동영상 검색을 위한 고속 장면 전환 검출
High-speed scene change detection for video indexing

나윤정, 하명환, 이상길
한국방송공사 기술연구소

요 약

동영상을 효율적으로 검색하기 위해서 동영상을 장면 단위로 분할하는 과정이 필요하다. 이 논문에서는 동영상에서 장면이 전환되는 지점을 빠르게 검출하는 방법을 제안한다. 먼저 MPEG 압축 영상에 대하여 시간적 표본화를 통하여 추출된 압축 영역의 데이터를 사용하여 장면 전환의 후보 구간들을 정하고, 다음으로 이들 구간 안에서 각 프레임의 화소값을 얻고, 이를 사용하여 정확한 장면 전환 지점을 찾아낸다. 제안된 방법은 빠른 검출 속도와 함께 높은 정확성을 갖는다. 또한 카메라 플래시 때문에 장면 전환으로 잘못 검출되는 것을 방지할 수 있는 방법을 제안한다.

I. 머릿말

디지털 동영상이 널리 사용되면서 종래의 텍스트를 중심으로 한 검색 환경에서 동영상의 내용에 바탕을 둔 검색 환경으로 발전하고 있다. 따라서 동영상을 그 내용에 따라 색인하는 작업이 필요하게 되었다. 이를 위하여 동영상 내용의 기본 단위가 되는 연속된 장면 단위로 동영상을 분할하는 작업이 선행되어야 한다. 동영상의 분할을 위해 장면이 전환된 지점을 자동으로 찾는 방법들이 많이 연구되었다[1].

장면 전환에는 인접하는 두 프레임 사이에서 갑자기 장면이 변하는 컷(cut)과 여러 프레임에 걸쳐 서서히 장면이 변하는 디졸브(dissolve), 와이프(wipe), 페이드 인(fade in), 페이드 아웃(fade out) 등과 같은 점진적 장면 전환이 있다. 컷은 비교적 간단하게 검출할 수 있으나, 점진적 장면 전환의 검출은 훨씬 더 어렵다.

장면 전환을 검출하는 방법들은 압축 영역에서 하는 방법과 화소 영역에서 하는 방법으로 구분될 수 있다. 압축 영역에서는 MPEG과 같은 압축 영상으로부터 추출된 DCT(Discrete Cosine Transform) 계수 또는 움직임 벡터를 사용하여 장면 전환을 검출하는 것이다. 화소 영역에서는 압축 영상을 디코딩하여 얻은 화소값을 사용하여 장면 전환을 검출한다. 일반적으로 압축 영역에서는 검출 속도를 빠르게 할 수 있지만 다룰 수 있는 정보의 한계 때문에 검출 성능을 높이는 데 많은 제한이 있다. 반면에 화소 영역에서는 화소값을 얻기 위해 디코딩을 해야 하므로 처리 속도가 느리지만 영상에 관한 모든 정보를 다룰 수 있으므로 검출 성능을 높이는 데 제한이 적다.

이 논문에서는 압축 영역과 화소 영역을 결합하여 양쪽의 장점을 살림으로써 정확성을 유지하면서 검출 속도를 빠르게 할 수 있는 두 단계로 이루어진 장면 전환 검출 구조를 설계하였다. 화소 영역에서는 명암 변화와 윤곽선 변화를 효율적으로 결합하여 정확성을 유지하면서 검출 속도를 향상시켰다. 아울러 화소의 밝기 변화를 이용하여 카메라 플래시 때문에 장면 전환으로 잘못 검출되는 것을 방지할 수 있는 방법을 설명한다. 이러한 방법들을 이용하여 컷 검출, 점진적 장면 전환 검출, 카메라 플래시 검출을 하나의 통합된 알고리듬으로 구성한 것을 보인다. 실험을 통하여 제안된 방법의 성능을 확인하였다.

