

내용 기반 비디오 인덱싱 및 검색 기술

김진태* · 김동욱**

1. 서 론

반도체 및 통신장비 기술의 발달과 더불어 사용자 세대의 변화에 따라 정보기기가 다루는 정보의 형태가 기존의 문자나 음성 위주에서 벗어나 영상이나 비디오, 오디오 등을 중심으로 한 종합적인 멀티미디어 형태로 전환되고 있다. 이러한 멀티미디어 산업의 급부상은 정보통신 분야의 핵심으로 자리매김하면서 현재 이를 위한 많은 연구가 국내외적으로 이루어지고 있으며 적용 분야도 점차 확대되는 추세이다.

즉, 지금까지는 방대한 양의 정보를 컴퓨터나 통신망, 단말기 등과 같은 정보기기에 어떻게 담을 것인가에 초점이 맞추어짐에 따라 정보의 전송 및 저장을 위한 연구가 활발히 논의되어 왔다. 그러나 그에 대한 문제가 많이 해결되고 또 상용화되고 있는 현 상황에서는 정보를 담는 그릇의 문제가 아니라 무엇을 담을 것인가라는 정보의 내용이 향후 연구 개발에 있어 승패의 핵심인 것이다.

전자의 경우 디지털 동영상 및 오디오 압축 부호화에 대한 국제 표준 활동을 연구 진행하여 현재 MPEG-1, MPEG-2 표준 기술을 제정하였고 MPEG-4에 대한 표준 작업이 마무리 단계에 있다. 후자의 경우 지금까지의 압축 부호화 개념으

로부터 벗어나 효율적인 데이터의 저장 및 관리, 검색 등을 위해 정보의 내용물을 갖고 대상을 표현하고자 하는 MPEG-7에 대한 표준 작업이 곧 진행될 예정이다.

<표 1>은 MPEG 표준 작업의 진행 과정 및 향후의 계획을 단계별로 정리한 것이다.

<그림 1>은 곧이어 국제 표준 제정을 위해 착수할 연구 대상인 MPEG-7의 구성 요소 및 범주를 도식한 것이다[1]. 즉, 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 제시해 주는 부분으로 구성된다.

2. 비디오 정보처리

인간의 정보 취득에 중요한 역할을 담당하는 시각 정보는 정보 전달 효과가 음성이나 문자에 비해 뛰어나기 때문에 영상 정보에 대한 수요는 날이 갈수록 급증하고 있다. 이는 멀티미디어 정보 가운데서도 특히 영상 정보가 통신, 교육, 오락 및 출판 등 우리 생활의 많은 부분에서 필수적인 요소로 등장하고 있음을 보더라도 확연히 알 수 있다.

*한서대학교 컴퓨터학과

**전주대학교 전기전자정보통신공학부

표 1 MPEG 표준 관련 일정표

	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	2000	2001
MPEG-1	Start			CD/DIS		IS								
MPEG-2		Start				CD	DIS/IS							
MPEG-4						Start		Proposal /Test	WD	CD	DIS/IS			
MPEG-7											Proposal	WD	CD	DIS /IS

* WD(Working Draft), CD(Committee Draft),
DIS(Draft of International Standard), IS(International Standard)

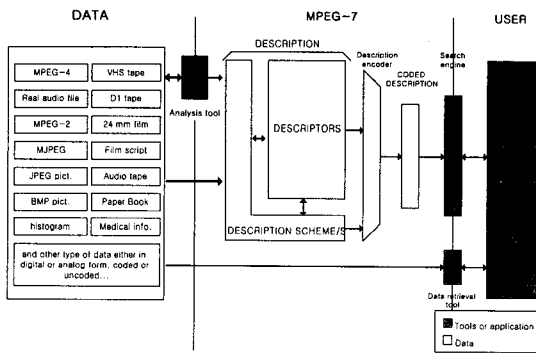


그림 1. MPEG-7의 구성 요소 및 범주

국제적인 한 통계자료[2]에 의하면 현재 만들어진 비디오 필름의 양은 약 6백만 시간 분량이 넘으며, 해마다 약 10%이상의 증가를 거듭하고 있다. 이는 디지털화 될 경우 1.8 millionGB의 MPEG 부호화된 비디오 데이터량에 해당한다. 따라서 데이터의 내용에 기반한 대상 표현을 위해 효율적인 정보 처리 체계 없이는 이러한 엄청난 양의 비디오 데이터의 증가에 효과적으로 대처할 수가 없다.

