

# 고속 key frame 추출 기법을 이용한 내용 기반 비디오 검색 기법

## Context-based Video Retrieval using Fast Key Frame Extraction

홍보현\*, 엄민영\*\*, 김명호\*\*\*, 최윤식\*\*\*\*

Bohyun Hong and Myoungho Kim and Minyoung Eum and Yoonsik Choe

**Abstract**-We propose efficient video retrieval scheme which use fast key frame extraction in DCT domain. Our scheme extract key frame using the edge histogram difference which is extracted in compressed domain for I-frames. And the video retrieval is implemented using Hausdorff distance function about edge histogram of key frame. This approach enables fast content-based video retrieval of the compressed video content without decompression process. Experimental results show our scheme is very fast and efficient.

**Key Words** : DCT domain, Edge Histogram, Key Frame, Video Retrieval

### 1. 서 론

최근 통신 네트워크와 멀티미디어 기술의 발전으로 멀티미디어 정보의 양이 급격하게 증가함에 따라 다양한 멀티미디어 정보를 검색하여 이용하고자 하는 사용자의 욕구가 커지고 있다. 특히 DVD, DMB 등의 비디오 영상과 관련된 기술이 발달함에 따라 다양한 멀티미디어 정보 검색 중 비디오 검색은 가장 중요한 분야 중의 하나로 인식되고 있다.

효율적인 비디오 검색을 위해, 비디오 시퀀스를 대표할 수 있는 key frame 집합을 추출하는 과정이 필요하다. 기존의 key frame 추출 방법에는 Kim et al 이 제안한 누적 히스토그램 방법 (CHM : Cumulative Histogram Measure) 을 이용한 key frame 추출 방법과 Janko et al 이 제안한 매크로 블록 타입 정보를 이용한 key frame 추출 방법 등이 있다[1][2].

멀티미디어 정보의 내용 기반의 효율적인 검색에 관한 표준인 MPEG-7 : Multimedia Content Description Interface 에서는 비디오의 특징 정보를 추출하는 기술자로 Group-of-Frame (GoF) or Group-of-Picture (GoP) 기술자를 소개하고 있다. GoF-GoP 기술자는 각 프레임의 영상을 나타내는 히스토그램을 추출한 다음 하나의 대표 히스토그램 빈을 생성한다. 생성된 대표 히스토그램 빈을 SCD (Scalable Color Descriptor) 를 이용하여 여러 프레임에 대표하는 컬러 기반 특징 정보를 추출하는 기술자이다.[3]

그러나 이러한 기존의 방법은 프레임의 히스토그램을 얻기 위하여 압축된 프레임을 복호화하는 과정 등의 몇몇 처리과정에 의한 시간의 지연과 계산량 증가가 불가피하다.

본 논문에서는 매우 빠르고 효과적인 MPEG2 동영상을 위한 비디오 검색 기법을 제안한다. 우선, 효율적인 비디오 검색을 위하여 복호화 과정 없이 압축 영역에서 바로 에지 히스토그램을 얻어내는 EHDiD (Edge Histogram Descriptoe in DCT domain) 에 대해 소개하고[4], 이렇게

얻어진 히스토그램의 차이 정보를 이용하여 key frame을 추출할 수 있는 방법 (KFE-E : Key Frame Extraction using EHDiD) 을 제안한다. 마지막으로 추출된 key frame의 에지 히스토그램에 대한 하우스도르프 거리 함수를 사용하여 영상을 검색하는 기법을 제시한다.

### 2. CHM 을 통한 key frame 추출 기법

일반적으로 사용된 비디오 인덱싱 방법은 히스토그램 비교 방법을 이용하는 것이다. 히스토그램 비교 방법은 하나의 shot 안에서 프레임 간의 변화에 덜 민감하고 히스토그램 추출이 쉽다. 히스토그램 차이는 다음과 같이 정의된다.

$$\sum_j |H_{t+1}(j) - H_t(j)| \quad (1)$$

위 식에서 표기된  $H_t(j)$ 는 비디오 시퀀스의 t 번째 프레임의  $0 \leq j \leq 255$ 의 범위를 갖는 j 번째 히스토그램 빈을 의미한다. 빈은 히스토그램의 그레이 레벨 값을 나타낸다.

