DBpia

논문 99-24-8B-20 한국통신학회논문지 '99-8 Vol.24 No.8B

MPEG 영상에서의 점진적 장면전환에 대한 효과적인 검출 기법

정회원 윤 석 중*, 지 은 석**, 김 영 로**, 고 성 제**

Effective Detection Techniques for Gradual Scene Change on MPEG Video

Seok-Joong Yoon*, Eun-Seok Ji**, Young-Ro Kim**, Sung-Jea Ko** *Regular Members*

요 약

점진적인 장면전환 종류로는 비디오 편집 과정에서 삽입되는 디졸브(dissolve), 페이드인/아웃(fade-in/out), 와이프(wipe) 등과, 카메라 움직임에 의한 팬(pan), 줌(zoom), 스크롤(scroll) 등이 있다. 본 논문에서는 점진적 장면전환 종류 중 편집과정에서 사용되는 디졸브, 그리고 카메라 움직임에 의한 팬과 줌의 검출을 위한 효과적인 방법을 제안하였다. 디졸브 구간의 검출을 위하여 각 프레임에 대한 공간적 영역 분할을 하고, 영역별 화소값의 평균치를 제곱하여 전환 프레임의 특징 요소로 사용하였다. 팬과 줌의 검출을 위하여 배경화면을 대표할 수 있는 4개의 국부 영역으로 선정하여 각 국부 영역의 움직임 벡터로부터 대표 움직임 벡터를 결정하고, 이들의 방향 성분을 사용하였다. 제안된 방식들은 검출의 정확도를 향상시키기 위해 모든 프레임(I, P, B)을 사용하였으며 아울러 검출 속도를 높이기 위해 완전 복원된 영상을 사용하는 대신 DCT DC 계수와 움직임 벡터를 이용한 축소영상을 사용하였다. 실제 MPEG 비디오에 적용하여 기존의 방법들에 비해 우수한 검출 결과를 확인하였다.

ABSTRACT

In this paper, we propose detection methods for gradual scene changes such as dissolve, pan, and zoom. The proposal method to detect a dissolve region uses scene features based on spatial statistics of the image. The spatial statistics to define shot boundaries are derived from squared means within each local area. We also propose a method of the camera motion detection using four representative motion vectors in the background. Representative motion vectors are derived from macroblock motion vectors which are directly extracted from MPEG streams. To reduce the implementation time, we use DC sequences rather than fully decoded MPEG video. In addition, to detect the gradual scene change region precisely, we use all types of the MPEG frames(I, P, B frame). Simulation results show that the proposed detection methods perform better than existing methods.

I. 서 론

비디오 정보를 효율적으로 활용하기 위해 자료의 검색, 색인, 그리고 편집 기능 등이 필요하다. 이러한 기능을 구현하기 위해서 비디오의 각 장면별로 색인과 특징을 부여하는 비디오 분류 작업이 선행되어야 한다. 비디오는 여러 장면으로 구성되어 있기 때문에 각 장면이 바뀌는 부분을 검출함으로써 비디오 분류를 효과적으로 할 수 있다 [1]-[5]. 장면전환 종류에 따라 급격한 장면전환, 점진적인 장면전환 등으로 구분할 수 있다. 급격한 장면전환은 한 프레임만에 순간적으로 장면이 변하는 것이고, 점진적인 장면전환은 여러 프레임에 거쳐 서서히 장면이 전환되는 것이다.

* 삼성전자 반도체총괄 시스템 LSI media팀 ** 고려대학교 전자공학과

논문번호 : 99026-0121, 접수일자 : 1999년 1월 21일

www.dbpia.co.kr

점진적인 장면전환 중에서 디졸브 특성 및 구간을 찾는 기존의 대표적인 방법들은 분산, 프레임간 화소값 또는 히스토그램 차, 움직임 벡터, 그리고 에지 성분 등을 사용하였다 [6]-[9]. Meng [6]은 B, P 프레임 움직임 벡터의 통계적인 분석과 I, P 프레임의 축소 영상의 분산 분포를 이용하는 방법을 제안하였다. Yeo [7]는 모든 프레임(I, P, B)에 대한 최소한의 복원 과정을 통해 생성된 축소 영상의 프레임간 화소값 차를 전환 특성으로 사용하였다. Zhang [8]은 화소값 대신 프레임별 화소값에 대한 히스토그램의 차를 사용하였다. Yeo와 Zhang의 방법은 GOP 단위가 클 경우 축소영상의 열화가 생기므로 부정확한 디졸브 구간 검출을 할 수 있다.

카메라 동작을 검출하는 기존의 대표적인 방법들은 프레임 전 영역에서 시공간(spatiotemporal) 투시, 움직임 벡터, Hough 변환 등을 이용하였다 [10]-[13]. 이 방법들은 물체의 움직임이나 디졸브, 와이프와 같은 비디오 편집 효과로 인하여 여러 방향의 움직임 벡터가 불규칙적으로 발생하면 잘못된 카메라 동작 검출 결과를 얻을 수 있다.

본 논문에서는 정확한 디졸브 전환 구간을 검출하기 위하여 공간적 영역 분할에 의한 특징 요소 추출하고, 이 특징 요소를 프레임 윈도우(window) 단위로 비교하는 방법을 제안하였다. 또한 카메라 동작 구간을 검출하기 위하여 배경화면을 대표할 수 있는 4개의 국부 영역으로 선정하여 각 국부 영역의 움직임 벡터로부터 대표 움직임 벡터를 결정하고, 이들의 방향 성분을 이용하는 방법을 제안하였다. 실험을 통하여 제안한 알고리즘과 기존 알고리즘을 비교 분석하였고 우수한 장면 변환 검출 성능을 가짐을 보였다.

