켜 분류하고자 한다.

1. 오디오 데이터에 대한 내용기반 검색

오디오 데이터에 대한 내용기반 검색 방법에는 크게 브라우징과 인덱스를 통한 검색방법이 있다. 인덱스를 통한 검색 방법은 다시 오디오의 음향이나 음악 등을 분석하여 특징벡터로 인덱스를 만든 후 사용자가 멜로디나 음향 효과로 질의를 하여 원하는 곡을 찾는 방법과 오디오내의 음성을 인식하여 키워드 기반의 인덱스를 만든 후 사용자가 질의를 음성이나 텍스트로 해주는 방법으로 구분할 수 있다.

[6]에서는 이야기 개요(skimming speech)를 만들기 위한 분할 방법과 사용자 인터페이스를 연구하였다. 이 시스템은 이야기에 대한 시간적인 압축과 이야기중의 불필요한 멈춤점을 제거함으로서 이야기 내용을 줄여서 들을 수 있게 하였다. 이 시스템은 3단계의 서로 다른 이야기 개요로 구성되었는데, 각 단계는 다시 시간적으로 압축을 하여 여분의 정보를 줄였다. 가장 낮은 단계는 원래의 이야기로 구성이 되어있고, 두 번째 단계에는 이야기중의 불필요한 멈춤점을 줄이거나 제거한 이야기 내용으로 구성이 되었으며, 마지막 단계에서는 소리가 전혀 없는 멈춤점과 의미 없는 소리로 생긴 멈춤점등을 구분하여 이야기 내용을 의미적으로 더욱 축소하며 분할하였다. 사용자는 이러한 서로 다른 이야기 단계를 브라우징하여 원하는 내용을 빨리 검색할 수 있다. 이 시스템은 또한 이야기 각 단계의 한 구역에서 다른 구역으로 이동하거나, 이야기를 앞으로 또는 뒤로 빨리 듣거나 천천히 들을 수 있는 인터페이스를 제공한다. 이러한 시스템은 강의 복습, 음성 사서함, 과거 이야기 기록 검토 등 다양한 응용에 사용될 수 있다.

음향의 분석을 통한 내용 기반 검색의 가장 대표적인 응용으로는 노래방을 들 수 있다. 현재 노래방에서 원하는 노래를 선택하려면 노래의 제목을 알아야한다. 하지만 원하는 노래의 제목을 기억하여 찾는 기존의 방법보다는 노래의 멜로디를 통해 원하는 곡을 찾는 방법이 더 자연스러운 방식일 것이다. [7]에서는 콧노래를 불러서 원하는 곡을 오디오 데이터베이스에서 찾는 방법을 연구하였다. 이 시스템에서는 음악의 멜로디를 표현하기 위해서 피치간의 관계성을 사용하였다. 켑스트럼 분석기법을 적용하여 피치를 추적하였고 이를 통하여 멜로디내의 연속된 음조의 관계성을 과거 음조와 같은 음, 낮은음, 또는 높은음(UDS)의 연속된 문자열로 바꾸는 방법을 사용하였다. 사용자가 콧노래로 질의를 하면 이 멜로디는 마찬가지 방식으로 UDS의 문자열로 바뀌고 데이터베이스에 같은 방법으로 저장되어있는 문자열들과의 매칭을 통하여 가장 근접한 멜로디들을 검색한다. 비슷한 연구로서 [8]에서는 잡음에 강하고 효율적인 인덱스를 만들기 위해 오디오 데이터를 먼저 주파수 영역으로 변환한 후 특정 계수만을 사용하여 인덱스를 만들었다. 오디오 사운드 질의를 수행하기 위해서 오디오 시그널을 일정한 크기의 블록으로 나누고 각 블록을 DCT(Discrete Cosine Transfer) 등의 직교 변환을 통해 주파수 영역으로 변환한 후 특정의 계수만을 선택하여 데이터베이스에 마찬가지 방식으로 저장된 오디오 파일들의 특징 벡터들과의 유사성 검색을 하였다. 오디오들을 오디오내의 음향 특징에 의해 분류하고 검색할 수 있다면 많은 오디오나 멀티미디어 응용에 활용할 수 있을 것이다. [9]에서는 오디오의 사운드를 소리의 크기(loudness), 피치, 밝음(brightness), 대역폭, 하모니 등의 음향효과로 분석하여 대응하는 특징벡터로 표현하였다. 이 연구에서는 동물소리, 기

계소리, 악기소리, 스피치 등 15초미만의 사운드들을 사용했으며 이러한 사운드들은 특징 벡터로 바꾸어져서 몇 가지 사용자가 정의하는 키들과 함께 오디오 데이터베이스에 저장된다. 사용자는 질의로 긁는 소리(scratchy sound), 사람의 목소리, 동물소리, 또는 이것들을 조합한 소리와 비슷한 소리들을 요청할 수 있다.

오디오의 음성을 인식하여 키워드 기반의 인덱스를 만드는 응용으로는 음성 사서함 시스템이 있다. [10]에서는 도착하는 음성 편지에 단어 검출(word spotting) 기법을 적용해 편지내의 단어들을 인식하여 인덱스를 만들고 사용자가 편지의 일부 단어들로 구성된 문자열로 질의를 하면 그 단어들을 포함한 편지를 돌려준다. 이 시스템은 300개의 음성 편지들을 포함하고 있는데 각 편지는 미리 정해진 35개의 단어 중 평균 7개의 단어로 구성되었다. 미리 정해진 단어들로 음성 훈련을 시켜 인식을 하도록 시도하였기 때문에 음성 인식은 이미 정해진 어휘 수에 한정될 수밖에 없다. 반면에 [11]에서는 오디오 데이터를 은닉 마코프 모델에 기반한 음소 인식기를 통하여 음소 번역물(phonetic transcription)로 바꾼 후 통계적인 분석을 통하여 인덱싱에 사용될 음소 열들을 직접 추출하였다. 이 시스템은 4.5 시간의 라디오 뉴스를 갖고 있으며 사용자가 문자열 질의를 하면 해당 정보를 포함하고 있는 이야기 문서들의 목차를 우선 순위로 나열해준다.

