

□ 기술애설 □

내용 기반 영상 및 비디오 검색 기술의 연구 현황

고려대학교 이미숙*·황본우·이성환**

1. 서 론

최근 컴퓨터와 통신 기술의 발달과 더불어 영상 및 비디오, 오디오 등을 중심으로 한 멀티미디어 정보 서비스에 대한 요구가 증가하고 있으며, 이를 제공하기 위한 멀티미디어 검색 기술에 대한 연구가 활발하게 진행되고 있다. 현재 보편적으로 사용되고 있는 멀티미디어 정보 검색 방법은 다음과 같이 두 가지로 분류될 수 있다.

첫번째 방법은 검색의 대상이 되는 모든 멀티미디어 데이터에 사람이 직접 색인을 첨가하고, 사용자 또한 주제어를 이용하여 원하는 정보를 검색하는 텍스트 기반 검색 방식이다. 이 방법은 비정형적인 멀티미디어 데이터에 사람이 직접 의미 정보를 기술하기 때문에 제한된 범위 내에서 효율적인 검색이 가능하다는 장점을 가지고 있으나, 대용량의 데이터에 대하여 사람이 일일이 색인을 첨가해야 하기 때문에 시간 비용이 많이 들며, 색인을 첨가하는 사람과 검색하는 사용자의 관점이 불일치할 경우 검색의 효율성이 크게 떨어지게 된다. 또한 멀티미디어 데이터가 가지는 복잡한 속성을 텍스트만으로는 정확하게 표현할 수 없다는 단점을 가지고 있다[1].

두번째 방법은 멀티미디어 데이터의 내용을 대표할 수 있는 특징을 추출하여, 이를 기반으로 색인과 검색을 수행하는 내용 기반 검색 방법이다. 이 방법은 멀티미디어 데이터로부터 특징을 자동으로 추출하여 색인 과정에 사용함

으로써 데이터베이스 구축에 필요한 시간 및 인력의 소모를 줄였다는 장점을 가지고 있으나, 멀티미디어 데이터로부터 정확한 내용을 추출하기가 어렵다는 단점을 가지고 있다. 그러나 최근에는 컴퓨터 비전이나 영상 처리, 컴퓨터 그래픽스와 같은 분야들의 연구 결과를 이용하여 멀티미디어 데이터로부터 보다 정확한 정보를 추출하려는 연구가 진행되고 있다.

현재까지 주로 연구되고 있는 내용 기반 검색 기술은 대상 데이터의 종류에 따라서 영상 검색과 비디오 검색의 두 가지로 나누어진다. 내용 기반 영상 검색 기술은 정지 영상을 대상으로 하는 검색 기술로서, 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 등이 영상의 주요 특징으로 사용된다[2]. 내용 기반 비디오 검색 기술은 대상으로 하고 있는 비디오 데이터가 영상이나 텍스트가 지닌 일반적인 속성 이외에 시간적 속성을 가지고 있으며 비정형화된 구조를 하고 있기 때문에 복잡한 처리 기술이 필요하다.

2. 내용 기반 영상 검색

영상을 사용한 정보 검색에 대한 시도는 이미 1980년대부터 시작되었다[3]. 영상을 사용한 초기의 검색 방법은 텍스트 정보를 이용한 검색 방법이었다. 그러나 텍스트 중심의 검색 방법은 멀티미디어 자료를 효율적으로 검색하는데 제약이 있고, 검색의 사전 작업으로 필요한 색인 작업에 있어서 장기간 숙달된 전문 작업자를 요구하므로 시간과 비용면에서 비효율적이다. 또한 현재는 멀티미디어 시대로 의학

* 학생회원

** 종신회원

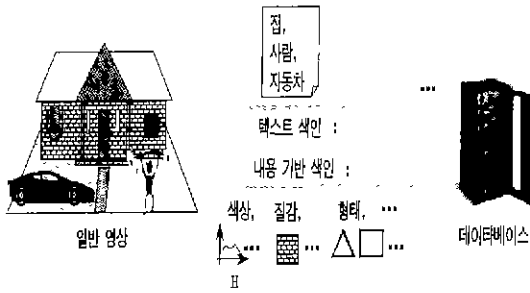


그림 1 일반 영상에 대한 색인 정보의 예

영상, 위성 사진, 과학 실험, 전자 문서 등 다양한 응용에서 발생하는 영상 데이터의 양이 급속도로 증가되고 있다. 이처럼 많은 영상 데이터들을 대상으로 사용자가 원하는 영상을 빠르고 정확하게 검색하기 위해서는 내용 기반 영상 검색 기술에 대한 연구가 필수적으로 요구된다. 일반적으로 사용되고 있는 영상의 내용 기반 특징으로는 색상, 질감, 형태와 공간상의 위치 관계 등이 있다. 예를 들어 그림 1에 대한 기존의 텍스트 중심의 색인은 자동차, 사람, 집, 정원 등의 사물을 지칭하거나, 주차, 운동 등의 간단한 추상 명사가 될 수 있다. 반면 이에 대한 내용 기반 색인은 텍스트 색인 정보로 표현하기 힘든 색상(붉은 색, 녹색, 검은 색 등), 질감(벽돌, 잔디, 금속 등), 형태(원형, 삼각형, 사각형, 사다리꼴 등)와 공간상의 위치 관계(집 앞의 왼쪽에 자동차, 오른쪽에 사람이 있다)에 대한 정보를 표현할 수 있다.