II. MPEG 비트스트림과 구간 복호 영상을 사용한 장면 전환 검출

MPEG과 같은 압축 영상의 비트스트림으로부터 직접 추출할 수 있는 DCT 계수 또는 움직임 벡터를 사용하여 장면 전환을 검출하는 여러 방법들이 제안되었다. 예를 들면, 각 프레임에서 표본으로 추출한 DCT 계수로 벡터를 구성하고 이 벡터들의 내적을 이용하는 방법[2], B 프레임의 경우에는 순방향 예측 움직임 벡터와 역방향 예측 움직임 벡터의 갯수를 비교하고, P 프레임에 대해서는 움직임 보상이 있는 매크로 블록과 움직임 보상이 없는 매크로 블록의 갯수를 비교하는 방법[3] 등이 있다. 이렇게 압축 영상의 비트스트림 데이터를 직접 이용하면 처리 시간을 줄일 수 있다. 그러나 이러한 방법으로는 다룰 수 있는 정보가 한정되고 윤곽선과 같은 영상의 특징에 관련된 정보를 얻기 어렵기 때문에 검출 성능을 높이는 데 많은 제한이 있다. 반면에 화소값을 사용하여 장면 전환을 검출하면 디코딩으로 인해 처리 시간이 많이 소요되지만 정확성을 높일 수 있다. 따라서 압축 영상 비트스트림의 데이터를 직접 이용하는 방법과 영상의 화소값을 이용하는 방법을 적당히 결합하면 두 방법의 장점을 모두 취할 수 있다.

MPEG 영상은 GOP(Group of Pictures) 단위로 압축된다. GOP의 첫 번째 프레임은 I 프레임이다. 예를 들면 GOP안에 있는 프레임 수가 6이고 I 또는 P 프레임이 나타나는 주기가 3인 경우, MPEG 비트스트림의 프레임 순서는 IBBPBBIBBPBBI...와 같다. 따라서 MPEG 영상에서 GOP 간격으로 GOP의 첫 번째 프레임을 추출하여 그 프레임들을 비교한다면 시간적 표본화를 하면서 장면 전환을 검출하는 효과를 낳는다.

압축 영역에서 I 프레임에 대해서는 DCT 계수를 직접 얻을 수 있지만, P와 B 프레임에 대해서는 본래 영상의 DCT 계수를 직접 얻을 수 없어서 그 값을 추정하는 방법이 이용된다. 곧, 압축 영상을 디코딩하지 않고는 P와 B 프레임에 대한 정확한 DCT 계수를 직접 구할 수 없다. 이 논문에서는 정확한 값을 얻을 수 있는 I 프레임의 DCT 계수만을 사용하였다. 곧, GOP의 첫 프레임만을 선택하고, 그 프레임에 포함된 DCT 블록들의 첫 번째 계수(DC 계수)만을 추출한다. 이 값을 사용하여 인접한 두개의 I 프레임을 비교한다. DC 계수를 사용하여 두 프레임을 비교하기 위해 DC 계수의 히스토그램 차이를 구하여 이 값이 만일 일정한 임계값 보다 작으면 두 프레임 사이에 장면 전환이 없는 것으로 보고, 그 값이 임계값을 넘으면 두 프레임 사이에 장면 전환이 있을 가능성이 있는 것으로 보고, 이 구간을 장면 전환 후보 구간으로 정한다. 다음 단계인 화소 영역에서는 이 후보 구간에 대해서만 부분 디코딩을 하여 화소값을 얻고 한 프레임씩 이동하면서 인접한 프레임들을 비교하여 그 구간 안에서 정확한 장면 전환 지점을 찾는다.

III. 화소 영역에서 장면 전환 검출

1. 명암과 윤곽선 변화를 결합한 장면 전환 검출

장면 전환을 검출하는 방법 가운데 화소 비교법, 히스토그램 비교법과 같이 영상의 명암 변화를 이용한 방법들이 있다. 이러한 방법들은 비교적 간단하여 구현하기 쉽기 때문에 많이 사용되고 있지만, 조명의 변화에 민감하고, 압축 영상을 복호화하여 그 방법들을 적용하는 경우에는 양자화 잡음의 영향을 많이 받는 단점이 있다. 이러한 문제점을 해결하기 위하여 영상의 특징을 이용하여 장면 전환을 검출하려는 방법이 연구되고 있다. 그 가운데 영상의 윤곽선 변화를 이용하는 방법이 제안되었다[4]. 이 방법은 비록 많은 처리 시간을 요구하지만, 명암 변화에 바탕을 둔 방법들보다 정확성이 높고 점진적 장면 전환 검출에서도 우수한 성능을 보인다.

