

압축 비디오 영상에서의 점진적인 장면전환 검출

Gradual Scene Change Detection in a Compressed Video Sequence

윤석준, 지은석, 고성재

고려대학교 공과대학 전자공학부 영상통신연구소

Seok-joong Yoon, Eun-Seok Ji, and Sung-jea Ko

Dept. of Electronic Engineering, Korea University

요 약

본 논문에서는 비디오 장면 전환 효과 중 디졸브(dissolve)와 카메라 동작에 의한 팬(pan), 줌(zoom)에 의한 점진적인 장면 전환 구간을 검출하는 알고리즘을 제안한다. 디졸브 검출의 처리 속도 향상을 위하여 MPEG 표준에 따라 압축된 영상을 최소한의 복원 과정을 거쳐서 얻은 DC 영상을 사용하였다. 디졸브 특성의 효과적 추출을 위하여 공간적 영역 분할을 하였고, 각 영역별 평균값을 제공하여 이에 대한 프레임별 평균을 전환구간의 특징 요소로 이용하였다. 추출된 요소들을 이용하여 정확한 디졸브 구간의 검출을 위하여 비중첩 윈도우 비교 방식을 사용하였다. 또한 효과적인 카메라 동작 검출을 위해 배경화면을 대표할 수 있는 지역 매크로 블럭의 움직임 벡터를 추출하였다. 이를 이용하여 카메라 동작에 의한 효과를 검출하였다.

I. 서 론

최근 비디오 정보의 급격한 증가에 따라 비디오 정보를 효율적으로 다루기 위하여 각 장면별로 특징을 부여하는 비디오 분류(segmentation) 작업이 필요하다[1]. 이를 위해 장면이 전환되는 경계를 효과적으로 검출할 수 있는 기술이 요구된다. 이러한 장면 전환은 순간적으로 다른 장면으로 변하는 급격한 장면 전환과 점차적으로 다른 장면으로 교체되는 점진적인 장면 전환으로 구분된다. 점진적인 장면 전환의 종류로는 비디오 편집과정에서 삽입되는 페이드(fade), 디졸브, 와이프(wipe)등이 있고, 카메라 움직임에 의한 팬, 줌, 스크롤(scroll)등이 있다. 본 논문에서는 점진적인 장면 전환 효과 중 디졸브, 카메라 동작 중 팬, 줌, 스크롤의 검출을 다루었다.

기존의 디졸브 검출 방법 중 완전 복원된 영상(fully decoded video)에서의 연속된 프레임의 분산 분포를 이용한 방법은 연산량이 많고[5], 축소 영상에서는 DC 계수를 이용한 복원과정에서 생긴 영상 열화에 따른 문제가 있다[4]. 따라서 계산량을 줄이면서 디졸브 특성을 보존하기 위하여 프레임 평균의 시간에 따른 분포를 이용한다. 이 때 전환되는 두 영상의 평균 차이가 큰 경우에는 이러한 특성이 뚜렷하게 나타나기 때문에 검출이 가능하다. 그러나 픽셀의 공간적 분포가 다른 두 영상에서 프레임 전체의 평균이 비슷한 경우, 해당 구간의 검출이 불가능해진다. 본 논문에서는

비슷한 평균을 갖는 두 영상이 교체되는 경우에도 프레임을 여러 영역으로 나누어 각 영역에 대한 평균 제곱의 합(sum of squared means)을 이용하여 비디오 영상에서 정확한 디졸브 특성을 추출하고, 비중첩 윈도우 비교(non-overlapping window comparison)를 통하여 원하는 디졸브 구간을 검출하는 방법을 제안하였다.

기존의 카메라 동작에 의한 점진적인 장면전환 검출은 프레임 전체의 움직임 벡터를 분석하기 때문에 물체의 움직임이나 디졸브, 와이프에 의한 오차로 인해 검출 오류를 발생할 수 있다[2]. 이를 개선하기 위해 본 논문에서는 배경화면의 특성을 잘 나타내는 프레임 모서리 영역에서 추출된 대표 움직임 벡터의 방향을 관찰하여 여러 가지 카메라 동작을 검출하는 방법을 제안하였다.