2. 정지 영상 데이터에 대한 내용기반 검색

정지영상에 대한 일반적인 내용기반 검색 기법들은 영상의 색이나 질감, 형태 등의 특징을 사용한다. 이러한 시스템에서는 처음에 영상의 특징들이 자동으로 추출되어 영상과 함께 인덱싱되어 데이터베이스에 저장된다. 사용자는 질의로 원하는 색이나 질감 등을 포함하는 영상을 요청하거나 원하는 영상의 모양을 스케치하는 방식으로 내용 기반 질의를 할 수 있다. 질의 결과는 정확한 영상이 반환되기보다는 요청한 영상과 비슷한 이미지들의 집합이 반환된다.

[12]에서는 사용자가 그린 대략적인 스케치 또는 복사본 이미지로써 이미지 데이터를 검색할 수 있는 연구를 하였다. 이 연구에서는 이미지 데이터의 특징값을 나타내기 위하여 아이콘 이미지를 사용하였다. 아이콘 이미지는 원래의 이미지 데이터를 작은 크기로 축소한 것을 의미하며, 데이터베이스에 저장된 각각의 이미지들은 모양과 색상에 대한 아이콘 이미지를 가지고 있으며 검색 시에 아이콘 이미지를 대상으로 검색을 하게 된다. 이러한 접근 방식은 다양한 형태의 이미지 데이터를 다룰 수 있다는 장점이 있으나 유사한 이미지를 검색하기 위해 모든 데이터를 픽셀 대 픽셀 비교로 순차적 검색방법을 사용하므로 매우 비효율적이다. [13]에서는 이미지 데이터로부터 형태적 특징과 색상 특징을 추출하여 원하는 이미지를 검색할 수 있도록 하였다. 컬러 히스토그램과 정적으로 분할된 이미지 영역으로부터 외곽선을 추출하여 공간상의 관계를 포함하는 형태 정보를 추출하여 이미지 비교에 사용하였으며, 이들을 효과적이고 빠르게 비교 검색하기 위해서 색상 성분에 대한 특징값은 R 트리, 형태의 특징값은 변형된 트라이로 인덱싱하는 방법을 제안하였다. 영상의 색, 질감, 형태뿐 아니라 스케치방식의 질의를 모두 처리해주는 대표적인 시스템은 IBM Almaden 연구소에서 개발된 QBIC(Query By Image Content)이다[14]. 이 시스템은 사용자에게 이미지의 색상이나 질감, 모양 등과 같은 다양한 속성에 기반한 시각적 질의를 제공할 뿐 만 아니라 제한된 범위내의 비디오 검색이 가능하다. QBIC은 또한 사용자로 하여금 데이터베이스내의 이미지에 대하여 적절한 키워드로 주석을 붙일 수 있도록 허용함으로써 제한적이나마 주석에 기반한 검색도 가능하다. 반면에 Columbia 대학에서 개발한 VisualSEEk은 WWW에 기반해서 원하는 영상을 검색하는 내용기반 영상 검색 시스템으로 QBIC에서와 유사한 검색을 제공할 뿐만 아니라 영상 내의 객체영역에 대한 색상과 객체영역의 공간정보 사이의 상관관계를 더 자세히 묘사할 수 있는 인터페이스를 제공한다[15]. 이 시스템은 객체영역의 크기와 공간적 위치정보에 대한 향상된 인덱싱 기법을 이용함으로써 복잡하게 결합된 다양한 색상/공간 및 색상/

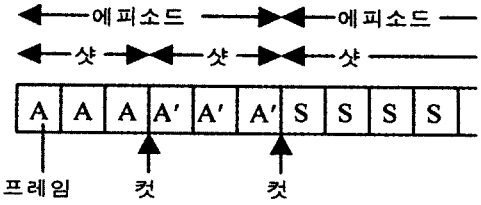
질감, 주석에 기반한 문자/영상 질의를 제공한다.