2.1 색상 정보를 사용한 검색 방법

영상의 색상 정보를 사용한 검색 방법은 내용 기반 영상 검색의 초기부터 사용되어 온 방법으로서, 영상 내 조명 변화나 관측 위치, 크기 변화 등에 어느 정도 무관하게 적용 가능하다는 장점을 갖고 있다. 색상 정보를 사용한 검색 방법은 주로 영상의 색상 히스토그램을 사용하며, 특징 추출 영역에 따라서 전역적 방법과 지역적 방법으로 나눌 수 있다. 전역적인 색상 히스토그램을 사용할 경우에는 영상의 전체적인 색상 분포만으로 유사성을 계산하기 때문에 영상내 객체들의 위치 정보는 전혀 고려하지 않는다는 단점을 갖고 있다[3, 4].

색상 정보를 사용한 검색 방법의 경우에는 색상 공간 상에서의 거리 측정을 정확하게 하기 위하여 일반적인 색상 공간인 RGB 모델을 사용하는 대신, HSV, YIQ, YUV 등과 같은 모델을 사용한다. 이 중에서도 특히 HSV 모델은 인간의 시각 능력에 유사한 색상 모델로서, 영상의 색상 정보 추출에 많이 사용된다[3].

2.2 질감 정보를 사용한 검색 방법

벽지나 옷감, 곡물, 대리석 영상 등과 같이 특정한 질감 특성을 갖는 영상을 대상으로 하는 검색에서는 영상의 색상 정보만으로는 정확한 검색 결과를 기대할 수 없다. 따라서, 영상의 질감 정보를 추출하여 검색에 사용하는 질감 정보 기반의 검색 방법이 필요하다. 일반적으로 영상 검색에 사용되는 질감은 통계적인 특징, 구조적인 특징과 스펙트럼 특징을 이용하여 표현된다. 통계적인 특징은 영상의 거친 정도, 부드러운 정도를 나타내며, 구조적인 특징은 영상 내의 수평선의 반복과 같은 규칙적인 배열을 표현한다. 스펙트럼 특징은 주파수 스펙트럼의 형태를 기반으로 스펙트럼의 높은 에너지 성분을 분석함으로써 영상 내의 전체적인 주기성을 알아내는데 사용된다. 최근 Manjunath와 Ma[5]는 Gabor 웨이블릿을 사용하여 영상의 질감 특성을 표현하고, 이를 기반으로 하는 영상 검색 방법을 제안하였다. 또한 Pentland 등[6]은 Wold 분해법을 사용한 질감 영상 검색 방법을 제안하였다. 이 방법에서는 영상의 질감 특성을 주기적(periodic), 방향적(evanescent), 복합적(complex) 질감의 세 종류로 분류하고, 입력 영상이 지닌 질감의 방향성을 기준으로 질감 특성을 분석하고 있다.

질감 정보를 사용한 검색 방법은 주로 일반 자연 영상에 적용되기보다는 영상 전체가 몇 개의 두드러진 질감 특징으로 구별되는 영상이나 전체가 동일한 질감 특징을 갖는 영상, 예를 들어 벽지 무늬나 옷감 등을 찍은 사진 영상 등을 대상으로 연구되고 있다.

2.3 형태 정보를 사용한 검색 방법

영상이 가지는 형태 특성은 크게 다음과 같이 두 가지로 분류될 수 있다. 첫번째는 영상

내 물체들의 윤곽선이 나타내는 형태로서, 형태 정보의 추출이 간단한 반면, 동일한 물체에 대하여 보는 각도에 따라 서로 다른 형태의 윤곽선이 나타날 수 있다. 두번째는 방향에 상관 없이 물체들이 지니는 고유한 형태로서, 보는 각도, 크기, 위치 변화 등에 무관한 형태 정보를 갖는 반면, 형태 추출과 검색 과정이 복잡하다는 단점을 지니고 있다[3]. 따라서 형태 정보를 사용하여 영상을 검색하는 경우에는 우선 영상의 특성에 따라서 위의 두 가지 형태중에 어느 것을 사용할 것인지 결정해야 한다. 또한 영상 내에 여러개의 물체나 복잡한 배경이 있을 경우에는 영상으로부터 객체를 분리하는 작업이 선행되어야 한다.

Pentland 등[7]은 유한 요소법(Finite element method)을 사용하여 물체의 형태 정보를 추출하고, KL(Karhunen-Loeve) 변환을 사용하여 형태간 유사도를 계산함으로써, 보는 각도와 위치 변화에 무관한 영상 검색 방법을 제안하였다. 제안된 방법은 배경이 없고 단일 물체만을 포함하고 있는 영상을 대상으로 실험한 결과 우수한 결과를 나타내었다. Scassellati [8]는 대수적 모멘트, 곡선 거리, 회전각의 누적값, 곡률을 이용한 방법 등과 같은 다양한 종류의 윤곽선 형태를 사용한 영상 검색 방법들을 분석하고 각 방법의 성능을 대용량 영상 데이터베이스 상에서 평가하였다.

형태 정보를 사용한 검색 방법은 다른 검색 방법에 비하여 사용자에게 고수준의 검색 환경을 제공할 수 있으나, 실제계에 존재하는 다양한 물체의 형태를 표현할 수 있는 수학적 모델이 없으며, 형태간 유사도를 측정할 수 있는 기준이 부족하다는 문제점을 가지고 있다. 실제로, 이러한 형태에 의한 유사성 검사는 모델 기반 비전 응용에서도 아직까지 어려운 분야로, 최근에는 많은 통계학적인 방법을 이용한 효율적인 방법들을 연구하고 있다.