본 논문의 구성은 다음과 같다. II장에서는 디졸브 검출에 관한 기존의 방법들과 제안한 전환 검출 방법에 관하여 기술하였고, III장에서는 팬, 줌에 의한 전환을 검출하기 위한 제안한 기법을 설명하고, IV장에서는 실험 영상을 통하여 기존의 기법들과 제안한 방법을 비교 분석하고 결론을 맺는다.

II. 압축 비디오에서의 디졸브 검출 기법

디졸브는 두 장면을 연결할 때 두 장면이 중첩된 상태로 기존 장면의 밝기는 점차 감소하고 새로운 장면의 밝기는 점차 증가하여 화면의 교체가 이루어지는 편집 효과이다. 전환 구간에서 교체되기 이전과 이후의 영상을 각각 $f_1(t)$, $f_2(t)$ 라 할 때 디졸브가 일어나는 구간은 다음과 같이 표현된다.

$$f(t) = [1 - \alpha(t)] \times f_1(t) + \alpha(t) \times f_2(t) \quad (1.1)$$

$$\alpha(t) = \begin{cases} 0, & t < t_1 \\ (t - t_1)/(t_2 - t_1), & t_1 \leq t \leq t_2 \\ 1, & t > t_2 \end{cases} \quad (1.2)$$

여기서 전환의 시작을 t_1 , 전환의 끝을 t_2 라 한다. 그리고 $\alpha(t)$ 는 전환구간에서 시간에 따라 0에서 1까지 증가하는 선형 함수이다. 예를 들면 $f(t)$ 는 $f_1(t)$ 가 0일 때 fade-in이 되고, $f_2(t)$ 가 0일 때 fade-out이 된다. 디졸브 특성을 찾

는 기존의 대표적인 방법은 프레임별 분산의 시간적 분포를 이용한 것과 프레임간 차를 이용한 방법이 있다. 우선, 분산 분포를 이용하는 방법은 디졸브 구간에서 명확한 [-자형의 프레임 분산 분포 곡선을 나타내어야 한다. 완전 복원된 영상에서는 복원 과정에서의 오차 발생이 적기 때문에 이러한 형태의 곡선이 잘 나타난다. 그러나 DC 계수와 움직임 벡터를 이용하여 복원한 DC 영상에서는 프레임간 예측 과정에서 발생하는 오차로 인하여 영상 왜곡 현상이 발생한다 [3]. 따라서 DC영상에 이 기법을 직접 적용하는 것은 부적당하다. 두 번째 방법은 프레임간 차를 이용하는 방법으로 식 (2)와 같이 연속되는 두 프레임간 픽셀 밝기 차의 절대값들의 합(sum of absolute differences between pixels: P_i)이 주변의 값보다 월등히 크게 나타나는 구간을 디졸브 전환 구간으로 간주하는 것이다[4].

$$P_i = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} |P_i(k, l) - P_{i-1}(k, l)| \quad (2)$$

이 때 $P_i(k, l)$ 는 $M \times N$ 의 크기를 갖는 영상에서 i 번째 프레임의 (k, l) 번째 픽셀의 밝기값을 의미한다. 그러나 두 프레임간 픽셀의 밝기 차를 이용할 경우 영상 내의 물체나 카메라의 움직임이 있을 때에도 프레임간 차이가 크게 나타나기 때문에 이를 그대로 적용할 수는 없다. 이를 보완하여 프레임간 차이를 구하기 위해 픽셀 대신 식 (3)(sum of absolute differences histogram bins: H_i)과 같이 프레임별 픽셀 밝기의 히스토그램을 사용하였다[6].

$$H_i = \sum_{k=0}^{L-1} |H_i(k) - H_{i-1}(k)| \quad (3)$$

위의 $H_i(k)$ 는 i 번째 프레임에서 L 개의 밝기 레벨을 갖는 픽셀 히스토그램이다. 프레임 내에 물체의 움직임이 있어도 전체적인 픽셀의 히스토그램은 크게 변하지 않기 때문에 픽셀을 직접 차분하는 방법보다 효과적이다. 그러나 이 방법도 DC 영상에 그대로 적용하면 프레임 내 픽셀에 대한 정보의 손실이 증가되어 검출 오류를 발생시킨다. 이러한 문제를 해결하기 위해 제안한 디졸브 검출 방법은 그림 1과 같이 프레임을 여러 영역으로 나누어 식 (4)와 같이 영역별 평균의 제곱을 평균하여 얻은 S 를 프레임에 대한 특징 요소로 사용한다.

