

논문-99-4-2-07

MPEG 압축 영역에서의 고속 특징 요소 추출을 이용한 장면 전환 검출과 키 프레임 선택

송병철*, 김명준*, 나종범*

Scene Change Detection and Key Frame Selection Using Fast Feature Extraction in the MPEG-Compressed Domain

Byung Cheol Song*, Myung Jun Kim*, and Jong Beom Ra*

요 약

본 논문은 새로운 장면 전환 검출과 키 프레임 선택 기법을 제안하였다. 이를 위해 본 논문에서는 MPEG 압축 동영상에서 직접 DC 영상 및 에지(edge) 영상을 추출하여 이용하는데, 공간 영역으로 변환 후 에지 영상을 추출할 경우 계산량이 많다는 문제점이 있다. 따라서 본 논문에서는 그 계산량을 줄이기 위해 DCT 블록 당 5개의 저 대역 AC 계수들만을 이용하여 축소된 에지 영상을 고속으로 추출하는 방법을 제안하고, 이를 바탕으로 AC 예측(prediction)을 이용한 고속 에지 추출 기법도 추가적으로 제안하였다. 화질 측면에서 전자가 후자보다 약간 우수하지만, 두 방법 모두 영상의 중요한 에지 특징들을 잘 추출할 수 있다. 이와 같이 얻어진 에지 영상 및 DC 영상을 이용하여 에지 에너지 다이어그램(edge energy diagram)과 히스토그램(histogram)을 구하여 점진적인 장면 전환 및 페이드(fade), 디졸브(dissolve) 같은 점진적인 장면 전환을 정확하게 검출함을 모의 실험을 통해 확인하였다. 또한 공간 영역에서 구한 에지 영상들에 비해 제안한 방법들에 의한 에지 영상들이 점진적인 장면 검출에 있어 훨씬 적은 계산량으로 비슷한 성능을 보임을 확인하였다. 마지막으로 HVS(human visual system)에 기반하여 각 장면에서 키 프레임을 선택하는 방법도 제안하였다. 위에서 얻어진 에지 및 DC 영상을 이용하기 때문에 optical flow를 이용하는 기존 방법에 비해 적은 계산량으로 의미 있는 키 프레임을 선택할 수 있었다.

ABSTRACT

In this paper, we propose novel scene change detection and key frame selection techniques, which use two feature images, i.e., DC and edge images, extracted directly from MPEG compressed video. For fast edge image extraction, we suggest to utilize 5 lower AC coefficients of each DCT. Based on this scheme, we present another edge image extraction technique using AC prediction. Although the former is superior to the latter in terms of visual quality, both methods all can extract important edge features well. Simulation results indicate that scene changes such as cut, fades, and dissolves can be correctly detected by using the edge energy diagram obtained from edge images and histograms from DC images. In addition, we find that our edge images are comparable to those obtained in the spatial domain while keeping much lower computational cost. And based on HVS, a key frame of each scene can also be selected. In comparison with an existing method using optical flow, our scheme can select semantic key frames because we only use the above edge and DC images.

I. 서 론

* 한국과학기술원 전기 및 전자공학과
Dept. of Electrical Engineering, KAIST

지난 몇 년 동안 디지털 저장 매체의 가격이 저렴해지고, 데이터 전송 속도 및 압축 효율이 급속도로 증가함에

따라 디지털 영상은 멀티미디어의 대표적인 예로써 그 사용 영역을 점점 넓히고 있다. 하지만 상대적으로 동영상의 검색을 위한 기술 개발은 아직 미미하다. 효과적인 동영상 검색을 위해서는 장면 전환 검출 및 키 프레임 선택이 필수적이며, 최근 이에 관한 연구가 활발하다.

동영상은 보통 여러 개의 의미 있는 객체(segment)들로 구성된다. 다시 각 객체는 몇 개의 장면(scene)들로 분할이 가능하다. 한 장면은 시공간적으로 연속성을 갖는 일련의 프레임들을 의미하며, 동영상 이해(video understanding)의 기본 단위로써 이용된다. 장면 전환은 그 연결 방식에 따라 급진적인 장면 전환과 점진적인 장면 전환으로 나눌 수 있다. 급진적인 장면 전환으로서 컷(cut)은 두 개의 장면들이 연속하여 연결되며, 점진적인 장면 전환은 두 개의 연속한 장면들이 한 프레임 이상 동안 서서히 전환하는 페이드, 디졸브 등을 말한다. 어떤 특정한 기준으로 장면 전환을 검출하는 과정을 동영상 분할이라고 한다.^{[3][4][5][6]}

모든 장면 전환이 검출되면 각 장면에서 키 프레임을 추출한다.^{[8][9][10]} 키 프레임은 해당 장면을 대표하는 프레임이다. 장면의 복잡도에 따라 키 프레임은 하나 이상이 추출될 수 있지만, 장면 당 하나의 키 프레임을 추출하는 것이 일반적이다. 몇 개의 키 프레임들로 동영상의 대략적인 내용을 파악할 수 있기 때문에 키 프레임은 동영상 인덱싱을 위한 기본 요소가 된다.