Ⅱ. 제안한 디졸브 특성 및 검출 방법

1. 장면전환 검출을 위한 DC 영상 추출

DCT 계수 중 DC 값은 8x8 블럭의 평균값이다. 따라서 DC 성분만으로 구성된 축소 영상은 원영상의 특성을 잘 나타낸다. 지금까지는 장면전환 검출 수단으로 I 프레임 또는 I, P 프레임으로부터 추출된 DC 영상이 주로 사용되었다 [6]. MPEG 영상은 GOP 단위로 구성되어 있기 때문에 GOP 단위가 작을 경우는 I 프레임만 이용하여 대부분의 장면전환 검출을 할 수 있다. 그러나 GOP의 단위가 클 경우(12 프레임 이상)는 B, P 프레임에서 장면전환이 일어날 경우 잘못된 프레임에서 장면전환 검출을 할 수 있다. 따라서 정확한 구간에서의 장면전환 검출을 하기 위해서 I 프레임뿐 아니라 B, P 프레임의 DC 영상을 이용하여 시간적 정밀도를 높일 필요가 있다. DC 영상은 수평, 수직 방향으로 각각 1/8 크기로 줄어 원 영상의 1/64 크기로 축소된 영상이다. I 프레임의 경우 모든 블럭이 블럭내 부호화(intra coding) 되기 때문에 각 블럭의 DC 값으로부터 직접 DC 영상을 만든다. B, P 프레임에서는 이전 기준 프레임(I 또는 P 프레임)의 DC 값과 움직임 벡터를 이용하여 DC 값을 구하여 DC 영상을 만든다 [14].

2. 디졸브 특성 및 구간 검출 기법

디졸브는 두 장면이 중첩된 상태로 기존 장면의 화소값은 점차 감소하고 새로운 장면의 화소값은 점차 증가하여 화면의 교체가 이루어지는 비디오 특수 편집 효과이다. 전환 구간에서 교체되기 이전과 이후의 영상을 각각 $f_1(t)$, $f_2(t)$라 할 때 디졸브가 일어나는 구간은 식(1a)와 같이 표현된다.

$$f(t) = [1-\alpha(t)] \times f_1(t) + \alpha(t) \times f_2(t) \tag{1a}$$

$$\alpha(t) = \begin{cases} 0, & t < t_1 \\ (t-t_1)/(t_2-t_1), & t_1 \le t \le t_2 \\ 1, & t > t_2 \end{cases} \tag{1b}$$

여기서 전환 시작 시간을 t_1, 전환의 끝 시간을 t_2라 할 때 $\alpha(t)$는 식 (1b)와 같이 전환구간에서 시간 t 에 따라 0에서 1까지 증가하는 선형함수이다. 한편 $f(t)$는 $f_1(t)$가 0일 때 fade-in이 되고, $f_2(t)$가 0일 때 fade-out이 된다.

완전 복원된 영상(fully decoded video)을 이용한 기존의 디졸브 검출 방법은 연산량이 많아 빠른 검출을 할 수 없고, 축소 영상에서 연속적인 프레임 분산 분포를 이용한 방법은 디졸브 특성을 나타내는 포물선 형태의 곡선이 뚜렷하게 나타나지 않는 문제가 있다. 따라서 계산량을 줄이면서 디졸브 특성을 보존하기 위하여 프레임 평균의 시간에 따른 변화를 이용할 수 있다. 이 방법은 전환되는 두 영상의 평균 차이가 큰 경우에는 전환 특성이 뚜렷하게 나타나기 때문에 정확한 디졸브 구간 검출이 가능하다. 그러나 두 영상의 프레임 전체 평균이 비슷한 경우 전환 구간에서도 평균값의 변화가 나타나지 않아서 정확한 디졸브 구간을 검출하지 못하는 단점이 있다. 이러한 문제를 해결하기 위해 그림 1과 같이 프레임을 여러 영역으로 나눈 다음,

www.dbpia.co.kr

식 (2)과 같이 영역별 화소 평균의 제곱을 합하여 얻은 S를 프레임에 대한 특징 요소로 사용하는 디졸브 검출 방법을 제안하였다.

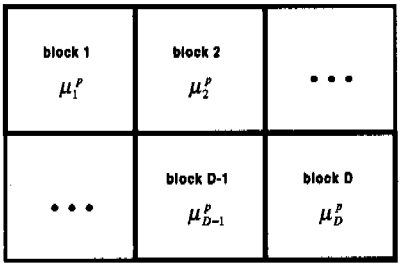


그림 1. 프레임의 공간적 분할.

$$S=\sum_{j=1}^{D}\mu_j^p \tag{2}$$

위의 식에서 D는 분할된 영역의 개수이고, μ_j^p는 해당 프레임의 j번째 분할 영역에서 구한 화소 평균값의 p승을 의미한다. 그림 2는 4개의 디졸브 전환 구간이 있는 영상에서 프레임 평균과 제안한 분할 영역의 평균 제곱의 합을 나타낸 것이다. 세로축은 각 특성 분포에서 첫 번째 프레임을 기준하여 정규화된 값(normalized value)이다. 영상의 프레임 평균을 보면 세 번째 구간을 제외한 나머지 전환 구간에서 장면전환에 따른 변화가 뚜렷하다. 세 번째 구간처럼 프레임내 화소값의 공간적 분포가 다름에도 불구하고 프레임 평균이 비슷한 두 영상의 경우, 전환 구간에서 평균의 변화가 거의 없어 검출이 불가능해진다. 반면에 제안한 방법에 의한 전환 특성을 보면 영역을 6, 24, 그리고 96개로 나누었을 때 4개의 전환 구간에서 뚜렷한 선형함수 형태를 나타난다. D개의 영역으로 분할된 프레임 A의 평균 M_A과 식 (2)의 p가 2인 경우 제안한 전환 특성 S를 분할 수로 나눈 T_A는 식 (3)과 같다.

