

디지털 비디오의 활용 기술 : 내용기반 비디오 분할

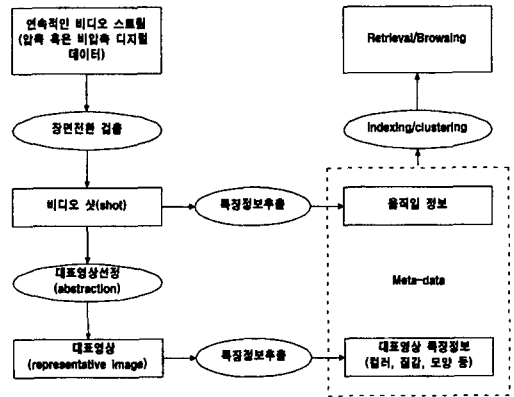
원 치 선

동국대학교 전자공학과

I. 서 론

컴퓨터의 급속한 가격대비 성능향상과 고속정보통신망의 보급으로 멀티미디어 시대의 도래가 가속화되고 있다. 즉, 점점 더 많은 문자, 음성, 영상, 비디오 등의 멀티미디어 정보가 디지털화되어 저장, 전송 및 처리되고 있다. 이들 멀티미디어 정보 중에 영상 및 비디오 정보가 차지하는 정보량은 다른 미디어의 그것과는 비교가 되지 않을 정도로 많다. 따라서 방대한 디지털 영상 및 비디오 데이터를 어떻게 효율적으로 전송, 저장, 처리하는가 중요한 문제로 대두되고 있다. 지금까지는 디지털 영상 및 비디오의 데이터 량을 감축하기 위한 영상압축 기술에 대한 연구가 집중되었다. 그러나 최근에 와서 디지털화되고 압축되어 저장된 수많은 영상 및 비디오 정보를 “활용”하는 측면을 생각하게 되었다. 즉, 하루에도 수없이 많은 비디오 데이터들이 제작되고 이들 데이터가 압축되어 저장되고 있는데 이는 미래의 재활용을 전제로 함이 틀림없다. 그렇다면 압축되어 저장된 수많은 비디오 데이터 중에서 사용자가 원하는 특별한 형태의 비디오 장면들을 빠른 시간 내에 어떻게 찾아올 수 있을까? 주어진 2~3시간 짜리 비디오의 내용을 단 몇 분 사이에 파악할 수 있는 방법은 없는가? 인터넷에 매일 수 없이 많이 등록되는 새로운 영상 및 비디오 물에 대해 사용자의 취향에 맞는 영상물들만 자동적으로 필터링하여 제공할 수 있는 컴퓨터는 없는가? 이와 같은 요구를 만족시킬 수 있는 알고리즘을 개발하기 위해 최근에 많은

사람들이 비디오의 내용기반 영상분석 문제에 관심을 갖게 되었다. 예를 들어, 멀티미디어 내용 서술 인터페이스(multimedia content description interface)를 표방하며 디지털 데이터의 내용을 부호화하여 표현함으로써 효율적인 검색을 가능케 하는 MPEG-7의 표준화가 시작되었다.



(그림 1) 비디오 정보처리 시스템을 위한 단계적 과정

디지털 비디오 정보를 원활히 활용하기 위해서는 그림 1에서 보는 바와 같이 단계적인 처리 과정이 필요하다^[1]. 우선 연속적인 비디오 시퀀스를 시간 축에서 분할하여, 카메라의 기록(record)과 멈춤(stop) 사이를 나타내는 샷(shot)의 경계를 검출해야 한다. 이때 적용되는 장면전환검출은 비디오 정보처리를 위한 가장 기본적이고 중요한 작업이다. 일단 샷이 검출되면 샷의 단위로 특징정보를 추출하거나, 샷을 대표하는 대표영상으로부터 특징정보를 추출하여 비디오의 내용을 검색하기

위한 인덱싱(indexing)이나 비디오의 내용을 한 눈에 살펴볼 수 있는 브라우징(browsing)에 활용 할 수 있다. 본 고에서는 이와 같은 비디오의 정보 처리를 위해 필요한 핵심기술들에 대해 살펴보고 앞으로의 발전방향을 조명해 본다.

