

에지 히스토그램을 이용한 압축영역에서의 고속 키 프레임 추출 기법

Fast Key Frame Extraction in the Compressed Domain using Edge Histogram

*박준형, **엄민영, ***김명호, ****최윤식

Jun-Hyung Park and Myoungho Kim and Minyoung Eum and Yoonsik Choe

Abstract-As multimedia data and huge-quantity video data having been increasingly and commonly used, the key frame algorithm, as one of the methods for manipulating these kinds of data, became an important matter and has been studied for many years. But the formerly proposed key frame extraction methods take much processing time or need complex calculations due to decoding processes. In order to solve these problems which the former methods have and to enhance the key frame extraction efficiency, a novel key frame extraction method in compressed domain is proposed in this paper. In this method we get an edge histogram for each I-frame in DCT domain and then extract the key frames by means of histogram difference metric. Experimental results show that our algorithm achieves fast processing speed and high accuracy.

Key Words : 에지 히스토그램, 압축영역, 키 프레임 추출

1. 서론

멀티미디어 데이터 및 대용량 비디오 데이터의 생산과 수요는 최근에도 빠르게 증가하는 추세에 있다. 그리고 이 방대한 양의 데이터를 효과적으로 관리하고 이용하는 일이 매우 중요한 문제로 부각되고 있다. 이러한 비디오 데이터를 효과적으로 다루기 위해서는, 비디오 내용 정보를 기반으로 하는 인덱싱 및 검색(CBIR: Content-based Indexing and Retrieval)과 같은 일들이 필요하다.

비디오의 내용 정보를 효과적으로 얻어내기 위해, 우선 그 비디오를 구성하는 전체 프레임 시퀀스를 가장 잘 대표하는 키 프레임 집합을 추출하는 과정이 필요하다. 그리고 키 프레임 추출은 실시간(real time) 처리가 가능하도록 복호화 과정이 없이 압축 영역에서 할 수 있어야 한다.

Cumulative histogram measure(CHM)[2]를 비롯한 몇몇 기존의 알고리즘들 [1], [3]은 처리 과정 때문에 시간 지연과 복잡한 계산이 불가피하고, 갑자기 바뀌는 장면이 있는 비디오에서는 효율적이지 못하였다. 본 논문에서는 처리 속도가 매우 빠르고, 검색에 효과적이면서도 강인한 MPEG-2 비디오 키 프레임 추출 기법을 제안한다. 이 기법은 DCT 영역에서 얻어진 각 I-프레임의 Edge Histogram Descriptor (EHD)를 사용한다. 우선 2절에서, 이전에 제안된 기법들에 대한 간략한 소개를 하고, 3절에서는 본 논문에서 제안하려는 EHDiD(EHD in DCT Domain)를 사용한 키 프레임 추출 기법(KFE-E: Key Frame Extraction using EHDiD)을 설명한다. 제안된 기법에 의한 실험 결과는 4절에 제시한다.

2. 기존의 키 프레임 추출 기법

* 연세大學 전기전자學課 碩士課程

** 연세大學 전기전자學科 博士課程

*** 연세大學 전기전자學科 碩士課程

**** 연세大學 전기전자學科 正教授·工博

2.1 CHM 통한 키 프레임 추출 기법

이 기법은 효율적으로 키 프레임을 추출하기 위하여 히스토그램 차이를 기초로 한 누적 히스토그램 측도를 사용하였다[2]. 누적 히스토그램 측도는 다음과 같이 정의된다.

$$C = \sum_t^{t+k} \left(\sum_j H_{t+1}(j) - H_t(j) \right) \quad (1)$$

위 식에서 표기된 $H_t(j)$ 는 j번째 히스토그램 빈을 의미하고 t는 비디오 시퀀스에서 t 번째 프레임을 의미한다. k는 누적된 프레임 수를 뜻한다. 현재 프레임과 이전 키 프레임간의 누적 값이 임계치보다 크고 이전 키 프레임과 현재 프레임 사이에 히스토그램 차가 임계치보다 크면 키 프레임으로 정의한다.