여기서 윤곽선 변화만을 사용하면 많은 시간이 걸리지만, 명암 변화와 윤곽선 변화를 결합하면

검출 시간을 단축시킬 수 있다. 곧, 인접한 두 프레임 사이를 순차적으로 비교할 때, 먼저 명암 변화를 계산하여 두 프레임 사이의 차이가 특정 임계값을 넘을 때에만 윤곽선을 구하여 두 프레임을 비교한다. 이와 같은 방법을 사용하면 윤곽선 변화만을 사용한 방법에 견주어 볼 때 검출의 정확성을 그대로 유지하면서 처리 시간을 줄일 수 있다.

윤곽선의 변화를 계산하기 위하여 기존의 윤곽선에 새로 추가되는 윤곽선의 양을 나타내는 ρ_{in}과, 기존의 윤곽선 가운데서 사라지는 윤곽선의 양을 나타내는 ρ_{out}을 아래 식 (1), (2)와 같이 각각 계산하고, 식 (3)과 같이 두 값 가운데 최대값을 취한다. 여기서 E와 E'은 각각 비교할 두 영상의 윤곽선 영상을 나타낸다. $\overline{E}$는 E를 일정한 반지름 r 만큼 팽창시킨 영상이다.

$$\rho_{in} = 1 - \frac{\sum_{x,y} \overline{E}[x,y]E'[x,y]}{\sum_{x,y} E'[x,y]} \tag{1}$$

$$\rho_{out} = 1 - \frac{\sum_{x,y} E[x,y]\overline{E'}[x,y]}{\sum_{x,y} E[x,y]} \tag{2}$$

$$\rho = \max(\rho_{in}, \rho_{out}) \tag{3}$$

컷에서는 ρ가 큰 값으로 나타나고, 점진적 변화 부분에서는 특수 효과에 따라 ρ가 그 구간에서 증가하거나 감소한다. 곧, 디졸브의 경우에는 ρ가 증가하다가 감소하는 모양을, 페이드의 경우에는 ρ가 증가하는 모양을 갖는다. 윤곽선을 검출하는 여러 알고리듬 가운데 비교적 간단한 Sobel 연산자에 의한 방법을 사용하여 실험하였다.

2. 밝기 변화를 이용한 카메라 플래시 검출

뉴스 영상의 경우 카메라 플래시가 터지는 장면이 자주 나오는데, 그 순간 화면의 밝기가 화면의 모든 영역에 걸쳐 크게 증가하므로 화소 차이(pixel difference)나 히스토그램 차이(histogram difference)가 크게 나타난다. 따라서 연속된 영상임에도 불구하고 장면 전환 지점으로 잘못 검출되는 결과를 낳는다.

이 논문에서는 이러한 문제점을 해결하기 위하여 화면의 밝기가 변하는 유형을 이용하였다. 카메라 플래시 한 개가 터지는 경우에는 화면 전체가 밝아졌다가 다음 프레임에서 어두워지면서 원래의 영상으로 되돌아온다. 여러 개의 카메라 플래시가 한 프레임 내의 시간차를 두고 연속적으로 발생하는 경우에도 플래시가 터지는 시작점에서 화면이 밝아지고 몇 프레임 안에 본래의 영상으로 되돌아오는 특성을 갖는다. 따라서 화면의 밝기가 일정 수준 이상 밝아지는 지점이 나타나면 이 지점을 표시해 두고, 그 이후 수 프레임 안에 화면의 밝기가 일정 수준 이상 어두워지는 장면이 나타나고, 이 장면이 밝아지기 직전의 본래의 영상과 큰 차이가 없으면, 밝아지기 시작한 지점과 어두워진 지점 사이에 있는 모든 예비 장면 전환 지점을 카메라 플래시 영상으로 판단할 수 있다. 이상의 방법은 아래와 같이 구현될 수 있다.