II. 비디오 데이터의 구조

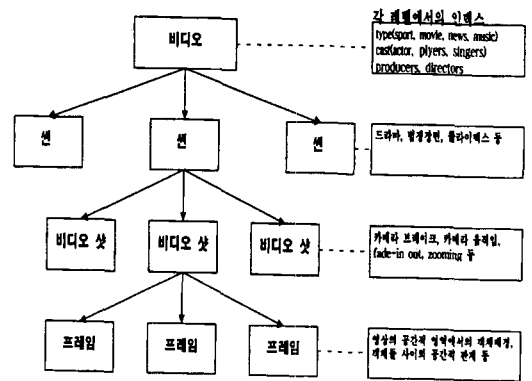
비디오 데이터는 표 1과 같이 다단계의 계층적 구조로 표현될 수 있다. 비디오 데이터의 가장 상위 레벨은 “비디오” 계층으로 제작자나 감독자에 의해 구별될 수 있는 단위로 도서관에서 책의 단위와 유사하다. 비디오의 바로 아래 단계는 “씬(scene)” 또는 “에피소드(episode)”로 불리는 비디오의 이벤트(event) 단위의 시간축 분할에 해당되는 것으로 책의 장(chapter)과 유사한 개념이다. 컴퓨터에 의해 비디오를 의미 있는 씬으로 나누는 것은 상당히 어려운 문제로 비디오의 핵심 단위인 샷을 클러스터링(clustering)하여 어느 정도 달성할 수 있다. “샷(shot)”은 비디오 제작시 카메라의 기록과 멈춤 사이에서 얻은 일련의 연속적 프레임으로 비디오의 내용 해석을 위한 가장 기본적인 단위이다. 샷의 검출은 카메라 브레이크(camera break) 혹은 장면전환검출(scene change detection)이라고도 부르며 씬의 검출보다

〈표 1〉 비디오 데이터의 계층적 구조

레벨	구조적 특징 및 정의
비디오	프로그램의 개념적 차이, 제작자, 감독자 등의 전체적 특징에 의해 분류된 계층
씬(에피소드)	공통적인 사건이나 시간적 연속성에 의해 결합된 연속된 샷의 집합.
샷 (클립)	비디오 제작시 카메라의 기록(record)과 멈춤(stop) 사이의 일련의 연속된 비디오 프레임의 집합
프레임	픽셀이 모여 이루어진 장면 특정 시간의 공간적 데이터

비교적 쉬운 문제이다. 비디오 구조의 가장 하위 단위는 프레임(frame)으로 샷을 이루는 기본 단위이다. 프레임은 비디오 제작시 자동으로 생성되어 별도의 프레임 검출을 위한 노력이 필요하지 않다.

비디오를 표 1과 같은 계층적 구조로 분할한 다음에는 각각의 계층에서 나타낼 수 있는 내용에 따른 라벨링을 부여하는 과정이 필요하며, 이를 비디오 인덱싱이라 부른다. 이 과정은 비디오를 체계적으로 데이터 베이스화하는데 중요한 역할을 한다. 각 계층별 구조에서 적용할 수 있는 라벨의 종류는 그림 2와 같다.



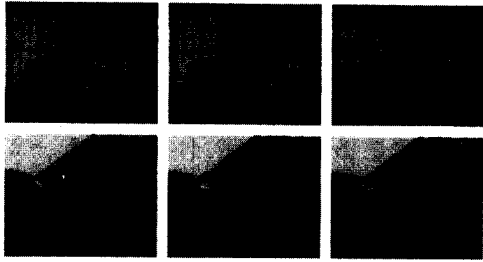
〈그림 2〉 각 계층에서의 인덱싱의 내용

III. 비디오 분할

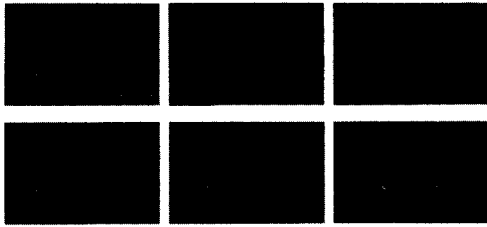
비디오 인덱싱을 위해 필요한 기술 중에 가장 중요한 것은 비디오 샷을 검출하는 일(장면전환 검출)과 유사한 샷을 묶어 씬을 만드는 일(클러스터링)이다. 이들 두 개의 과제를 성공적으로 달성하기 위해 비디오의 핵심적인 내용을 추출하여 유사도에 반영하는 내용기반 비디오 분할 기법이 필요하다.