2.2 픽셀간 강도차이에 의한 키 프레임 추출 기법

픽셀 차 거리함수를 이용한 방법은 비교하려는 두 프레임 사이에서 같은 위치에 있는 픽셀 들 간의 강도 차이에 기반을 둔다. 픽셀간의 강도 거리함수는 다음과 같이 정의된다[3].

$$\Delta_1 = \sum_x \sum_y |I(x, y; j) - I(x, y; k)| \quad (2)$$

위 식에서 j, k는 비디오 시퀀스에서 각각 j번째와 k번째 프레임임을 의미하고 I는 x, y위치에서의 픽셀의 강도를 의미한다.

2.3 MPEG거리함수를 이용한 키 프레임 추출 기법

이 기법은 MPEG시퀀스에서 DCT계수들을 이용하여 IDCT를 거치지 않고 직접 키 프레임을 추출한다[3].

$$\psi = 1 - \frac{|V_j \cdot V_k|}{|V_j| |V_k|} \quad (3)$$

위 식에서 \overline{V}_j 와 \overline{V}_k 는 j 번째와 k 번째 프레임의 각 블록으로부터 추출된 DCT 계수들이다. 이 추출 기법은 한 프레임 안에 있는 DCT 계수들을 1차원 벡터로 간주하고 두 영상 사이의 유사도를 측정하기 위하여 한 벡터를 다른 벡터에 투영시키는 (두 벡터의 내적) 것이다.

3. Key Frame Extraction using EHDiD

3.1. 두 AC 계수의 의미

에지 방향을 결정하는데 사용할 계수들은 AC 계수들로서, 이들은 특정 방향으로의 gray level의 변화와 그 변화 정도를 담고 있다. 입력 영상 $f(i, j)$ 에 대한 DCT 식에 $(u, v) = (1, 0)$ 을 대입하면 계수 AC_{10} 은 :

$$AC_{10} = \frac{C_1 C_0}{4} \sum_{i=0}^7 \sum_{j=0}^7 \cos \frac{(2i+1)\pi}{16} f(i, j) \\ = \frac{C_1 C_0}{4} \sum_{i=0}^7 \cos \frac{(2i+1)\pi}{16} \sum_{j=0}^7 f(i, j) \quad (4)$$

여기서 $\cos(\pi - \theta) = -\cos\theta$ 임을 이용하면, 식 (4)는 다음과 같이 표현될 수 있다 :

$$AC_{10} = \frac{C_1 C_0}{4} \left[\cos \frac{\pi}{16} \left(\sum_{j=0}^7 f(0, j) - \sum_{j=0}^7 f(7, j) \right) \right. \\ + \cos \frac{3\pi}{16} \left(\sum_{j=0}^7 f(1, j) - \sum_{j=0}^7 f(6, j) \right) \\ + \cos \frac{5\pi}{16} \left(\sum_{j=0}^7 f(2, j) - \sum_{j=0}^7 f(5, j) \right) \\ \left. + \cos \frac{7\pi}{16} \left(\sum_{j=0}^7 f(3, j) - \sum_{j=0}^7 f(4, j) \right) \right] \quad (5)$$

식 (5)를 보면 AC_{10} 은 블록 내 상단의 픽셀들과



그림 1. AC 계수들의 물리적 의미

하단의 픽셀들의 수직 방향으로의 밝기 값의 차이를 나타낼 수 있다. 이는 그림 1.에서 볼 수 있듯이 수평 방향으로의 에지 강도를 뜻한다. 비슷한 관점에서 AC_{01} 은 수직 방향으로의 에지 강도를 뜻한다.