대부분의 동영상들은 MPEG으로 압축되어 저장된다. 압축된 동영상들을 완전 복호화(decoding) 한 후 장면 전환 검출을 하면 많은 계산량이 필요하다. 그러므로, 완전히 복호화된 데이터보다는 최소한의 복호화만으로 얻어진 적은 양의 데이터를 이용하여 처리하는 것이 더 효과적이다.

본 논문에서는 장면 전환 검출 및 키 프레임 선택을 위해 DC 영상과 에지 영상이 이용된다. 먼저 Motion JPEG이나 MPEG 압축 동영상에서 두 특징 요소들을 고속으로 추출하는 두 가지 기법들을 제안한다. 공간 영역에서의 에지 영상 추출에 비해 두 방법 모두 훨씬 적은 계산량을 가지며, 화질 측면에서도 큰 차이를 보이지 않는다. 다음으로 위에서 얻어진 DC 영상 및 에지 영상을 장면 전환 검출에 적용하여 높은 장면 전환 검출률을 검증하였다. 또한 공간 영역에서 얻어진 정확한 에지 영상을 이용할 때와 비교하여 성능 차이가 거의 없음을 모의 실험을 통해 확인하였다. 따라서 제안된 장면 전환 검출 기법은 페이드와 디졸브같은 점진적인 장면 전환 검출에 빠르며 효과적

이다. 마지막으로 두 가지 특징 요소를 이용한 키 프레임 선택 기법을 제안한다. 제안한 방식은 optical flow를 이용한 기존의 방식에 비해 고속으로 의미 있는 키 프레임들을 선택할 수 있다.

II장에서 기존의 방법에 대해서 검토하고, III장에서는 MPEG 동영상에서 에지 영상을 추출하는 방법을 제안한다. IV장에서는 에지 영상과 DC 영상을 이용한 장면 전환 검출 기법을 설명한다. 다음으로 V장에서는 키 프레임 검출을 위한 기법을 제안한다. VI장에서는 제안된 방법에 대한 모의 실험 결과를 보일 것이며 마지막으로 VII장에서 결론을 맺는다.

II. 기존의 방법들

1. DC 영상 추출 기법

MPEG의 I프레임에서 DC 영상을 추출하는 것은 비교적 쉽다[1]. DC 계수 $F(0,0)$ 는 화소값 $f(i,j)$ 들로부터
$$F(0,0) = \frac{1}{8} \sum_{i=0}^7 \sum_{j=0}^7 f(i,j)$$
 과 같이 주어진다. 이는 8x8 블록의 평균 화소값의 8배이다. 이와 같이, I프레임의 DC 영상은 블록 단위로 원영상 블록 단위 평균을 이용함으로써 쉽게 구할 수 있다.

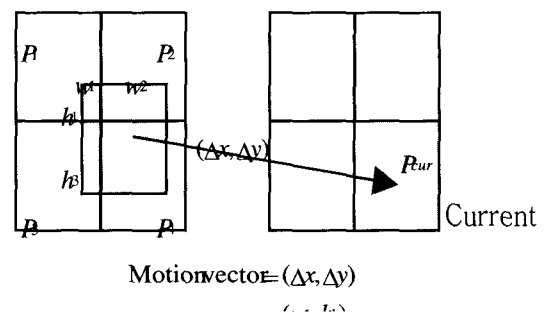


그림 1. 현재 블록 P_{cur} 와 그 움직임 벡터
Fig. 1. Current block P_{cur} and its motion vector

P_i , B프레임의 DC 영상 추출은 움직임 벡터를 이용한다. 그림 1에서 h_1 와 w_1 는 $P_{cur} \cap P_i$ 의 높이와 너비이다. 움직임 벡터가 $(\Delta x, \Delta y)$ 라면, $(\Delta x, \Delta y) = (w_1, h_1)$ 이다. DCT의 특성에 의해, P_{cur} 의 DC 계수 $\hat{P}_{cur}(0,0)$ 는 다음과 같이 주어진다. \hat{P}_{cur} 는 P_{cur} 의 DCT결과를 의미한다.

$$\hat{P}_{cur}(0,0) = \sum_{i=1}^4 \left(\sum_{m=0}^7 \sum_{l=0}^7 w_i(m,l) \hat{P}_i(m,l) \right) \quad (1)$$

여기에서 $w_i(m,l)$ 는 $\hat{P}_i(m,l)$ 의 가중치^[2]이며, $w_i(m,l) = S_{i1}(0,m) \times S_{i2}(l,0)$ 으로 얻어진다. 또한, S_{i1} 과 S_{i2} 는 $\begin{pmatrix} 0 & 0 \\ I_{h_i} & 0 \end{pmatrix}$ 혹은 $\begin{pmatrix} 0 & I_{h_i} \\ 0 & 0 \end{pmatrix}$ 행렬이다. I_n 는 크기가 n 인 단위 행렬을 의미한다. 행렬 S_{ij} 는 각 P_i 의 해당 영역을 P_{cur} 위치로 이동시키는 역할을 한다. 그림 2는 P_{cur} 에 대한 S_{ij} 의 역할을 보여주는 예이다.

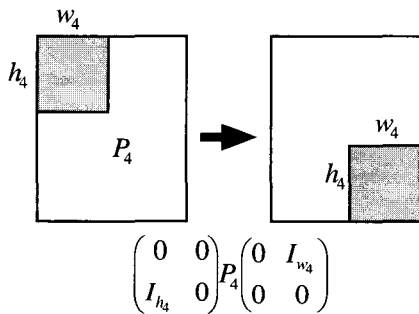


그림 2. 회색 영역의 화소 값들을 얻기 위한 전/후 행렬곱
Fig. 2. Pre- and post-matrix multiplications to move a subblock.