1. 샷의 종류

샷은 영상 제작시 카메라의 기록(record)과 멈



(그림 3) 급진적인 장면전환



(그림 4) 점진적 장면전환

춤(stop) 사이에서 얻은 일련의 프레임으로 구성 되어 있으며, 샷과 샷의 연결은 그림 3과 같이 급진적인 변화(abrupt scene change)를 나타낼 수도 있고, 그림 4와 같이 점진적인 변화(gradual scene change)를 나타낼 수도 있다. 그림4의 점진적인 변화는 비디오 제작의 특수효과 중 디졸브(dissolve)에 해당된다. 특수효과를 이용한 또 다른 점진적 장면전환은 페이드인(fade-in), 페이드아웃(fade-out), 와이프(wipe)등이 있다. 비디오 시퀀스내에 샷을 정의하기 위해서는 급진적인 변화를 갖거나, 점진적인 변화를 갖는 장면전환점을 검출해야 한다. 쉽게 짐작할 수 있듯이 점진적인 변화를 갖는 경우의 장면전환 검출이 급진적인 변화를 갖는 경우보다 훨씬 어려운 문제가 된다. 또한 샷 내부에 페닝(panning), 틸팅(tilting), 줌링(zooming), 트래킹(tracking) 등의 카메라 움직임이나 플래쉬(flash)에 의한 갑작스런 조명변화가 존재하는 경우 샷이 아닌 부분에서 샷을 잘 못 검출하는 경우도 있다.

2. 공간영역에서의 장면전환 검출

공간영역에서 장면전환을 검출하기 위해 사용할 수 있는 기법은 연속적인 프레임 사이의 차이 값

을 이용하는 방법^[2,3]과 영상 히스토그램의 변화를 이용하는 방법^[2]이 있다. 첫 번째 방법은 연속적인 두 프레임의 같은 위치에 있는 픽셀의 밝기세기 차이 값의 절대치를 모든 픽셀에 대해 구하여 누적한 값이 주어진 기준치를 넘으면 장면전환으로 간주하는 방법이다. 이 방법은 적용이 간단한 반면 영상내 잡음, 밝기의 일시적 변화, 그리고 카메라의 움직임 등의 변화에 매우 민감하다는 단점을 갖는다. 이와 같은 단점을 해결하기 위해 연속적인 두 영상 사이의 히스토그램 변화를 측정하여 변화 값이 일정한 값을 초과하는 경우 장면전환으로 판단하는 히스토그램 기법을 사용한다. 히스토그램은 영상의 전체적인 특징을 표현하는 것으로 잡음, 밝기의 미세 변화, 그리고 카메라 움직임 등에 비교적 강한 장점을 갖는다. 연속적인 프레임 사이의 픽셀단위 차이 값만 이용하는 방법은 영상의 국부적인 변화만 고려하고, 반면에 히스토그램 기법은 영상의 전체적인 특징만 이용하는 결과가 된다. 따라서 영상의 국부적인 특징 변화와 전체적인 특징 변화를 모두 고려하는 방법이 제안되고 있다. 예를 들어, 영상을 일정 크기의 블록으로 나눈 뒤 연속적인 영상에서 같은 위치의 블록에 대한 히스토그램을 얻고 이들 독립적인 블록들의 히스토그램 데이터로부터 장면전환 검출을 시도하였다^[4,5].

대부분의 급격한 장면전환은 히스토그램과 밝기 차이를 이용한 함수로 잘 검출된다. 그러나 점진적 장면전환이나 영상내 특수효과가 있는 경우에 대한 장면전환 검출은 아직 검출율이 낮으며 이에 대한 많은 연구가 진행 중에 있다. 디졸브등 특수 효과에 의한 점진적 장면전환이 발생하는 경우에 연속적인 두 프레임 사이의 변화는 미소한 반면 이들 변화가 누적되어 결국 점진적 변화가 시작된 장면과 끝나는 장면은 큰 차이를 보이는 것이 보통이다. 지금까지 제안된 대부분의 점진적 장면전환 방법은 바로 이런 특징을 이용하고 있다. 예를 들어, 연속적인 두 프레임 사이의 밝기의 변화의 차이 값이 장면전환이라고 판단할 수 있을 만큼 큰 값인지를 결정하는 기준 값을 두 개를 설정하는 투인비교방법(twin-comparison approach)이 제안되었다^[2]. 즉, 비교적 낮은 기준 값은 점진적