정지영상에 대한 내용기반 검색의 다른 방법은 영상의 원래 데이터 (raw data)로부터 직접 특징을 추출하는 것이 아니라 일단 직교변환 등을 시켜 변환된 데이터로부터 특징들을 추출하는 방법이다. 하나의 예로 웨이브릿 변환(Wavelet Transform)을 통해 특징을 추출하는 방법을 들 수 있다. 웨이블릿 변환은 기본함수로 사인, 코사인 함수뿐만 아니라 좀더 복잡한 웨이브릿 모함수를 사용할 수 있고, 푸리에 변환에는 없는 공간에 대한 특성, 즉 저주파 밴드는 영상 전체의 특징을 잘 나타내고 고주파 밴드는 국부적인 특성을 갖고 있다는 장점이 있다. [16]에서는 이미지 데이터로부터 외곽선을 추출한 후 웨이브릿 변환을 이 외곽선 데이터에 적용하여 그 결과로 나온 특징 벡터를 이미지 전체의 모양에 대한 특징으로 표현하였다. 따라서 사용자는 그래픽 도구를 이용하여 원하는 이미지의 전체적인 그림을 그리거나 스캐너 등을 통해 읽어들인 예제 이미지를 선택함으로써 원하는 이미지를 검색할 수 있다. 반면에 [17]에서는 영상의 색, 질감, 형태의 모든 특징을 함께 웨이브릿 변환에 적용하였다. 또한 영상의 웨이브릿 계수를 통한 내용 표현 문제 뿐 아니라 데이터 압축 문제도 하나의 통합된 프레임웍(framework)에서 연구하였다. 즉 영상들은 압축이 되면서 색, 질감, 모양 등의 특징에 의해서 인덱싱 되며, 사용자는 색, 질감, 모양 또는 이들의 조합을 통한 질의를 할 수 있도록 하였다. MIT의 Photobook은 영상의 통계적인 성질에 기반을 둔 KL (Karhunen-Loeve)변환을 적용한 내용 기반의 이미지 검색 시스템이다[18]. 이 시스템은 KL 변환을 사용하여 영상을 몇 개의 주성분 값으로 표현하였으며, 영상을 공간으로 변환하기 위한 기저 벡터로 영상의 벡터로부터 구한 공분산 행렬의 고유벡터를 사용했다. 이 방식의 특징은 영상 구별에 필요한 성분만을 추출하여 압축을 하였고 다시 원래의 영상으로 복원이 가능하다는 것이다. 이 시스템은 얼굴인식 분야 등에 응용되었는데, 기존의 얼굴 윤곽선 등의 특징점들을 통한 매칭 방법과는 다르게 얼굴의 표정 등 얼굴에 약간의 변형을 주어도 같은 얼굴을 찾아낼 수가 있다.

3. 동영상 데이터에 대한 내용기반 검색

동영상에 대한 내용기반 검색을 위해서는 동영상 데이터를 색인하기 위한 비디오 파싱기법과 사용자가 원하는 데이터를 쉽게 검색할 수 있는 사용자 인터페이스 뿐만 아니라 제한된 저장공간에 대용량 비디오를 효율적으로 저장하기 위한 비디오 데이터 압축 및 저장 방법 등의 기술들이 필요하다. 동영상에 대한 내용기반 검색기법은 보는 관점에 따라 여러 가지로 분류할 수 있다. 본 절에서는 동영상에서 내용검색을 하기 위해 사용할 수 있는 정보들, 즉 영상정보, 문자정보, 오디오 정보 중 어떤 정보들을 사용하였는가에 따라 내용기반 검색 기법들을 분류하고자 한다.

동영상의 내용기반 검색을 위하여 가장 일반적으로 사용할 수 있는 정보는 영상정보이다. 영상정보는 주로 비디오를 장면 분할할 때에 사용되며 이를 통하여 구조적인 비디오 브라우징을 할 수 있다. 비디오를 구성하는 최소단위는 필름 한 장에 해당하며 하나의 영상을 나타내는 프레임이다. 비디오에서 장면의 전환이 이루어지는 부분을 컷(cut)이라고 하고, 컷으로 구분되며 하나의 카메라 동작에 의해 촬영된 작은 비디오 단위를 샷(shot), 논리적인 내용이 같은 연속된 샷으로 이루어진 단위를 에피소드(episode)라 한다. 따라서 구조화된 비디오는 그림 3에서와 같이 내용전환으로 구분되는 연속된 에피소드로 구성되고 각 에피소드는 장면 전환의 단위인 샷으로 구성된다.



〈그림 3〉 비디오 구조

비디오를 샷으로 구분하는 작업을 비디오 분할(video segmentation)이라고 하며, 비디오 분할을 위해 장면의 전환점인 컷을 검출하는 작업을 컷검

출(cut detection)이라고 한다. 비디오는 연속된 프레임의 집합이므로 연속된 장면에서는 인접한 프레임 사이의 유사성이 강하고 장면의 전환이 이루어지는 부분에서는 프레임 사이의 유사성이 상대적으로 약하다. 따라서 컷을 추출하기 위해서는 비디오 요소의 프레임간의 차이를 이용하여 그 요소의 연속성을 계산하고 불연속 지점을 컷으로 간주한다. 지금까지 컷검출을 위한 다양한 알고리즘이 연구되었다. 자동으로 컷을 검출하기 위한 방법으로는 히스토그램의 차이 비교, 화소간의 차이 비교, 에지 변화 비교, 압축 상관 계수 비교, 유사율 측정법 , 그리고 움직임 벡터 비교 등이 있다. 히스토그램 기반의 방법은 같은 장면으로 분류해야 할 프레임들의 색상 분포는 거의 비슷하다는 성질을 이용하여 각 프레임간의 히스토그램 차이를 계산해 정해진 임계값을 넘을 경우 컷으로 판단한다. 화소간의 차이 비교방법은 화면을 구성하는 화소들은 히스토그램과 마찬가지로 동일한 장면 내에서 변화가 적다는 성질을 이용하여 각 프레임의 화소들을 비교해 차이가 임계값을 초과하면 그 프레임간에는 장면 전환이 있다고 본다. 유사율 측정법은 개개의 화소를 비교하는 것이 아니라 연속된 프레임들에서 대응하는 일정 영역의 통계치를 비교하는 방법이며, 에지 기반의 방법에서는 주요 성분 에지를 파악하고 프레임간에 에지 단위의 비교를 수행함으로써 특정 객체가 다음 프레임에 포함되어 있는지 또는 변화가 어떻게 이루어지는지를 판단한다. 압축 상관 계수의 비교법에서는 연속된 프레임의 DCT 계수의 차이를 구하고, 이러한 차이를 이용해 장면 변화를 검출하고자 하는 방법이다. 움직임 벡터의 비교 방법에서는 MPEG 데이터로부터 얻어지는 움직임 벡터를 사용하여 이동물체의 움직임 분석뿐 아니라 카메라의 움직임, 예를 들면 줌(zoom)또는 패닝(panning)등과 같은 카메라의 연산을 인식함으로써 장면 변화를 보다 정확하게 검출하고자 하였다.