식 (1)은 최고 256개 곱을 요구한다. 계산량의 부담을 줄이기 위해, Yeo와 Liu는 식 (2)와 같이 1차 근사 기법을 이용하였다. 인접한 4개의 DC 값에 의한 현재 DC 값에 대한 기여 정도는 P_{cur} 이 각 P_i, K, P_4 블록에서 차지하는 면적비로부터 얻어진다.

$$\hat{P}_{cur}(0,0) = \sum_{i=1}^4 \frac{h_i w_i}{64} \hat{P}_i(0,0) \quad (2)$$

식 (2)의 계산에 4개의 인접 블록들의 DC 계수들과 움직임 벡터들만이 이용되며, 이러한 정보는 MPEG 비트열에서 쉽게 얻을 수 있다는 점에 주목해야 한다. DC 영상은 동영상 브라우징 및 편집 등 다양한 분야에 적용이 가능하며 특히, 색상 히스토그램을 이용한 컷 검출에 유용하게 이용되고 있다.

2. 장면 전환 검출 기법

지난 몇 년 동안 히스토그램, 프레임 차, 움직임 벡터, 비트 발생률 등을 이용한 많은 컷 검출 기법들이 개발되

었다^{[3][4][5][6]}. 그러나, 디졸브나 페이드같은 점진적인 장면 전환의 성공적인 검출은 여전히 관심의 대상이다.

Hampapur는 인접한 두 영상으로부터 색차(chromatic) 영상을 추출하여 이용했다^[3]. 색차 영상은 페이드 구간에서 0이 아닌 일정한 값으로 유지되는 경향이 있다. 이 방법은 객체나 카메라의 움직임이 장면 전환 구간 동안 작다는 것을 기본 가정으로 하고 있다. 그러나, 이러한 가정은 실제 동영상에서는 잘 맞지 않는다.

Zabih는 다양한 장면 전환의 검출을 위해 에지 변화율을 이용하는 기법을 제안했다^[4]. 점진적인 장면 전환은 사라지거나 생성되는 에지의 비율에 의해 확인이 가능하다. 그러나 이 기법은 연속하는 프레임에 대한 움직임 추정이 필요하며, 성능이 움직임 추정 방법에 상당히 좌우되는 단점을 지니고 있다.

최근 Yu는 점진적인 장면 전환 검출을 위한 새로운 영상 특징 요소에 기반한 알고리즘을 제안했다^[5]. 이 기법은 영상 특징 요소의 시간적 변화를 이용하여 정확하게 컷뿐만 아니라 페이드와 디졸브 등을 검출한다.

3. 키 프레임 선택 기법

Zhang은 여러 개의 시각적인 기준들을 도입하여, 하나 이상의 키 프레임을 선택하는 방법을 제안했다^[8]. 먼저 각 장면의 첫 프레임을 항상 첫번째 키 프레임으로 선택한다. 색상 특성에 근거하여 현재 프레임과 바로 이전 키 프레임을 비교한 후 두 프레임 사이에 기준 이상의 변화가 있으면 현재 프레임도 키 프레임으로 선택한다. 마지막으로, 몇 개의 핵심 장면들이 움직임에 근거하여 선택된다. 줌(zoom)의 경우, 첫번째와 마지막 프레임들이 전체적인 변화를 나타내므로 선택된다.

Wolf는 optical flow에 기반한 간단한 움직임 metric을 도입하여 키 프레임 검출을 시도하였다^[9]. 이 방법은 동영상의 강조할 구간에서 카메라나 객체의 움직임이 적다는 즉, stillness 기준에 의해 키 프레임을 결정한다는 면에서 신뢰할 만하다. 그러나 이 방식은 움직임의 분석에 상당히 의존할 뿐만 아니라 계산량의 부담이 크다.

III. MPEG 동영상에서의 에지 영상 추출

1. AC 계수를 이용한 에지 영상 추출^[11]

먼저 I 프레임의 에지 영상을 공간 영역에서 Sobel 연

산자^[7] 를 도입하여 만드는 과정을 설명한다. 입력 영상의 (m, n) 에서의 밝기값을 $u(m, n)$ 라고 할 때, $u(m, n)$ 의 x , y 방향 gradient $g_x(m, n)$ 와 $g_y(m, n)$ 는 두 직교 방향의 Sobel 연산자 H_x, H_y 에 의해 $g_x(m, n) \equiv \langle U, H_x \rangle_{m,n}$ 와 $g_y(m, n) \equiv \langle U, H_y \rangle_{m,n}$ 로부터 구해진다. (m, n) 에서 에지 에너지는 식 (3)과 같다.