변화가 의심되는 모든 영상들을 검출하고 검출된 영상들에 대해 더 세밀한 분석을 진행하여 오판된 장면전환인지 특수효과에 의한 점진적 장면전환인지를 구분한다. 물론 급격한 장면전환은 높은 기준값에 의해 검출된다. 지금까지 소개한 방법들은 데이터의 해석결과를 장면전환 검출에 이용하였다. 반면에 비디오의 제작시 적용한 제작과정을 모델화하고 이 모델을 기반으로 컷(cut)과 디졸브등의 특수효과(장면전환)를 분류해 내는 모델기반 방법도 제안되었다^[6].

연속된 모든 비디오 프레임에 장면전환 검출을 위한 변화를 측정하는 것은 상당한 계산시간을 요구한다. 계산시간을 줄일 수 있는 방법은 비디오 데이터의 공간적 혹은 시간적 해상도를 줄이는 것이다. 즉, 공간적 해상도를 줄이기 위해 픽셀을 표본화하여 일부 데이터만 사용하고 장면전환 검출을 위한 변화를 측정하거나, 초당 30 프레임의 비디오를 초당 15 프레임으로 표본화하여 장면전환을 조사하는 등 시간적 해상도를 줄일 수 있다. 실제의 경우 비디오 데이터는 시간적인 중복성이 공간적인 것보다 크므로 시간축 상의 해상도를 줄이는 방법이 더 선호되고 있다. 시간적 공간에서 해상도를 줄이는 방법은 비디오의 점진적 장면전환을 검출하는 데도 유용하게 사용되고 있다. 즉, 비디오 시퀀스를 일정한격으로 건너뛰면서 선택된 연속된 두 프레임들 사이의 변화를 측정한다. 이때, 변화의 기준 값을 비교적 낮게 설정하여 장면전환으로 의심되는 모든 프레임을 찾은 다음에 각각의 프레임의 좌우로 건너뛰는 프레임의 간격을 좁혀가면서 정확한 장면전환의 위치와 진위를 결정하는 다중경로 방법(multi-pass approach)도 제안되고 있다^[7].

3. 압축영역에서 장면전환 검출

비디오 신호를 디지털화하면 데이터 량이 급격히 증가하므로 전송 및 저장에 큰 부담이 된다. 따라서 대부분의 디지털 비디오 데이터는 MPEG-2와 같은 국제 압축 표준에 의해 압축된 상태로 저장/전송된다. 이와 같은 상황을 고려해 볼 때 압축된 데이터 비트 스트림으로부터 직접 장면전환을

검출하는 것은 상당히 설득력을 갖는다. 즉, 다음과 같은 세 가지 이유 때문에 압축영역에서 직접 장면전환 검출 등의 비디오 신호처리를 수행하는 연구가 필요하다.

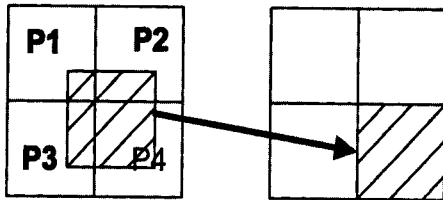
- 계산시간의 절약: 압축된 비트 스트림을 원 영상으로 복원하여 공간영역에서 장면전환 검출 등의 비디오 처리를 한 후에 저장/전송을 위해 다시 압축하는 것은 상당한 계산을 요구한다. 또한, 압축된 데이터 열은 데이터의 양 자체가 감소된 상태이므로 처리하는데 그 만큼 시간이 덜 걸린다.
- 압축의 보편적 사용: 대부분의 비디오 데이터는 MPEG과 같은 국제표준에 의해 압축된 상태로 저장 및 전송된다.
- 화질열화의 방지: 풀디코딩(full-decoding) 및 재압축(recompression)을 반복하면 화질의 열화가 누적된다.

