

MPEG 압축된 비디오의 자동 분할 기법

김 가 현^{*} · 문 영 식^{**}

요 약

본 논문에서는 MPEG으로 압축된 비디오(MPEG-compressed video)를 대상으로 내용기반 색인(content-based indexing)에 기초가 된 동영상 자동 분할에 관한 효과적인 방법을 제안한다. 제안하는 방법은 MPEG 시퀀스의 I(Intra), P(Predictive), B(Bidirectional) 픽처 구성에 구애받지 않고 장면 전환점(scene change)을 검출해 낸다. 컷(cut) 검출을 위해서는 I 픽처의 dc 계수와 P, B 픽처의 매크로 블록 참조 특성을 이용하여 차이 측도(difference measure)를 설정한다. 그리고 점진적인(gradual) 장면 전환에서는 P, B 픽처의 참조 블록 비율을 이용하여 정확하게 장면 전환 지점을 검출한다. 이때 MPEG 시퀀스를 완전히 복원하지 않고 필요한 데이터만을 추출해 내어 전체 데이터 처리 과정을 좀 더 효율적으로 구성한다. 차이 척도의 성능과 검출 결과는 정확도(precision)와 완전추출도(recall)를 기준으로 비교분석하고, 제안한 방법을 다양한 MPEG 시퀀스에 적용시켜 검출 결과와 수행 시간 측면에서 그 효율성을 확인하였다.

Automatic Parsing of MPEG-Compressed Video

Ga-Hyun Kim^{*} · Young Shik Moon^{**}

ABSTRACT

In this paper, an efficient automatic video parsing technique on MPEG-compressed video that is fundamental for content-based indexing is described. The proposed method detects scene changes, regardless of IPB picture composition. To detect abrupt changes, the difference measure based on the dc coefficient in I picture and the macroblock reference feature in P and B pictures are utilized. For gradual scene changes, we use the macroblock reference information in P and B pictures. The process of scene change detection can be efficiently handled by extracting necessary data without full decoding of MPEG sequence. The performance of the proposed algorithm is analyzed based on precision and recall. The experimental results verified the effectiveness of the method for detecting scene changes of various MPEG sequences.

1. 서 론

분산 환경하의 정보 검색 기술이 계속적으로 발전하고 있는 가운데 멀티미디어 데이터의 효율적인 관리의 중요성이 더해지고 있다. 얼마나 빠른 시간 내에

얼마나 정확한 정보를 찾아낼 수 있는냐는 정보 검색과 멀티미디어 데이터베이스 구축에 궁극적인 목적이 된다. 특히, 가장 효과적인 정보 전달 매체라 꼽히는 동영상의 경우, 데이터베이스 구축시 편리한 사용자 인터페이스와 더불어 내용기반 색인 구조는 반드시 필요하다. 효율적인 동영상 데이터의 색인을 위해서는 동영상 데이터가 한번의 카메라 조작(shot)이나 편집(edit)단위 또는 동영상내의 일련의 시공간 단위(scene)로 분할 될 필요가 있다. 본 논문에서는 동영상 분할

* 본 연구는 한국과학재단 '97 핵심전문연구(과제번호 971-0902-022-2)지원으로 수행되었음.

^{*} 정 회 원 : 공학기술연구소

^{**} 성 회 원 : 한양대학교 전자계산학과 교수

논문접수 : 1998년 10월 26일, 심사완료 : 1999년 1월 12일

(video parsing) 단위를 샷(shot) 또는 장면(scene)이라 칭한다.

최근 연구되고 있는 장면 전환 검출 (scene change detection) 방법은 대체로 처리의 고속화를 위해서 압축된 데이터의 정보를 직접 이용하는 방법[1]-[5]이 많다. 또 장면 전환 검출을 위해 동영상 데이터의 활용범위, 예를 들면 뉴스, 스포츠, 영화, 광고 등의 분야에 따른 사전 지식(knowledge)을 이용하여 응용에 적합한(application specific) 정교한 장면 전환 검출 방법이 제시되고 있다[6]. 비압축 동영상 데이터의 장면 전환 검출 분야에서는 다양한 차이 척도(difference measure)가 제안되었고, 데이터베이스의 시스템적인 측면과 응용적 측면을 고려한 다양한 검출 방법이 소개되었다[7][8]. 그러나 인터넷 시대에 멀티미디어 데이터 처리에 관심이 되고 있는 데이터 검색 시간과 데이터 저장 공간의 한계를 극복하기 위해 압축 상태의 데이터를 대상으로 한 장면 전환 검출 방법에 관한 연구는 더욱 그 중요성이 커지고 있다.

많은 동영상 데이터는 동영상 압축 표준인 MPEG으로 압축 저장되어 있다. MPEG은 공간적 중복성을 제거하기 위해 DCT(discrete cosine transform)를, 시간적 중복성으로 제거하기 위해 움직임 보상(motion compensation)을 한다. MPEG 압축 데이터를 대상으로 장면 전환 검출을 수행하는 이유는, 비압축 동영상 데이터에 적용시켰던 장면 전환 검출 방법을 사용할 경우 이미 압축 저장된 데이터를 복원해야 하는데 있다. 복원된 영상은 원영상에 비해 화질이 떨어지기 때문에 검출 효과를 떨어뜨릴 수 있다. 그리고 MPEG 데이터 복원시 IDCT와 움직임 보상 등 복잡한 계산을 거쳐야 하기 때문에 시간이 많이 걸리는 문제가 있다.