2.4 위치 정보를 사용한 검색 방법

영상내 객체들의 위치 정보를 사용한 검색 방법은 색상이나 형태, 질감 정보를 사용한 검색 방법에 비하여 색인에 필요한 처리 과정이 복잡하기 때문에 현재까지는 많은 연구가 이루어

어지고 있지 못하다. 일반적으로 영상이 갖는 위치 정보는 다음과 같이 두 가지로 분류할 수 있다[3].

첫번째는 영상 내 객체들이 갖는 위치로서, 이러한 정보를 추출하기 위해서는 우선 정확한 객체 분할이 선행되어야 한다. 이는 한 영상안에 있는 객체들간의 공간적(spatial), 위상적(topological) 관계에 의한 질의 방법으로 객체들간의 인접(adjacency), 겹침(overlap), 포함(containment) 등과 같은 관계에 대한 질의가 가능하다. 이 방법은 각 객체간 위치 정보를 정확하게 나타낼 수 있다는 장점이 있으나, 현재 기술로는 정확한 객체 분할과 인식이 불가능하기 때문에 수작업에 의한 색인 과정이 필요하며, 영상내 객체수에 따라서 각 객체간 위치 정보의 양이 기하급수적으로 늘어난다는 단점을 갖는다.

두번째는 객체와 무관하게 영상의 기하학적인 위치를 표현하는 방법이다. 즉, “영상의 왼쪽 아래에 검은색 물체가 있고, 오른쪽 위에는 흰색 물체가 있다.”와 같이 영상을 표현하는 방법으로서, 객체 분할 없이도 영상의 위치 정보를 추출할 수 있는 반면에 객체에 대한 정보는 표현하지 않기 때문에 정확한 검색이 불가능하다는 단점을 갖는다.

본 장에서는 색상, 질감, 형태, 위치 등과 같은 영상의 하위 수준의 특징을 사용한 내용 기반 영상 검색에 관하여 살펴보았다. 이러한 하위 수준의 특징들은 영상의 일부 속성만을 포함하고 있기 때문에 단일 특징만을 사용할 경우 사용자의 요구를 모두 만족시킬 수 없으며 대용량의 영상 데이터베이스에 적용시킬 경우에는 검색 효율이 현저하게 감소한다는 문제점을 가지고 있다. 따라서 현재 상용화 단계에 있는 대부분의 영상 검색 시스템들은 텍스트 기반의 검색 방법을 병행하여 수행하고 있으며 검색 성능의 향상을 위해서 여러 종류의 특징들을 복합적으로 사용하고 있다. 그러나 진정한 의미의 내용 기반 영상 검색 시스템의 개발을 위해서는 영상내의 객체들에 대한 인식 기술과 고수준의 사용자의 질의로부터 적용이 가능한 적절한 검색 조건을 생성할 수 있는 기술이 필요하며 이와 함께 대용량 데이터베이스에

대한 효율적인 검색을 위해 색인 정보의 차원 변화에 따라 색인 구조를 융통성 있게 변화시킬 수 있는 색인 구조의 개발이 반드시 병행되어야 한다.

3. 내용 기반 비디오 검색

현재까지 주로 사용되어온 비디오 검색 방법은 비디오의 내용을 요약해 놓은 줄거리를 보거나, 비디오를 빠르게 돌려보는 방법이었다. 그러나 텍스트 형태의 줄거리만으로는 비디오의 내용을 정확하게 알기 어렵고, 빠르게 돌려보는 방법은 시간이 오래 걸리기 때문에 빠른 시간안에 비디오의 전체적인 내용을 쉽게 알아볼 수 있는 내용 기반의 비디오 검색 방법이 절실히 요구되고 있다. 내용 기반 비디오 검색은 비디오 데이터를 색인하기 위한 비디오 파싱, 제한된 저장 공간에 대용량 비디오 및 이에 관련된 메타 데이터를 효율적으로 저장하기 위한 비디오 데이터 압축 및 저장, 그리고 사용자가 원하는 비디오를 쉽게 검색 할 수 있는 환경을 제공하기 위한 검색 및 브라우징 등을 핵심 요소 기술로 하고 있다. 이 중에서 본 사례연구에서는 특히 비디오의 내용을 정확하게 표현하기 위한 비디오 파싱 기술을 중심으로 소개하고자 한다.

비디오 파싱은 대용량의 비디오 데이터를 효율적으로 저장, 관리하고 사용자가 쉽게 검색 할 수 있는 환경의 제공을 목적으로 하고 있다. 이러한 비디오 파싱은 대용량 비디오를 보다 적은 용량의 요약형 비디오로 만들기 위한 비디오 분할과 분할된 비디오의 내용을 효과적으로 표현하기 위한 비디오 색인 과정으로 구성된다.

3.1 비디오 분할

비디오는 그림 2와 같이 프레임(frame), 샷(shot), 장면(scene)의 세 가지 구성 요소로 이루어진다.

프레임이란 비디오를 구성하는 최소 단위로서, 필름 한장에 해당하는 하나의 영상을 나타낸다. 샷은 하나의 카메라로 촬영한 영상들을 나타내며, 샷 내에서는 필름이 끊기지 않고 연

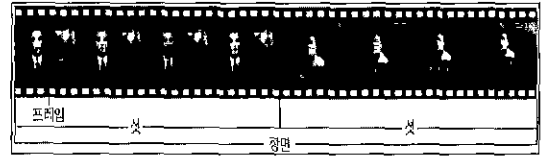


그림 2 비디오 데이터의 구성

속적으로 연결되어 있다. 샷은 비디오 분할의 기본 단위로 사용된다. 장면은 연속된 일련의 샷으로 구성되며, 주인공이나 특정한 장소와 같이 하나의 대상을 연속하여 촬영한 영상을 나타낸다.