1) 장면 전환 검출을 위해 순차적으로 프레임들을 비교하다가 화소의 밝기 변화(luminance difference)가 일정한 임계값(T_f)을 넘는 프레임(a)을 발견하면 이 지점을 표시해 두고, 이전 영상의 히스토그램($H_{a-1}[j]$)을 저장한다. 여기서 i번째 프레임과 그 다음에 인접하는 프레임 사이의 화소의 밝기 변화는 다음과 같이 표시된다.

$$LD_i = \frac{SPD_i - SND_i}{M \times N} \times 100 \tag{4}$$

여기서, $M \times N$은 한 프레임에 포함된 화소의 갯수이고, SPD_i 와 SND_i 는 각각 밝기가 증가한

화소의 갯수, 밝기가 감소한 화소의 갯수를 나타낸다. SPD_i는 다음과 같이 표시된다.

$$SPD_i = \sum_{x,y=1}^{M,N} PDP_i(x,y) \tag{5}$$

여기서 $P_i(x, y)$를 i번째 프레임의 화소값, t를 일정한 임계값이라고 하면,

$$PDP_i(x,y) = \begin{cases} 1 & \text{if } P_{i+1}(x,y) - P_i(x,y) > t \\ 0 & \text{otherwise} \end{cases}$$

마찬가지 방법으로 SND_i는 다음과 같이 표시된다.

$$SND_i = \sum_{x,y=1}^{M,N} NDP_i(x,y) \tag{6}$$

여기서,

$$NDP_i(x,y) = \begin{cases} 1 & \text{if } P_i(x,y) - P_{i+1}(x,y) > t \\ 0 & \text{otherwise} \end{cases}$$

2) 검색을 계속 수행하면서 두세 프레임 안에 화소의 밝기 변화가 일정한 음의 임계값($-T_f$)보다 작은 프레임(b)이 나타나고, 이때 $H_{a-1}[j]$과 $H_b[j]$를 비교하여 이것이 일정 임계값보다 작다면, a와 b 사이에 있는 모든 예비 장면 전환 지점들을 카메라 플래시 영상으로 판단한다.

3) 위의 조건을 만족하지 못하면 b 프레임 다음으로 장면 전환 검출을 계속한다.

IV. 장면 전환 검출의 통합 구현

컷, 점진적 변화 영상, 그리고 카메라 플래시 영상의 검출을 하나로 통합한 구조를 그림 1에 나타냈다. 압축 영역에 바탕을 둔 첫 번째 단계에서 I 프레임에 대하여만 DC 계수를 추출하고 이것을 인접한 I 프레임과 서로 비교하여 그 차이가 큰 구간을 장면 전환이 예상되는 후보 구간으로 선정한다. 화소 영역에서 진행되는 두 번째 단계에서는 비디오열의 처음으로 돌아가 후보 구간에 대해서만 디코딩을 하여 화소값을 얻고, 이 값을 사용하여 정확한 장면 전환 지점을 찾는다. 여기서는 히스토그램 비교법과 윤곽선 비교법을 함께 사용하였고, 윤곽선 변화를 사용하여 점진적 변화 영상을 검출하는 부분과 카메라 플래시 영상을 검출하는 부분을 모두 통합하였다.

그림 1(b)에 보인 것과 같이 화소 영역에서는 히스토그램 차이(HD)를 계산하고, 이것을 임계값과 비교한다. 임계값으로는 컷을 판별하기 위한 Tl과, 이것보다 작은 값으로서 점진적 변화 영상을 판단하기 위한 Ts가 사용된다. 작은 임계값은 점진적 변화가 시작되고 계속되는 것을 판단하는 기준으로 사용된다. HD가 Tl보다 큰 경우에는 윤곽선을 검출하고, 윤곽선 변화가 임계값보다 크면 컷으로 판단한다. 아울러 화소의 밝기 변화를 계산해 두어 카메라 플래시 영상을 검출하는 데 사용한다. 다음으로 HD가 Ts와 Tl 사이에 있으면 점진적 변화가 시작되거나 점진적 변화 구간 안에 있는지 여부를 판단하기 위해, 윤곽선을 검출하여 윤곽선 변화의 패턴을 저장한다. HD가 Ts보다 작으면 장면 전환이 없는 것으로 판단하거나, 앞부분이 점진적 변화 구간이라면 그 구간이 끝나는 지점으로 판단한다. 점진적 변화 영상에 대한 판단은 그 구간의 시작과 종료를 위와 같이 판단하고 그 구간에서 얻은 윤곽선 변화의 패턴을 분석하면 된다. 그런데 점진적 변화가 일어나는 구간은 여러 GOP 구간에 걸쳐 있을 수 있으므로 점진적 변화가 시작되었는데도 불구하고 후보 구간의 끝까지 진행했을 때 점진적 변화가 종료되지 않은 것으로 판단되면, 다음 GOP 구간을 계속 확인할 필요가 있다. 실험에서는 GOP 길이가 15이고 점진적 변화가 15프레임 이상 계속되지 않는다고 가정하고 후보 구간의 바로 다음에 있는 GOP 구간까지만 디코딩을 하면서 점진적 변화 여부를 확인하였다.