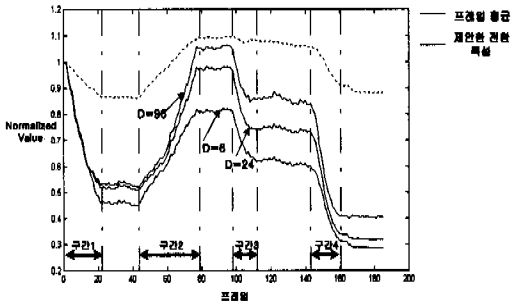


그림 2. 프레임 평균과 제안한 전환 특성.

$$M_A=\frac{1}{D}\sum_{j=1}^{D}u_{Aj} \tag{3a}$$

$$T_A=\frac{1}{D}\sum_{j=1}^{D}u_{Aj}^2 \tag{3b}$$

여기서 식(3a)를 제곱하면 식 (4)가 된다.

$$M_A^2=\left(\frac{1}{D}\sum_{j=1}^{D}u_{Aj}\right)^2=\frac{1}{D}\left\{\frac{1}{D}\left(\sum_{j=1}^{D}u_{Aj}\right)^2\right\} \tag{4}$$

제안한 전환 특성 S는 T_A의 분할 수로 곱한 것으로 프레임 평균과 평균의 제곱 M_A, M_A^2보다 크거나 같을 경우 $(D\times T_A\geq M_A^2\geq M_A)$, 전환 특성 더 잘 나타냄을 알 수 있다. 이 관계를 만족시키기 위하여 프레임 평균 제곱과 T_A의 다음 비교식에서 D가 2라면 식 (6)과 같이 된다.

$$\sum_{j=1}^{D}u_{Aj}^2 \geq \frac{1}{D}\left(\sum_{j=1}^{D}u_{Aj}\right)^2 \tag{5}$$

$$(u_{A1}^2+u_{A2}^2) \geq \frac{1}{2}(u_{A1}+u_{A2})^2 \tag{6}$$

식 (6)를 다시 쓰면 식 (7)과 같이 전개된다.

$$(u_{A1}^2+u_{A2}^2)-\frac{1}{2}(u_{A1}+u_{A2})^2=\frac{1}{2}(u_{A1}^2+u_{A2}^2 -2u_{A1}\times u_{A2})=\frac{1}{2}(u_{A1}-u_{A2})^2 \geq 0 \tag{7}$$

디졸브가 있는 연속하는 두 프레임을 I, J라 할 때 $u_{I1}=u_{I2}=\cdots=u_{ID}$, $u_{J1}=u_{J2}=\cdots=u_{JD}$가 아니라면 식 (7)의 관계식에 따라 다음 식이 성립이 된다.

$$S_I \gg M_I^2 \geq M_I,\ S_J \gg M_J^2 \geq M_J \tag{8}$$

따라서 제안한 방법과 평균차를 비교하기 위하여 식 (9)에 양변에 제곱하면 식 (10)을 얻는다.

$$|S_I-S_J| \geq |M_I-M_J| \tag{9}$$

$$S_I^2+S_J^2-2S_I\times S_J \geq M_I^2+M_J^2-2M_I\times M_J \tag{10a}$$

$$(S_I^2-M_I^2)+(S_J^2-M_J^2)-2(S_I\times S_J-M_I\times M_J) \geq 0 \tag{10b}$$

식 (10b)는 식 (8)의 관계식에 의하여 연속하는 디졸브 영상에서의 S_I, S_J가 같지 않다면 식 (9)은

www.dbpia.co.kr

항상 성립한다. 따라서 연속하는 두 프레임의 평균차 $|M_I - M_J|$보다 프레임을 여러 영역으로 분할하여 각각의 평균의 제곱의 합의 차 $|S_I - S_J|$가 디졸브 전환 특성을 잘 나타냄을 보였다. 또한 식 (2)에서 분할 영역의 평균 p제곱 파라메터가 클수록 식 (8)의 관계식은 더욱 차이가 나 식 (9)에 의한 특성 차이는 더욱 뚜렷하게 나타난다.