MPEG 압축 데이터를 대상으로 한 장면 전환 검출 알고리즘으로, Yeo[1]등은 MPEG 데이터로부터 dc 성분만을 추출하여 원래 영상 크기를 1/64로 줄인 dc 시퀀스를 구성하였고 히스토그램 비교법을 적용시켜 장면 전환을 검출해 내었다. I 픽처로부터 dc 성분을 추출해 내는 것은 쉬운 일이나 P나 B 픽처에서는 움직임 벡터를 이용하여 dc성분을 따로 계산해 내는 과정이 필요하다. Shen[2]은 dc 성분과 움직임 벡터를 이용하여 컬러 히스토그램 차이로 장면 전환을 검출하였다. Meng[3]등은 CVEPS(Compressed Video Editing and Parsing System)을 소개하였는데, 동영상 데이터베이스의 효율적인 구축을 위해 동영상 데이터의 분할

과 더불어 동영상 데이터의 편집적 측면을 함께 고려한 방법을 제시하였다. Kuof[4]는 I, P, B 픽처 각각에 것이 발생하는 때를 컷 마스크(cut mask)로 구성하였고 데이터 시퀀스와 마스크 매칭을 통해 장면 전환점을 검출하였다. 그러나 이 경우는 MPEG의 IPB 픽처 비율이 1:2:6인 경우만을 고려한 컷 마스크를 설정하였기 때문에 모션 JPEG이나 다른 IPB 구성을 갖는 MPEG 시퀀스에 적용시키기 어렵다. Feng[5]은 매크로 블록(macro block)에서 발생하는 비트량(bit rate)을 측정하고 두 연속된 픽처 사이의 비트 발생량의 차이로 컷을 검출하였다. 이와 같은 기존의 압축 데이터상의 장면 전환 검출 연구들은 IPB 픽처의 구성에 따라 그 적용 범위가 제한적이었다.

본 논문에서는 임의의 IPB 픽처로 구성된 MPEG 압축된 동영상 데이터에 직접 적용 가능한 장면 전환 검출 방법을 제안한다. 2장에서는 MPEG 비트 열의 구조와 MPEG 시퀀스상의 장면 전환 검출 방법을 설명하고 3장에서 dc계수와 매크로 블록 참조 특성(macro block reference information)을 이용한 장면 전환 검출 알고리즘을 기술한다. 4장에서는 다양한 MPEG 데이터를 대상으로 한 실험 결과를 보여주고 5장에서 결론을 맺는다.

2. 압축 상태에서의 장면 전환 검출

2.1 MPEG 비트 열의 구조

MPEG에서는 동영상의 시공간적인 중복성을 제거하기 위해서 MC-DCT(motion compensation-discrete cosine transform)압축 방식을 이용한다. 이때 동영상 프레임들은 픽처 단위로 부호화 된다. 간단히 MPEG의 구조를 본다면 가장 상위 레벨에는 비디오 시퀀스(video sequence)가 있고 그 아래 동영상 편집을 고려한 GOP(group of picture)가 있다. GOP는 픽처들의 집합을 나타내며 각각 I(intra picture), P(predictive picture), B(bidirectionally predictive picture)로 구분되어 있고 각 픽처마다 다른 방식으로 부호화된다. 픽처는 16*16 크기의 매크로 블록을 기준으로 움직임 벡터 정보를 포함한다. 매크로 블록은 DCT 변환 단위인 8*8 블록으로 구성된다. 한 매크로 블록은 밝기 정보를 포함하고 있는 4개의 Y 블록과 색 정보를 포함하고 있는 2개의 Cr, Cb 블록을 갖고 있다. MC-DCT 압축 방식에 따라 압축된 MPEG 비트 열은 각 계층적

인 구조를 나타내는 헤더 정보, 각종 파라미터, DCT 계수, 움직임 벡터 등이 VLC(variable length coding)된 결과이다.

장면 전환 검출은 프레임간에 컷(cut)이나 점진적 변환(gradual transition)이 존재하는지를 판단하는 것이므로 픽처 단위에 대한 이해가 필요하다. 픽처는 I, P, B 픽처 세 종류로 되어 있는데 각각 다른 부호화 기술이 적용된다. I 픽처는 복호화시 임의 접근(random access)을 위한 기준이 되는 픽처인데 P, B 픽처로 인해 발생하는 에러를 줄이기 위해서 DCT와 양자화를 통해 공간 중복성만을 제거한다. P, B 픽처는 참조 픽처와의 차영상(residual error image)과 움직임 벡터(motion vector)로 시간적 중복성을 제거한다.

2.2 DCT 계수에 기반한 장면 전환 검출

MPEG의 I 픽처는 전체 영상을 DCT 변환하고 P 픽처 및 B 픽처는 참조 영상과의 차영상을 블록별로 DCT 변환한다. DCT 변환 후 영상은 저주파 영역과 고주파 영역으로 신호가 분리되고 실제 압축 과정인 양자화를 거치면 영상의 대부분 정보를 포함한 저주파수의 신호들은 그대로 남게 된다. 블록 내의 64개의 DCT 계수 중 정방향 2차원 배열의 첫 번째 DCT(0,0)에 해당하는 dc 계수는 영상의 저주파수 영역의 정보를 가장 많이 포함하고 있는 성분으로 다음과 같이 원 영상 블록 f(x,y)의 평균 밝기의 8배가 된다[1].

$$DCT(0,0) = \frac{1}{8} \sum_{x=0}^7 \sum_{y=0}^7 f(x,y) \quad (1)$$

이런 dc 계수를 MPEG 비트 열로부터 추출하여 dc 시퀀스를 구성한 후 픽셀 밝기 차이나 히스토그램 비교법을 이용하면 모션 JPEG의 컷 검출에 유리하다. 그리고 dc 계수의 특성상 평활화 효과가 있으므로 잡음이 존재하는 영상이나 약간의 움직임이 존재하는 영상 시퀀스에 적용시키는 것이 적합하다. 그러나 dc 계수는 영상의 에지 영역 정보를 포함하고 있지 않으므로 응용에 따라 에지 정보가 중요시되는 경우는 dc 계수 뿐만 아니라 ac 계수도 함께 사용할 필요가 있다.