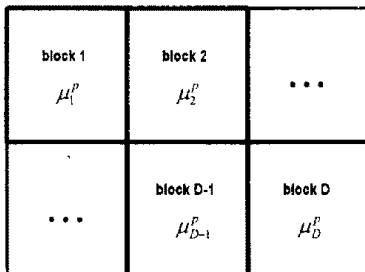


그림 1. 프레임의 공간적 분할

$$S = \frac{1}{D} \sum_{j=0}^{D-1} \mu_j^p \quad (4)$$

위의 식에서 μ_j^p 는 해당 프레임을 D 개의 영역으로 분할하였을 때 j 번째 분할 영역에서 구한 픽셀 평균값의 p 승을 의미한다. 디졸브 구간에서 식 (4)에 의한 S 의 변화는 선형적인 증가 또는 감소 형태로 나타난다. 전환이 일어날 때 S 는 선형 함수가 되고 이를 시간에 대하여 일차 미분하면 해당 구간에서의 펄스파 형태의 디졸브 크기를 나타내게 된다. 실제 영상에서 전환이 발생하는 구간에서도 카메라 또는 피사체 움직임의 영향으로 이상적인 펄스파 형태가 나타나지 않는다. 따라서 정확한 검출을 위하여 전환 특성이 나타난 구간을 디졸브 프레임으로 결정하는 과정이 필요하다. 그림 2는 제안된 비중첩 윈도우 비교 방식에 의한 전환 구간 검출 과정을 나타낸 순서도이다.

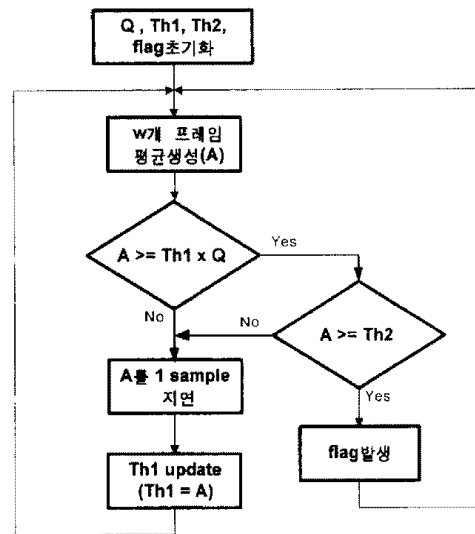


그림 2. 디졸브 구간 검출도

우선 해당 구간이 디졸브임을 나타내는 신호(flag), 두 임계값(Th1, Th2), 그리고 임계값(Th1)과의 비교를 위한 Q 값을 초기화한다. 여기서 Th1은 윈도우 비교를 위한 임계값으로 검출 과정동안 디졸브 구간이 아닌 경우에 계속 변경되는 값이다. Th2는 최소 디졸브 크기를 나타내는 전역 임계값이다. 디졸브 특성 검출 과정에서 나온 결과로부터 w 개씩 프레임 평균(A)을 구한다. 이와 같이 얻은 d 번째 윈도우의 평균의 크기(A_d)가 두 번째 이전의 평균(A_{d-2})의 Q 배보다 클 경우 디졸브 신호를 발생한다. 만약 임계값보다 작을 경우 디졸브 신호는 발생하지 않고, 그 때의 평균값을 한 샘플(sample) 지연시킨다. 지연된 평균값은 새로운 임계값(Th1)으로 대체한다. 즉, 현재 평균과 비교할 임계값은 항상 디졸브 구간을 제외한 두 번째 이전 평균값이 된다. 이는 w 프레임씩 평균을 취할 때 디졸브 구간의 일부가 포함되어 있을 경우 임계값이 높아져 발생하는 검출 오류를 방지하기 위해서이다.