$$g^2(m, n) = g_x^2(m, n) + g_y^2(m, n) \quad (3)$$

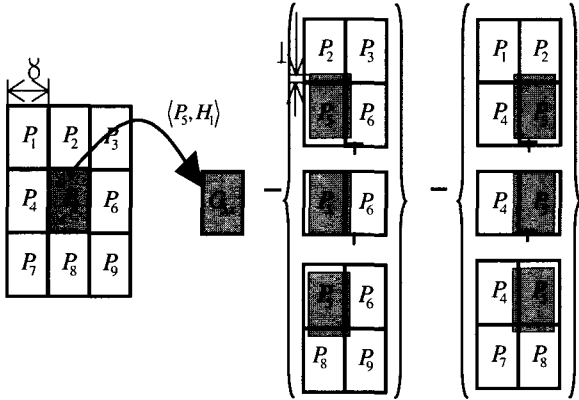


그림 3. 현재 블록 P_5 의 $G_{5,x}$ 행렬 계산

Fig. 3. Computation of gradient $G_{5,x}$ of the target block P_5

P_i 블록의 Sobel 연산자에 의한 에지 블록 G 도 유사한 방법으로 구할 수 있다. 그림 3처럼 현재 블록 P_5 에 대응하는 G 의 x 성분, 즉 $G_{5,x} \equiv \langle P_5, H_x \rangle$ 는 6개의 어두운 블록들의 합과 차 연산을 수행함으로써 얻어진다. 그러나 P_5 의 주변 8개 블록들은 $G_{5,x}$ 의 가장자리에만 영향을 준다는 점에 주목하자. 따라서 P_5 만을 사용함으로써 가장자리를 제외한 $G_{5,x}$ 내부에 있는 6x6성분을 정확하게 계산할 수 있다. 가장자리를 0으로 채운 $G_{5,x}$ 를 $G'_{5,x}$ 라 하자. $G'_{5,x}$ 는 다음과 같이 유도된다.

$$G'_{5,x} = L_x P_5 R_x \quad (4)$$

식 (4)의 양변에 DCT를 취하면,

$$\hat{G}'_{5,x} = \hat{L}_x \hat{P}_5 \hat{R}_x \quad (5)$$

여기서 \hat{L}_x 와 \hat{R}_x 는 다음과 같다.

$$\hat{L}_x = \begin{bmatrix} 3 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \hat{R}_x = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

$G'_{5,y}$ 도 유사하게 얻어진다. 따라서 파스발 정리에 의해 $G'_{5,x}$ 와 $G'_{5,y}$ 로부터 P_5 의 에지 에너지를 근사적으로 구할 수 있다. 예를들어, \hat{P}_5 의 5개의 저주파 AC 계수를 사용하면, P_5 의 에지 에너지 $EE(P_5)$ 는

$$EE(P_5) \approx 2(P_5^2(1,0) + P_5^2(0,1)) + 4(P_5^2(2,0) + P_5^2(0,2)) + 3P_5^2(1,1) \quad (6)$$

P, B프레임의 에지 영상은 DC 영상의 경우처럼 움직임 벡터를 사용하여 구한다. P_{cur} 에 x , y 방향에 Sobel 연산자를 취한 후, DCT를 하면 다음과 같은 식이 성립한다.

$$\hat{G}'_{cur,x} \approx \hat{L}_x \hat{P}_{cur} \hat{R}_x = \sum_{i=1}^4 \hat{L}_{ix} \hat{P}_i \hat{R}_{ix} \quad (7)$$

$$\hat{G}'_{cur,y} \approx \sum_{i=1}^4 \hat{L}_{iy} \hat{P}_i \hat{R}_{iy}$$

파스발 정리에 의해 $G'_{cur,x}$ 와 $G'_{cur,y}$ 로부터 P_{cur} 의 에지 에너지를 대략적으로 구할 수 있다. 지금까지 I, P, B 모든 종류의 프레임에 대해 적은 계산량으로 축소된 에지 영상을 얻을 수 있음을 보였다^[11].

2. AC 예측을 이용한 에지 영상 추출

실제 계산된 AC계수대신 예측을 통해 얻어진 AC계수들로 에지 영상을 추출하는 기법을 제안한다. 2차 평면 모델(quadratic surface model)을 이용하여 몇 개의 저주파 AC 계수들을 예측할 수 있다. 여기서 우리는 JPEG에서 채택하고 있는 널리 알려진 AC prediction 기법을 사용한다^[12]. 현재 블록의 AC 계수들은 주변의 3x3 블록의 DC 값들로부터 다음과 같이 예측된다.

$$\begin{aligned}
 AC_{01} &= (1.13885/8)(DC_4 - DC_6), \\
 AC_{10} &= (1.13885/8)(DC_2 - DC_8), \\
 AC_{20} &= (0.27881/8)(DC_2 + DC_8 - 2DC_5), \\
 AC_{02} &= (0.27881/8)(DC_4 + DC_6 - 2DC_5), \\
 AC_{11} &= (0.16213/8)((DC_1 - DC_3) - (DC_7 - DC_9))
 \end{aligned}
 \tag{8}$$

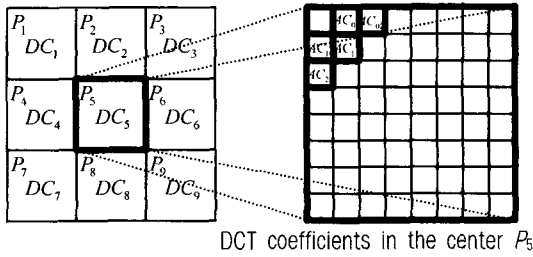


그림 4. 중심 블록 P_5 의 AC 예측을 위해 사용된 3x3 DC계수 행렬
 Fig. 4. 3x3 array of DC coefficients used for AC prediction of the center block P_5 .