2.3 움직임 벡터에 의한 장면 전환 검출

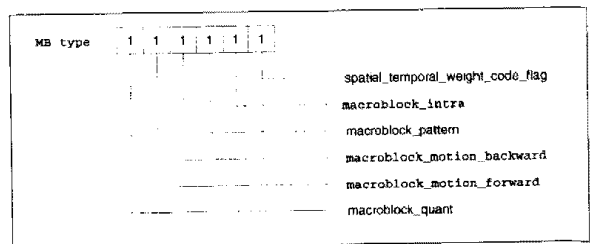
움직임 벡터(motion vector)는 프레임간의 참조 관계를 나타낸다. 일반적으로 장면 전환 검출시 프레임간의 특징 차이를 계산하게 되는데 움직임 벡터 정보

를 이용하면 프레임간 특징 차이를 계산하지 않고도 동영상 시퀀스의 흐름을 파악할 수 있다. 움직임 벡터는 매크로 블록 단위로 계산되며 전방향 예측 벡터(forward prediction vector)와 후방향 예측 벡터(backward prediction vector) 두 종류로 나뉜다. P 픽처는 이전의 I 픽처나 P 픽처를 참조 영상으로 하여 움직임을 예측하므로 전방향 움직임 벡터를 갖게 된다. 만약 동일 컷 내의 픽처를 참조하는 P 픽처라면 픽처 내에 전방향 움직임 벡터의 개수가 많다. B 픽처의 경우에는 전방향 움직임 벡터뿐만 아니라 후방향 움직임 벡터 개수도 많다.

움직임 벡터의 방향성을 판단하기 위해서 사용된 움직임 벡터의 종류는 비트 열 파싱 과정에서 매크로 블록의 타입(macro block type, MB type)을 보면 알 수 있다. (그림 1)은 매크로 블록의 비트 열 구조를 나타내고 (그림 2)는 그 중 매크로 블록 타입(MB type)의 내용을 나타낸다.



(그림 1) 매크로 블록 비트 열 구조 (Fig. 1) Bit stream structure of a macro block

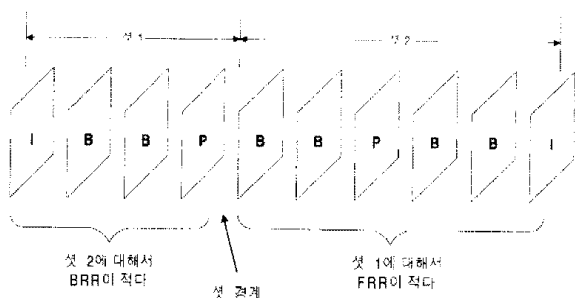


(그림 2) 매크로 블록 타입의 의미 (Fig. 2) Semantic diagram of a macro block type

I, P, B 픽처 내의 매크로 블록 타입(MB type)은 서로 다른 비트의 조합으로 되어 있지만 그 중 전향 타입(forward type), 후향 타입(backward type), 방향성 정보를 포함하지 않은 인트라 타입(intra coded type)등이 장면 전환 검출의 관심 대상이다. I 픽처는 모든 매크로 블록 타입이 인트라 타입이다. 따라서 I 픽처에 대해서는 참조 관계를 얻을 수 없다. 그러나 나머지 P 픽처나 B 픽처의 경우에는 전향 타입과 후향 타입의 매크로 블록 개수로부터 참조 픽처(refer-

ence picture)와의 유사성을 파악할 수 있다. P 픽처에는 전향 타입과 인트라 타입 매크로 블록이 존재할 수 있는데, 만약 현재 P 픽처에 전향 타입 매크로 블록 개수가 적다면 참조 프레임과 유사한 영역이 많지 않다는 것을 의미한다. 이러한 P 픽처는 결국 새로운 셋(shot)의 시작 프레임이 된다.

B 픽처에는 전향 타입, 후향 타입, 인트라 타입이 모두 존재할 수 있는데 B 픽처가 동일 셋 내부에 존재하는 픽처를 참조하였다면 전향 및 후향 타입 매크로 블록의 개수가 동시에 많아진다. 그렇지 않고 B 픽처가 새로운 셋의 시작 프레임이 된다면 전향 타입(forward type) 매크로 블록 수는 적어지고 후향 타입(backward type) 매크로 블록 수는 상대적으로 많아진다. (그림 3)과 같이 IPB 비율이 1:2:6인 시퀀스에서 연속된 두 B 픽처중 첫 번째 B 픽처에서 컷이 발생하였다면 이전의 I 픽처나 P 픽처를 참조할 가능성이 적으므로 전향 타입의 개수는 아주 적어지고 반면 후향 타입의 개수는 아주 많아진다. 즉, 셋 2의 B 픽처들은 셋 1에 대해 FRR(Forward Reference Ratio)이 적어지고, 셋 1의 B 픽처들은 셋 2에 대해 BRR(Backward Reference Ratio)이 적어진다.