III. 카메라 동작 검출 기법

프레임 전역에서 움직임 벡터의 분포를 이용하여 카메라 동작을 검출하는 기존의 방법은 물체의 움직임이나 디졸브, 화이트와 같은 비디오 편집 효과가 있는 경우 불규칙적인 움직임 벡터가 있기 때문에 잘못된 결과가 발생 할 수 있다. 또한 모든 매크로 블록의 움직임 벡터 방향을 찾기 위해 복잡한 계산식이 필요하며, 좀 효과를 검출하는 기법은 정확도가 매우 낮다.

본 논문에서 제안한 방법은 전술한 단점을 개선하기 위하여 MPEG 비디오에서 모든 움직임 벡터를 추출하지 않고 피사체의 움직임 영향을 크게 받지 않는 영역의 움직임 벡터만을 추출하여 이용하였다. 그림 3은 제안한 카메라 동작에 의한 팬, 줌을 검출하는 블록도이다. I 프레임에는 움직임 벡터가 없기 때문에 이전 GOP의 마지막 B 프레임 움직임 벡터의 역을 취해 얻었고, 피사체는 일반적으로 카메라 중앙에 있으므로 물체의 움직임으로 인한 잘못된 대표 움직임 벡터의 검출을 방지하기 위해 프레임 모서리 부분만을 관찰하여 배경화면의 움직임 특성을 나타낼 수 있도록 하였다.

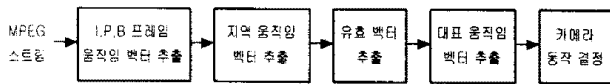


그림 3. 카메라 동작에 의한 팬, 줌 검출도

매크로 블록별 움직임 벡터는 이전화면에서 변화가 없을 경우 이전 매크로 블록에서의 움직임 벡터를 가져오고, 변화가 클 경우 부호화 과정에서 정한 움직임 추정 범위를 벗어나기 때문에 그 매크로 블록에서는 블록 내 부호화를 하게 되어 움직임 벡터가 없다. 따라서 각 영역별로 움직임 벡터가 있는 매크로 블록의 유효 움직임 벡터를 식 (5)와 같이 구한다.

$$V_n = \{ (x_1, y_1), (x_2, y_2), \dots, (x_a, y_a) \} \quad (5)$$

$V_n (n=1, 2, 3, 4)$ 은 4개의 영역에 대한 유효 움직임 벡터의 집합이다. (x_a, y_a) 는 각 영역의 유효 움직임 벡터이고, a 는 유효 움직임 벡터의 개수이다. 각 영역별 유효한 움직임 벡터들을 영역별 수평, 수직방향으로 각각 합하고 유효 벡터 수로 나누어 식 (6)과 같이 4개의 영역에 대한 각각의 대표 움직임 벡터(local motion vector, l)를 구한다.

$$l_{n,x} = \frac{1}{a} \sum_{i=0}^a V_n(x_i) \quad (6a)$$

$$l_{n,y} = \frac{1}{a} \sum_{i=0}^a V_n(y_i) \quad (6b)$$

여기서 $l_{n,x}$, $l_{n,y}$ 는 네 영역에 대한 수평, 수직 움직임 벡터를 나타내는 것이다. 이러한 대표 움직임 벡터의 수평, 수직 방향 성분을 분석하여 그림 4와 같이 팬, 스크롤, 줌 등의 카메라 동작을 구별할 수 있다.

