

# 효율적인 비디오 검색 및 브라우징을 위한 대표 프레임 군집화

김 영민, 이 성환

고려대학교 컴퓨터학과/인공시각연구센터  
E-mail: {ymkim, swlee}@image.korea.ac.kr

## Key Frame Clustering for Efficient Video Retrieval and Browsing

Young-Min Kim and Seong-Whan Lee

Dept. of Computer Science and Engineering/Center for Artificial Vision Research,  
Korea University

### 요 약

효율적인 비디오 검색과 브라우징을 위해서는 비디오를 장면 단위로 나누는 비디오 분할과 더불어 분할된 비디오의 썸을 대표하는 대표 프레임을 군집화하는 기술이 필요하다. 이는 내용 기반 비디오 검색 및 브라우징의 바탕이 되는 핵심 기술로써, 국내외적으로 많은 연구가 요구되고 있는 실정이다.

본 논문에서는 주파수 정보를 이용한 대표 프레임 군집화 방법을 제안하고 실험 비디오 데이터에 대하여 그 성능을 평가해 본다. 제안된 방법에서는 웨이블릿 변환을 통하여 대표 프레임의 주파수 정보를 구한 후, 고주파 영역과 저주파 영역에 가중치를 두어 대표 프레임을 군집화하였다. 제안된 방법을 드라마 비디오 데이터에 대하여 실험한 결과 군집화의 정확도가 우수할 뿐 아니라 군집화 정도를 조절할 수 있어 다양한 수준의 군집화를 수행할 수 있음을 확인할 수 있었다.

## 1. 서론

하드웨어와 통신 기술의 발달로 비디오 데이터의 사용이 급증하였으며, 이에 따라 주문형 비디오, 디지털 라이브러리 등 다양한 비디오 서비스에 대한 요구가 증가하고 있다. 그러나 비디오 데이터는 시간적, 복잡적, 대용량적 특성을 지니고 있어 효율적인 처리와 이용이 매우 어렵다. 따라서 이와 같은 비디오 데이터의 효율적인 저장, 관리를 위해서는 장면 전환 검출을 통한 비디오 분할 작업이 선행되어야 한다.

비디오 분할은 대용량의 비디오 데이터를 다루기 쉬운 크기로 잘라내는 것을 의미하며, 장면이 전환되는 장면 전환점을 기준으로 분할하는데, 일반적으로 비디오 장면 전환 검출의 가장 기본 단위로는 썸을 사용한다. 본 논문에서는 여러 장면 전환 검출 방법 중 비교적 정확하다고 평가된 영도 히스토그램을 이용한 방법으로 장면 전환 검출을 수행하였다[1]. 한편, 분할된 비디오 데이터의 효율적 사용을 위해서는 비디오 검색과 브라우징에 관한 연구가 필요한데, 많은 비디오 중에 원하는 비디오를 찾는 작업을 비디오 검색이라 하고 비디오의 요약된 내용을 짧은 시간 안에 살펴보는 것을 비디오 브라우징이라 한다. 분할된 비디오의 썸을 대표하는 대표 프레임은 일반적으로 비디오 검색을 수행하기 위한 기본 단위로 많이 쓰이고 있으며, 비디오 브라우징도 대표 프레임을 나열하는 방식이 기본적으로 일반적으로 가장 많이 쓰이고 있는 방법이므로 대표 프레임을 사용하기 편리한 형태로 표현하는 기술은 효율적인

비디오 검색과 브라우징을 위하여 매우 중요하며 연구되어야 할 분야 중 하나이다.

일반적인 영화나 드라마 비디오는 한시간에서 두시간 정도의 분량을 지니게 되며, 약 한시간 분량의 비디오 데이터에는 수백개의 대표 프레임들이 존재하게 된다. 따라서 모든 대표 프레임을 이용하여 검색과 브라우징을 수행하게 되면 효율이 매우 떨어지게 되므로 효율적인 비디오 검색과 브라우징을 위해서는 비슷한 대표 프레임을 같은 부류로 묶는 대표 프레임 군집화에 관한 연구가 필요하다. 그러나, 너무 많은 대표 프레임을 군집화하면 비디오 브라우징시 비디오의 전체적인 내용을 파악할 수 없게 되며, 너무 적게 군집화하게 되면 검색의 효율이 떨어지게 되므로 이 두 가지를 조절할 수 있는 방법 또한 고안되어야 한다.

본 논문에서는 효율적인 비디오 검색과 브라우징을 위해 주파수 정보를 이용한 대표 프레임 군집화 방법을 제안하였다. 제안된 방법은 웨이블릿 변환 계수를 사용하여 대표 프레임의 각 주파수 대역별 에너지 정보를 구하고, 고주파 부분과 저주파 부분에 가중치를 부여하여 군집정도를 조절할 수 있도록 하였다.

## 2. 관련연구

유사한 대표 프레임들을 군집화 하기 위해서는 프레임 간의 정합을 통해 대표 프레임간 유사도를 측정, 계산하는 과정이 필요하며, 따라서 대표 프레임 정합에 대한 여러 가지 방법들이 제안되고 있다. Yeung 등[2]은 대표 프레임

의 색상 정보를 구하기 위하여 RGB 색상 모델을 이용하였는데, 이를 위하여 RGB 색상 공간을 n개의 히스토그램 빈으로 나눈 후 화소의 값에 해당하는 히스토그램의 빈의 빈도수를 계산하여 유사도를 측정하는 방법을 사용하였다. Zhong 등[3]은 먼저 RGB 히스토그램 빈 값을 구하여 이것을 CIE-L\*u\*v\*, HSV 등 인간의 시각과 비슷한 색상 모델 히스토그램 값으로 변환, 색상 정보를 구해냈으며, 이를 이용하여 대표 프레임간의 유사도를 측정하였다. 또한 Yeung 등[2]은 물체의 위치나, 크기 회전 변화에 무관한 결과를 나타내기 위하여 명도값의 2차 모멘트를 통해 형태 정보를 구한 후, 대표 프레임간의 유사도를 계산하는 방법을 사용하였다. 그림 1은 색상 모델을 사용한 대표 프레임 군집화 과정의 예이다.