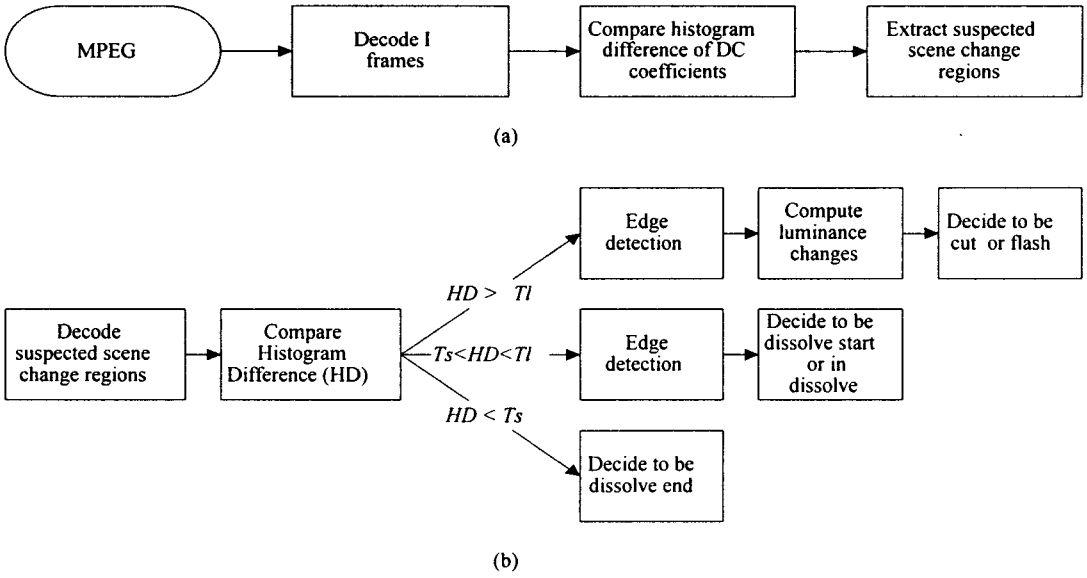

그림 1. 장면 전환 검출을 위한 통합 구조
(a) 첫 번째 단계, (b) 두 번째 단계

표 1. 실험 결과

동영상	화소영역만 사용한 방법의 검출시간(t_1)	제안된 방법의 검출시간(t_2)	검출 속도비 (t_1/t_2)	실제 장면전환 갯수	맞게 검출한 갯수	검출하지 못한 갯수	잘못 검출한 갯수
뉴스(5분)	312초	151초	2.07	63	62	1	4
뉴스(15분)	838초	368초	2.28	214	212	2	12
드라마(15분)	903초	337초	2.68	112	110	2	10
스포츠(15분)	897초	476초	1.88	185	184	1	29

V. 실험 결과

실험은 윈도 NT가 설치된 컴퓨터에서 수행하였다. 뉴스, 드라마 등의 동영상을 352×240 크기의 MPEG-1 파일로 압축하였다. 표 1에서 제안된 결합 방법의 검출 시간과 화소 영역만 사용한 방법의 검출시간을 비교하였다. 여기서 화소 영역만 사용한 방법이란 압축 영상을 처음부터 끝까지 디코딩하여 화소값을 얻고, 여기에 이 논문에서 제안한 화소 영역의 장면 전환 검출 알고리듬을 적용한 것이다. 제안된 결합 방법의 검출 시간이 화소 영역만 사용한 방법의 검출 시간보다 작다는 사실을 확인할 수 있다. 여러 영상에 대해 실험한 결과 검출 속도가 평균적으로 2배 정도 향상되는 것으로 나타났다. 움직임이나 변화가 많은 영상에서는 장면 전환의 후보 구간이 많아지므로 검출 시간이 길어지게 된다. 따라서 제안된 방법은 움직임이나 변화가 적은 영상에 적용될 때 검출 속도가 향상되는 효과가 크다.