비디오에서의 장면 전환은 컷(cut)과 점진적 전환(gradual transition)의 두 가지로 분류된다. 컷은 그림 2와 같이 하나의 샷이 끝난 후, 바로 다음 샷이 시작하는 것을 말하며, 점진적 전환은 끝과 시작이 뚜렷하게 구분되지 않는 장면의 전환을 나타낸다. 비디오 분할 기술은 장면 전환 효과에 의해서 발생하는 샷 사이의 경계를 검출하는 방법으로서, 분할 대상 비디오의 종류에 따라서 비압축 비디오에 대한 분할 기술과 압축 비디오에 대한 분할 기술로 나누어진다.

비압축 비디오에서의 비디오 분할 방법은 사용하는 특징의 크기에 따라서 화소 단위의 분할 방법, 부분 영역 단위의 분할 방법, 프레임 단위의 분할 방법으로 나누어진다[1]. 화소 단위의 분할 방법은 연속된 두 프레임에서 상응하는 화소 사이의 특징차를 구하여, 그 차이가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 검출하는 방법으로, 주로 화소의 명도값이나 에지 등을 특징으로 사용한다[9, 10]. 화소 단위의 분할 방법은 처리 시간이 빠른 반면, 잡영이나 영상내 물체 이동에 민감하다는 단점을 가지고 있다. 부분 영역 단위의 분할 방법은 한장의 프레임을 N개의 부분 영역으로 나누고, 연속된 두 프레임에서 상응하는 부분 영역간의 특징차가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 검출하는 방법이다. 부분 영역 단위의 분할 방법에서는 부분 영역별 평균 명도값이나 명도 히스토그램, 색상 히스토그램, 이동 벡터 등을 특징으로 사용한다[11]. 가장 일반적으로 사용되는 프레임 단위의 분할 방법에서는 한장의 프레임에 대해

여 대표적인 특징값을 구하고, 각 프레임간 특징값의 차이가 임계값 이상이 되는 프레임은 장면의 경계 프레임으로 검출한다. 프레임 단위의 분할 방법에서는 색상 히스토그램, 명도 히스토그램, 차영상의 히스토그램 등을 특징으로 사용한다[12].

일반적으로 비디오 데이터는 용량이 매우 크기 때문에 데이터베이스의 효율적인 관리를 위해서 압축된 형태로 저장, 관리된다. 압축 비디오에 비압축 비디오에 대한 분할 기술을 적용하기 위해서는 우선 복호화(decompression) 작업이 필요하다. 복호화는 단순한 작업이기는 하지만, 처리 시간이 오래 걸리기 때문에 매우 비효율적인 방법이다. 따라서 최근에는 압축 비디오에서 직접적으로 셋의 경계 프레임을 추출하는 방법이 연구되고 있다. 압축 비디오의 분할 방법은 사용되는 특징에 따라서 DCT 계수를 사용한 방법과 이동 벡터를 사용한 방법으로 나누어진다. DCT 계수를 사용한 방법은 연속된 프레임의 DCT 계수의 차가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 추출하는 방법으로써, 대부분의 압축 비디오 분할 방법이 DCT 계수를 사용하고 있다[13]. 이 방법은 MPEG 형식으로 압축된 비디오에 적용할 경우 처리 시간이 빠른 반면, 점진적인 장면 전환은 검출하지 못한다는 문제점을 가지고 있다. 이동 벡터를 사용한 분할 방법은 장면의 경계에 해당하는 프레임에서는 대부분의 이동 벡터값이 0이 된다는 사실에 기반하여 양의 값을 갖는 이동 벡터의 개수가 임계값 이하인 프레임을 장면의 경계 프레임으로 추출하는 방법이다[14, 15]. 이동 벡터를 사용할 경우에는 컷뿐만이 아니라 panning, zooming과 같이 카메라의 이동에 의해서 발생하는 장면의 변화도 분석할 수 있다는 장점을 갖는다.

이상에서 설명한 비디오 분할 방법은 두 프레임간 특징값의 차이를 사용하여 장면의 전환 유무를 결정하기 때문에 프레임 사이의 차이가 확실한 컷을 검출하는 데는 효율적이다. 그러나 페이드(fade), 디졸브(dissolve), 와이프(wipe) 등과 같은 점진적인 장면 전환은 연속된 프레임간 특징값의 차이가 크지 않기 때문에 이러한 방법으로는 정확한 검출이 어렵다. 따라서

점진적인 장면의 전환이 발생하는 경우에 생기는 프레임간 변화를 정확하게 정의할 수 있는 모델이 필요하다. Wei와 Zhong은 점진적인 장면 전환이 발생하는 경우에는 영상간의 차이가 일정하다는 사실에 기반하여 프레임간 변화 정도의 비율을 나타내는 상수 차영상(Constant Difference Image)을 정의하고, 점진적 장면 전환의 시작 프레임과 끝 프레임을 검출하였다[15]. 그러나 필름 편집에 의한 점진적 장면 전환은 종류가 매우 다양하고, 효과도 복잡하기 때문에 컷 검출에 비하여 정확성이 떨어지며, 촬영이나 물체의 움직임에 민감하다. 또한 컷 검출에 비하여 처리 속도가 느리다는 문제점을 가지고 있다.