디졸브 전환 구간에서 S는 선형함수 분포를 나타내고 연속되는 두 프레임간 S값의 차는 펄스파 형태로 나타난다. 그러나 장면전환이 발생하는 구간에서 카메라 또는 피사체 움직임의 영향이 있을 경우에 이상적인 펄스파 형태가 나타나지 않는다. 따라서 정확한 검출을 위하여 특성 S에 대한 w개 프레임 단위로 윈도우를 정하여 이들을 비교하는 방식을 제안하였다. 그림 3은 제안한 비 중첩 윈도우 비교 방식을 이용하여 전환 구간을 검출하는 과정을 나타내는 순서도이다. 우선 프레임의 디졸브를 나타내는 신호 *flag*, 최소 디졸브 크기를 나타내는 임계값 T, 영상의 특성에 따라 변하는 임계값 V, 비교 파라메터 Q를 초기화한다. 여기서 V는 윈도우 비교를 위한 임계값으로 검출과정에서 계속하여 이전의 평균값(A)으로 바뀐다. 그리고, w개 프레임 단위로 S들의 평균 A를 구한다. 현재 얻어진 평균 A를 이용하여 그림 3과 같이 전환 구간을 검출한다. 현재 얻어진 평균 A를 이용하여 그림 5와 같이 전환 구간을 검출한다. 여기서 A가 $V \times Q$보다 크면 A와 최소 디졸브 크기 T를 비교하여 안정적인 디졸브 구간을 검출한다. A가 $V \times Q$ 보다 작으면 A는 다음에 오는 새로운 프레임의 임계값 V가 된다. 즉, 새로운 프레임의 윈도우 평균에 대한 디졸브를 나타내기 위한 비교값을 고정 파라메터를 사용하는 대신 영상의 특성에 따라 변하는 임계값을 얻기 위해서다. 따라서 V 는 임의의 초기값을 사용한다.

Ⅲ. 제안한 카메라 동작 검출 기법

기존 방법들은 프레임 전 영역에서 모든 움직임 벡터를 추출하여 이용함으로써 처리시간이 많이 요구되고 부정확한 카메라 동작 검출을 하는 단점을 가지고 있다. 이러한 단점을 개선하기 위하여 본 논문에서는 프레임의 영역별 대표 움직임 벡터를 이용하여 카메라 동작 검출을 하는 방법을 제안하였다. 그림 4는 제안한 카메라 동작에 의한 팬, 줌을 검출하는 블럭도이다.

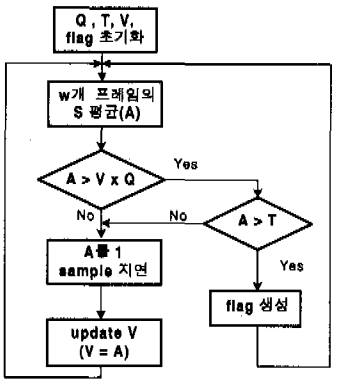


그림 3. 디졸브 구간 검출 방법.

그림 4. 카메라 동작에 의한 팬, 줌 검출도

카메라 동작을 검출하기 위하여 우선 MPEG 스트림에서 매크로 블럭별 움직임 벡터를 추출한다. 여기서 B와 P 프레임의 움직임 벡터는 MPEG 스트림에서 직접 얻을 수 있다. 그리고 I 프레임은 부호화할 때 움직임 벡터가 없기 때문에 다른 프레임으로부터 참조해야 한다. I 프레임은 P, B 프레임으로부터 피 예측 대상이 되는 기준 프레임이기 때문에 전후 모든 예측 프레임에서 움직임 벡터를 얻을 수 있다. 본 논문에서는 I 프레임의 움직임 벡터를 다음과 같이 구하였다. GOP의 마지막 프레임인 B 프레임(I 프레임의 바로 이전 프레임)의 역방향 움직임 벡터(backward motion vector)를 구한 다음, 이 움직임 벡터의 방향 부호를 바꾸어 I 프레임 움직임 벡터를 구한다. 각 프레임에 대한 모든 움직임 벡터를 사용하는 기존의 방법들과는 달리 제안한 방법은 그림 5와 같이 배경화면을 잘 나타내는 국부 움직임 벡터(local motion vectors, LMVs)를 사용하였다.

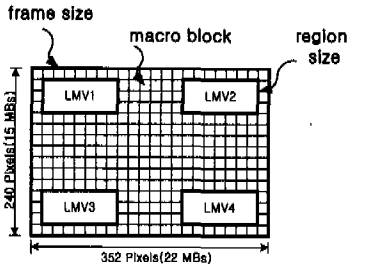


그림 5. 352x240 영상의 매크로 블럭과 국부 영역.

www.dbpia.co.kr

이는 비디오 촬영시 피사체는 일반적으로 프레임의 중앙 부분에 있으므로 중앙화면 변화보다 배경화면의 변화가 카메라 동작을 잘 나타낼 수 있기 때문이다. 본 논문에서는 각 국부 영역마다 수평 7개, 수직 3개 매크로 블럭을 정하여 모두 21개를 사용하였다. 이렇게 얻어진 영역별 움직임 벡터들에서 움직임 벡터가 있는 블럭만을 선별하여 식 (11)와 같이 얻는다.

$$V_1 = \{(x_{11}, y_{11}), (x_{12}, y_{12}), \cdots, (x_{1M}, y_{1M})\} \quad (11a)$$

$$V_2 = \{(x_{21}, y_{21}), (x_{22}, y_{22}), \cdots, (x_{2M}, y_{2M})\} \quad (11b)$$

$$V_3 = \{(x_{31}, y_{31}), (x_{32}, y_{32}), \cdots, (x_{3M}, y_{3M})\} \quad (11c)$$

$$V_4 = \{(x_{41}, y_{41}), (x_{42}, y_{42}), \cdots, (x_{4M}, y_{4M})\} \quad (11d)$$

여기서 (x_{ij}, y_{ij}), $(i=1,2,3,4,\ j=1,2,\ \ldots,\ M)$는 각 영역의 선별된 움직임 벡터들이고, M는 각 영역별 움직임 벡터의 개수이다. 각 영역별로 움직임 벡터들을 수평, 수직방향으로 구분하여 각각 합하고 해당 움직임 벡터수로 나누어 식 (12)와 같이 4개의 영역에 대한 각각의 대표 움직임 벡터 (representative local motion vector, l)를 구한다.