제안된 방법의 검출 성능을 알아보기 위해 실제 장면 전환 갯수, 맞게 검출한 장면 전환 갯수, 검출하지 못한 장면 전환 갯수, 잘못 검출한 장면 전환 갯수를 측정하고 그 결과를 표 1에 보였다. 화소 영역만 사용한 방법을 채택했을 때에도 검출 성능은 동일하게 나왔다. 따라서 압축 영역과 화소 영역의 결합을 통해 검출 성능 또는 정확성에 영향을 주지 않으면서 검출 속도를 향상시

킬 수 있다.

카메라 플래시 영상의 검출 실험을 위하여 사용한 뉴스 영상의 일부분을 보면, 이 영상은 프레임 번호 4640부터 4697까지 계속되면서 여러 대의 카메라가 임의의 순서로 동일 장면 위에 플래시를 터뜨린다. 그림 2에 그 구간에서 프레임 사이의 히스토그램 차이를 보였다. 플래시가 터지거나 본래의 영상으로 돌아올 때, 히스토그램 차이가 매우 큰 것을 알 수 있다. 그림 3은 프레임 사이의 밝기 변화를 나타낸 것이다. 플래시가 터졌을 때에는 밝기가 크게 증가하고, 본래의 영상으로 돌아올 때에는 밝기가 크게 감소하는 것을 볼 수 있다. 이러한 특성을 이용하여 제안된 카메라 플래시 검출 방법을 적용한 결과, 그 구간 안에 있는 모든 카메라 플래시를 검출할 수 있었다.

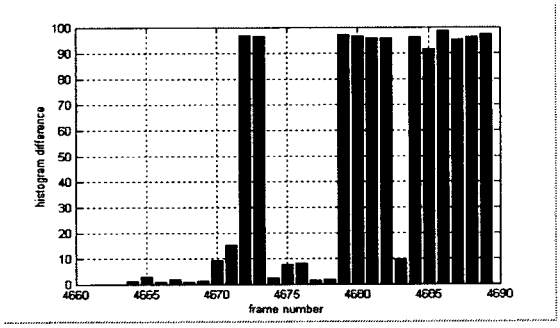

그림 2. 카메라 플래시 영상의 히스토그램 차이

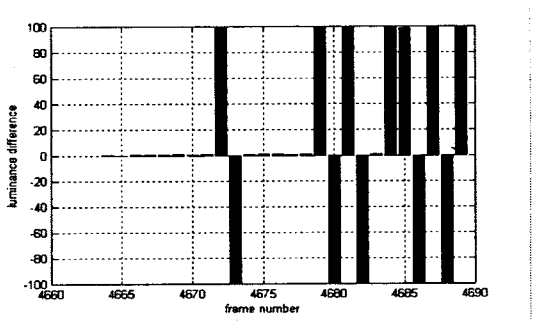

그림 3. 카메라 플래시 영상의 밝기 변화

VI. 맺음말

본 논문에서는 압축 영역과 화소 영역을 결합하여 장면 전환을 검출하는 방법을 제안하였다. 실험을 통하여 이 방법이 검출의 정확성을 유지하면서 검출 속도를 향상시킨다는 사실을 확인하였다. 화소 영역에서는 명암 변화가 일정한 임계값을 넘는 경우에만 윤곽선 검출을 하도록 하여 모든 프레임에 대해 윤곽선 검출을 하는 방법보다 처리 시간을 빠르게 하였다. 화소의 밝기 변화를 분석하여 카메라 플래시를 검출하는 방법은 뉴스 영상 등에 효과적으로 적용될 수 있을 것이다.

참고문헌

[1] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," *ACM Multimedia System*, Vol. 1, No. 1, pp. 10-28, 1993.

[2] F. Arman, A. Hsu, and M. Y. Chiu, "Image processing on compressed data for large video databases," *Proceedings 1st ACM International Conference on Multimedia*, pp. 267-272, Aug. 1993.

[3] J. Meng, Y. Juan, and S. F. Chang, "Scene change detection in a MPEG compressed video sequence," *Proc. IS&T/SPIE 2419*, pp. 14-25, Feb. 1995.

[4] Ramin Zabih, Justin Miller, and Kevin Mai, "A feature-based algorithm for detecting and classifying scene breaks," *Proceedings 4th ACM International Conference on Multimedia*, pp. 189-200, Nov. 1995.