3.2 비디오 색인

비디오 데이터는 비정형적인 구조와 시간적인 속성을 가지고 있기 때문에 모든 내용을 정확하게 색인한다는 것은 매우 어려운 일이며, 대부분의 경우에는 수작업에 의한 텍스트 기반의 색인 방법을 사용한다. 그러나 이 경우에는 색인에 드는 시간과 노력이 많이 필요하며, 비디오의 내용을 텍스트만으로는 정확하게 표현하기가 어렵기 때문에 텍스트 이외의 특징을 사용하여 비디오의 내용을 표현하고자 하는 노력이 진행중에 있다.

내용 기반 비디오 검색에서 텍스트 이외에 사용자가 주로 이용하는 검색 형태는 대표 프레임으로 표현되는 영상이며, 이러한 영상들은 2장에서 살펴보았던 내용 기반 영상 검색에서 사용하는 특징인 색상, 질감, 형태 정보 등을 색인값으로 사용한다. 그러나 대표 프레임만으로는 비디오의 시간적인 특성을 나타낼 수 없기 때문에 카메라 동작 분석, 이동 물체의 움직임 분석 등을 통하여 비디오의 시간적인 특성을 표현하고자 하는 연구가 진행 중이다. 비디오 상에서 가장 일반적으로 발생하는 카메라 동작은 그림 3과 같이 일곱 가지로 분류할 수 있다[16].

이중 panning, tilting, zooming은 카메라의 몸체가 고정된 상태에서 발생하는 동작이고, booming, dollying, tracking은 카메라의 몸체가 함께 이동하는 상태에서 발생하는 동작이

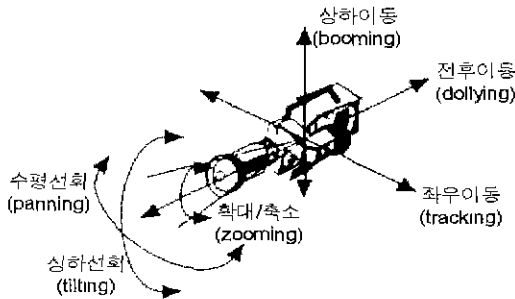


그림 3 카메라의 기본 동작[16]

다. 그러나 비디오 상에서 이 일곱가지 카메라 동작을 정확하게 구별하는 것은 매우 어려운 일이기 때문에, 수평 이동과 상하 이동, 그리고 확대, 축소의 세 종류를 검출하는 방법에 관한 연구가 주로 이루어지고 있다. 카메라 동작 분석 방법은 크게 수학적 모델을 사용한 방법 [16]과 패턴을 사용한 방법 [12, 17]으로 나눌 수 있다.

카메라 동작 이외에 비디오 상에서의 동적인 정보를 나타내는 이동 물체의 움직임에 대한 분석은 배경이 고정된 상태와 이동하는 상태로 분류하여 이루어진다. 배경이 고정된 상태에서는 물체 부분의 이동 벡터만이 양수값을 갖는다는 사실에 기반하여 이동 물체 영역과 이동 방향을 분석하게 된다. 배경이 이동 물체와 함께 움직이고 있는 영상에서는 우선 카메라 동작 분석 과정을 통하여 배경 부분을 제외한 후, 물체의 이동 벡터를 추출하여 이동 방향을 분석하게 된다. Wei와 Zhong은 Zhang 등이 제안한 이동 벡터의 패턴을 사용하여 배경이 이동하고 있는 상태에서 이동 물체를 추출하는 방법을 제안하였다 [15].

현재까지 연구된 카메라 동작과 이동 물체의 움직임 분석 방법들이 가지는 공통적인 문제점은 영상의 변화 속도가 빠른 경우 정확한 분석이 어렵다는 것과 영상 내에서 커다란 물체가 이동하는 경우에는 배경과 이동 물체를 구별하지 못한다는 것이다.

카메라 동작과 이동 물체의 움직임 이외에 비디오 색인에 사용되는 특징값은 비디오 상에 나타나는 객체의 이름이나 자막 및 오디오의 내용과 같은 보다 고수준의 특징값으로써, 이러한 정보를 자동으로 색인하기 위해서는 객체

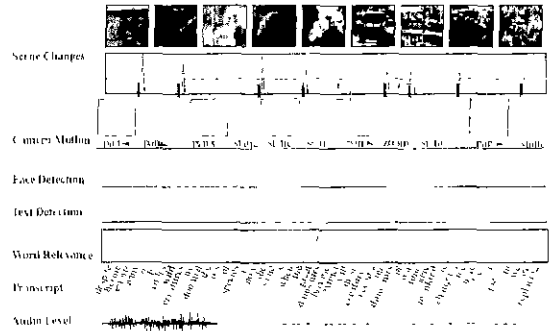


그림 4 Carnegie Mellon 대학에서 제안한 비디오 파상 과정의 예 [18]

및 문자, 음성 인식 기술 등이 필요하다. 인식 기반의 특징값을 사용한 대표적인 예로 미국 Carnegie Mellon 대학의 Informedia가 있다 [18]. 그림 4는 Informedia 색인 과정의 예로써, 비디오 상에서 객체 인식 및 음성과 문자 인식, 카메라 동작 분석 등의 다양한 인식 과정을 통하여 생성된 결과를 사용한 색인 과정에 사용함으로써, 데이터베이스 구축 및 검색의 효율성을 높였다.