$$l_{n,x} = \frac{1}{M} \sum_{i=1}^{M} x_{ni} \quad (12a)$$

$$l_{n,y} = \frac{1}{M} \sum_{i=1}^{M} y_{ni} \quad (12b)$$

여기서 $l_{n,x}$, $l_{n,y}$, $(n=1,2,3,4)$는 네 개의 각 영역에 대한 수평, 수직 대표 움직임 벡터를 나타내는 것이다. 이렇게 얻어진 각 영역별 대표 움직임 벡터의 방향 성분을 이용하여 카메라 동작을 검출하였다. 표 1은 여러 가지 카메라 동작에 따른 영역별 대표 움직임 벡터의 방향을 움직임 벡터의 부호로 써 나타낸 것이다. 수직 방향의 대표 움직임 벡터의 부호와는 관계없이 수평 방향 움직임 벡터 부호가 동일할 경우 카메라의 팬 동작을 나타낸다. 반대로, 수평 움직임 대표 벡터와는 무관하게 수직 방향 움직임 벡터의 부호가 같을 경우 카메라의 스크롤 동작을 나타낸다. 줌 동작은 대표 움직임 벡터의 영역별 수평, 수직 성분을 나타내는 부호가 표 1과 같이 나타나게 된다. 카메라 동작이 연속적으로 나타날 때에는 팬, 줌 신호가 연속적으로 발생하지만, 손떨림과 같은 불규칙한 카메라 운동이 있을 경우 불연속적인 신호가 발생할 수 있다. 또한, 팬, 줌에 의한 카메라 동작 중 하늘, 호수면과 같은 단조로운 부분이 프레임 내 상당 부분을 차지할 경우 불연속적인 신호가 발생한다. 이러한 오류를 방지하기 위해 본 논문에서는 프레임마다 발생한 검출 신호에 대하여 윈도우 사이즈가 5인 미디안 필터를 사용하여 보다 정확한 카메라 동작구간을 검출하였다.

표 1. 카메라 동작에 따른 대표 국부 움직임 벡터의 방향.

	$l_{1,x}$	$l_{2,x}$	$l_{3,x}$	$l_{4,x}$	$l_{1,y}$	$l_{2,y}$	$l_{3,y}$	$l_{4,y}$
pan	+ / −	+ / −	+ / −	+ / −				
scroll					+ / −	+ / −	+ / −	+ / −
zoom out	+	−	+	−	+	+	−	−
zoom in	−	+	−	+	−	−	+	+

Ⅳ. 실험 결과 및 분석

디졸브 검출 실험을 위하여 4번의 디졸브를 포함하는 DC 영상을 사용하였다. 원 영상의 크기가 352x240이고, 길이가 총 186 프레임이며 그림에서는 홀수 번째 프레임을 나타내었다. 이 비디오 데이터는 1~20, 43~78, 98~112, 그리고 145~160번째 프레임에서 디졸브가 일어난다. 세 번째 구간에서는 느린 카메라 팬 효과가 포함되어있다.

그림 6의 그래프는 기존 방법에 의한 디졸브 영상의 검출 결과들이다. 완전하게 복원된 영상의 분산 분포는 그림 6(a)와 같이 뚜렷한 "U"자형 곡선을 이루기 때문에 이를 이용한 프레임간 분산차는 그림 6(b)에서 나타나듯이 전환 구간의 가운데 부분에 나타나는 극소점을 중심으로 전환 시작과 끝 부분의 극대점이 있어 디졸브에 의한 전환 구간을 검출할 수 있다 [6]. 그러나, DC 영상에서의 분산 분포는 그림6(c)와 같이 "U"자형 곡선이 잘 나타나지 않기 때문에 분산차를 나타내는 그림 6(d)에서와 같이 두 번째 구간 외에는 극대점과 극소점의 경계가 불분명하여 전환구간 검출이 불가능하다. 완전하게 복원된 영상의 화소간 차의 합은 디졸브 구간에서 고원(plateau)이 잘 나타나지만, 카메라 움직임에 의한 팬 구간에서도 고원이 나타나 검출상의 오류를 발생시킨다. 반면 DC 영상으로부터 얻은 결과에서는 축소영상을 만드는 과정이 공간적 저역 필터 역할을 하기 때문에 팬에 의한 오차가 약간 감소한다.

www.dbpia.co.kr

또한, 프레임별 히스토그램 차의 합을 이용할 경우 완전 복원된 영상에서는 화소 차의 합에 의한 방법에 비해 물체나 카메라의 움직임에 의한 영향을 비교적 적게 받으며 디졸브 특성이 나타난다. 그러나 그림 6(e),(f)를 보면 축소 영상을 만드는 과정에서 누적되는 오차로 인하여 열화된 B 프레임과 열화가 없는 I 프레임 사이에서 급격한 프레임간 차이를 나타내고 있다. 이러한 현상은 화소간 차를 이용한 방법에서도 생기는 것으로 이를 제거하기 위해 후처리과정이 필요하다.