과거의 비슷한 예로서 현재 일반적인 문서의 정형화된 표현 수단으로 색인 및 검색에 널리 이용되는 문자 위주의 데이터베이스 관리 시스템(DBMS : Database Management System)의 필

요성이 대두되었던 것과 일맥상통한다고 볼 수 있다. 그러나 기존 DBMS의 경우 대상 자체가 비디오 데이터가 아닌 단순히 구조화된 형태의 텍스트 데이터 타입을 위해 설계되었는데, 비디오 데이터의 경우 그 특성이 기본적으로 기존의 문자 위주와는 다르고, 데이터 모델링, 삽입(insertion), 특징 추출 및 인덱싱(indexing) 등 여러 가지 측면에서 새로운 처리 기법을 요한다.

예를 들어, 검색에 있어서 기존 DBMS의 경우 비교 데이터간의 '일치' 혹은 '불일치'의 관계만을 고려한 정확한 매칭을 요구하는 반면, 비디오 데이터의 경우 유사성 기반(similarity-based)의 검색 알고리즘을 요한다. 한편 그래픽 사용자 인터페이스(GUI) 설계 및 비디오 데이터 브라우징은 기존의 DBMS보다 내부적으로는 처리 과정이 더욱 복잡해진 반면 성능에 있어서는 문자 이외의 비디오 데이터 내용을 대상으로 표현 및 검색을 수행하므로 보다 유연하고 융통적인 결과를 제시할 수 있으므로 그 기능 및 역할이 더욱 중요하게 되었고, 훨씬 더 편리함을 사용자에게 제공할 수 있다.

비디오 데이터의 처리 및 구성은 데이터의 특성상 대용량의 저장 공간과 빠른 처리 속도를 요

구하므로 매우 복잡하고 어려운 문제이다. 이는 비디오가 시공간적 구조로 구성된 비정형화된 형태, 즉 숫자나 문자 형식이 아닌(non-alphanumeric) 데이터이기 때문에 기인한다. 데이터의 유형에 따른 이와 같은 여러 가지 특성을 <표 2>에 나타내었다.

표 2 데이터 유형에 따른 특성 비교

characteristic type	Textual Data	Video Data
Resolution	Low	High
Information	Poor	Rich
Interpretation Ambiguity	Low	High
Interpretation Efforts	Low	High
Data Volume	Small	Large
Similarity Measure	Well defined	Ill defined
Symbol Set	Finite	Infinite

<표 2>에서 보듯이 정보 전달 측면에서는 텍스트 형식에 비해 비디오 데이터가 시각적 특성이 크기 때문에 보다 효율적이다. 그러나 구성 형식에 있어서는 기존 텍스트 데이터 형식이 '주제-본문'이라는 단어의 체계적인 나열로 이루어진 반면, 비디오 데이터는 비정형화된 형태 즉, '상황(scenario)에 따른 복합적인 장면'으로 이루어져 있으므로 데이터의 내용을 모델링하고 표현하는데 있어 훨씬 복잡한 처리 과정을 거친다. 또한 데이터간에 관계 연산은 텍스트 데이터인 경우 주어진 단어가 '같다' 혹은 '같지 않다'라는 일치성 여부로 판별을 하므로 판단 과정이 단순하고 명확한 반면, 비디오 데이터의 경우에는 주어진 정보의 내용이 얼마만큼 서로 유사한가에 비취 판별을 하므로 유연하고 융통적인 결과를 제시할 수 있으나 판정에 대한 정의가 명확하지가 않다.