[1]에서는 효율적으로 key frame을 추출하기 위하여 히스토그램 차이를 기초로 한 CHM 를 사용하였다. CHM 은 다음과 같이 정의된다.

$$C = \sum_T^{t+k} (\sum_j |H_{t+1}(j) - H_t(j)|) \quad (2)$$

위 식에서 k는 누적된 프레임 수를 나타낸다. key frame을 추출하기 위하여 식 (1)과 식(2)를 이용하여 두 가지 기준이 제시된다. 첫 번째 기준은 식(2)처럼 현재 프레임과 이전 key frame간의 누적 값이 임계값보다 커야 한다. 두 번째 기준은 식 (1)처럼 이전 key frame과 현재 프레임 사이에 히스토그램 차가 임계값보다 커야 한다. 이처럼 두 가지 기준을 모두 만족하는 프레임을 key frame으로 정의한다.

### 3. KFE-E 와 비디오 검색을 위해 제안된 방법

이 논문에서 우리는 MPEG2 로 압축된 비디오를 복호화 과정을 거치지 않고 DCT 계수를 이용하여 빠른 에지 특징 정보를 추출하고자 한다. 그리고 추출된 에지 특징 정보를 이용하

\* 연세大學 전기전자學科 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

여 고속으로 key frame을 추출하고 비디오 검색을 효과적으로 수행하고자 한다.

### 3.1 DCT 영역상의 에지 히스토그램

#### 3.1.1 DCT 계수들의 에지 특성

MPEG에서 사용되는 2차원 8\*8 DCT는 다음과 같다.

$$AC_{u,v} = \frac{1}{4} C_u C_v \sum_{i=0}^7 \sum_{j=0}^7 \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} f(i,j) \quad (3)$$

$$C_u, C_v = \begin{cases} \sqrt{\frac{1}{2}}, & \text{for } u, v = 0 \\ 1, & \text{otherwise} \end{cases}$$

각 DCT 계수  $AC_{u,v}$  는 블록내의 모든 픽셀들의 선형 조합으로 만들어진다. 이 DCT 블록을 가지고 에지 방향에 대한 정보를 추출해 내기 위한 방법은 공간영역에서 블록의 픽셀 값들과 DCT 계수들 간의 관계를 살펴보는 것이다.[5]  $AC_{0,0}$ 은 DC를 의미하고, 이것은 블록의 평균 밝기를 의미한다. 또 나머지 AC계수들의 값은 특정 방향과 특정 변화율을 가지고 그레이 레벨 값에서의 변화를 반영한다.  $AC_{1,0}$ 의 경우를 살펴보자.

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

위 식은  $AC_{1,0}$ 의 값이 공간영역에서의 블록의 위쪽 부분과 아래쪽 부분간의 수직 방향으로의 차이에 의존한다는 것을 의미한다.[5] 즉,  $AC_{1,0}$ 을 통해 에지의 수평 성분의 크기를 알 수 있는 것이다. 이와 유사하게  $AC_{0,1}$ 을 통해 에지의 수직 성분의 크기를 알 수 있다.



그림 1. AC 계수들의 물리적인 의미

#### 3.1.2 AC 계수를 이용한 에지 방향 추출

DCT 계수 블록으로부터 에지의 방향을 추출하기 전에 이 블록이 에지를 포함하고 있는지 아닌지를 검사한다. 한 블록의 에지를 포함 여부는 그 블록의 분산을 척도로 판별할 수 있다.  $N*N$  DCT 계수 블록에서 DC값은 공간영역에서의 평균값의 N배가 된다는 사실과 Parseval의 정리에 의해 분산은 다음과 같이 표현할 수 있다.[5] 여기서  $X_{u,v}$ 는  $u,v$ 번째

DCT 계수를 의미한다.

$$\sigma^2 = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} X_{u,v}^2, \quad (u,v) \neq (0,0) \quad (5)$$

위의 검사를 통해 8\*8 DCT 계수 블록이 에지를 포함하고 있는 블록이라면  $AC_{1,0}$ 과  $AC_{0,1}$ 의 비를 가지고 그 블록이 가지고 있는 에지의 방향을 대략적으로 알 수 있다. 그림 3에서는 블록을 네 가지 에지 영역으로 분류하고,  $AC_{1,0}$ 과  $AC_{0,1}$ 의 비로 에지가 어느 영역에 포함될 수 있는지를 나타낸 것이다. 1번 영역에 포함되면 vertical, 2번 영역에 포함되면 horizontal, 3번 영역에 포함되면 45° diagonal, 4번 영역에 포함되면 135° diagonal 이라고 판단하며 에지의 강도가 특정 임계값보다 작을 경우에는 non-directional 으로 분류한다. 여기서  $R1 = \frac{|AC_{0,1}|}{|AC_{1,0}|}$  와  $R2 = \frac{|AC_{1,0}|}{|AC_{0,1}|}$  을 나타낸다.