(그림 3) 셋 경계와 셋 간의 참조 특성
(Fig. 3) Shot boundary and reference types between shots

3. 제안한 MPEG 시퀀스 장면 전환 검출

장면 전환에 사용되는 MPEG의 DCT 계수들은 원영상이 가지고 있던 밝기 정보 대부분을 포함하고 있다. 또한 MPEG은 참조 픽처와의 영상간 관계를 매크로 블록 단위마다 움직임 벡터로 부호화하게 되는데 여기서 움직임 벡터와 벡터의 참조 방향 정보는 장면 전환 검출에 유용하게 사용된다. 본 논문에서는 위 두 가지 중요한 MPEG 데이터 정보를 이용하여 임의의 IPB 픽처 구성에 적용시킬 수 있는 장면 전환 검출 기

법을 제시한다.

3.1 차이 척도 설정

MPEG은 부호화기를 따로 규정하고 있지 않기 때문에 IPB 픽처 비율이 일정하지 않다. 예를 들면, 모션 JPEG 시퀀스 구성은 HIL... B 픽처가 존재하지 않는 시퀀스는 IPPPPP..., 두 P 픽처 사이의 B 픽처의 개수가 적은 시퀀스는 IBBPBB..., P 픽처 사이에 B 픽처의 개수가 많은 시퀀스는 IBBBBPBBB... 등 MPEG 시퀀스는 다양한 IPB 비율로 구성된다. 이런 픽처의 구성을 미리 예측할 수 없기 때문에 픽처의 종류에 따라 선택적으로 차이 척도를 설계하여 장면 전환 검출 방법의 확장성(flexibility)을 높인다.

매크로 블록 타입(macro block type)은 매크로 블록을 파싱하는 과정에서 얻게 되는데 복원될 움직임 벡터의 방향성을 내포한다. 즉, 매크로 블록 타입으로부터 움직임 벡터를 복원하지 않더라도 참조 방향을 알 수 있고 픽처의 참조 경향을 파악할 수 있게 된다. 전향 타입(forward type)은 이전의 픽처를 참조해서 움직임 보상을 수행하였다는 정보로 픽처간 유사도를 파악할 수 있도록 해준다. 전향 타입의 매크로 블록은 P 픽처와 B 픽처를 구성하는 한 매크로 블록 타입이다. 반대로 후향 타입(backward type)은 움직임 보상시 시간적으로 현재 프레임 이후의 픽처를 참조해서 움직임 보상을 수행하였다는 의미를 갖고 B 픽처를 구성하는 한 매크로 블록 타입이다. P 픽처나 B 픽처에는 인트라 타입의 매크로 블록도 존재할 수 있는데 이는 참조 픽처와의 유사성이 없는 관계로 움직임 보상을 수행하지 않고 블록의 정보를 그대로 유지하고 있음을 나타낸다. 이렇듯 매크로 블록 타입으로부터 각 타입에 해당되는 블록 수를 조사하여 참조 픽처와 현재 픽처 간 유사 정도를 알아낼 수 있다. 컷 검출을 위해 매크로 블록 타입으로 참조 방향을 알아내고 P 픽처와 B 픽처간에 존재하는 각 타입의 블록 개수를 조사한다. 그리고 잠정적인 장면 전환 검출을 위해서 인트라 타입의 블록 수를 조사한다. 장면이 서서히 변하는 디졸브와 페이드 장면 일 때는 이전 셋과의 연속성이 많이 떨어지므로 매크로 블록이 인트라 부호화 될 확률이 높다.

따라서 I, P, B 픽처의 구성에 제한을 두지 않고 I 픽처에 대해서는 이전의 I 픽처와의 dc계수의 차이를 이용하고 P와 B 픽처에 대해서는 매크로 블록 타입에 근거한 픽처 참조 정보를 이용한다.

3.2 I 픽처에서 발생한 장면 전환 검출

MPEG 비트 열을 파싱하여 DCT 계수를 얻을 수 있는데 이 중 dc 계수는 원영상 블록의 평균 밝기 값이라는 점을 이용하여 dc 계수만을 이용한다. 장면 전환 검출을 위한 차이 척도는 다음과 같다.

$$D_I = \sum_{n=1}^N |DCT_{I_n}(0,0) - DCT_{I_{n-1}}(0,0)| \quad (2)$$

$DCT_{I_n}(0,0)$ 는 바로 이전의 I 픽처 dc 계수를 나타내고 $DCT_{I_{n-1}}(0,0)$ 은 현재 I 픽처 dc 계수를 의미한다. n 은 한 픽처 내에 존재하는 전체 휘도(Y) 블록의 개수를 나타낸다. 여기서 n 개의 휘도 블록이 존재하는 I 픽처의 블록간 dc 계수 차이의 합 D_I 가 임계치 T_I 보다 큰 경우 현재 I 픽처에서 장면 전환이 발생했으므로 판단한다. 일반적으로 동영상 시퀀스에서 컷 발생 빈도는 GOP 단위 내에서 두 번 이상 발생하는 경우는 거의 없으므로 I 픽처에 대해서만 dc 계수의 차이를 계산한다. 따라서 움직임 벡터를 이용하여 P, B 픽처의 dc 계수를 재구성 할 필요가 없기 때문에 정확하게 장면 전환점을 검출해낸다. 특히, I 픽처로만 구성된 모션 JPEG 데이터에 유리하며, P 픽처간 B 픽처의 개수가 적을 경우에 효과적이다.