다음으로 동영상의 내용검색을 위하여 사용될 수 있는 정보는 영상 내에 있는 문자정보이다. 영상 내에 있는 문자정보는 우선 비디오 촬영 시 장면과 같이 찍혀 장면 내에 포함되어 있는 문자와, 특정한 내용 또는 설명을 부여하기 위해 비디오 영상에 부가적으로 인코딩된 시그널 형태의 주석(closed caption)이나 인위적으로 삽입된 문자 등으로 구분할 수 있다. 일반적으로 장면 내에 포함되어 있는 문자들은 해상도가 낮고 글자 형태와 크기가 다양하기 때문에 추출과 인식이 쉽지 않을 뿐더러 의도하지 않은 배경화면의 문자인 경우도 많기 때문에 내용검색 기법에서 잘 사용되지 않는다. 반면에 주석이나 삽입된 문자는 문자의 크기나 위치가 일정하고 영상의 내용을 설명하기 때문에 내용기반 인덱싱에 효율적으로 사용될 수 있다.

마지막으로 동영상의 내용검색을 위하여 사용될 수 있는 정보는 영상 내에 있는 오디오 정보이다. 오디오에 있는 음향정보나 음악 등을 분석하여 구조적인 비디오의 장면 분할을 할 수 있을 뿐만 아니라 오디오내의 음성정보를 인식하여 대응하는 문자 정보로 변환한 후 인덱스를 만들어 내용기반 검색에 사용할 수 있다. 따라서 동영상 데이터에 대한 내용기반 검색 기법은 동영상에서 어떤 정보들을 사용했는가에 의존하게 된다. 만약 동영상에서 영상 정보만을 사용하였다면 영상 분할을 통한 브라우저나 비슷한 색상이나 형태 등의 내용을 찾는 하위 레벨의 내용기반 질의를 할 수 있는 반면, 문자정보나 음성정보를 함께 사용한다면 인식 기술을 통한 인덱스를 만들 수 있어 자연어나 음성 등의 상위 레벨의 내용기반 질의를 가능케 해준다.

1) 영상의 정보만을 사용한 내용기반 검색

동영상에서 영상의 정보만을 사용했을 때는 주로 동영상의 장면 분할을 통한 구조화된 브라우징 검색 기법이 많이 사용된다. 비디오 분할기법은 다시 분할 대상 비디오 종류에 따라서 비압축 비디오에 대한 분할과 압축비디오에 대한 분할 기술로 구분할 수 있다. 동영상에서 영상의 정보를 검색에 사용하는 다른 방법은 영상의 정보들, 예를 들면 움직임, 색, 모양 등의 특징들을 사용해 일정한 영역의 움직임을 찾아내는 방식이다. 사용자는 원하는 특정 영상을 찾기 위해 정적인 스케치를 하는 것이 아니라 원하는 동영상의 일부를 찾기 위해서 동영상내의 특정 객체의 움직임을 스케치할 수 있다.

[19]에서는 기존의 비디오를 구조화된 논리적 비디오로 재구성하여 검색 및 브라우징을 효과적으로 할 수 있는 비디오 데이터베이스 시스템을 구현하였다. 비디오 구조화의 전체적인 구성은 비디오 분할과 비디오 인덱싱의 두 단계로 이루어진다. 입력된 비디오는 비디오 분할에 의해서 장면단위의 샷으로 구분된 후 사용자는 각 샷의 물리적 정보 및 논리적 정보를 추출하여 샷단위로 인덱싱할 수 있다. 샷의 물리적 정보는 카메라 효과와 장면의 특수효과로 구성되며 샷의 논리적인 정보는 자연의 배경, 물체, 상황으로 이루어진다. 또한 사용자는 내용상으로 연관성이 있는 연속된 샷들을 다시 연결하여 에피소드를 구성할 수 있다. 즉, 비디오는 내용상의 구분인 에피소드 단위와 장면 구분인 샷단위의 계층적 구조로 재구성된다. 비디오 브라우징 시스템은 정의된 대표 프레임들을 시간순서로 화면에 나열하고 비디오 플레이 기능을 통해 샷의 내용을 빠르게 파악할 수 있다. 비디오 검색을 위한 사용자 질의는 비디오의 논리적 정보에 해당하는 설명, 장면 구성 정보와 비디오의 물리적 특성을 나타내는 장면 특성 정보, 비디오 상영시간 등으로 구성되며 각각의 질의는 부울연산으로 연결될 수 있다.

MPEG으로 압축된 비디오는 압축정보로 DCT 계수와 이동 벡터를 가지고 있다. 따라서 압축비디오의 분할방법은 사용되는 특징에 따라서 DCT 계수를 사용한 방법, 이동 벡터를 사용한 방법, 또는 이 둘을 함께 사용하는 방법 등이 있다. DCT 계수는 원래 영상의 화소의 세기와 색차에 해당하는 정보를 갖고 있기 때문에 장면 분할하는데 효율적으로 사용될 수 있다. DCT계수를 사용하는 일반적인 방법은 연속된 프레임의 DCT 계수의 차가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 추출하는 방법이며, 이때 임계값으로는 전체 영상의 통계적 특성을 이용하여 평균, 표준편차, 또는 분산값에 가중치를 적용하여 사용한다. 장면 전환을 검출하기 위해 기존에는 경험적으로 임계점을 결정하였는데 반해 [20]에서는 현재 조사하려는 장면을 장면전환이 일어난 부류와 장면전환이 일어나지 않은 부류 중 하나로 결정하는 패턴 인식 문제로 모델링하고 신경망 모델의 하나인 학습 가능한 퍼셉트론 패턴인식기를 이용하여 보다 정확한 장면전환을 검출하기 위한 방법을 제안하였다. 이 연구에서는 MPEG으로 압축된 데이터에서 DC 영상만을 추출하여 프레임간의 차이값을 결정하고 구해진 차이값을 이용하여 퍼셉트론을 학습시키고 학습된 퍼셉트론을 이용하여 장면전환을 검출하였다. 제안한 방법이 지역적이나 전역적인 임계점을 이용하여 장면전환을 검출하는 방법보다 효율적임을 보였다.