현재 비디오 데이터의 증가 추세로 볼 때 비디오 데이터 내용에 대한 특징 정보를 얼마만큼 잘 추출하고 이를 체계적으로 정리하여 표현할 것인가는 멀티미디어 정보화 사회에 있어 매우 심각하고도 중요한 사안이다.

비디오 데이터베이스 구축 및 검색 시스템은 대용량의 비디오 데이터를 다루기 때문에 비디오 파싱, 압축 등을 비롯해서 그래픽 사용자 인터페이스를 위한 브라우징 등에 이르기까지 많은 기술 요소들을 필요로 한다. 그러나 현재 이러한 분야에 대해 세계 각국의 연구 단계는 비디오 분할 등에 관한 기술을 어느 정도 확보하였을 뿐 인덱싱 및 검색 등에 있어서는 매우 미약한 연구 결과를 얻고 있으며, 국내의 경우에도 일부 대학을 중심으로 기초적인 초기 단계의 연구가 진행되고 있으나 결과는 미미한 실정이다.

21세기는 '정보화 사회'라고 일컬을 정도로 비디오 데이터를 중심으로 한 멀티미디어 분야는 매우 중요한 위치를 차지한다. 우리 정부도 초고속 정보통신망 구축 계획을 범국가적인 차원에서 추진하여 서비스망을 구축, 성공적으로 시행하고 있으나 실제 서비스망에서 제공되는 비디오 데이터를 포함한 멀티미디어 전반에 대한 처리 부분은 매우 취약한 것이 현실이다. 따라서 날마다 기하급수적으로 늘어나고 있는 비디오 데이터를 효율적으로 관리하고 처리하기 위한 데이터베이스 구축 및 검색 기술을 개발하는 것은 매우 중요한 일이라 판단된다.

3. 비디오 인덱싱

(1) 비디오 데이터의 모델링

검색을 위한 효과적인 데이터베이스의 조직 및 운영을 위해서 고려해야 할 요소들은 다음과 같다. 첫째, 다단계의 비디오 구조를 갖추어야 한다.

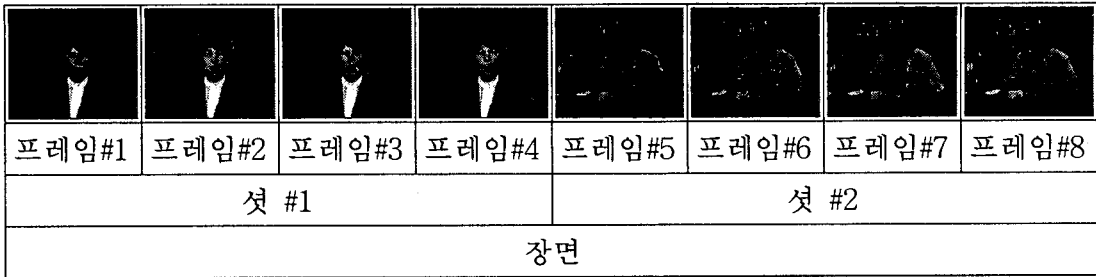


그림 2. 비디오 데이터의 구성

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 여러 가지 서로 다른 경우의 비디오 검색시에도 참조 대상으로 사용될 수 있어야 한다.

예를 들어, 비디오는 <그림 2>와 같이 프레임(frame), 셋(shot), 장면(scene)의 세가지 구성 요소로 이루어진다. 프레임이란 비디오를 구성하는 최소 단위로 필름 한 장에 해당하는 하나의 개별 영상을 나타낸다. 셋은 하나의 카메라로 촬영한 영상을 나타내며, 셋 내에서는 필름이 끊기지 않고 연속적으로 연결되어 있다. 장면은 연속된 일련의 셋으로 구성되며, 주인공이나 특정한 장소와 같이 하나의 대상을 연속하여 촬영한 영상을 나타낸다. 비디오 분할 기술은 장면 전환 효과에 의해 발생하는 셋 사이의 경계를 검출하는 방법으로, 셋은 비디오 분할의 기본 단위로 사용된다.