AC계수들은 인접한 DC 값들의 선형 합으로 구해지므로, AC계수 예측을 위한 추가적인 메모리는 없다. 식 (8)을 식 (5)나 식 (7)에 적용하여, III.1절에서처럼 축소된 에지 영상을 만들 수 있다. 실제 AC계수를 사용하는 방법과는 달리, P나 B프레임에서도 DC영상만 있으면 AC 예측이 가능하기 때문에 메모리뿐만 아니라 계산량의 부담도 크게 덜 수 있는 장점이 있다.

IV. 장면 전환 검출

점진적인 장면 전환 검출을 위해 먼저 컷을 검출한 후 컷과 컷 사이에 존재하는 점진적인 장면 전환을 검출한다. 본 논문에서는 컷의 검출을 위해 널리 사용되는 히스토그램 기법들 중 Ye0의 방법^[6]을 택했다. 즉, 각 DC 영상으로부터 얻어진 히스토그램 $\{h_i(n)\}$ 들을 구한 후, 인접한 프레임들 간의 히스토그램 차를 계산하여 적응적인 임계값과의 비교를 통해 컷을 검출한다.

여기서 $h_i(n)$ 는 i 번째 프레임의 n 번째 bin값을 의미한다. 적응적인 임계값은 주변 프레임들의 히스토그램 차 평균이다

2. 점진적인 장면의 전환 검출

페이드는 두번째 장면이 나타나기(페이드 인: fade in)

전에 첫번째 장면이 사라지는(페이드 아웃: fade out) 장면 전환이다. 디졸브는 두번째 장면이 나타남과 동시에 첫번째 장면이 사라지는 전환을 말한다. 두 장면 전환의 차이는 두 번째 장면이 나타나는 시점이다. 대개의 경우 디졸브의 검출이 페이드의 검출에 비해 어려우며, 디졸브 검출 기법으로 페이드의 검출도 가능하다. 디졸브의 한가지 중요한 특성은 장면 전환이 시작될 때 그림 5와 같이 에지 에너지가 감소하여 영상이 흐릿해지면서 물체의 윤곽선이 모호해진다는 점이다. 따라서 각 프레임의 에지 에너지를 점진적인 장면 전환 검출에 이용할 수 있다.

III장에서 추출된 에지 영상을 기반으로 각 프레임 당 에지 에너지를 계산하여 그림 6과 같이 각 프레임 순서로 도시함으로써 에지 에너지 다이어그램을 만들 수 있다. 디졸브 장면 전환은 U자 형 에지 에너지 변화를 보인다. 이 유형 검출을 위해 먼저 median 필터를 사용하여 의미 없는 작은 변화들을 제거한 후, 각 프레임의 에지 에너지를 그 위치에서의 지역 평균 (local mean) 으로 나눈다. i 번째 프레임의 에지 에너지를 E_i 라고 하면, 주변 $2M+1$ 개의

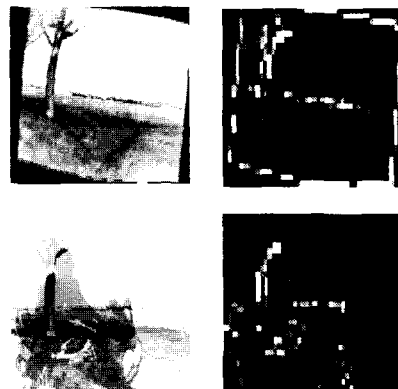


그림 5. (a) 디졸브 전환의 한 예. (b) 대응되는 축소된 에지 영상.
 Fig. 5. (a) An example of dissolve transition and (b) the corresponding reduced edge intensity images

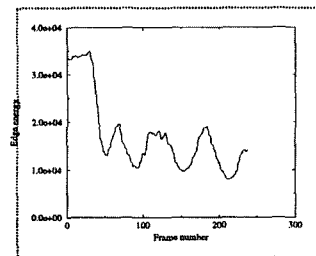


그림 6. 에지 에너지 다이어그램
 Fig. 6. Edge energy diagram.

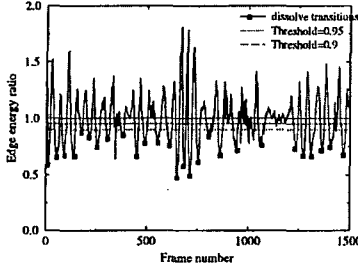


그림 7. *MissingU* 영상의 에지 에너지 비 다이어그램
Fig. 7. Edge energy ratio diagram for *missingU* sequence.

에지 에너지들의 평균과의 비 RE_i 는 다음 식에 의해 얻어진다.

$$RE_i = \frac{E_i}{\frac{1}{2M+1} \sum_{j=-M}^M E_j} \quad (9)$$

그림 7은 위 식을 이용한 한 MPEG1 영상 시퀀스의 에지 에너지 비 다이어그램이다. 그림 7의 그래프를 이용하여 다음과 같은 규칙에 의해 디졸브를 검색했다.

- ⊙ 일정 프레임 구간 τ 이상, 모든 E_i 는 임계값 θ^{dis} 보다 작으면서, 하나의 지역 최소를 중심으로 U자형을 이룬다.