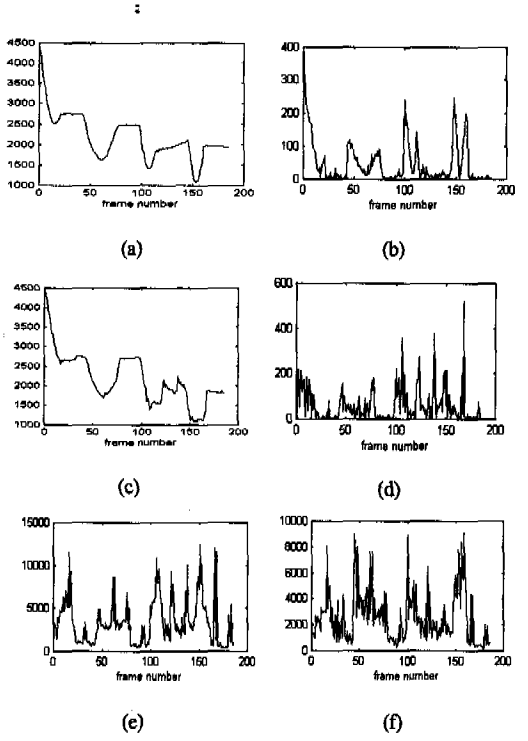


그림 6. 기존 방법에 의한 디졸브 검출 결과, (a) 복원된 영상의 분산 분포, (b) 복원된 영상의 프레임간 분산차, (c) DC 영상의 분산분포,(d) DC 영상의 프레임간 분산차, (e) DC 영상의 화소값차의 합, (f) DC 영상의 히스토그램차의 합.

그림 7은 제안된 방식에 의해 얻어진 디졸브 검출 결과이다. 그림 7(a)는 식 (3)을 사용하여 구한 S 곡선이고 그림 7(b)는 그림 7(a)의 프레임간 S 곡선의 차를 구한 결과이다. 이는 그림 6(e)와 (f)의 결과보다 전환 구간에서 뚜렷한 특성을 가짐을 보여주고 있다. 그림 7(c)는 본 논문에서 제안한 방법과 완전 복원 영상의 분산분포 결과를 비교한 것으로 4개의 디졸브 전환 구간을 모두 검출하는 결과를 나타내고 있다. 검출된 디졸브 구간이 오른쪽으로 약간 이동한 것은 비 중첩 윈도우 평균을 만드는 과정에서 생긴 오차이다.

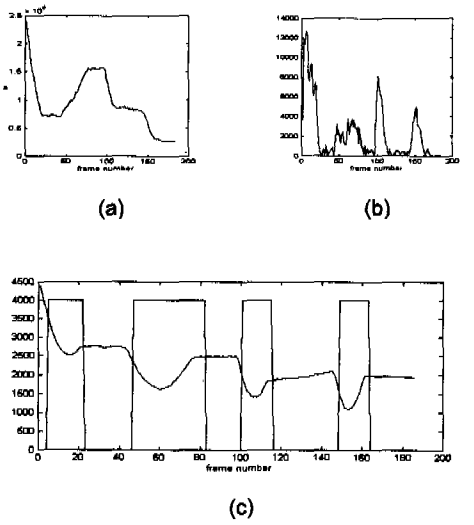


그림 7. 제안한 방법에 의한 실험 결과, (a) $S(t)$, (b) $S(t)$의 일차미분. (c) 완전 복원 영상의 분산과 제안한 기법의 결과.

표 2는 실험에 사용했던 3개의 MPEG 비디오 영상에서 제안한 알고리즘을 통한 디졸브 전환 검출 결과이다. 첫 번째와 두 번째 실험영상(sample #1, 2)은 각각 두 번의 디졸브가 있는 일반 영상이다. 세 번째 실험 영상(sample #3)은 4개의 디졸브를 포함하고 있다. 여기서 사용한 파라메터는 전환 특성을 구할 때 분할 영역 D는 6으로 하였고, 평균의 p제곱에서 p는 2, 전환 구간 검출을 위한 프레임 윈도우 개수 w는 3, 비교 파라메터 Q는 2, 최소 윈도우 평균값인 T는 600을 사용하였다.

표 2. 제안한 방법에 의한 디졸브 구간 검출 결과표.

Sample Video	디졸브 구간	실제 검출된 구간
sample #1	5~14, 25~34	4~15, 25~35
sample #2	13~21, 55~60	13~21, 55~60
sample #3	1~20, 43~78, 98~112, 145~160	1~19, 43~78, 97~111, 145~159

카메라 동작 검출을 위하여 프레임 크기가 352x240인 MPEG 비디오 영상 3개를 사용하였다. 그림 8은 실험 영상을 사용하여 Gerek [12]의 방법과 제안한 방법의 비교한 결과이다. Gerek은 움직임 벡터의 방향 θ와 그 히스토그램을 카메라 동작 검출에 사용하였다. 여기서 θ는 $\arctan(y/x)$이고 방

www.dbpia.co.kr

향 히스토그램은 9개의 구간으로 나뉜다. 이 방법에서 팬 동작인 경우는 대부분의 움직임 벡터가 일정한 방향으로 향하기 때문에 주 방향의 수가 그 외 방향의 수보다 월등하게 크게 나타나고, 줌인 경우는 방향 성분과 그 수가 골고루 분포하는 성질을 이용한 것이다. 첫 번째 실험 영상은 첫 프레임부터 마지막 프레임까지 전 구간에 걸쳐 팬이 있는 "flowergarden.mpg" 영상이다. 그림 8(a)와 (b)는 주 방향(dominant theta)의 수와 방향 히스토그램(theta histogram)을 나타낸 것으로 첫 프레임부터 140 프레임까지 한쪽 방향으로 팬 동작을 잘 나타내고 있다.