MPEG-2로 압축된 비트 스트림으로부터 영상의 특징정보를 추출하여 장면전환 검출 등의 비디오 신호처리에 적용되는 알고리즘의 성능 평가는 아마도 다음의 두 가지가 고려되어야 할 것이다. 우선, 압축된 비트 스트림으로부터 추출된 특징 정보가 공간영역에서 추출한 특징 값들에 비해 얼마나 유효하고 정확하게 영상의 특징을 나타낼 수 있는가를 평가해야 하고, 두 번째는 그런 특징정보를 추출하기 위해 필요한 계산 량을 바탕으로 평가해야 할 것이다. 즉, 필요한 특징 정보를 추출하기 위해 어느 정도의 레벨까지 디코딩(Decompression)을 진행했는가에 의해 계산 량이 산정될 것이다. 결국 압축영역에서의 비디오 신호처리 알고리즘의 개발은 위의 두 가지 성능평가 요소를 향상시키는 방향으로 진행되고 있다.

MPEG-2 비트 스트림으로부터 영상의 특징정보를 추출하기 위해 활용할 수 있는 정보는 다음과 같다.

- I, B, P 픽처의 DC 혹은 AC 계수
- 움직임 벡터
- 데이터 발생량
- 매크로블록(macroblock)의 각종 모드(mode)

DC 성분을 활용하는 경우 MPEG의 비트 스트림으로부터 DC 성분만을 디코딩하여 DC 성분으로 구성된 작은 크기의(가로 세로가 각각 원래 크기의 1/8로 줄어든) 엄지손톱영상(thumbnail image)을 구성하고 이 엄지손톱영상으로부터 공간영역에서 적용되었던 기법들을 그대로 적용하는 알고리즘이 제안되고 있다^[8]. P와 B 픽처에 대한 엄지손톱영상을 구성하기 위해서는 움직임 벡터를 디코딩하고 움직임 보상된 데이터와 원영상의 에러 값을 IDCT 해야 하나, IDCT 연산 자체에 계산시간이 많이 필요하고 IDCT를 위한 계수 값의 디코딩 과정도 부담스러우므로 움직임 벡터만 디코딩하고 해당 블록의 밝기세기 값은 그림 5에서 보는 바와 같이 이전 프레임의 움직임 벡터에 의해 겹쳐진 블록들의 DC 성분을 겹쳐진 면적에 비례한 가중치만큼 만큼 고려하여 현재의 P와 B 픽처의 DC 값을 대략 구하므로써 약 1/4 가량의 곱셈계산을 줄일 수 있다^[8].

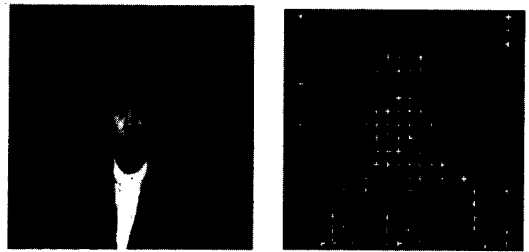


〈그림 5〉 P 픽처의 DC 성분 복원

MPEG 비트 스트림으로부터 각 픽처의 데이터 발생량을 측정하므로써 장면전환 등의 특징을 추출할 수 있다^[9]. 예를 들어, 두 개의 연속된 I 픽처의 매크로블록 레벨에서의 데이터 변화율이 크면 장면전환으로 간주할 수 있으며, P 픽처의 경우 연속적인 두 개의 P 픽처에서 장면전환이 발생하면 움직임 예측 에러가 증가하여 데이터 발생량이 증가하고 인트라(intra) 부호화된 블록의 수도 증가하게 된다. B 픽처에서 장면전환이 발생하면 이후에 오는 P나 I 픽처의 움직임 벡터에만 의존하게 되어 양방향 예측이 아닌 후방향 예측 움직임에 의한 단방향 예측 효과를 나타낸다. 따라서 양방향 예측의 경우보다 데이터 발생량이 증가하고 화면내 후방향 예측 모드가 많아지면 장면전환

을 의심할 수 있다^[10].

디코딩 과정을 최소화하기 위해 가변 길이로 부호화된 MPEG 비트 스트림으로부터 영상의 에지(edge) 정보를 추출해 내는 알고리즘이 개발되었다^[11]. 즉, DCT의 DC 성분은 부호화를 위해 매크로블록내 이전 DCT 블록의 DC 성분 값과의 차이 값에 비례하여 가변길이 부호화 된다는 사실을 이용하여 MPEG-2의 비트 스트림을 단지 파싱만 하고 부호화된 DC 성분의 길이로부터 DCT 블록 사이의 에지 유무를 판정한다. 그림 6에 MPEG-2 I 픽처로 부호화된 비트열로부터 추출해낸 에지 성분을 보이고 있다. 이렇게 얻은 에지 성분은 장면 전환 검출에 응용되었다^[11].