V. 키 프레임 선택

키 프레임을 선택하기 위해 DC 영상과 에지 영상을 이용해 *stillness*를 관찰한다. 또한, 큰 에지 에너지를 가진 프레임이 HVS에 더 민감하다는 점을 보조적인 선택 기준으로 택한다. 즉, 이 선택 기준은 이미 *stillness* 기준에 만족하는 프레임들에 대해 적용된다. 두 가지 일련의 과정을 통해, 각 장면 당 하나의 키 프레임이 선택된다.

IV장에서 설명한 장면 전환 검출 기법으로부터 얻어진 장면의 프레임 시퀀스 $s = \{f_1, f_2, \dots, f_K\}$ 가 주어지면, 다음의 *stillness* 기준을 만족하는 키 프레임이 검출된다. 여기에서 K는 장면 길이이다. 임의의 f_i 에 대한 *stillness* 기준은 다음과 같다.

1. 현재 프레임 f_i 을 중심으로 한 $2N+1$ 프레임들 $f_{i-N}, \dots, f_i, \dots, f_{i+N}$ 을 고려한다. $i < N$ 이거나 $i > K - N$ 인 f_i 들만이 고려 대상이다.
2. 컷 검출을 위해 사용된 DC 히스토그램 차의 분산을

다음과 같이 계산한다.

$$\sigma_i^D = \frac{1}{2N} \sum_{m=i-N}^{i+N-1} \left(\sum_{j=0}^{255} |h^m(j) - h^{m+1}(j)| - \mu_i^D \right)^2, \quad (10)$$

$$\mu_i^D = \frac{1}{2N} \sum_{m=i-N}^{i+N-1} \left(\sum_{j=0}^{255} |h^m(j) - h^{m+1}(j)| \right)^2.$$

3. 에지 에너지 차이의 분산을 계산한다. 분산 σ_i^E 를 다음과 같이 정의 한다.

$$\sigma_i^E = \frac{1}{2N} \sum_{m=i-N}^{i+N-1} (E_m - \mu_i^E)^2, \quad \mu_i^E = \frac{1}{2N} \sum_{m=i-N}^{i+N-1} E_m, \quad (11)$$

여기에서 E_i 는 f_i 의 에지 에너지이다. 만약 σ_i^D 와 σ_i^E 가 각각 임계치 θ^D, θ^E 보다 작을 경우, f_i 는 키 프레임 후보 집합인 C 에 포함된다. 임계치 θ^D, θ^E 를 다음과 같이 정의한다.

$$\theta^D = 0.8 \cdot \min_{i=1}^K \{\sigma_i^D\} + 0.2 \cdot \max_{i=1}^K \{\sigma_i^D\} \quad (12)$$

$$\theta^E = 0.8 \cdot \min_{i=1}^K \{\sigma_i^E\} + 0.2 \cdot \max_{i=1}^K \{\sigma_i^E\} \quad (13)$$

stillness 기준을 만족하는 핵심 장면이 선택된 후, 앞서 설명한 보조적인 기준을 C 의 각 후보에 적용한다. 후보들 중 에지 에너지가 최대인 프레임을 그 장면의 키 프레임 f^{Key} 으로 선택한다.

$$f^{Key} = \arg \max_{i \in C} \{E_i\}, \quad (14)$$

VI. 모의 실험 결과

1. 에지 영상 추출

Susie (720x480) 영상 시퀀스를 4Mbps로 MPEG1압축한 후, 첫 번째 I 프레임에 대해 III장에서 제안된 두 가지 방법을 적용했다. 공간 영역에서 구한 정확한 에지 영상과 비교할 때, 제안한 방법에 의한 축소된 에지 영상들은 여전히 영상 전체의 특징을 가지고 있다(그림 8). 또한 AC 예측을 이용한 에지 영상과 실제 AC 계수를 이용해 얻은 에지 영상이 유사함을 알 수 있다. 한편 각 에지 영상 추출 방법들의 수행 시간 (run-time)을 표 1에 보였다. 이

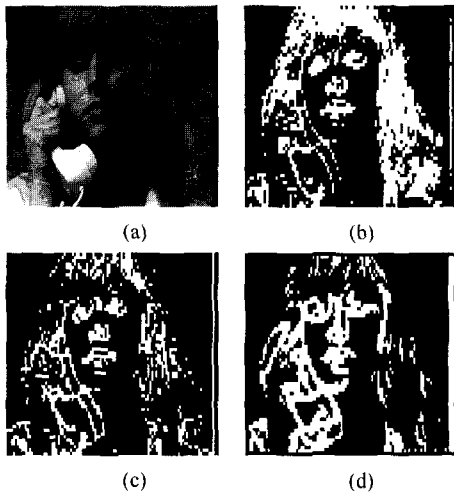


그림 8. (a) Susie 원 영상; (b) 실제 AC 계수들을 이용해 얻어진 에지 영상; (c) 공간 영역에서 얻어진 에지 영상; (d) AC 예측 기법을 이용해 얻어진 에지 영상.
 Fig. 8. (a) Original image of Susie; (b) the edge image using actual AC coefficients; (c) the edge image obtained in the spatial domain after full decoding; (d) the edge image using the AC prediction.