그러나 140 프레임 이후부터 마지막 프레임까지 하늘의 단조로운 영상에 의하여 주 방향의 수가 줄어들어 팬 구간을 나타내지 못하고 있다. 반면 제안한 방법의 결과 그림 8(c)는 처음부터 마지막 프레임까지 정확하게 팬 동작을 나타내고 있다. 두 번째 영상은 3개의 팬 구간이 있는 "lions. mpg" 영상이다. 이 영상에서는 프레임 중앙 부분의 물체 움직임으로 인하여 주 벡터의 수, 그림 8(d)가 팬 구간이 아닌 곳에서 높은 값을 나타내므로 잘못된 팬 구간 검출을 할 수 있다. 반면 제안한 방법의 결과 그림 8(f)는 세 구간에 대해서 비교적 정확한 팬 구간을 검출하였다.

세 번째 영상 "lake.mpg"는 60~98 프레임 구간에 카메라 줌, 그리고 183~마지막 프레임 구간에 연속된 카메라 팬과 스크롤을 가지고 있다. 주 방향의 수를 나타내는 그림 8(g)와 (h)의 분포에서는 호수 물결의 불규칙적인 움직임 벡터로 인하여 1~20 프레임 구간과 155~178 프레임 구간에서 부정확한 팬 검출 결과를 나타낸다. 그리고 60~134 프레임 구간에서 줌 구간이 끝난 후에도 그림 8(g)와 (h)의 방향 분포가 일정한 값을 유지하고 있어 잘못된 줌 구간을 검출할 수 있다. 또한 이러한 방법은 줌을 검출하기 위한 방향 히스토그램에 대한 임계값을 정하기가 어렵다. 그림 8(i)는 제안한 카메라 동작 검출 기법을 통하여 얻은 결과로써 줌 구간을 정확하게 검출하였다. 114~124 프레임 구간에서 호수의 물결로 인하여 잘못된 구간 검출 결과를 나타내었으나 나머지 팬 구간에서는 비교적 정확한 구간 검출 결과를 보였다. 그림 8(c), (f), 그리고 (i)의 결과는 제안한 카메라 동작 검출 신호값에 따라 팬과 스크롤은 10을 곱하고, 줌인 경우는 20을 곱하여 나타낸 것이다. 따라서 Y축의 크기가 20이면 줌을 나타내고, 10이면 팬을 나타낸다.

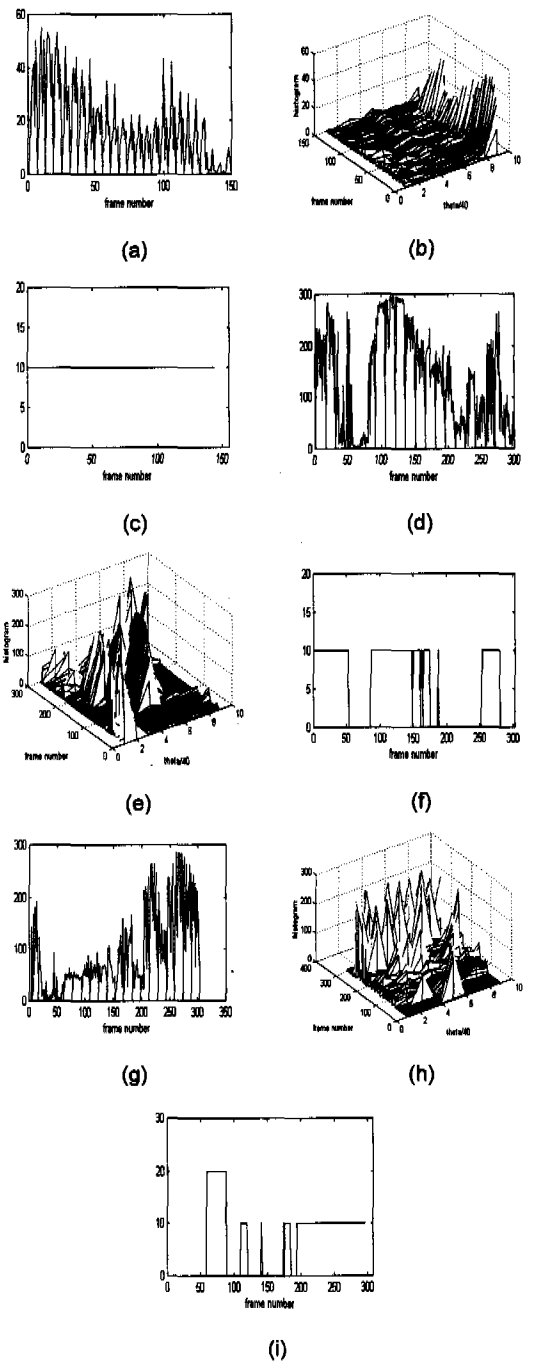


그림 8. 기존 방법과 제안한 방법의 결과, (a),(d),(g) 주 방향 수, (b),(e),(h) 방향 히스토그램, (c),(f),(i) 제안한 결과.

V. 결 론

본 논문에서는 디졸브와 카메라 동작인 팬, 줌에

www.dbpia.co.kr

의한 점진적인 장면전환을 검출하는 알고리즘을 제안하였다. 공간적 특성과 비 중첩 윈도우 비교 방식을 이용하여 디졸브 구간을 검출하는 기법을 제안하고 네 개의 영역별 대표 움직임 벡터를 이용하는 카메라 동작 검출 기법을 제안하였다. 시간적인 검출 정밀도를 높이기 위해 모든 종류(I, P, B)의 MPEG 프레임을 사용하였고, 빠른 검출을 위하여 I 프레임의 DC 계수와 B, P 프레임의 움직임 벡터를 사용하여 1/64 크기로 축소된 DC 영상을 사용하였다.