비디오내의 특정 객체의 움직임을 찾아내는 연구로 VideoQ 시스템에서는 사용자로 하여금 비디오내의 특정 영역의 움직임을 표현해주는 움직이는 스케치(animated sketch) 질의를 하게 해준다[21]. 이 시스템은 먼저 비디오를 분할하여 샷으로 구분한 후 샷 단위로 특정 영역들의 움직임을 추출하여 하나의 객체로 만든다. 시스템은 각 객체로부터 객체단위의 색이나 질감, 모양, 움직임 등의 특징들을 추출하여 데이터베이스에 저장한 후 검색시에 이용한다. 이 시스템은 웹기반 비디오 검색 시스템으로 영상내의 특정 영역이 배경과 구분이 잘되고 스케치로 표현하기 쉬운 경우, 예를 들면 축구 사건, 스키 사건들의 경우에 좋은 검색 결과를 보였다.

2) 문자정보 또는 영상과 문자정보를 함께 사용한 내용기반 검색

동영상에서 영상 정보 외에 문자정보를 사용하였을 때는 문자정보 인식을 통해서 추가적인 인덱스를 만들 수 있기 때문에 키워드나 자연어 질의를 통한 검색도 가능하게된다. 영상에서 문자정보를 이용하는 방법들은 주로 영상에 부가적으로 추가된 주석이나 영상에 삽입된 문자로부터 키워드들을 추출한 후 해당 장면들과 연관시켜 문자열 질의 검색과 구조적인 브라우징을 가능케 해준다.

문자 정보만을 내용기반 검색에 사용한 예로 [22]에서는 상업광고나 뉴스의 비디오 프레임내의 문자를 인덱싱에 사용하였다. 이 시스템은 영상에 삽입된 문자의 일반적인 특징들을 분석하고 영상의 내용을 여러 칼라영역으로 분리한 후 크기, 칼라, 모양 변화와 움직임 분석 등을 통하여 문자 영

역만을 분할해 내었다. 추출된 각 문자로부터 획득한 이진 영상은 표준 OCR 시스템에 의해 인식되어 인덱스로 바뀌어지고, 사용자가 문자열 질의를 하면 질의 문자열을 포함하는 모든 영상들을 반환해준다. 반면에 [23]에서는 영상내의 삽입된 문자정보를 비디오 분할의 근거로 활용하였다. 이 시스템은 문자열의 출현과 소멸을 지역적인 컷이나 디졸브로 정의하였고, 비디오의 샷 경계를 추출하기 위해 비디오 내에 포함된 문자 정보의 유무를 감지하여 키 프레임을 검출한 후 검출된 프레임으로부터 문자영역을 추출하는 방법을 사용하였다. 추출된 문자영역 내의 각 문자는 인식하지 않고 단지 문자를 포함하고 있는 프레임과 시간정보를 결합하여 뉴스방송을 분할하였고 이를 비디오 인덱싱에 사용하였다.

영상과 문자 정보를 함께 사용한 예로, [24]에서는 비디오 뉴스의 특정 단어들과 특정 영상들을 서로 연관시켜서 의미 있는 장면들을 추출하고 이를 통하여 비디오를 사건별로 분할하여 브라우징할 수 있도록 하였다. 의미 있는 사건들을 찾아내기 위하여 비디오 주석에서 이미 정해진 키워드와 관련 문장들을 분석하여 언어 근거를 뽑아내고, 확대된 사람의 얼굴은 자동으로 추출하고, 사람들 장면, 야외 장면들은 수동으로 추출하여 장면 근거들을 뽑아냈다. 동적 프로그래밍 기법을 통하여 관련되는 언어근거와 영상 근거들을 연관시켜 의미 있는 장면들 예를 들면, 대화장면, 모임장면, 여행장면, 장소장면 등의 몇 가지 사건들로 구분한 후 사건별로 비디오를 분할할 수 있도록 하였다. 반면에 [25]에서는 뉴스 비디오의 시간적, 공간적 특성을 이용한 실시간 뉴스 비디오 파서를 통한 장면분류와 삽입된 문자를 인식하여 구조적 뉴스검색 도구를 구현하였다. 장면분류에서는 히스토그램 차를 이용해 키프레임들을 먼저 추출한 후 얼굴인식과 사전지식 등을 사용해 앵커장면과 뉴스 아이콘 장면을 추출하여 샷과 에피소드로 분류하였다. 자막 문자열 인식에서는 자막이 갖는 공간적 사전 정보를 사용해 먼저 문자열을 추출한 후 문자를 분할하여 인식하였다. 이와 같이 추출된 아이콘과 인식된 문자열은 인덱스 파일의 구성정보로 사용되어 사용자가 문자열을 통해 원하는 프레임을 쉽게 검색할 수 있도록 하였다.

3) 오디오 정보 또는 영상과 오디오 정보를 함께 사용한 내용기반 검색

동영상에서 오디오 정보를 사용하였을 때는 문자정보와 마찬가지로 동영상에 대한 장면 분할뿐 아니라 음성 인식을 통한 인덱스를 만들어 키워드나 자연어 질의를 통한 내용기반 검색도 가능케 된다.