3.2 AC 계수를 이용한 에지 방향 결정

AC_{01} 와 AC_{10} 두 계수의 비율을 계산하면 그 블록의 주된 에지 방향을 수직, 수평 뿐 아니라 45°, 135° 등으로도 판단

할 수 있다. 에지 방향을 결정하기에 앞서, 8x8 블록이 에지를 가지고 있는지 아닌지를 먼저 결정해야 한다. Parseval의 정리에 의해 AC 계수들의 제곱의 평균의 근사 값을 블록의 분산 σ^2 로 하고 이 값이 특정 임계치보다 크면 에지 블록이라 판단한다.

$$\sigma^2 \approx \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |X_{u,v}|, (u, v) \neq (0, 0) \quad (6)$$

블록이 에지를 가지면 에지의 방향을 결정하기 위해 AC_{01} 와 AC_{10} 의 비율을 계산한다. 그림 2.는 두 AC 계수의 비율을 다음과 같이 R1, R2로 정의했을 때, 이에 따른 에지 구분을 보여준다.

$$R1 = \left| \frac{AC_{01}}{AC_{10}} \right|, \quad R2 = \left| \frac{AC_{10}}{AC_{01}} \right| \quad (7)$$

이 값들은 블록이 정확히 수직 또는 수평 방향의 에지를 가질 때에는 무한대가 되고, 정확히 45- 또는 135- 대각 방향의 에지를 가질 때에는 1이 된다. 1이나 무한대가 아닌 비율에 대해서는 실험적인 임계치와 비교해서 에지를 구분한다. 두 계수의 부호가 같으면 45-대각 이고, 다르다면 135-대각이 된다.

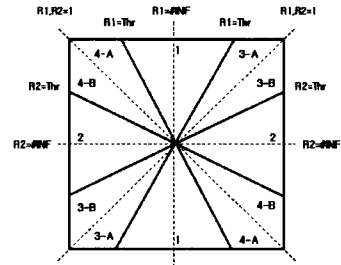


그림 2. AC_{01} 와 AC_{10} 의 비율에 의한 에지 영역 구분

네 방향 중 하나로 결정된 에지의 강도가 실험적 임계치보다 작은 경우 non-directional 에지가 된다. 수직과 수평 방향 에지의 임계치가 같고, 45-와 135- 대각 방향 에지의 임계치가 같다.

3.3 EHDiD를 이용한 프레임 추출

EHDiD는 전체 이미지를 16등분하여 'sub-image'로 나누고, 그 안의 8x8 블록들의 에지를 판단한다. 각 sub-image 당 5개의 히스토그램 빈이 할당되어 그 sub-image 안에 각 빈에 해당되는 에지를 가진 블록들의 개수를 세어 빈값에 넣는다. 이렇게 총 80개의 히스토그램 빈이 전체 이미지에 대해 만들어지고, 그 이미지를 표현하는 정보가 된다. 앞서 말했듯이, DCT 영역 상에서 비디오 시퀀스의 각 I-프레임에 대해 80-빈 에지 히스토그램을 구한다. 이러한 각각

의 K 번째 I-프레임에 대한 에지 히스토그램을 $EHB(K)$ 로 정의하고, 히스토그램 차이정보에 의해 키 프레임울 결정한다. 히스토그램 차이 정보는 다음과 같이 정의한다 :

$$EHBD(K) = \sum_{i=1}^{80} |EHB_i(K) - EHB_i(K-1)| \quad (8)$$

여기서 $EHB_i(K)$ 는 $EHB(K)$ 의 i 번째 빈값을 뜻한다. 현재 I-프레임과 이전 I-프레임의 에지 히스토그램을 가지고 계산한 히스토그램 차이 정보가 특정 임계치보다 크면 현재 I-프레임을 키 프레임으로 결정한다.