4. 내용 기반 검색의 응용 사례

정지 영상을 대상으로 하는 내용 기반 영상 검색 시스템은 이미 상용화되어 인터넷을 통하여 서비스가 제공되고 있다. 다음은 현재 월드 와이드웹을 통하여 서비스가 제공되고 있는 대표적인 내용 기반 영상 검색 시스템들이다.

- IBM사의 QBIC(Query By Image Content, <http://www.qbic.almaden.ibm.com/~qbic/qbic.html>)
- Virage사의 Virage(<http://206.169.1.90/cgi-bin/query-c>)
- Informix Software의 Visual Intelligence (<http://www.illustra.com/virdemo>)
- Interpix Software사의 Image Surfer (<http://isurf.interpix.com>)
- Columbia 대학교의 ADVENT(<http://www.ctr.columbia.edu/advent/demos.html>)
- Chicago 대학교의 WebSeer(<http://infolab.cs.uchicago.edu/webseer/>)

- Excalibur사의 Excalibur Visual RetrievalWare(<http://www.excalib.com/rev2/demos/vrw/cstdemointro.html>)

이중 IBM의 Almaden 연구소에서 개발된 QBIC은 내용 기반 영상 검색뿐만 아니라 제한된 범위 내에서의 비디오 검색이 가능한 시스템으로 현재 여러 가지 면에서 가장 뛰어난 검색 시스템으로 평가받고 있다. QBIC은 사용자에게 색상이나 질감, 형태 등과 같은 다양한 속성에 기반한 시각적 질의를 제공하여 검색의 폭을 확대시켰으나, 복잡한 질의 인터페이스가 체계적으로 통합되어 있지 않다는 단점을 가지고 있다.

내용 기반 영상 검색의 특수한 응용 예로 얼굴 영상 검색 시스템을 들 수 있다. 얼굴 영상 검색 시스템은 질의로 들어온 얼굴 영상과 가장 유사한 영상을 구축된 얼굴 영상 데이터베이스로부터 검색하는 시스템으로써, 주로 범죄자 관리나 미아 및 가출인 관리를 위해서 쓰여지며, 아직까지 상용화된 시스템은 없으나, 미국과 유럽을 중심으로 활발한 연구가 진행중이다. 얼굴 영상 시스템의 대표적인 예로는 MIT의 Photobook이 있다(<http://vismod.www.media.mit.edu/vismod/demos/facerec/>). 이 얼굴 영상 시스템은 3,800명의 얼굴 영상 데이터베이스에 대하여 90% 이상의 정확한 검색률을 보여주었으나, 데이터베이스 구축에 사용된 얼굴 영상이 배경이 단순하고 조명 상태가 균일한 영상이고, 얼굴의 크기 변화에 민감하다는 단점을 가지고 있다. 그림 5는 Photobook에서 얼굴 영상을 검색한 결과에 대한 예를 보여주고 있는데, 다양한 표정의 동일 인물에 대

한 영상들이 정확하게 검색이 되었음을 알 수 있다.

현재 진정한 의미의 내용 기반 비디오 검색 기술을 사용한 응용이나 서비스는 아직 프로토타입에 대한 실험이 수행되고 있을 뿐이고, 실용화 단계에는 이르지 못하고 있다. 인터넷이나 방송을 통하여 이미 뉴스, 영화 등을 대상으로 하는 VOD 서비스가 제공되고는 있으나, 텍스트 기반의 색인을 사용하고 있기 때문에 비디오 데이터베이스 구축에 드는 비용이 매우 크고, 검색에 사용되는 질의 형태도 텍스트만을 허용하고 있다. 그러나 미국이나, 유럽, 일본, 싱가포르 등과 같은 선진국에서는 정보 고속도로의 구축과 더불어 다양한 통신 서비스의 기반이 되는 디지털 라이브러리 구축을 진행하고 있으며, 이에 가장 바탕이 되는 기술로써 내용 기반 비디오 검색 기술을 활용하고 있다.

내용 기반 비디오 검색 기술을 사용한 대표적인 응용 예로 뉴스 비디오 검색 시스템을 들 수 있다. 뉴스는 사회, 경제, 정치, 문화 등 다양한 정보를 포함하고 있을 뿐 아니라, 다른 비디오에 비하여 구조적인 형태를 지니고 있기 때문에 분석하기가 용이하다는 장점을 가지고 있다. 내용 기반 뉴스 비디오 검색에 관한 대표적인 연구는 독일 GMD, 캐나다의 CITR, 싱가포르의 ISS(Institute of Systems Science) 등에서 수행하고 있으며, 국내에서는 고려대학교에서 연구가 진행되고 있다. 이 중에서 독일 GMD, 캐나다의 CITR에서는 비디오 자동 분할 기술을 사용한 텍스트 기반의 색인 방법을 사용하고 있기 때문에, 데이터베이스 관리자에 의한 색인 과정이 필요하다. 싱가포르의 ISS는 뉴스에 대한 사전 지식을 바탕으로 뉴스 비디오를 자동 분할하고, 앵커 프레임을 기준으로 계층적으로 재구성함으로써 검색의 효율을 높였다. 고려대학교의 내용 기반 뉴스 비디오 파서는 얼굴 인식과 문자 인식 기술을 사용함으로써 색인의 정확성을 높였다[19]. 그림 6은 고려대학교에서 제안한 뉴스 비디오 파서의 구조를 보여주고 있다.