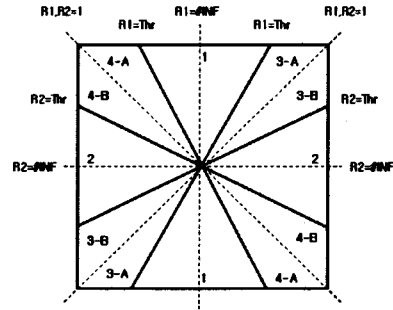


그림 2. 에지 영역과  $AC_{0,1}$ 과  $AC_{1,0}$ 의 비

즉,  $AC_{1,0}$ 과  $AC_{0,1}$ 의 비가 무한대에 가까우면 수평 또는 수직 방향의 에지를 가질 것이고, 1에 가까운 경우 45° 혹은 135°의 에지를 가질 것이다. 둘 중 어느 방향의 에지인지를 구별하는 방법은 두 계수의 부호가 같은 경우에는 45° 방향의 에지일 것이고, 다를 경우에는 135°의 에지일 것이다.

#### 3.1.3 EHDiD

MPEG-7 에지 히스토그램 기술자(EHD)는 영상의 로컬 에지 분포를 나타내는 다섯 가지 종류의 에지 bin값, 80개로 영상을 대표하도록 한다.[3] 이 논문에서는 위에서 설명한 에지 추출 방법을 사용하여 압축 영역에서 직접적으로 80개의 bin값을 추출할 수 있다. [4]

#### 3.2 KFE-E

MPEG2로 압축된 비디오 영상의 키 프레임을 추출하기 위하여 DCT 영역에서 추출한 에지 히스토그램 차이 정보를 사용한다. EHDiD를 사용하여 추출된 80개의 bin값을 이용하여 DCT 영역에서 비디오 영상의 I-프레임의 에지 히스토그램을 생성한다. 에지 히스토그램 bin의 차이 정보, EHD (Edge Histogram Bin Difference)는 아래와 같이 정의된다.

$$EHD(K) = \sum_{i=1}^{80} |EHB_i(K) - EHB_i(K-1)| \quad (6)$$

위 식에서  $EHB_i(K)$ 는 K번째 프레임에서 i번째 에지 히스토그램 bin을 나타낸다. EHD가 임계값보다 크면 두 프레

임간의 에지 변화가 큰 것이므로 key frame으로 정의한다.

### 3.3 비디오간의 유사도 측정 검색

비디오간의 유사도를 측정하기 위하여 위에서 정의된 EHB를 도구로 하여 Hausdorff distance measure를 이용하였다. 키 프레임들로 이루어진 두 집합 사이에 유사도를 효율적으로 측정하기 위하여 수정된 Hausdorff distance를 사용한다. 수정된 Hausdorff distance는 다음과 같이 정의된다.

$$D(S, R) = \max \{ \min \{ d(s_1, r) \}, \dots, \min \{ d(s_m, r) \} \} \quad (7)$$

위 식에서  $S = \{s_1, \dots, s_n\}$ 은 질의 시퀀스의 키 프레임 집합을 나타내고,  $R = \{r_1, \dots, r_m\}$ 은 매칭되는 시퀀스의 키 프레임 집합을 나타낸다. 거리 함수  $d(s, r)$ 은 다음과 같이 계산할 수 있다.[3]

$$d(s, r) = \sum_{i=1}^{80} |h_s(i) - h_r(i)| + 5 \times \sum_{i=1}^5 |h_s^g(i) - h_r^g(i)| + \sum_{i=1}^{65} |h_s^s(i) - h_r^s(i)| \quad (8)$$

여기서  $h_s(i)$ 와  $h_r(i)$ 는 프레임 s와 r의 에지 히스토그램 bin을 의미하며,  $h_s^g(i)$ 와  $h_r^g(i)$ 는 프레임 s와 r의 영상 전체에 대한 에지 히스토그램 bin을,  $h_s^s(i)$ 와  $h_r^s(i)$ 는 프레임 s와 r의 특정 영역별로 분류한 에지 히스토그램 bin을 의미한다.