3.3 P 픽처에서 발생한 장면 전환 검출

P 픽처에는 전향 타입(forward type)과 인트라 타입(intra coded type)의 매크로 블록만이 존재한다. P 픽처가 참조 프레임과 동일 샷 내에 존재한다면 전향 움직임 벡터에 의한 움직임 보상을 한다. 만약 장면 전환이 발생하였다면 이전 샷과 연속성이 없기 때문에 이때는 상당수의 매크로 블록이 인트라 부호화된다. 따라서 다음과 같은 척도를 사용하여 이 값이 임계치 T_P 이상이면 장면 전환이 발생하였다고 판단한다.

$$D_P = \frac{Intra_n}{fw_n} \quad (3)$$

위 척도에서 $Intra_n$ 은 인트라 타입의 매크로 블록 수를 나타내고 fw_n 은 전향타입의 매크로 블록 수를 나타낸다.

3.4 B 픽처에서 발생한 장면 전환 검출

B 픽처의 경우에는 양방향 움직임 보상을 하기 때

문에 전향 타입(forward type), 후향 타입(backward type), 인트라 타입(intra coded type)의 매크로 블록이 존재한다. IPB 비율이 1:2:6인 시퀀스 상에서 연속된 두 B 픽처 중 앞선 B 픽처에서 장면 전환이 발생되었을 경우 이전 샷(shot) 내의 픽처를 참조할 경향은 거의 없기 때문에 전향 타입의 매크로 블록 수는 매우 적다. 이때 후향 타입 매크로 블록 수는 전향 타입 매크로 블록 수에 비해 훨씬 많기 때문에 적절한 임계치를 설정하면 컷의 존재 여부를 알아낼 수 있다. 연속된 두 B 픽처중 두 번째 픽처에서 장면 전환이 발생하였다면 전향 타입 매크로 블록수가 적고 후향 타입 매크로 블록 수는 상대적으로 많다. 이때 앞선 B 픽처는 반대로 전향 타입의 매크로 블록수가 많고 후향 타입 매크로 블록 수는 상대적으로 적다. 다음과 같은 척도를 설정하여 이 값이 임계치 T_B 이상이면 장면 전환이 발생하였다고 판단한다.

$$D_{B_c} = \frac{bw_n}{fw_n} \quad (4)$$

위 척도에서 fw_n 은 전향타입의 매크로 블록 수를 나타내고 bw_n 은 후향타입의 매크로 블록 수를 나타낸다. 척도 D_{B_c} 는 현재(current) B 픽처에 대한 척도를 말한다. 일반적인 MPEG 시퀀스에서는 P 픽처 사이에 두개 이상의 B 픽처가 픽처 열을 이룬다. 따라서 IPB 비율이 1:2:6인 MPEG 시퀀스의 B 픽처에 대한 척도는 다음과 같다.

$$\text{case1} : \left\{ (D_{B_c} = \frac{bw_n}{fw_n}) \cdot (D_{B_n} = \frac{bw_n}{fw_n}) \right\}$$

$$\text{case2} : \left\{ (D_{B_n} = \frac{fw_n}{bw_n}) \cdot (D_{B_c} = \frac{bw_n}{fw_n}) \right\}$$

위에서 D_{B_c} 는 두 연속된 B 픽처 중 장면 전환 여부가 판단될 현재 B 픽처에 적용시킬 척도가 된다. case1의 D_{B_c} 는 바로 이어지는 B 픽처(next B picture)에 적용시키기 위한 척도이고 case2의 D_{B_c} 는 바로 앞선 B 픽처(pre B picture)에 적용시키기 위한 척도이다.

위의 척도는 두 가지 상황을 고려한 것이다. 첫 번째 경우는 두 연속된 B 픽처중 앞선 B 픽처에서 장면 전환이 발생했을 때이다. 두 연속된 B 픽처 모두 척도는 $\frac{bw_n}{fw_n}$ 이고 D_{B_c} 와 D_{B_n} 가 동시에 T_B 이상이어야 한

다. 두 번째 경우에는 두 연속된 B 픽처중 두 번째 B 픽처에서 장면 전환이 발생했을 때이다. 이 경우 앞선 B 픽처의 척도는 $\frac{fw_n}{bw_n}$ 이고, 두 번째 B 픽처의 척도는 $\frac{bw_n}{fw_n}$ 이다. 또한 D_{B_1} 와 D_{B_2} 가 동시에 T_B 이상이어야한다.

3.5 점진적인 장면 전환 검출

장면의 점진적 변화(gradual transition)는 두 셋이 편집 과정에 의해 겹쳐져 있는 상태로 장면이 서서히 전환됨을 말한다. 따라서 점진적 전환중인 P, B 픽처에는 인트라 타입 매크로 블록이 상당히 많을 것으로 기대된다. 따라서 매크로 블록을 파악하는 과정에서 연속된 P, B 픽처마다 인트라 타입 매크로 블록 개수를 조사하여 장면 전환 여부를 알아낸다. 그리고 정확한 장면 전환점을 찾아내기 위해 시간 창(temporal window)에 해당하는 임계치 T_{local} 를 설정하고 점진적 변화가 진행되는 동안 가장 변화가 큰 지점을 검출하기 위해 임계치 T_{noise} 를 설정한다. 척도는 다음과 같고 이 값이 T_{local} 과 T_{noise} 보다 클 경우 점진적 장면 전환점이 된다.

$$D_{gt} = \frac{Intra_n}{all_n} \tag{5}$$

위 척도에서 $Intra_n$ 은 인트라 타입의 매크로 블록 수를 나타내고 all_n 은 픽처 내에 존재하는 모든 매크로 블록 수를 나타낸다.