<그림 3>은 각 프레임에 있어서 물체 개별 모델링을 위해 영상내 분포한 주파수 성분을 이용하여 다해상도로 분해시키는 웨이브릿 변환(Wavelet Transform)을 이용하여 하나의 프레임을 표현한 예이다. 이 외에도 여러 가지 다양한 기법들을 이용하여 비디오 데이터의 모델링에 적용시킬 수 있다.

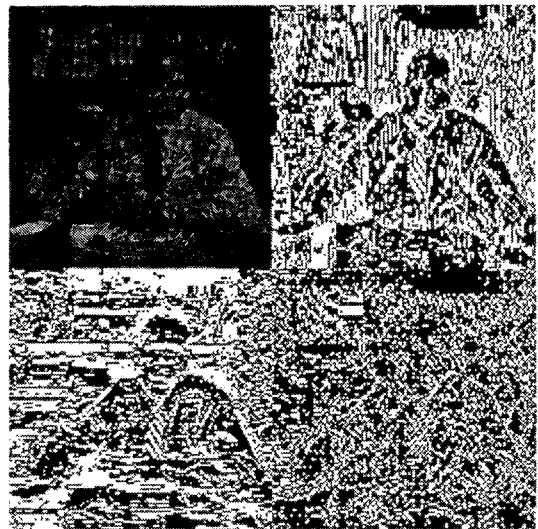


그림 3. 영상의 다해상도 분해

(2) 비디오 데이터의 특징 추출

비디오 데이터의 처리를 위한 특징으로 색상, 텍스춰, 형태, 위치 정보 등을 들 수 있다. 칼라

특징은 색상에 대한 히스토그램을 구하고 이를 적용할 수 있다. 칼라 히스토그램 분포는 영상내 조명 변화나 관측 위치, 크기 변화나 회전 등에 무관하게 적용할 수 있다는 장점을 갖고 있다. 그러나 현재 제시된 기법상의 문제점은 화면 전체에 대한 칼라 정보만을 고려한다는 점이다. 이런 경우 영상내의 국부적인 특징 즉, 각 객체에 대한 고려가 없기 때문에 한정된 범위의 적용만이 가능하다.

텍스춰 영상의 경우 통계적인 특징, 구조적인 특징, 스펙트럼 특징 등을 고려할 수 있으며, 이러한 특징들은 각기 개별적인 처리 및 적용보다는 서로 융합되고 보완적인 형태의 적용이 바람직하다.

<그림 4>는 텍스춰 영상의 예를 나타낸 것이고, <그림 5>는 이들 텍스춰 영상이 갖고 있는

일정한 패턴 형식을 이웃한 화소들간의 관계로부터 추출한 것으로 공간 도메인상에서의 인접 화소간의 미분값의 크기와 위상 성분을 히스토그램으로 정량화시켜 그 특징을 나타내 보인 것이다.

<그림 6>은 영상내 객체들의 형태 정보를 추출하기 위해 현재 영상처리에서 많이 사용하고 있는 소벨 연산자(Soble operator)를 이용하여 영상에 대한 윤곽선을 나타낸 예이다. 그러나 이와 같은 형태 정보를 특징으로 이용하기 위해서는 먼저 영상내에 여러개의 객체나 복잡한 배경이 있을 경우 영상으로부터 객체를 분리하는 작업이 선행되어야 한다.