### 4. 실험 방법 및 결과

실험에 사용된 비디오 영상은 20~100초 사이의 길이로 329개의 MPEG2로 압축된 비디오 영상으로 구성되어있다. 그림 3에서는 실험에 사용된 영상의 중 배 영상을 보여준다.

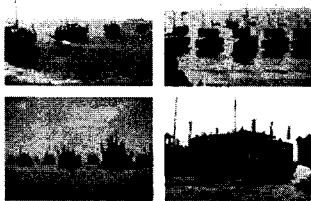


그림 3. 배 영상들의 일부

비디오 검색 성능의 평가 기준으로는 MPEG-7 기술자간의 검색 성능을 평가하는데 주로 사용되는 NMRR (Normalized Modified Retrieval Rank)와 ANMRR (Average NMRR)를 이용하였다.[3]

표 1은 네 개의 각 질의 영상에 대한 NMRR과 검색 기법에 따른 ANMRR을 보여주고 있고 이것들에 대한 결과를 그림 5에서 그래프로 도식화하여 보여주고 있다. 질의 영상에 대한 NMRR 값이 작을수록 검색 효율이 좋은 것이다. 대부분의 질의 영상에 대해서 제안된 검색 기법이 기존의 검색 기법보다 낮은 NMRR과 ANMRR 값을 갖는다. 즉, 제안된 검색 기법이 조금 더 좋은 성능을 보여준다. 또한 DCT 영역에서 바로 추출된 에지 히스토그램 차이를 이용하여 key frame을 추출하는 KFE-E의 사용으로 약 90% 이상의 연산량을 줄일 수 있다.

	Query	KFE-E	CHM
NMRR	배	0.570776	0.586758
	축구	0.587692	0.435385
	뉴스	0.41866	0.487241
	토크쇼	0.497768	0.646577
ANMRR		0.518724	0.53899

표 1. KFE-E를 사용한 비디오 검색과 CHM을 사용한 비디오 검색의 NMRR과 ANMRR 비교

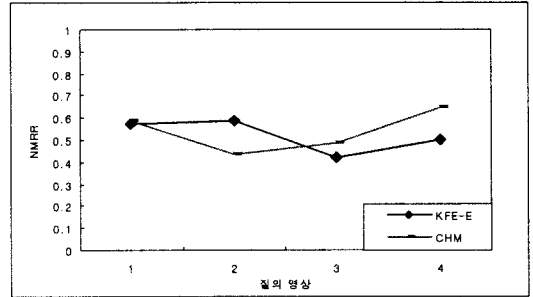


그림 4. 4개의 질의 영상에 대한 KFE-E를 사용한 비디오 검색과 CHM을 사용한 비디오 검색의 NMRR 비교

### 5. 결론

본 논문에서는 MPEG2 동영상상을 위한 효과적인 비디오 검색을 위하여 복호화 과정 없이 압축 영역에서 바로 얻어진 에지 히스토그램의 차이 정보를 이용하여 key frame을 추출하는 방법을 제안한다. 이것은 역양자화 과정, IDCT를 포함한 복호화 과정 등이 생략됨으로써 연산량을 약 90% 이상 줄일 수 있다. 또한 NMRR 측정 결과가 기존의 비디오 검색 기법과 비교하여 낮은 값을 갖는다. 즉, 제안한 비디오 검색 기법은 기존의 비디오 검색 기법과 비교하여 key frame을 찾기 위한 특징 추출의 처리 시간과 연산량에서 장점을 가지고 높은 검색 효율을 보인다.

### 참고 문헌

- [1] 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, ISCAS Vol. 2, pp.57-60 2004
- [2] Calic, J.; Izquierdo, E., "A Multiresolution Technique for Video Indexing and Retrieval" Image Processing. International Conference vol.1, pp.952-955, Sept. 2002
- [3] B. S. Manjunath, Philippe Salembier, Thomas Sikora, "Introduction to MPEG-7 Multimedia Content Description Interface", Wiley
- [4] 전용준, 엄민영, 최윤식, "DCT 영역에서의 에지 히스토그램 고속 추출 기법", 제 17회 영상처리 및 이해에 관한 워크샵 발표 논문집, 2005
- [5] Bo-Shen, Ishwar K. Sethi, "Direct Feature Extraction from Compressed Images" IS&T, Vol. 1995-1996, CA, pp.33-49