표 1. 각 추출 방법 당 수행 시간 비교
 Table 1. Run-time comparison for various edge extraction methods

추출 방법	수행 시간
실제 AC 계수를 이용한 에지 영상 추출	2.42
AC 예측을 이용한 에지 영상 추출	1.03
공간 영역에서의 에지 영상 추출	7.98
MPEG1 복호화	3.5

표 2. 테스트 영상 당 컷 및 디졸브의 개수
 Table 2. The number of cuts and dissolves for every test video sequence

	컷	디졸브
<i>MissingU</i>	2	28
<i>News</i>	98	17
<i>Beati</i>	2	6
<i>MLB</i>	7	2

수행 시간은 SUN ULTRA1 WorkStation에서 DC영상 추출 시 수행 시간을 1로 했을 때의 상대적인 값이다. 표 2에 사용한 영상들을 나열했다. 참고로 단순한 MPEG1 복

호화 수행 시간을 보였다. 표 1의 결과에서 알 수 있듯이 제안한 방법의 계산량이 공간 영역에서의 에지 영상 추출 시 수행 시간보다 훨씬 적다.

2. 장면 전환 검출

장면 전환 검출 및 키 프레임 선택의 성능 평가를 위해 352x240 크기의 네 가지 유형의MPEG-1 동영상 시퀀스들(30Hz)을 이용했다: 뮤직 비디오 *missingU*, 뉴스 영상 *News*, CF 영상 *Beati*, 야구 경기 영상 *MLB*이다. 처음 2개의 영상은 1500프레임(50초)으로 구성되어 있다. 광고 영상은 450프레임(15초), *MLB* 영상은 900프레임(30초)이다. 표 2는 각 시퀀스 당 컷과 디졸브의 개수를 보인다. 디졸브 검출 성능 평가를 위한 척도로써 다음의 두 가지 기준을 이용하였다. #correct는 옳게 검출된 경우의 수, #false negative는 검출이 안된 경우의 수, 그리고 #false positive는 잘 못 검출된 경우의 수를 의미한다. 두 값이 클수록 성능은 좋다.

$$recall = \frac{\#correct}{\#correct + \#false\ negative}, \tag{15}$$

$$precision = \frac{\#correct}{\#correct + \#false\ positive}$$

먼저 IV.1절에서 설명한 것처럼, 컷들이 Yeo의 방법으로 검출된다. 컷의 검출은 항상 95%가 넘는 recall 및 precision을 보임을 실험을 통해 알 수 있었다. 다음으로, 두 컷 사이에 존재하는 디졸브 검출을 수행한다. 그림 7는 제안된 방법에 의해 얻어진 *missingU* 영상의 에지 에너지 다이어그램이다. 여기서 AC 예측에 의해 얻어진 에지 영상을 이용한다. 사용한 median 필터의 탭 수는 5이다. M은 15로 했다. 즉 한 디졸브 후 0.5초 이내에는 다른 디졸브가 발생하지 않는다고 가정한다. τ 및 θ^{dis} 는 실험적으로 각각 7 프레임, 0.95로 정했다. 표 3과 4는 모든 테스트 영상에 대한 실험 결과이다. 제안한 방법들에 의해 얻어진 에지 영상들을 사용할 때 동일하게 평균 88.7%의 recall과 81%의 높은 정확도를 얻을 수 있었다. 또한 공간 영역에서 추출된 에지 영상을 사용했을 때와의 성능을 비교해도 차이 거의 없음을 알 수 있다. 따라서 제안한 방법은 고속으로 점진적인 장면 전환을 찾기에 적합한 방법이다.

표 3. 제안된 방법의 recall과 precision

Table 3. Recall and precision of the proposed scheme

	Dissolve detection	
	Recall	Precision
<i>MissingU</i>	92.9%(26/28)	86.7%(26/30)
<i>News</i>	82.3%(14/17)	82.3%(14/17)
<i>Beati</i>	83.3%(5/6)	71.4%(5/7)
<i>MLB</i>	100%(2/2)	50%(2/4)
<i>Overall</i>	88.7%	81.0%

표 4. 공간 영역에서 추출한 에지 영상을 이용할 때 recall과 precision

Table 4. Recall and precision when edge images obtained in the spatial domain are used

	Dissolve detection	
	Recall	Precision
<i>MissingU</i>	92.9%(26/28)	83.9%(26/31)
<i>News</i>	88.2%(15/17)	88.2%(15/17)
<i>Beati</i>	83.3%(5/6)	71.4%(5/7)
<i>MLB</i>	100%(2/2)	50%(2/4)
<i>Overall</i>	90.6%	81.4%

3. 키 프레임 선택

키 프레임 선택을 위해 V장에서 설명한 방법을 News1 영상에 대해 적용시켰다. 그림 9는 선택된 키 프레임들 중 일부를 도시한 것이다. Optical flow를 이용한 Wolf의 방법^[9]과 비교하여, 제안된 키 프레임 선택 기법은 계산량 면에서 효율적일 뿐만 아니라, 의미 있는 키 프레임들을 비교적 정확하게 선택함을 확인하였다.