운동 경기 비디오 검색은 주로 축구 비디오를 중심으로 연구가 진행중에 있다. 그러나 운동 경기는 장면의 전환과 영상내 이동 물체의

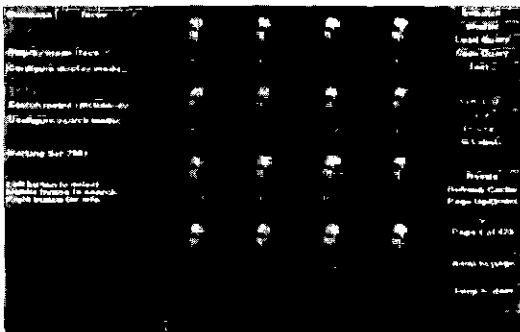


그림 5 Photobook의 얼굴 영상 검색 결과의 예

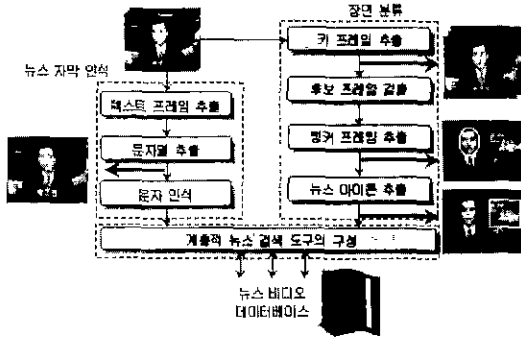


그림 6 고려대학교의 뉴스 비디오 파서

속도가 다른 비디오에 비하여 빠르기 때문에 중요 장면에 대한 분석이 어려워져 실용화에 많은 문제점을 가지고 있다. 축구 비디오 분석에 관한 대표적인 연구는 미국의 Princeton 대학교의 축구 경기 분석 및 파노라마 영상 재구성에 관한 연구가 있으며, 국내에서는 학계를 중심으로 축구 비디오 분석 시스템에 대한 연구가 진행 중에 있다.

5. 결 론

본 사례연구에서는 내용기반 검색 기술의 최근 연구 현황에 대하여 영상과 비디오를 중심으로 살펴보았다. 내용 기반 검색은 디지털 라이브러리, 방송 프로그램 관리, 의료관리 시스템, 교육용 시스템, 주문형 비디오, 홈 쇼핑, 웹 캐스팅 등과 같은 광범위한 응용 분야에서 활용될 수 있으며, 정보 고속도로의 구축과 더불어 현재 미국, 유럽, 일본 등과 같은 선진국을 비롯한 각 국에서 진행되고 있는 대용량 멀티미디어 데이터베이스 구축을 위한 핵심 기술로서 활발히 연구되고 있다.

내용 기반 영상 검색은 색상, 질감, 형태, 위치 정보 등을 사용하여 영상을 표현하고, 질의 영상을 통하여 사용자가 원하는 영상을 검색하는 방법으로서, 대용량 데이터베이스에서는 검색의 효율성을 높이기 위하여 두 가지 이상의 특징을 함께 사용하고 있다. 그러나 색상이나 질감과 같은 하위 수준의 영상 특성만으로는 영상의 내용을 정확하게 표현하기가 어렵기 때문에 보다 진정한 의미의 내용 기반 영상 검색을 위해서는 사람이나 동물, 문자 등과 같이

영상 내에 존재하는 객체를 추출하고 인식하는 기술이 필요하다.

내용 기반 비디오 검색은 대용량 비디오 데이터를 다루고 있기 때문에, 비디오 파싱, 압축, 검색 및 브라우징과 같은 요소 기술을 필요로 하고 있으며, 현재에는 디지털 라이브러리 구축의 가장 중요한 멀티미디어 데이터인 비디오의 내용을 자동으로 추출하기 위한 파싱 방법에 대한 연구가 활발히 진행 중이다. 초기 단계의 내용 기반 비디오 파싱에 관한 연구는 수작업으로 이루어지던 색인의 자동화에 초점을 맞추어 이루어졌으나, 현재는 비디오의 내용과 장면의 의미 등 보다 높은 수준의 특징을 추출하기 위한 연구가 수행되고 있다. 이를 위하여 비디오에서 중요한 장면을 추출함은 물론, 비디오 상에 존재하는 주요 객체나 문자에 대한 인식 기술이 개발되고 있으며, 내용의 정확성을 위하여 오디오 데이터의 분석에 관한 연구도 활발히 진행되고 있다. 이러한 내용 기반 검색은 미국, 싱가포르, 유럽, 일본 등의 선진국에서는 이미 1990년대 초부터 연구가 진행되어 현재는 시제품을 제작하여 인터넷이나 방송을 통하여 시범 운영에 들어갔으나, 국내에서는 아직도 일부 학계를 중심으로 초기 단계의 연구가 진행되고 있을 뿐이다.

다가오는 21세기를 비디오의 세기라고 할만큼 영상과 비디오를 비롯한 멀티미디어 데이터는 이미 우리 삶에서 막대한 비중을 차지하고 있다. 따라서 날마다 기하급수적으로 늘어나고 있는 대용량의 멀티미디어 데이터를 빠르고 효율적으로 처리할 수 있는 내용 기반 영상 및 비디오 검색 기술을 개발하는 것이 정보 선진국으로 발돋움하기 위한 발판을 구축하는 일이라고 사료된다.