한편, 비디오 데이터의 경우 가장 효과적인 특

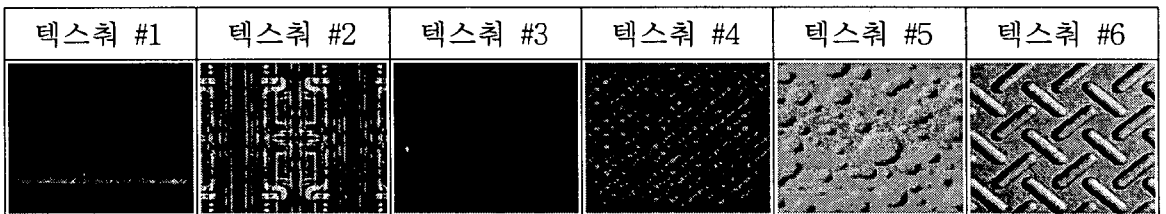


그림 4. 텍스춰 영상의 예

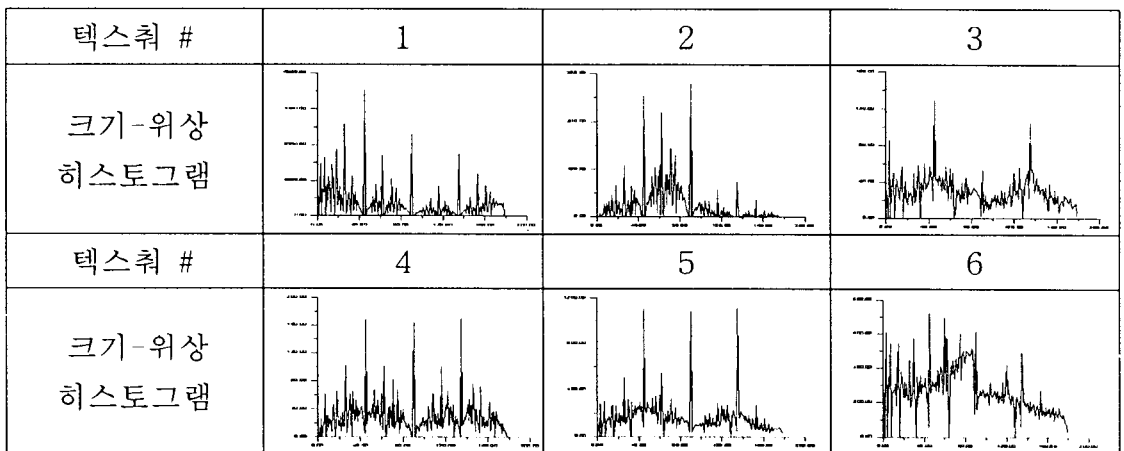


그림 5. 공간 도메인에서의 인접 화소간의 미분값의 크기-위상 히스토그램 분포도



그림 6. 영상의 윤곽선 형태

징 추출은 위치 정보이다. 위치 정보는 압축된 데이터의 경우 복호화를 통해 움직임 벡터를 얻을 수 있으며, 2차원적 위치 정보에 대한 시간적 관계를 고려하여 3차원적 이동 관계를 얻음으로써 더욱 효과적인 특징추출이 가능하리라 본다. 가령 <그림 7>에서와 같이 영상내 객체들의 위치 정보에 따라 그래프적인 특징을 이용할 수 있으며, 연속된 일련의 비디오인 경우에 시간축상으로 확대하여 움직임 궤적을 추정함으로써 이를 객체의 위치 정보로 활용할 수 있다.

(3) 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 경우 무엇을

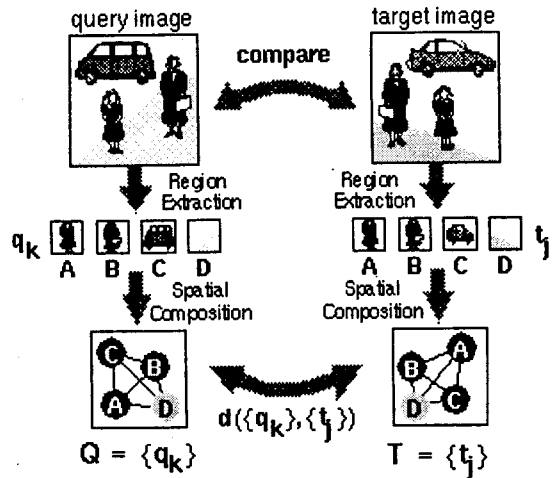


그림 7 위치 정보를 고려한 영상 특징 추출

특징 정보로 사용할 것인지에 대한 구분부터 불 명확하다.