〈그림 6〉 부호화된 DC 비트열로부터 추출한 에지 성분

IV. 비디오 브라우징

샷의 경계가 결정된 다음에는 각 샷을 대표하는 대표영상이 결정되어야 한다. 대표영상의 용도는 다음과 같다. 첫째, 대표영상들을 이용하여 비디오의 전체적인 내용을 한눈에 파악할 수 있는 브라우징 영상을 구성할 수 있다. 둘째, 비디오 시퀀스 중에 특정영상을 검색할 때 질의 영상(혹은 질의어)에 대한 비교영상으로 대표영상을 활용할 수 있고 검색의 결과 검색된 영상을 제시할 때도 이들 대표영상들을 이용할 수 있다. 좀 더 자세하게 해당 샷의 연속적인 프레임을 보고싶을 때는 해당 대표영상을 클릭하여 볼 수 있도록 한다.

대표영상은 샷 내부의 첫번째, 가운데, 혹은 맨 마지막 장면을 선택하여 쓸 수 있다. 그러나, 긴

카메라 패닝(panning)을 갖는 샷의 경우 하나의 프레임이 샷 전체의 내용을 나타낼 수 없다. 따라서 샷내의 모든 프레임을 참조하여 모자이크 방식으로 합성한 영상을 대표영상으로 쓰는 경우도 있다^[12,13]. 각 샷에 대해 대표 영상이 선정되면 이들로부터 색깔, 질감, 모양 등의 특징정보를 추출하여 인덱싱에 사용하고 내용기반 비디오 검색을 가능하게 할 수 있다. 또한, 비슷한 샷들을 통합하여 샷의 구조보다 한 단계 높은 씬(scene) 단위의 내용 요약 구조를 만들 수 있다. 그러나 샷 보다 한 단계 위의 씬 단위로 비디오를 분할하는 일은 상당히 어려운 문제다. 이 문제는 샷을 대표하는 영상들의 특징벡터에 클러스터링 알고리즘을 적용하여 어느 정도 해결 할 수 있으나 비디오 내용의 유사도는 주관적인 측면이 많아 경우에 따라 그 결과가 상당히 다를 수 있다. 현재까지 샷 클러스터링을 위해 적용된 방법은 크게 두 가지로 나눌 수 있다^[14]. 첫 번째 방법은 비디오 프로그램의 성격을 미리 알고 있는 제한적인 경우에 사용하는 모델기반 방법으로 TV 뉴스프로그램에 활용될 수 있다. 예를 들어, 아나운서의 왼쪽 (혹은 오른쪽) 위의 작은 상자 속의 영상, 그리고 아래 혹은 상자 내의 자막들은 뉴스 프로그램에서 자주 나타나는 장면이며, 이와 같은 현상을 모델화하여 샷 클러스터링에 활용할 수 있다. 두 번째 샷 클러스터링 방법은 비슷한 특징정보를 갖는 대표영상들이 서로 한정된 시간간격 내에 존재하는 경우 동일 씬으로 클러스터링하는 방법이다^[15]. 앞서서도 언급했듯이 씬의 경계를 검출하는 것은 성능 평가시 주관적인 측면이 있어 분명한 기준을 설정하기 어렵지만 다음과 같은 단서가 있을 경우 씬의 경계를 의심할 수 있다^[16].

- 칼라: 유사한 칼라를 갖는 연속적인 샷은 공통적인 배경을 가질 가능성이 높고 따라서 같은 씬으로 인정할 수 있다.
- 오디오: 서로 다른 씬은 대개 상당히 다른 내용의 오디오를 가지므로 유사한 오디오가 연속되는 경우에는 씬의 경계를 정의할 수 없다.
- 주기적 장면전환: 연속적인 샷이 대화환경

(Dialog scene)임이 확인된 경우 같은 씬으로 간주할 수 있다.