3.6 장면 전환 검출 과정

앞서 제시한 판단 척도를 통해 전체 MPEG 압축 데이터의 장면 전환점을 검출하는 과정은 다음과 같다. 먼저 MPEG 비트 열을 파악하여 픽처의 종류와 매크로 블록 타입에 관한 정보를 추출한다. 그리고 I 픽처에 대해 매크로 블록에서 휘도 성분(Y) 블록만을 추출하여 dc계수간 차이를 계산한다. 픽처의 종류에 따라서 식(2)~(5)를 이용하여 장면 전환을 검출한다. 이를 요약하면 (그림 4)와 같다.

```

parse picture header
  obtain picture code_type
parse macro block type
  obtain all_n = # of all parsed macro blocks
    
```

```

obtain  $fw_n$  = #of forward macro block type
obtain  $bw_n$  = #of backward macro block type
obtain  $intra_n$  = # of intra coded macro block type
parse block and obtain dc coefficients
if(code_type = I type){
  if(  $D_I > T_I$  ) detect cut;
  return;
}
if(code_type = P type){
  if(  $D_P > T_P$  ) detect cut;
  if( (  $D_{gt} > T_{local}$  ) · (  $D_{gt} > T_{noise}$  ) )
    detect gradual transition and extract high peak point;
  return;
}
if(code_type = B type){
  if(  $D_B > T_B$  ) detect cut;
  if( (  $D_{gt} > T_{local}$  ) · (  $D_{gt} > T_{noise}$  ) )
    detect gradual transition and extract high peak point;
  return;
}
    
```

(그림 4) 장면 전환 검출 알고리즘
(Fig. 4) Algorithm for scene change detection

4. 실험 결과

MPEG으로 압축된 데이터의 장면 전환 검출을 위해서 Berkeley MPEG1 복호화기를 사용하였다. MPEG 데이터는 인터넷에서 얻는 다양한 종류의 데이터를 대상으로 한다. 기존에 많이 사용되고 있는 픽셀차이 방법과 히스토그램 비교법을 통해 검출 결과를 비교한다.

4.1 임계치 설정

내용기반(content-based) 데이터베이스에서 데이터 검색 및 장면 전환 검출 알고리즘 평가 지표로서 정확도(precision)와 완전 추출도(recall)를 사용한다. 실험 대상의 장면 전환 검출 결과 오검출(false detection) 개수와 미검출(miss detection) 개수는 서로 역관계에 있다. 동영상 데이터의 활용범위와 데이터베이스 응용에 따라서 정확도(precision)가 더 중요시되는 경우도 있고 완전추출도(recall)가 더 중요시되는 경우도 있으므로 임계치는 응용에 따라 실험적으로 결정한다.

$$\begin{aligned}
 \text{precision} &= \frac{n_c}{n_c + n_f} & \text{recall} &= \frac{n_c}{n_c + n_m} \\
 \text{error rate} &= \frac{n_m + n_f}{all} & &
 \end{aligned} \tag{6}$$

n_c 는 장면 전환 지점을 제대로 검출해 낸 개수이고,

n_s 는 오검출 개수, n_m 은 미검출 개수를 나타낸다. 에러율은 장면 전환 지점에 대해 오검출 개수와 미검출 개수가 차지하는 비율을 말한다. 본 논문에서는 다양한 장면에 대해서 실험적으로 고정 임계치를 설정하였다. T_s 값은 50으로 설정하였고 T_p 는 2.3-3.5로 T_D 는 6으로 설정하였다.

4.2 컷 검출 결과

컷 검출 실험을 위해서 <표 1>과 같은 세 가지 테스트 시퀀스 pirates.mpg, wg_gdo_2.mpg, energizer.mpg를 사용하였다. 데이터의 종류는 광고 클립과 영화 클립 데이터로 장면 내에 큰 움직임(object motion)과 스트로브 라이트(strobe light)등이 첨가되어 있다.

<표 1> 컷 검출을 위한 테스트 시퀀스
<Table 1> Test sequences for cut detection

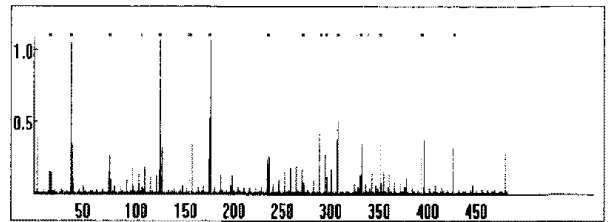
MPEG	종류 및 특징	크기	전체 프레임	셋수	컷수
pirates	ad., motion	160*120	279	19	18
wg_gdo_2	movie, light	304*224	358	14	13
energizer	commercial	192*144	487	18	18

데이터 종류가 서로 다른 장면에 대하여 제시한 검출 기법이 비압축 데이터에 적용시킨 알고리즘에 비하여 그 검출 효과가 떨어지지 않음을 다음과 같은 정확도와 완전 추출도를 지표로 비교하였다. 세 종류의 테스트 시퀀스의 검출 결과와 정확도 및 완전추출도, 에러율 측정 계산 결과는 <표 2>와 같다.