실험을 통하여 제안한 디졸브 검출 방법이 기존의 프레임간의 분산, 히스토그램 등을 이용하는 방법들보다도 정확한 디졸브 구간의 검출 성능과 적은 계산량을 가짐을 보였다. 그리고, 제안한 카메라 동작 검출 방법이 기존의 전역 움직임 벡터를 사용하는 방법들보다도 효과적으로 카메라 움직임을 검출하고 빠른 처리 속도를 가짐을 보였다. 본 연구의 결과는 향후 MPEG-7에서 지향하는 멀티미디어 데이터 베이스 구축과 비디오의 색인 및 검색 기능에 활용될 수 있을 것으로 사료된다.

참 고 문 헌

[1] S. Smoliar and H. Zhang, "Content-based video indexing and retrieval," IEEE Multimedia, vol. 1, pp. 62-72, 1994.

[2] S. Abe, Y. Tonomura, and H. Kasahara, "Scene retrieval method for video database applications using temporal condition changes," Proc. Conf. Machine Intelligence and Vision, Tokyo, Japan, pp. 355-359, April 1989.

[3] H. H. Yu and W. Wolf, "Multi-resolution video segmentation using wavelet transform," SPIE, vol. 3312, pp. 176-187, 1997.

[4] N. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," Proc. 2nd Working Conf. Visual Database Systems, Budapest, Hungary, pp. 119-133, October 1991.

[5] F. Arnam, 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.

[6] 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.

[7] 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, December 1995.

[8] H. Zhang, A. Kankanhalli, and S. Smoliar, "Automatic partitioning of full-motion video," Multimedia Systems, vol. 1, pp. 10-28, 1993.

[9] R, Zabih, J. Miller, and K. Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks," Proc. ACM Multimedia 95, San Francisco, CA, pp. 189-200, November 1995.

[10] E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of motion," Journal of the Optical Society of America, pp. 284-299, February 1985.

[11] A. Akutsu et.al. "Video indexing using motion vectors," in Visual Communications and Image Processing '92, SPIE, pp. 1522-1530, Boston, MA, November 1992.

[12] O. N. Gerek and Y. Altunbasak, "Key Frame Selection from MPEG Video Data," SPIE, vol. 3024, pp.920-925, 1997.

[13] H. Zhang, C.Y. Low, Y. Gong, and S. Smoliar, "Video parsing using compressed data," in Proc. SPIE Conf. Image and Video Processing II, vol. 2182, pp.142-149, 1994.

[14] B. L. Yeo and B. Liu, "On the extraction of DC sequence from MPEG compressed video," IEEE International Conference on Image Processing, October 1995.

윤 석 중(Seok-Joong Yoon) 정회원

1988년 2월 : 단국대학교 전자공학과 졸업(공학사)

1999년 2월 : 고려대학교 대학원 전자공학과(공학석사)

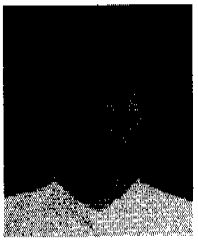
1988년 3월~현재 : 삼성전자 반도체총괄 시스템 LSI Media팀 선임

www.dbpia.co.kr

<주관심 분야> 영상처리 및 압축, Media Processor VLSI 설계

E-mail : souky@samsung.co.kr

지 은 석(Eun-Seok Ji) 정회원

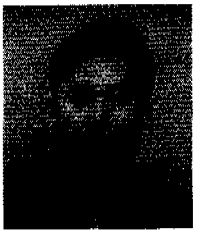


1998년 2월 : 고려대학교 전자공학과 졸업(공학사)

1998년 3월~현재 : 고려대학교 대학원 전자공학과 석사과정

<주관심 분야> 신호 및 영상처리, 영상압축

김 영 로(Young-Ro Kim) 정회원



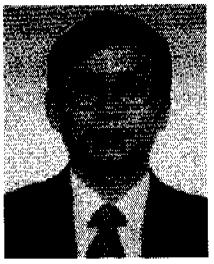
1993년 2월 : 고려대학교 전자공학과 졸업(공학사)

1996년 2월 : 고려대학교 대학원 전자공학과 졸업 (공학석사)

1996년 3월~현재 : 고려대학교 대학원 전자공학과 박사과정

<주관심 분야> 신호 및 영상 처리, 멀티미디어 통신

고 성 제(Sung-Jea Ko) 정회원



1980년 2월 : 고려대학교 전자공학과 졸업(공학사)

1986년 5월 : State Univ. of New York at Buffalo, 전기 및 컴퓨터공학과 (공학석사)

1988년 8월 : State Univ. of New York at Buffalo, 전기 및 컴퓨터공학과(공학박사)

1981년 8월~1983년 12월 : 대한전선 중앙 연구소 연구원

1988년 8월~1992년 5월: The University of Michigan Dearbon, 전기 및 컴퓨터 공학과 조교수

1992년 3월~현재 : 고려대학교 전자공학과 교수

현재 : IEEE Senior member, IEE member

1996년 11월 : IEEE APCCAS best paper award

1997년 12월 : 대한 전자공학회 해동 논문상 수상

<주관심 분야> 신호 및 영상 처리, 영상 압축 및 통신, 멀티미디어 통신 등

E-mail : sjko@dali.korea.ac.kr

www.dbpia.co.kr