현재 연구되고 있는 비디오 인덱싱 기법으로는 특징 정보를 어떻게 유도하는가에 따라 주석 기반 인덱싱(annotation-based indexing), 특징 기반 인덱싱(feature-based indexing), 그리고 영역 국한 인덱싱(domain-specific indexing) 등으로 분류할 수 있다.

첫째, 주석 기반 인덱싱인 경우 비용 효율면에서 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.

4. 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 정보

는 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.

현재 정지 영상을 대상으로 하는 내용 기반 영상 검색 시스템이 인터넷을 통하여 시범적으로 서비스가 제공되고 있으며, 아래의 <표 3>은 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 서로 간의 특징 정보를 비교하여 유사도가 높은 순서대로 결과를 제공하도록 되어 있다. 또한 시간적 특성을 고려한 비디오 데이터의 경우에는 서비스 내용이 배제되어 있는 상황이다.

(2) 카메라 동작 모델링

내용 기반 비디오 인덱싱 및 검색에서 텍스트 이외에 사용자가 주로 이용하는 형태는 대표 프레임으로 표현되는 영상이며, 이러한 영상들은 영상내의 색상, 질감, 형태 정보, 위치 정보 등을 특징값으로 사용한다. 그러나 대표 프레임만으로는 비디오의 시간적 특성을 나타낼 수 없기 때문에 카메라의 동작 분석, 이동 물체의 움직임 분석 등을 통하여 비디오의 시간적인 특성을 표현하고자 하는 연구가 현재 진행중이다.

비디오에서 가장 일반적으로 발생하는 카메라 동작은 상하, 좌우, 전후 화살표로 표시한 것과 같이 6가지로 크게 분류할 수 있다[10]. 이 중에서 틸팅(tilting), 팬닝(panning), 줌링(zooming)은 카메라의 몸체가 고정된 상태에서 렌즈가 각각 상

표 3 내용 기반 검색의 응용 사례

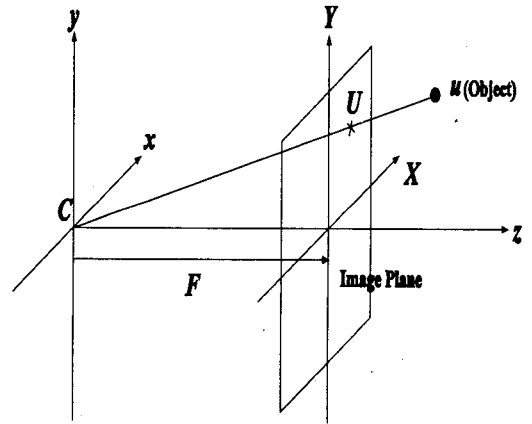
개발 기관	시스템 명칭	웹사이트 도메인 주소
IBM	QBIC	http://www.qbic.almaden.ibm.com
Virage	Virage	http://www.virage.com
Informix Software	Visual Intelligence	http://illustra.com
Interpix Software	Image Surfer	http://isurf.interpix.com
Columbia 대학	ADVENT	http://www.ctr.columbia.edu/advent
Chicago 대학	WebSeer	http://infolab.cs.uchicago.edu/webseer

하, 좌우, 전후 위치로 움직이는 상태에서 발생하는 동작이고, 붐밍(booming), 트래킹(tracking), 돌링(dolly)은 각각 카메라의 몸체가 상하, 좌우, 전후로 함께 이동하는 상태에서 발생하는 동작이다. 그러나 비디오 상에서 이 6가지 카메라 동작을 모두 정확하게 구별하는 것은 매우 어려운 일이다.