오디오 정보만을 사용하여 비디오의 장면 분할을 시도한 연구가 있다[26]. 오디오 정보의 중요한 특징은 오디오의 종류에 따라 특정 사건이 발생했음을 알 수 있다는 것이다. 예를 들어 음악은 클라이맥스나 의미가 있는 사건일 경우에 주로 사용된다. 따라서 음악과 음성은 비디오에 포함된 오디오에서 가장 의미 있는 정보라고 할 수 있으며 비디오에서 이러한 오디오 정보를 근거로 하여 장면 분할을 시도하였다. 오디오 데이터의 사운드 스펙트로그램(sound spectrogram)을 분석하여 배경음악과 음성을 추출한 후 전체의 비디오를 몇 개의 의미 있는 구역, 즉 오디오 특징이 존재하지 않는 구역, 음악만 존재하는 구역, 음성만 존재하는 구역, 음악과 음성이 함께 존재하는 구역 등으로 비디오를 분할하여 구조적인 검색을 할 수 있도록 하였다.

영상과 음성 정보를 함께 사용하여 내용기반 검색을 시도한 연구로 특정 응용분야인, 미식축구를 대상으로 한 연구가 있다. [27]에서는 음성처리 모듈을 통하여 먼저 비디오 데이터에서 예비 장면 구간을 찾아낸 후 영상 처리 모듈을 적용하여 비디오에서 터치다운하는 장면을 자동으로 찾는 연구를 하였다. 먼저 중요한 사건을 탐지하기 위해서 화자 종속의 단어검출 기법을 적용해 터치다운의 단어들을 인식하고 오디오 신호 분석을 통해 중요한 사건이 일어날 때 발생하는 요란한 갈채소리를 찾아낸다. 이를 통해 탐지된 비디오의 구간에 장면 분할 기법을 적용하여 예비 장면들을 추출한 후 모델에 기반한 특징점들을 매칭하는 방법으로 터치다운하는 장면을 찾아내었다.

4) 영상, 문자, 오디오정보를 사용한 내용기반

검색

동영상에서 영상정보, 문자정보, 오디오정보를 모두 사용하였을 때는 가장 효율적인 내용기반 검색 엔진을 만들 수 있다. 오디오 정보 중 특히 음향정보나 신호는 장면 분할에 도움이 되며, 문자정보나 음성정보는 인식 기술을 적용하여 대응하는 문자정보로 만들어 의미적인 장면 분할뿐 아니라 정보검색 기술이나 자연어 처리 기술을 적용하여 인덱스를 만들어 여러 가지 내용기반 검색을 가능케 해준다.

영상, 문자, 음향정보를 사용하여 특정한 응용에서 비디오 장면을 구조적으로 분할하는 시도가 있었다. [28]에서는 CNN 뉴스 방송에서 발생하는 여러 단서들에 대한 정보를 이용하여 뉴스 방송을 여러 개의 개별 뉴스 사건이나 광고로 분할하여 웹 기반 검색을 가능케 하였다. 이 시스템에서는 비디오 방송을 체계적으로 분할하기 위하여 로고나 앵커 등의 영상정보, 주석에서 규칙적으로 사용되는 문자정보나 기호들, 상업광고 전후 약 .7초간의 공백이 존재한다는 음향정보 등을 단서로 이용하였다. 이러한 단서들을 이용하여 시간에 대한 사건의 변화를 알아내기 위하여 FSA(Finate State Automata) 형태의 사건 변화 정보를 관계형 데이터베이스에 저장하여 사용자의 질의시 사용하였다.

반면에 카네기멜론 대학교에서 수행하고 있는 Informedia 프로젝트는 일반적인 동영상에 영상처리 기법, 음성인식기술, 문자인식 기술 등을 모두 사용해 내용기반 검색과 추출을 가능하게 하는 디지털 비디오 라이브러리를 만들고 있다[29]. 이 연구에서는 특히 Sphinx-II 음성인식 기술을 사용해 비디오 내의 오디오 데이터를 인식하여 대응하는 문자열로 바꾼 후 자연어처리 기술이나 정보 검색 기술을 적용해 샷이나 에피소드를 분리해내고 인덱싱을 위한 키워드들을 추출해낸다[30]. 또한 비디오를 분할하고 인덱싱하는데 영상처리, 음향 분석, 문자인식, 얼굴인식, 자연어 처리 기술 등 모든 관련 기술들을 적용하였다. Informedia 시스템은 기존의 시스템에서 제공하는 구조적인 브라우징, 자연어 질의, 음성 질의 등 다양한 인터페이스를 제공할 뿐 아니라 개요(skim)라는 영화의 예고편과 같은 개념의 인터페이스를 제공하는데, 개요는 각 에피소드 중 중요한 내용만을 발췌하여 이미지와 오디오를 함께 보여주는 기술이다[31]. 카네기멜론 대학에서 개발한 Informedia를 상품화 한 것이 ISLIP Media에서 개발한 MediaKey Digital Library System이다[32]. 이 제품은 실시간에 비디오 라이브러리를 만들고 검색까지 가능한 MediaKey Logger와 오프라인에서 동영상을 MPEG 데이터로 바꾼 후 인덱스를 만들어주는 MediaKey Builder, 그리고 브라우징, 탐색, 검색 등의 인터페이스를 제공하는 MediaKey Finder로 구성되어 있다. MediaKey Logger와 MediaKey Builder의 기능상의 큰 차이점은 Logger의 경우 에피소드의 구분을 사용자가 직접 해주어야 하나 Builder의 경우 음성인식과 자연어 처리 기술등을 사용하여 시스템이 자동으로 해주는데 있다. MediaKey Finder를 통해 간단한 키워드나 자연어 질의, 음성질의, 혹은 특정 이미지 자체를 질의로 사용할 수 있다. 이 제품은 통신이나 클라이언트/서버, 또는 웹 환경에서 사용할 수 있으며 800시간 분량의 비디오 라이브러리에서 원하는 데이터를 10초안에 찾을 수 있다.