위에서 각 장면 전환 검출시 임계치는 고정 임계치를 사용하였다. 결과적으로 픽셀 밝기차이에 의한 방법과 제안된 알고리즘이 정확도와 완전 추출도가 높았

고 에러율은 적었다.

energizer.mpg 장면은 상업적 광고 클립으로 단순 컷만 포함하고 있는 장면이다. 장면의 특징은 컷 발생 빈도가 잦고 물체의 움직임을 많이 포함하고 있다. (그림 5)는 제안된 방법의 차이 측정 결과를 프레임 시퀀스에 따라 디스플레이 한 결과이다. 그래프의 가로축은 프레임을 나타내고 세로축은 프레임간 차이 척도 결과를 정규화 시킨 값을 나타낸다. 그리고 그래프에 나타난 점은 컷 지점을 말한다.



(그림 5) 컷 검출에 대한 차이 측정 결과
(Fig. 5) Difference values for cut detection

4.3 점진적인 장면 전환 검출 결과

점진적 장면 전환 검출을 위해 clapton.mpg와 hill.mpg, stone.mpg등 뮤직 비디오와 스포츠 장면을 사용하였다. 이 세 가지 데이터의 특징은 <표 3>과 같다.

<표 3> 점진적 전환 검출을 위한 테스트 시퀀스
<Table 3> Test sequences for gradual change detection

MPEG	종류 및 특징	크기	전체 프레임	점진컷 종류	점진컷 개수
clapton	music video	160*120	150	dissolve	2
hill	sport,motion	240*180	199	dissolve	1
stone	music video	160*120	453	dissolve	3

<표 2> 컷 검출 결과
<Table 2> Results of cut detection

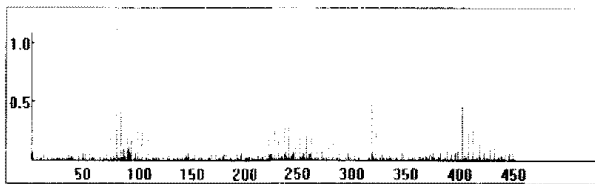
MPEG	방법	all	correct	false	miss	precision	recall	error
pirate	pixel	18	17	1	1	0.94	0.94	0.11
	histogram	18	14	10	4	0.41	0.73	0.78
	proposed	18	17	2	1	0.89	0.94	0.16
wg_gdo_2	pixel	13	9	2	4	0.82	0.69	0.46
	histogram	13	5	2	8	0.37	0.71	0.76
	proposed	13	10	3	3	0.76	0.76	0.46
energizer	pixel	18	11	0	7	1.0	0.61	0.39
	histogram	18	15	2	3	0.88	0.83	0.27
	proposed	18	16	1	2	0.94	0.89	0.16

이 표 5에 나타난 결과에 대해 식(5)에 따라서 차이 정도와 완전추출도율 계산한 결과를 <표 4>에 나타내었으며, 제안된 혼합형 알고리즘의 경우 다른 두 알고리즘에 비해 좋은 결과를 보이는 것을 알 수 있다.

<표 4> 점진적 장면 전환 검출 결과
<Table 4> Results of gradual change detection

MPEG	방 법	all	cor- rect	false	miss	preci- sion	recall	error
clapton	pixel	2	0	0	2	0.0	0.0	1.0
	histogram	2	0	0	2	0.0	0.0	1.0
	proposed	2	1	0	1	0.0	0.5	0.5
hill	pixel	1	0	0	1	0.0	0.0	1.0
	histogram	1	1	0	0	1.0	1.0	0.0
	proposed	1	1	0	0	1.0	1.0	0.0
stone	pixel	3	0	0	3	0.0	0.0	1.0
	histogram	3	2	0	1	1.0	0.67	0.33
	proposed	3	3	0	0	1.0	1.0	0.0

stone.mpg 장면은 뮤직 비디오 클립으로 점진적 장면 전환인 디졸브를 포함하고 있고 장면 내에 물체의 움직임이 다양하게 포함하고 있다. 픽셀 밝기 차이법과 히스토그램 비교법에서는 디졸브 변화 지점을 제대로 검출하지 못했다. 그러나 제안된 방법에서는 (그림 6)에서와 같이 3개의 디졸브 장면을 정확하게 검출해 내었다.



(그림 6) 점진적 장면 전환 검출에 대한 차이 측정 결과
(Fig. 6) Difference values for gradual change detection

4.4 처리 시간

본 논문에서 제안한 장면 전환 검출 기법은 MPEG 압축 데이터에 적용시킨 방법으로 압축 복원을 위한 불필요한 시간 소비를 줄일 수 있다. MPEG 복호화기를 통한 영상 복원 연산에 관한 연구[9]에 따르면 움직임 보상과 화면 디스플레이에 소비되는 시간이 전체 복원 과정의 50% 이상을 차지한다. 따라서 dc 계수 추출과 매크로 블록 타입만을 파악하여 참조 방향 정보

를 파악한 제안된 알고리즘은 효율적인 방법이라 할 수 있다.

실험에서는 차이 추정 척도에 따른 차이 계산과 장면 전환점 검출 시간을 다양한 크기의 장면에 대해 적용시켜 초당 처리 프레임 수를 측정하였으며, 그 결과는 <표 5>와 같다. 초당 처리 프레임 수는, 제안한 방법을 사용함으로써 복원 후 처리하는 방법보다 약 2배 정도 향상됨을 볼 수 있다.