참고문헌

- [1] 이미숙, 이성환, "내용 기반 비디오 검색을 위한 장면 전환 검출 방법의 성능 분석," 한국정보과학회 인공지능연구회 춘계학술대회 발표논문집, 서울, 1997년 3월, pp. 103-112.
- [2] 황본우, 이성환, "내용 기반 영상 검색을

- 위한 이산 웨이브렛 변환 영역에서의 프랙탈 영상 압축 기법,” 한국정보과학회 춘계학술발표 논문집, 춘천, 1997년 4월, pp. 361-364.
- [3] P. Aigrain, H. Zhang and D. Petkovic, “Content-Based Representation and Retrieval of Visual Media : A State-of-the-Art Review,” *Multimedia Tools and Applications*, Vol. 3, pp. 179-202, 1996.
- [4] M. Stricker and M. Orengo, “Similarity of Color Images,” *Proc. of SPIE - Storage and Retrieval for Image and Video Databases III*, San Jose, CA, USA, Vol. 2420, pp. 381-392, 1995.
- [5] Y. Gong, C. H. Chuan and G. Xiaoyi, “Image Indexing and Retrieval Based on Color Histograms,” *Multimedia Tools and Applications*, Vol. 2, pp. 133-156, 1996.
- [6] B. S. Manjunath and W. Y. Ma, “Texture Features for Browsing and Retrieval of Image Data,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp. 837-842, 1996.
- [7] A. Pentland, R. W. Picard and S. Sclaroff, “Photobook : Tools for Content-Based Manipulation of Image Databases,” *Proc. of SPIE - Storage and Retrieval for Image and Video Databases*, Vol. 2185, pp. 34-47, 1994.
- [8] B. Scassellati, “Retrieving images by 2D shape : a comparison of computation methods with human perceptual judgments,” *Proc. of SPIE-Storage and Retrieval for Image and Video Databases II*, San Jose, CA, USA, Vol. 2185, pp. 2-14, 1994.
- [9] H. Zhang, A. Kankanhalli and S. W. Smoliar, “Automatic Partitioning of Full-motion Video,” *Multimedia Systems*, Vol. 1, No. 1, pp. 10-28, 1993.
- [10] R. Zabih, J. Miller, and K. Mai, “A Feature-Based Algorithm for Detecting and Classifying Scene Breaks,” *Proc. of ACM Multimedia '95*, San Francisco, CA, pp. 189-200, 1995.
- [11] 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*, San Jose, CA, USA, Vol. 1908, pp. 13-24, 1993.
- [12] H. Zhang, and S. W. Smoliar, “Developing Power Tools for Video Indexing and Retrieval,” *Proc. of SPIE-Storage and Retrieval for Image and Video Databases II*, San Jose, CA, Vol. 2185, pp. 140-149, 1994.
- [13] 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*, San Jose, CA, USA, Vol. 1908, pp. 2-12, 1993.
- [14] H. Zhang, C. Y. Low, Y. Gong and S. W. Smoliar, “Video Parsing Using Compressed Data,” *Proc. of SPIE-Image and Video Parsing II*, San Jose, CA, USA, Vol. 2182, pp. 142-149, 1994.
- [15] Q. Wei and Y. Z. Zhong, “Content-Based Parsing in Video Database,” *Proc. of the First International Conference on Multimodal Interface*, Beijing, China, pp. 93-96, 1996.
- [16] A. Akutsu, Y. Tonomura, H. Hashimoto and Y. Ohba, “Video indexing using motion vectors,” *Proc. of SPIE-Visual Communications and Image Processing '92*, San Jose, CA, USA, Vol. 1818, pp. 1522-1530, 1992.
- [17] J. Maeda, “Method for Extracting Camera Operations to Describe Sub-scenes in Video Sequences,” *Proc. of SIPE - Digital Video Compression on Personal Computers : Algorithms and Technologies*, San Jose, CA, USA, Vol.

2187, pp. 56-67, 1994.

- [18] H. D. Wactlar, T. Kanade, M. A. Smith and S. M. Stevens, "Intelligent Access to Digital Video: Informedia Project," *IEEE Computer Magazine*, May, 1996, pp. 46-52.
- [19] 이미숙, 방진, 임영규, 홍영기, 김두식, 이성환, "내용 기반 색인 및 검색을 위한 실시간 뉴스 비디오 파서의 설계 및 구현," 한국정보과학회 봄 학술대회발표논문집, 춘천, 1997년 4월, pp. 365-368.




이 미 숙

1996 고려대학교 컴퓨터학과 학사

1996 한국정보과학회 논문경진대회 우수 논문상 수상

1996~현재 고려대학교 컴퓨터학과 석사과정 재학 중

관심분야: 영상 처리, 내용 기반 비디오 검색 등



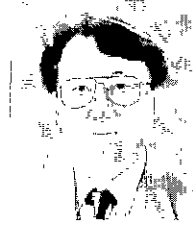
황 본 우

1995 성균관대학교 전자공학과 학사

1997 성균관대학교 전자공학과 석사

1997~현재 고려대학교 컴퓨터학과 박사과정 재학 중

관심분야: 영상 압축, 내용 기반 영상 검색 등



이 성 환

1984 서울대학교 계산통계학과 학사

1986 한국과학기술원 전산학과 석사

1989 한국과학기술원 전산학과 박사

1989~1994 충북대학교 컴퓨터학과 조교수

1995~현재 고려대학교 컴퓨터학과 부교수

관심분야: 패턴 인식, 컴퓨터 시각, 신경망 등

● 제24회 정기총회 및 추계학술발표회 ●

- 일 자 : 1997년 10월 24일(금)~25일(토)
- 장 소 : 이화여자대학교
- 문 의 처 : 한국정보과학회 사무국

Tel. 02-588-9246