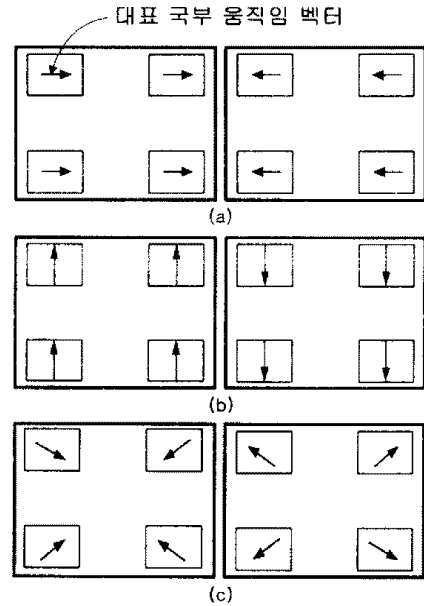


그림 4. 대표 국부 움직임 벡터에 따른 카메라 동작 (a) 팬, (b)스크롤, (c)줌

IV. 실험 결과 및 분석

디졸브 검출 실험에 사용된 영상은 352 X 240의 크기에 총 186프레임 중 4번의 디졸브가 포함되어 있다. 세 번째 전환의 시작부터 네 번째 전환이 시작하기 전까지(98-145) 비교적 느린 카메라 팬 효과가 있다.

그림 5의 그래프는 기존 방법에 의한 디졸브 영상의 검출 결과이다. 완전하게 복원된 영상의 분산은 뚜렷한 U-자형 곡선을 이루기 때문에 Meng[3]이 제안한 분산 차에서 전환 구간의 가운데 부분에 극소점을 중심으로 시작과 끝 부분에서의 극대점이 그림 5(a)에서처럼 일정 거리차이가 있어 디졸브에 의한 전환 구간임을 알 수 있다. 그러나, 축소 영상에서는 이러한 U-자형 곡선 형태가 잘 나타나지 않기 때문에 그림 5(b)와 같이 두 번째 구간 외에는 극대점과 극소점의 경계가 불분명하여 전환구간 검출이 불가능하다. 복원된 영상의 픽셀간 차의 합은 디졸브 구간에서 고원(plateau)이 잘 나타나지만, 카메라 움직임에 의한 팬 구간에서도 고원이 나타나 검출 오류를 발생시킨다. 반면 DC 영상으로부터 얻은 결과에서는 그림 5(c)처럼 축소영상을 만드는 과정이 공간적 저역 필터 역할을 하기 때문에 팬에 의한 오차가 약간 감소한다. 프레임별 히스토그램 차의 합을 이용할 경우 완전 복원된 영상에서는 픽셀 차의 합에 의한 방법에 비해 물체나 카메라의 움직임에 의한 영향을 비교적 받게 디졸브 특성이 잘 나타난다. 그러나 그림 5(d)와 같이 축소 영상을 예측 복원하는 과정에서 누적되는 오차로 인해 열화된 B프레임과 열화가 없는 I 프레임의 사이에서 급격한 차이를 나타내고 있다. 이러한 현상은 픽셀간 차를 이용한 방법에서도 생기는 것으로 이를 제거하기 위해 복잡한 후처리과정이 요구된다.

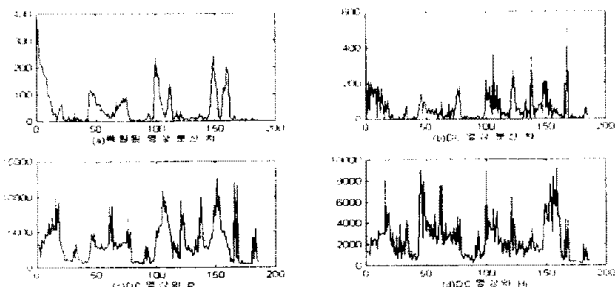


그림 5. 기존 방법에 의한 실험 결과

그림 6은 제안된 방식에 의해 얻어진 디졸브 검출 결과이다. (a)는 전환 특성을 나타내는 S이고 (b)는 (a)를 일차 미분한 결과이다. (c)는 비중첩 윈도우 비교 방법을 이용하여 얻은 디졸브 신호를 원신 복원 영상의 분산분포와 비교한 것이다. 검출된 디졸브 구간이 오른쪽으로 약간 이동한 것은 비중첩 윈도우 평균을 만드는 과정에서 생긴 오차이다.