일반적인 카메라 시스템의 수학적 모델링은 <그림 8>의 파라미터 관계로부터 유도할 수 있다. 예를 들어 줌은 카메라의 위치는 변하지 않고 초점거리(F)가 변함으로써, 영상면 위에 맺히는 상의 위치가 변하는 경우이다. 줌에 의해 영상면 위에 맺힌 대상체의 위치가 U에서 U'로 변했다면 다음과 같은 식으로 표현할 수 있다.

$$U' = \frac{F'}{F} \times U, \quad f = \frac{F'}{F}, \quad U' = f U$$

위 식으로부터 영상의 화소 변화량을 계산하여 히스토그램 분포로 나타내면 연속된 일련의 비디오 데이터 중에서 줌이 발생한 경우의 프레임에 대해 그 분포값이 달라 나타나는 것을 알 수 있고, 이로부터 카메라 이동에 의한 점진적인 장면 변화를 검출할 수 있다.



- 여기서,
 C : 카메라의 중심
 (x,y,z) : 영상면(Image Plane)의 좌표
 u : 대상체의 카메라 좌표상에서의 위치
 U : 대상체의 영상면(Image Plane) 상에서의 위치
 F : 초점거리

그림 8. 카메라 시스템의 수학적 모델링

(3) 정보 검색 기술

멀티미디어 정보의 검색에 있어 사용되는 방법으로 다음과 같은 텍스트 기반 검색과 내용 기반 검색의 두가지 방식으로 크게 분류할 수 있다.

첫 번째 방식으로 검색의 대상이 되는 모든 멀티미디어 데이터에 사람이 직접 색인을 첨가하

고, 사용자 또한 주제어를 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 분야들의 연구 결과를 복합적으로 이용하여 멀티미디어 데이터로부터 보다 정확한 특징 정보를 추출하려는 연구가 진행 중이다.

위의 두가지 방식 중 현재 주로 연구되고 있는 내용 기반 검색 기술은 대상에 따라 정지 영상 검색과 비디오 검색의 두가지로 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 세

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[3][4]. 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, 이동 벡터 등을 특징값으로 처리하며, 히스토그램 기법이 가장 많이 사용된다[5].

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[6].

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 여러 가지 제시되고 있다. DCT 계수를 이용하는 방법[7], 움직임 벡터를 이용하는 방법[8], 이 두가지를 혼용하는 방법[9] 등이 제시되었다.

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.

<그림 9>는 대상에 따른 검색 기법을 도해적으로 나타낸 것이다.

비디오 데이터는 비정형적인 구조와 시공간적인 특성을 가지고 있기 때문에 정확한 비디오 분할 및 인덱싱 등이 매우 어려우며, 따라서 효과적인 처리 방법에 대한 연구가 절실히 요구된다.

5. 결론

인터넷과 월드와이드웹(WWW), 그리고 초고속 정보 통신망 등은 우리가 살고 있는 현실 사회

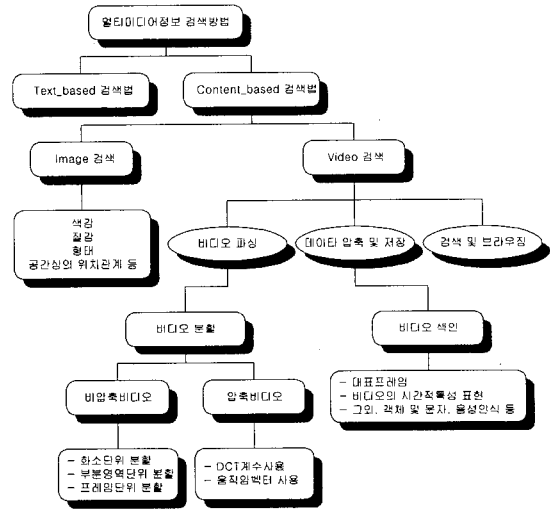


그림 9. 멀티미디어 정보 검색 방법

속에 정보화가 깊숙히 자리잡고 있다는 것을 실감하게 한다. 정보 사회의 가장 큰 이슈는 역시 정보이며 날마다 기하급수적으로 늘고 있는 대용량의 멀티미디어 데이터를 효과적으로 처리할 수 있는 기반 기술은 매우 중요하다.