샷과 씬의 경계를 검출하고 해당 비디오 단위를 대표할 수 있는 대표영상들을 선정하면 이 결과를 이용하여 비디오의 내용검색이나 브라우징에 활용할 수 있다. 비디오의 내용을 요약된 형태로 볼 수 있는 방법은 앱스트랙팅(abstracting)과 브라우징(browsing)의 두 가지 형태가 있다.

- 앱스트랙팅: 원래의 비디오보다는 훨씬 짧지만, 필수적인 비디오의 내용을 포함하는 일련의 동영상
- 브라우징: 샷이나 씬을 대표하는 대표 정지 영상들을 그래프나 계층적인 형태로 제시

비디오 앱스트랙트에 포함될 동영상을 선택하기 위해서는 샷이나 씬 단위로 앱스트랙트 시퀀스에 포함될 영상을 선정해야 한다. 선정시 고려해야 할 사항은 첫째, 전체 비디오 프로그램의 내용을 균형 있게 포함해야 하며, 둘째, 주연배우의 클로즈업(close-up) 이나 폭발, 혹은 문자자막이 나타나는 이벤트 씬들은 대개 비디오의 내용전환을 나타내는 중요한 장면들로 인식되므로 앱스트랙트에 포함시키는 것이 좋다^[16]. 대표 정지영상들을 사용하여 브라우징하는 방법은 아래의 두 가지 방법이 있다^[14].

- STG(Scene Transition Graph)^[17]: 샷의 집합을 노드(node)로 샷의 시간적 전개를 예시로 나타낸 그래프
- 비디오 포스터(Pictorial Summary)^[18]: 대표영상의 레이아웃과 상대적인 크기로 해당 샷의 중요성을 나타내는 방법으로 쿼드트리(quad-tree) 형태의 디스플레이도 이 방법에 포함된다.

V. 상품화 동향

샷이나 씬의 검출을 통해 얻은 비디오의 특징 값은 비디오 신호의 검색이나 비선형 편집기 등에 활용될 수 있다. 지금까지 개발된 내용기반 비디오

검색기 중에 가장 널리 알려진 것은 IBM의 QBIC (Query by Image and Video Content)^[13]이다. QBIC는 다음과 같은 질의 영상에 대해 대용량의 비디오 데이터로부터 내용기반 검색을 수행한다.

- 예제 영상
- 사용자가 그린 그림
- 사용자에 의해 선택된 색깔이나 질감 패턴
- 카메라나 물체의 움직임
- 기타 그래픽 정보

QBIC의 특징 중의 하나는 인간과 컴퓨터의 역할 분담을 통한 대화식 검색기의 실현이다. 즉, 인간은 영상으로부터 내용적 의미(semantic meaning)의 추출 능력이 탁월한 반면, 컴퓨터는 특징 값을 수치화하여 측정하고 이들 값들을 장기적으로 보관하는 능력이 우수하다는 사실을 심분 활용하고 있다. 그 밖에 인터넷 상에서의 비디오 검색을 위해 WebSeek^[19]이 개발되었다. 디지털 비디오의 비선형 편집기도 상품화되어 출시되고 있는데, 특히 MPEG 압축된 비트 스트림으로부터 직접 편집이 가능한 소프트웨어가 개발되고 있다. 예를 들어, Mediaware Solutions Pty Ltd^[20]에서는 MPEG-1 비트 스트림에서의 기본적인 cut&paste 기능이 가능한 WebFlix Pro를 출시하였으며 곧 MPEG-2에서도 적용될 수 있고 장면전환, 줌, 패닝 등의 성능을 향상시킨 WebFlix2를 출시할 예정이다.

VI. 향후 발전 방향

압축 또는 비압축 디지털 비디오 데이터로부터 특징정보를 추출하고 이를 이용하여 비디오의 내용 검색이나 편집 등에 활용하는 기술은 비디오 데이터 베이스의 급속한 확장과 함께 해결해야 할 문제로 대두되고 있다. 특히, MPEG-7 시스템의 전처리 단계인 비디오의 해석은(즉, 특징 값 추출) 표준화의 대상에 포함되지 않을 정도로 미래의 연구개발 결과에 상당히 의존하고 있다. 따라서 압축 또는 비압축 데이터로부터 특징정보를 추출하는

기술과 이를 응용하여 비디오의 내용검색에 적용하는 기술의 개발이 필요하다. 구체적으로 디지털 비디오 해석 문제에서 해결해야 할 문제로 꼽고 있는 내용들은 다음과 같다.