<표 5> 처리 시간
<Table 5> Execution time

장면	영상 크기	방법	초당 처리 프레임 수
Tennis	352*240	픽셀 밝기 차이법	6.8
		히스토그램 비교법	7.1
		제안된 알고리즘	10.6
Terminator2	240*180	픽셀 밝기 차이법	12.5
		히스토그램 비교법	13.6
		제안된 알고리즘	21.7
Energizer	192*144	픽셀 밝기 차이법	16.8
		히스토그램 비교법	18.7
		제안된 알고리즘	23.2

5. 결 론

본 논문에서는 MPEG 데이터를 부분적으로 파싱하여 얻은 dc 계수와 매크로 블록 타입의 비율에 근거한 장면 전환 검출 방법을 제안하였다. 픽처의 종류에 따라 데이터가 어떠한 IPB 픽처 구성을 갖더라도 장면 전환 검출이 가능하게 하였다. 그리고 매크로 블록 타입으로부터 인트라 타입의 매크로 블록 수를 조사하여 점진적인 장면 전환 지점을 정확하게 찾아내었다. 기존의 원 영상 시퀀스를 대상으로한 픽셀 밝기 차이법이나 히스토그램 비교법에서는 압축된 데이터를 복원할 때 IDCT 과정을 거치기 때문에 데이터 복원 과정에 많은 시간을 할애해야 했다. 반면 제안된 방법은 MPEG 데이터에 직접 적용시키는 방법으로 장면 전환 검출 과정이 복잡하지 않고, 장면 전환 지점을 정확히 검출해 내었다. 기존의 압축 데이터 상에서의 장면 전환 검출 알고리즘과의 성능을 비교한다면, 모션 JPEG 데이터에 컷이 존재하는 경우에 Yeo의 알고리즘은 픽셀 밝기 차이법이나 히스토그램 차이 비교법에 의한

실험 결과로 대변될 수 있겠고, Meng의 알고리즘은 본 논문에서 제안한 방법과 비슷한 성능을 보일 것으로 예측된다. 단, 기존의 알고리즘들은 IPB 픽처 비율이 다양한 데이터에 대해서는 적용하기 어려운 반면에 제안하는 알고리즘은 IPB 픽처 구성에 관계없이 적용할 수 있다. 또한, 움직임 벡터를 이용한 P, B 픽처의 dc계수 재계산 시간이나 움직임 벡터를 얻어내는 과정을 없애고 매크로블록 타입으로부터 얻은 정보로 차이 척도를 계산하며 점진적 변환 검출도 매크로블록 타입에 근거하였기 때문에 계산 과정이 매우 간단하다는 장점이 있다.

향후 연구 방향으로서는 압축 데이터 상에서 움직임 벡터를 분석하여 영상 내에 존재하는 카메라 및 물체의 움직임을 효과적으로 처리하는 방법에 관한 연구가 진행되어야 하며 압축 데이터 상에서 다양한 편집 효과를 검출하는 연구가 필요하다. 그리고 다양한 종류의 동영상 데이터에 관한 사전 지식을 습득함으로써 장면 전환 검출의 효과를 증대시키는 연구가 필요하다.

참 고 문 헌

[1] Boon-Lock Yeo and Bede Liu, "Rapid Scene Analysis on Compressed Video," *IEEE Transaction on Circuit and System for Video Technology*, Vol.5, No.6, pp.533-544, December 1995.

[2] Ke Shen and Edward J. Delp, "A Fast Algorithm for Video Parsing Using Compressed Sequences," *ICIP 95*, pp.252-255, 1995.

[3] Jianhao Meng and Shih-Fu Chang, "CVEPS - A Compressed Video Editing and Parsing System," *Proc. of ACM Multimedia 96*, November 1996.

[4] Tony C. T. Kuo, Y. B. Lin and Arbee L. P. Chen, "Efficient Shot Change Detection on Compressed Video Data," *Proc. of IEEE Workshop on Multimedia Database Management System*, August 1996.

[5] Jian Feng, kwok-Tung Lo and Hassan Mehropour, "Scene Change Detection Algorithm for MPEG Video Sequence," *ICIP 96*, pp.821-824, 1996.

[6] Qi Wei, HongJiang Zhang and Yuzhuo Zhong, "A Robust Approach to Video Segmentation

Using Compressed Data," *SPIE 97 Storage and Retrieval for Image and Video Database*, pp.448-456, 1997.

[7] John S. Boreczky and Lawrence A. Rowe, "Comparison of Video Shot Boundary Detection Techniques," *Storage and Retrieval Image and Video Database IV, Proc. of IS&T/SPIE 1996 Symp. on Elec. Imaging: Science and Technology*, February 1996.

[8] Gulrukh Ahanger and Tomas D. C. Little, "A Survey of Technologies for Parsing and Indexing Digital Video," *Journal of Visual Communication and Image Representation Vol.7, No.1*, pp.28-43, March 1996.

[9] William I. Grosky, Ramesh Jain and Rajiv Mehrotra, *The Handbook of Multimedia Information Management*, Prentice Hall, 1997.



김 가 현

e-mail : ghkim@cse.hanyang.ac.kr
 1996년 한양대학교 전자계산학과 졸업(학사)
 1998년 한양대학교 전자계산학과 대학원 졸업(석사)
 1998년~현재 공학기술연구소



문 영 식

e-mail : ysmoon@cse.hanyang.ac.kr
 1980년 서울대학교 전자공학과 졸업(학사)
 1982년 한국과학기술원 전기및전자공학과 졸업(석사)
 1990년 The University of California at Irvine 전기 및 컴퓨터공학과 졸업(박사)

1982년~1985년 한국전자통신연구소 연구원
 1989년~1990년 InnoVision Medical 선임연구원
 1990년~1992년 생산기술연구원 선임연구원
 1992년~현재 한양대학교 전자계산학과 부교수
 관심분야 : 컴퓨터비전, 패턴인식, 멀티미디어 응용