내용 기반 비디오 데이터 인덱싱 및 검색 기술은 다가오는 21세기의 필수적인 기술임과 동시에 현재 국내적으로는 매우 취약한 분야이기도 하다. 따라서, 이러한 분야에 대한 연구 결과는 향후 디지털 라이브러리, 방송 프로그램 관리, 의료 관리 시스템, 교육용 시스템, 주문형 비디오(VOD : Video on Demand), 홈 쇼핑, 웹 캐스팅 등의 각종 응용 부문에서 질 높은 서비스 구축을 위한 기반 기술이 될 수 있으며, 급속한 컴퓨터 하드웨어의 보급 및 기술 발전을 바탕으로 소프트웨어화를 통한 상품화도 기대 가능한 기술이다. 또한 현재 웹에서 통용되고 있는 텍스트 위주의 검색 시스템을 비디오 검색 시스템으로 교체할 경우 그 대안으로 충분히 활용할 수 있으리라 판단된다.

참 고 문 헌

[1] <http://drogo.cselt.stet.it/mpeg/standards/mpeg-7.htm>

[2] A. Hampapur, "Design video data management systems," Ph. D thesis, The University of Michigan, 1994.

[3] H. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," Multimedia Systems, vol. 1, no. 1, pp. 10-28, 1993.

[4] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks," Proc. of ACM Multimedia, pp. 189-200, 1995.

[5] D. Swanberg, C. F. Shu, and R. Jain, "Knowledge guided parsing in video databases," Proc. of SPIE-Storage and Retrieval for Image and Video Databases, vol. 1908, pp. 13-24, 1993.

[6] H. Zhang and S. W. Smoliar, "Developing Power Tools for Video Indexing and Retrieval," Proc. of SPIE - Storage and Retrieval for Image and Video DatabaseII, vol. 2185, pp. 140-149, 1994.

[7] F. Arman, A. Hsu, and M. Y. Chiu, "Feature management for large video databases," Proc. of SPIE - Storage and Retrieval for Image and Video Databases, vol. 1908, pp. 2-12, 1993.

[8] H. Zhang, C. Y. Low, Y. Gong, and S. W. Smoliar, "Video parsing using compressed data," Proc. of SPIE - Image and video Parsing II, vol. 2182, pp. 142-149, 1994.

[9] N. Dimitrova and M. Abdel-Mottaleb, "Content-based video retrieval by example video clip," Proc. of SPIE-Storage and Retrieval for Image and Video Databases V, vol. 3022, pp. 59-70, 1997.

[10] A. Akutsu, Y. Tonomura, H. Hashimoto, and Y. Ohba, "Video indexing using motion vectors," Proc. of SPIE-Visual Communications and Image Processing, vol. 1818, pp. 1522-1530, 1992.



김진태

- 1987년 중앙대학교 전자공학과 졸업
- 1989년 중앙대학교 대학원 전자공학과 졸업(공학석사)
- 1993년 중앙대학교 대학원 전자공학과 졸업(공학박사)
- 1995년~현재 한서대학교 컴퓨터학과 조교수
- 주관심분야 : 영상압축, MPEG, 비디오 인덱싱 및 검색



김동욱

- 1987년 성균관대학교 전자공학과 졸업(공학사)
- 1992년 중앙대학교 대학원 전자공학과 졸업(공학석사)
- 1996년 중앙대학교 대학원 전자공학과 졸업(공학박사)
- 1998년~현재 전주대학교 전기전자정보통신공학부 조교수
- 주관심분야 : 통신신호처리, 영상부호화, 비디오 인덱싱 및 검색