4. 실험결과

실험은 CNN 방송의 다섯 개의 비디오 클립을 가지고 수행하였다. 각 비디오 클립은 1200 프레임으로 이루어져있다. 제안된 키 프레임 추출 기법에 의한 비디오 시퀀스 분할을 평가하기 위해, 비디오 시퀀스의 시간적인 분할에 대한 통계적인 성능 평가에 자주 사용되는 척도인 missed detection의 수(MD's)와 false alarm의 수(FA's)를 사용한다. Recall과 Precision을 다음과 같이 정의하면[4] :

$$Recall = \frac{Detects}{Detects + MD's} \times 100 \quad [\%] \quad (9)$$

$$Precision = \frac{Detects}{Detects + FA's} \times 100 \quad [\%]$$

식 (9)를 보면 Recall과 Precision의 값은 MD's와 FA's가 작을수록 100%에 가까워지며 좋은 성능을 나타낸다고 할 수 있다. KFE-E가 CHM보다 낮은 Recall과 Precision 확률을 가진다. 또한, KFE-E로 90퍼센트 이상 연산 량을 절감시킬 수 있다.

다음의 그림들은 실험에 사용된 영상의 예와 누적 히스토그램 기법과 제안된 기법의 Recall과 Precision 비교 그래프를 보여준다.



그림 3. CNN 비디오 클립

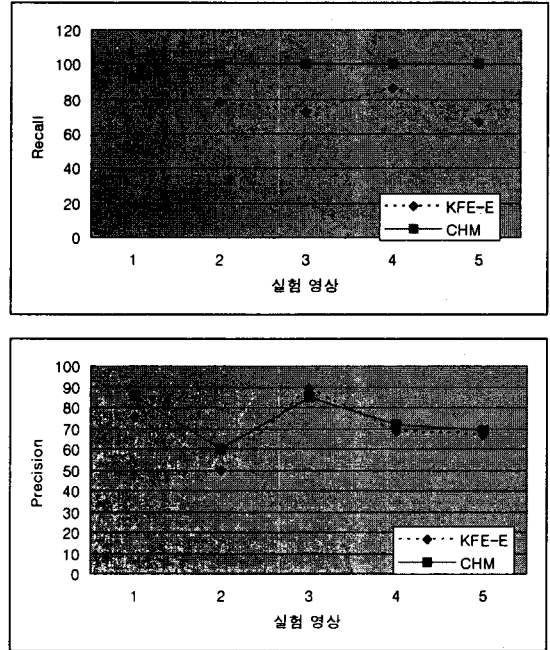


그림 4. KFE-E와 CHM의 Recall, Precision 비교

5. 결론

본 논문에서 제안된 압축 영역에서의 비디오 시퀀스의 키 프레임 추출 기법은 매우 효과적이고 강인하다. 그리고 에지 히스토그램을 복호화 과정 없이 비트열 상에서 바로 얻어내기 때문에 빠른 처리 속도와 적은 계산량을 가지게 된다. 또한, 컬러 히스토그램을 사용하는 기존의 방법은 장면 전환이 매우 빠른 페이드-아웃 장면이나 플래시가 터지는 장면에 약하지만, 우리의 방법은 각 I-프레임의 에지 정보를 사용하기 때문에 이렇게 갑작스런 장면 변화가 있는 비디오에 대해서도 매우 강하다.

참고 문헌

- [1] Boon-Lock Yeo; Bede Liu, "Rapid Scene Analysis on Compressed Video", Circuits and Systems for Video Technology, IEEE Transactions on Vol. 5, Issue 6, Dec. 1995 pp.533 - 544
- [2] Sang Hyun Kim; Rae-Hong Park, "A Novel Approach To Video Sequence Matching Using Color And Edge Features With The Modified Hausdorff Distance", Circuits and Systems, 2004. ISCAS Vol. 2, pp.II - 57-60
- [3] Ralph M. Ford, Craig Robson, "Metrics for shot boundary detection in digital video sequence", Multimedia Systems, Vol. 8, pp. 37 - 46, 2000
- [4] Gargi, U., Kasturi, R., Strayer, S.H., "Performance Characterization of Video-Shot-Change Detection Methods", Circuits and Systems for Video Technology, IEEE Transactions on Volume 10, Issue 1, pp. 1 - 13, Feb. 2000