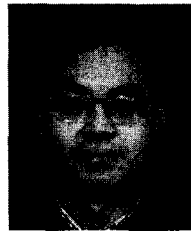
- MPEG 등 압축영역에서의 비디오 해석 및 처리
- 디지털 비디오 처리 시스템에서의 인간의 개입 및 역할
- 고차원의 의미적 (semantic) 개념과 저차원의 특징벡터값 사이의 연결
- 다양한 비디오 포맷과 다차원 특징 벡터의 처리

참 고 문 헌

- [1] H. Zhang, et. al., "An Integrated system for content-based video retrieval and browsing," Pattern Recognition, vol.30, no.4, pp.643-658, Sep. 1995.
- [2] H. Zhang, et. al., "Automatic partitioning of full motion video," Multimedia Systems, vol.1, pp.10-28, 1993.
- [3] K. Otsuji, Y. Tonomura, and Y. Ohba, "Video browsing using brightness data," Proc. of SPIE VCIP, pp.980-989, Nov. 1991.
- [4] A. Nakasaka and Y. Tanaka, "Automatic video indexing and full video search for object appearances," Proc. of 2nd Working Conf. Visual Database Systems, pp.119-133, Oct. 1991.
- [5] H. Ueda, et. al., "IMPACT: An interactive natural motion picture dedicated multimedia authoring system," Proc. of CHI'91, pp.343-350, April 1991.

- [6] A. Hampapur, R. Jain, and T. Weymouth, "Digital video segmentation," ACM multimedia, pp. 357-364, 1994.
- [7] B. Furht, S.W. Smoliar, and H. Zhang, Video and image processing in multimedia systems, Kluwer Academic Publisher, 1995.
- [8] B.-L. Yeo and B. Liu, "Rapid scene analysis on compressed video," IEEE Tr. on CSVT, vol.5, no.6, pp. 533-544, Dec. 1995.
- [9] J. Feng, K-T. Lo, and H. Mehrpour, "Scene change detection algorithm for MPEG video sequence," IEEE, pp.821-824, 1996.
- [10] Q. Wei H. Zhong, and Y. Zhong, "A robust approach to video segmentation using compressed data," Proc. of SPIE vol. 3022, pp.448-456, 1997.
- [11] C.S. Won, D.K. Park, S.-J. Yoo, "Extracting image features from MPEG-2 compressed stream," Proc. of SPIE, vol. 3312, pp.426-435, 1998.
- [12] L.A. Teodosia and W. Bender, "Salient video stills: content and context preserved," ACM Int'l Conf. Multimedia, ACM, 1993.
- [13] M. Flickner, et. al., "Query by image and video content: The QBIC system," IEEE Computer Magazine, vol.28, no.9, pp.23-32, Sep. 1995.
- [14] B.-L. Yeo and M.M. Yeung, "Retrieving and visualizing video," Communications of the ACM, vol.40, No.12, pp.43-52, 1997.
- [15] M. Yeung, and B.-L. Yeo, "Time-constrained clustering for segmentation of video into story units," Proc. of ICPR'96, pp.25-29, 1996.
- [16] R. Lienhart, S. Pfeiffer, and W. effelsberg, "Video abstracting," Communications of ACM, vol.40, No.12, pp.55-62, 1997.
- [17] M. Yeung, et. al., "Video browsing using clustering and scene transition on compressed sequences," Proc. of Multimedia Clustering and Networking, pp.399-413, 1995.
- [18] M. Yeung and B.-L. Yeo, "Video visualization for compact presentation and fast browsing of pictorial content," IEEE Tr. CSVT, vol.7, No.5, pp.771-785, 1997.
- [19] J. Smith and S.-F. Chang, "Visual search the Web for content," IEEE Multimedia, pp.12-20, July-Sept. 1997.
- [20] <http://www.mediaware.com.au/>

저자 소개



元 致 善

1982년 2월 고려대학교 전자공학과 학사, 1986년 2월 Univ. of Massachusetts/Amherst 석사, 1990년 2월 Univ. of Massachusetts/Amherst 박사, 1989년 11월-1992년 8월 금성사 선임연구원, 1992년 9월-현재 동국대학교 전자공학과 부교수, <주관심분야: 영상분할기반 영상압축, 내용기반 영상검색, 영상정보보호>