IV. 결 론

본 고에서는 멀티미디어 정보의 효율적인 검색을 지원하기 위하여 새로 시작된 MPEG-7 국제표준화의 현재 진행 상황 및 이와 관련된 내용 기반 정보 검색 기술의 연구 동향을 살펴보았다. MPEG-7 표준화는 오디오비쥬얼 데이터가 갖고 있는 다양하고 풍부한 특징 및 내용을 어떻게 표준화된 방식으로 표현할 것인가에 주된 목표를 두고 있으며, 디지털 AV 신호 처리, 컴퓨터 비젼, 자연어 처리, 멀티미디어, 음성 인식 및 데이터베이스 분야를 포함한 다양한 영역의 지식들이 복합적으로 활용될 것으로 보인다.

MPEG-7에서 요구하는 멀티미디어 정보의 내용 표현 및 내용 기반 검색에는 두 가지 방향의 접근

및 해법이 제시될 것으로 예상된다. 하나는 상위 레벨 내용에 기반한 검색과 다른 하나는 그보다 하위 레벨 내용에 기반한 검색이다. 하위 레벨의 내용에 의한 검색은 사용자 질의가 부자연스러운 점은 있으나 시스템이 자동으로 검색할 수 있다는 장점이 있는 반면, 상위 레벨의 내용에 의한 검색은 좀더 자연스런 질의를 할 수 있으나 현재의 컴퓨터 기술로는 사용자의 개입을 필요로 한다는 단점이 있다. MPEG-7의 유용성 및 표준으로서의 성공은 이러한 두 가지 측면을 어떻게 적절히 수용하는가 하는데 크게 좌우될 것이다. 본 고에서는 이러한 관점에서 멀티미디어 데이터를 오디오 데이터, 정지 영상 데이터 및 동영상 데이터로 분류하여 현재까지 연구되거나 개발된 내용 표현 및 내용 기반 정보 검색 기법들에 관해서 설명하였다.

MPEG-7 표준화는 다른 MPEG 표준들과는 독립적으로 사용될 예정이지만 MPEG-7의 내용 표현 기술들은 과거 MPEG 표준안들의 기능을 향상하기 위해서 사용될 수도 있다. MPEG-7 기술의 응용 분야는 광범위하며 그 시장성 또한 매우 클 것으로 전망되므로, 국내의 좋은 연구 결과들을 적극적으로 제안하여 표준안에 채택되도록 함이 요구된다.

참 고 문 헌

[1] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Context and Objective", MPEG98/N2207, Tokyo, March 1998.

[2] Jinwoong Kim, Munchurl Kim, Kyu-Won Lee, Hankyu Lee, Jae-Gon Kim and Donghyun Kwon, "Suggestions on the improvements of Draft of MPEG-7 Requirements", MPEG98/M3404, Tokyo, March 1998.

[3] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Applications Document V.5", MPEG98/N2209, Tokyo, March 1998.

[4] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Requirements Document V.5", MPEG98/N2208, Tokyo, March 1998.

[5] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Evaluation Ad Hoc Group : Paris Meeting Report", MPEG98/M3785, Dublin, July 1998.

[6] Barry Arons, "SpeechSkimmer: Interactively Skimming Recorded Speech", Proceedings of UIST '93: ACM Symposium on User Interface Software and Technology, ACM Press, pp. 187-196, Nov 3-5, 1993.

[7] A. Ghias, J. Long, D. Chamberlin and B.C. Smith, "Query By Humming: Mugical Information Retrieval in An Audi o Database", ACM Multimedia 95 Proceedings, pp.231-236, November 1995.

[8] S. Subramanya, R. Simha, B. Narahari, A. Youssef, "Transform-Bassed Indexing of Audi o Data for Multimedia Databases", Proceedings of the International Conference on Multimedia Computing and System, pp. 211-218, June, 1997.

[9] T. Blum, D. Keislaer, J. Wheaton, and E. Wold, "Audi o Databases with Content-based Retrieval", 1995 Int'l Joint Conf. on Artificial Intelligence, 1995.

[10] G. J. F. Jones, J. T. Foote, K. Sparck Jones, and S. J. Young, "The Video Mail Retrieval Project: Experiences in Retrieving Spoken Documents", Intelligent Multimedia Information Retrieval, 1996, Editor: M. T. Maybury, Menlo Park CA: AAAI Press, Cambridge MA: MIT Press, pp.191-214, 1997.

[11] P. Schauble, M. Wechsler, "First Experiences with a System for Content Based Retrieval of Information from Speech Recordings", IJCAI-95 Workshop on Intelligent Multimedia Information Retrieval, Maybury, M. T., (chair), working notes, pp. 59-69, August, 1995.

[12] Kyoji Hirata, Toshikazu Kato, "Rough Sketch-Based Image Information Retrieval", NEC Res. & Develop., Vol. 34, No.2, 1993.

[13] 염성주, 김우생, "형태와 컬러성분을 이용한 내용 기반의 이미지 검색 데이타베이스 시스템", 한국정보처리학회 논문지, 제3권 4호, pp.733-744, July 1997.

[14] M.Flickner, H.Sawhney, et al, "Query by Image and Video Content: The QBIC System", IEEE Computer, Vol.28, No. 9, pp.23-31, September 1995.