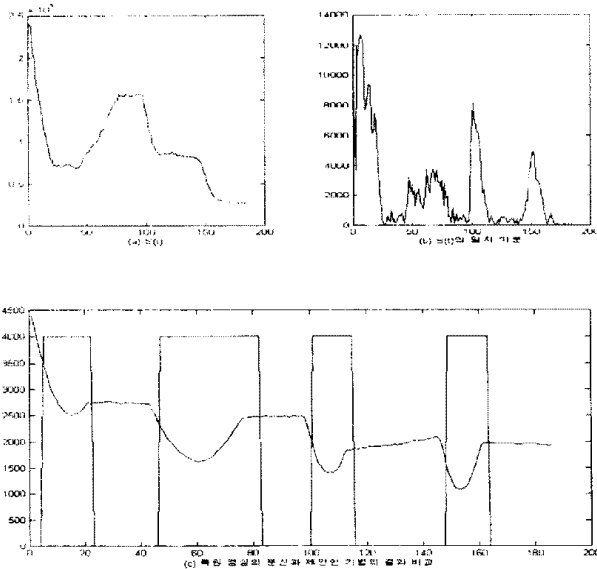


그림 6. 제안한 방법에 의한 실험 결과

카메라 동작 검출을 위해 총 303프레임의 길이를 갖는 실험 영상을 사용하였다. 실험 영상 내에는 60~101프레임 사이에 카메라 줌, 105~178프레임은 손 떨림에 의한 카메라 움직임이 있다. 또한 179프레임부터 끝까지의 카메라 팬과 스크롤이 연속하여 일어나는 구간이다. 여기서 그림 7(a)는 제안한 카메라 검출기를 통하여 얻은 결과를 그대로 나타낸 것이다. 그래프에서 Y축 방향의 크기가 20이면 줌을 나타내고, 10이면 팬을 나타낸다. 그림 7(b)는 (a)의 결과를 3택 미디언 필터 처리한 결과이다.

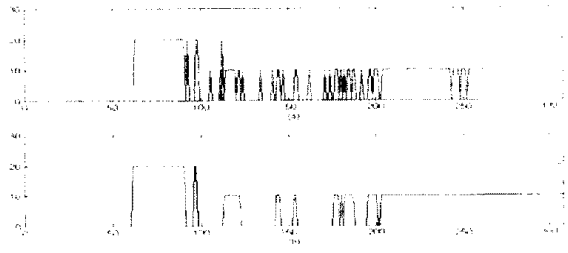


그림 7. 제안한 카메라 동작 검출 기법의 결과

V. 결론

본 논문에서는 디졸브와 카메라 동작에 의한 점진적인 장면 전환 검출 방법을 제안하였다. 시간적인 검출 정밀도를 높이기 위해 모든 종류의 프레임을 사용하였고 빠른 검출을 위해 I 프레임의 DCT DC계수와 P, B 프레임의 움직임 벡터를 통해 복원된 DC 영상을 이용하였다. 제안한 디졸브 특성 검출기법과 비중첩 윈도우 비교 방식의 구간 검출 기법은 기존의 방법들에 비해 물체나 카메라의 움직임에도 정확한 디졸브 구간을 검출하였다. 카메라 동작 검출을 위하여 배경 화면의 움직임만을 분석하기 위해 프레임의 가장자리에 네 개의 영역으로 나누어 영역별 대표 움직임 벡터의 방향을 구하고, 이를 이용하여 카메라 동작 중 팬, 줌의 구간을 검출하였다.

이 논문은 1998년 학술진흥재단의 공모과제 연구비에 의하여 연구되었음

참고 문헌

- [1] S. Smoliar and H. Zhang, "Content-based video indexing and retrieval," *IEEE Multimedia*, vol. 1, pp. 62-72, 1994.
- [2] O. N. Gerek and Y. Altunbasak, "Key Frame Selection from MPEG Video Data," *SPIE*, vol. 3024, pp. 920-925, 1997.
- [3] J. Meng, Y. Juan, and S. F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence," *Digital Video Compression: Algorithms and Technologies, SPIE*, vol. 2419, pp. 14-25, 1995.
- [4] B. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 533-544, Dec. 1995.
- [5] A. M. Alattar, "Detecting and Compression Dissolve Regions in Video Sequences with DVI Multimedia Image Compression Algorithm," *ISCAS*, pp. 13-16, 1993.
- [6] F. Armam, A. Hsu, and M. Y. Chu, "Image Processing on Compressed Data for Large Video Databases," *Proc. of ACM Multimedia '93*, pp. 267-272, June 1993.