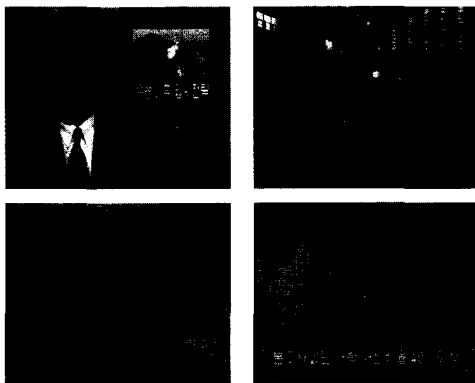


그림 9. News에서의 키 프레임들

Fig. 9. Examples of key frames for the News sequence

VI. 결 론

본 논문에서는 새로운 고속 에지 영상 추출 기법과 이를 이용한 장면 전환 검출 및 키 프레임 선택 기법을 제안하였다. 이를 위해 먼저 에지 영상 추출 기법으로서 소수의 AC계수들을 이용한 추출 방법과 AC 예측을 이용한 추출 기법 등 두 가지를 제안하였다. AC 예측을 이용한 추출 기법은 계산량 뿐만 아니라 메모리의 비용을 상당히 감소시킨다. 이와 같이 얻어진 DC 및 에지 영상을 이용하여 MPEG 동영상으로부터의 디졸브 검출 기법을 제안하였다. 이 방법에서는 보편적인 히스토그램을 이용한 컷 검출 후, 두 컷 사이에 존재하는 디졸브를 에지 에너지에서 나타나는 특정 유형을 이용하여 검출한다. 공간 영역에서 추출한 에지 영상을 이용했을 때와 비교하여 제안된 장면 전환 검출 기법은 적은 계산량으로 상당한 정확도를 보였다. 마지막으로, *stillness*와 *impressiveness*를 기반으로 하여 계산량이 적고 구현하기가 용이한 키 프레임 선택 방법을 제안하고, 방법의 우수성을 보였다.

참 고 문 헌

- [1] B. L. Yeo and B. Li, "On the extraction of DC sequences from MPEG compressed video", *Proc. IEEE ICIP*, pp. 260-263, 1995.
- [2] S. F. Chang and D. G. Messerschmitt, "Manipulation and composing of MC-DCT compressed video", *IEEE J. Sel. Areas Commun.*, vol. 13, pp. 1-11, Jan. 1995.
- [3] A. Hampapur, R. Jain, and T. Weymouth, "Digital video segmentation", *Proc. ACM Multimedia*, pp. 357-364, 1994.
- [4] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks", *Proc. ACM Multimedia*, pp. 189-200, San Francisco, CA, 1993.
- [5] H. H. Yu and W. Wolf, "Multi-resolution video segmentation using wavelet transformation", *Proc. SPIE Storage and Retrieval for image and video databases VI, SPIE* vol. 3312, pp. 176-187, San Jose, CA, Jan. 1998.
- [6] B. L. Yeo and B. Liu, "Rapid scene analysis on compressed video", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp.

553-544, Dec. 1995.

[7] A. K. Jain, *Fundamentals of digital image processing*, Englewood Cliffs: Prentice-Hall, Chapt. 9, pp. 346-357, 1989.

[8] H. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing", *Pattern recognition*, vol. 30, no. 4, pp. 643-658, 1997.

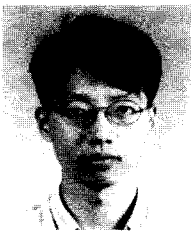
[9] W. Wolf, "Key frame selection by motion analysis", *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, 1996.

[10] Y. Zhuang, Y. Rui, T. S. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering", *Proc. IEEE ICIP*, pp. 866-870, 1998.

[11] B. C. Song and J. B. Ra, "Fast edge map extraction from MPEG compressed video data for video parsing", *Proc. SPIE Storage and Retrieval for Image and Video Databases VII*, pp. 710-721, Jan. 1999.

[12] W. B. Pennebaker and J. L. Mitchell, "JPEG still image data compression." New York: Van Nostrand Reinhard, 1993.

저 자 소 개



송 병 철

1990. 3.~1994. 2. : 한국과학기술원 전기 및 전자공학과 졸업 (공학사)
 1994. 3.~1996. 2. : 한국과학기술원 전기 및 전자공학과 졸업 (공학석사)
 1996. 3.~현재 : 한국과학기술원 전기 및 전자공학과 박사과정 재학 중
 주관심분야 : 동영상 부호화, 영상 표현 및 탐색



김 명 준

1992. 3.~1997. 3. : 경북대학교 전자공학과 졸업 (공학사)
 1997. 3.~1999. 2. : 한국과학기술원 전기 및 전자공학과 졸업 (공학석사)
 1999. 3.~현재 : 한국과학기술원 전기 및 전자공학과 박사과정 재학 중
 주관심분야 : 동영상 부호화, 영상 표현 및 탐색



나 종 범

1971. 3.~1975. 2. : 서울대학교 전자공학과 졸업 (공학사)
 1975. 3.~1977. 2. : 한국과학기술원 전기 및 전자공학과 졸업 (공학석사)
 1979. 3.~1983. 2. : 한국과학기술원 전기 및 전자공학과 졸업 (공학박사)
 1983. 7.~1987. 6. : 미국 Columbia 대학교 연구 조교수
 1987. 7.~현재 : 한국과학기술원 전기 및 전자공학과 교수
 주관심분야: 디지털 영상 처리, 동영상 신호 처리, 3차원 시각화