[15] J. R. Smith and S.-F. Chang, "VisualSEEk: a fully automated content-based image query system", In Proc. ACM Intern. Conf. Multimedia, Boston MA, pp.87-98, November 1996, http: / / www.ctr.columbia.edu / VisualSEEk.

[16] 이동호, 송용준, 김형주, "SCARLET: 웨이브릿 변환을 이용한 내용기반 이미지 검색 시스템의 설계 및 구현", 한국정보과학회 논문지(C), 제3권 4호, pp.353-364, August 1997.

[17] Kai-Chieh Liang and C.-C. Jay Kuo, "WaveGuide: A Joint Wavelet Image Description and Representation System", http://biron.usc.edu/~kliang/research/waveguide.html, March 1998.

[18] A.Pentland, R.W.Picard, S.Schlaroff, "Photobook: Content-based manipulation of image databases", International Journal of Computer Vision, Vol.18, No.3, pp.233-254, 1996.

[19] 권성복, 조영우, 김영모, "구조화된 논리적 비디오를 이용한 비디오 브라우징 및 검색 시스템", 한국정보처리학회 논문지, 제4권 6호, pp.1443-1452, June 1997.

[20] 이충훈, 이홍규, "패턴인식을 이용한 MPEG 비디오 스트림상에서의 장면전환 검출", '98 봄 한국정보과학회 학술발표논문집, 제25권 1호, pp.619-621, April 1998.

[21] Shih-Fu Chang, William Chen, Horace J. Meng, Hari Sundaram, Di Zhong, "VideoQ: An Automated Content Based Video Search System Using Visual Cues", ACM MULTIMEDIA 97 The Fifth ACM International Multimedia Conference, Seattle, USA, 9-13, November 1997.

[22] R. Lienhart, "Automatic Text Recognition for Vido Indexing", ACM Multimedia 96, Boston MA USA, pp.11-20, November 1996.

[23] Boon-Look Yeo, Bede Liu, "Visual Content Highlighting via Automatic Extraction of Embedded Captions on MPEG Compressed Video", In Digital Vido Compression: Algorithm and Technologies Proc. SPIE, 2668-07, 1996.

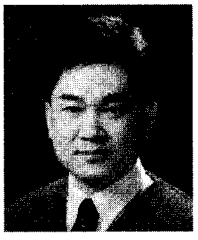
[24] Yuichi Nakamura, "Semantic Analysis for Video Contents Extraction", ACM Multimedia 97 Electronic Proceedings, pp.8-14, November 1997.

[25] 이미숙, 방건, 임영규, 홍영기, 김두식, 이성환, "내용 기반 색인 및 검색을 위한 실시간 뉴스 비디오 파서의 설계 및 구현", 한국정보과학회 봄 학술 발표논문집, Vol.24, No.1, April 1997.

[26] K. Minami, A. Akutsu, H. Hamada, Y. Tonomura, "Enhanced Vide o

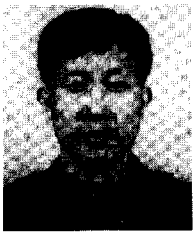
Handling based on Audio Analysis", Proceedings of the International Conference on Multimedia Computing and System, pp.219-226, June, 1997.
〔27〕 Y-L. Chang, W.Zeng, I.Kamel, and R. Alonso, "Integrated Image and Speech Analysis for Content-Based Video Indexing", 1996 International Conference on Multimedia Computing and Systems (Multimedia '96), 1996, www.ee.princeton.edu /~wzeng/ indexing.html.
〔28〕 A. Merlino, D. Morey, M. Maybury, "Broadcast News Navigation using Story Segmentation", ACM MULTIMEDIA 97 The Fifth ACM International Multimedia Conference, Seattle, USA, 9-13, November 1997.
〔29〕 A. Hauptmann, M. Smith, "Text, Speech, and Vision for Video Segmentation: The Informedia Project", AAAI Fall 1995 Symposium on Computational Models for Integrating Language and Vision, 1995.
〔30〕 A. and M. Witbrock, "Informedia News-On-Demand: Using Speech Recognition to Create a Digital Video Library", Working Notes for AAAI-97 Spring Symposium on Intelligent Integration and Use of Text, Image, Video and Audio Corpora, pp.24-26, March 1997.
〔31〕 M.A. Smith, Tanade Kanade, "Video Skimming and Characterization through the Combinatin of Image and Language Understanding Technique", CMU-CS-97-111, February 1997.
〔32〕 http://www.islip.com/

저 자 소 개



金 雨 生

1955년 10월 30일생, 1982년 서울대학교 수료, 1985년 Univ. of Texas at Austin 전산학 학사, 1987년 Univ. of Minnesota 전산학 석사, 1991년 Univ. of Minnesota 전산학 박사, 1987년 6월~1988년 7월 현대전자, Zeus Computer 과장, 1992년 3월~현재 광운대학교 부교수, 〈주관심 분야 : 멀티미디어, 화상처리 및 화상인식, 데이터베이스〉



金 鎭 雄

1959년 12월 23일생, 1981년 서울대학교 전자공학과 학사, 1983년 서울대학교 전자공학과 석사, 1993년 미국 Texas A & M 대학교 전기전자공학 박사, 1983년 3월~현재, 한국전자통신연구소 책임연구원, 〈주관심 분야 : 디지털 신호처리, 영상통신, 디지털 라이브러리, VLSI 신호처리〉

林 文 哲

1968년 10월 29일생, 1994년 순천대학교 전자계산학과 학사, 1996년 광운대학교 전자계산학과 석사, 1996년~현재 광운대학교 전자계산학과 박사과정, 1997년 11월~현재 광운대학교 전자계산학과 조교, 〈주관심 분야 : 멀티미디어, 화상처리 및 화상인식, 패턴인식〉