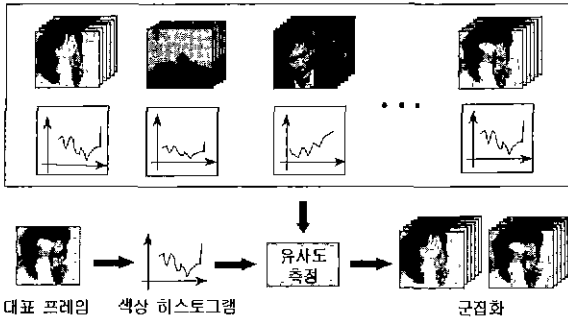


그림 1. 색상 모델을 사용한 대표 프레임 군집화 과정의 예

### 3. 주파수 정보를 이용한 군집화 방법

본 논문에서 제안한 대표 프레임 군집화 방법의 전체적인 구조는 그림 2와 같으며,  $R_{frame1}, R_{frame2}, \dots, R_{frameN}$ 은 군집화 하고자하는 비디오 셋의 대표 프레임들을 나타낸다

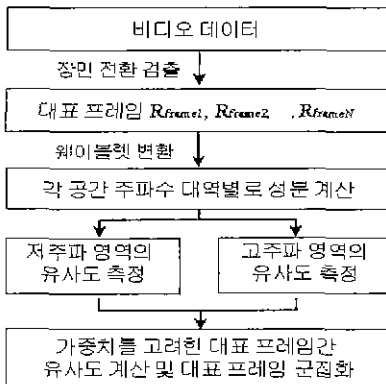


그림 2 제안된 주파수 정보를 이용한 정합 알고리즘

#### 3.1 컷 검출 단계

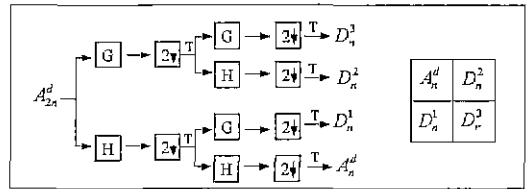
효율적인 비디오 검색과 브라우징을 위한 대표 프레임 군집화를 수행하기 위해서는 대표 프레임 추출 작업이 선행되어야 하며 본 논문에서는 성능이 우수하다고 알려진 명도 히스토그램[1]을 사용하여 장면 전환점을 검출하고 대표 프레임을 추출하였다. 아래 식 1은 두 프레임간 명도 히스토그램의 차를 계산하기 위한 수식으로  $G_{fn}$ 은 n번째

프레임의 명도 히스토그램의 값을 나타내며,  $Q_C$ 는 명도 히스토그램의 단계들,  $DOG$ 는 명도 히스토그램의 누적값을 나타낸다.

$$DOG(f_n, f_{n+1}) = \sum_{i=0}^{Q_C-1} |G_{f_n}(i) - G_{f_{n+1}}(i)| \quad (1)$$

#### 3.2 웨이블릿 변환 단계

대표 프레임 내의 주파수 성분을 구하기 위하여 주어진 대표 프레임을 Wavelet 변환한 후 옥타브-밴드 분할 방식을 사용하여 그림 3과 같이 주파수 대역별로 분할하였다.



$\square$  x 칸터른 사용한 수평방향 필터링  
 $\square$  인지가 2인 부요분화 T. 전치

그림 3. 주파수 분할 과정 [4]

그림 3과 같은 방식으로 프레임을 주파수 대역으로 분할하면, 그림 4와 같은 형태를 띠게 되는데, (a)는 저주파 영역이며, (b)-(f)는 고주파 영역으로 분류된다.

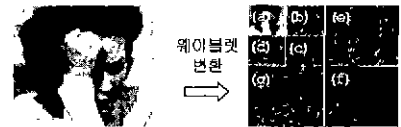


그림 4 웨이블릿 변환된 대표 프레임의 예

#### 3.3 주파수 대역별 유사도 측정 단계

두 대표 프레임의 주파수 대역별 유사도를 측정하기 위해서 본 연구에서는 저주파 영역과 고주파 영역의 유사도를 측정한 후 두 영역에 가중치를 부여하여 계산함으로써 최종 유사도를 결정하는 방법을 사용하였으며, 식 4의 유사도  $W(I_1, I_2)$ 가 일정 임계치를 보다 작은 경우에 두 대표 프레임이 유사하다고 간주하였다. 저주파 영역의 유사도를 측정하기 위하여 명도 히스토그램을 사용하였고, 고주파 영역의 유사도를 측정하기 위하여 각 주파수 대역의 에너지를 대역별로 비교하는 방법을 사용하였다.

$$W_L(d^i, d^j) = \frac{\sum_{k=1}^n \min(d_k^i, d_k^j)}{\sum_{k=1}^n d_k^i} \quad (2)$$

$$W_H(I^i, I^j) = \sum_{k=1}^n \begin{cases} 1 & \text{if } |l_k^i - l_k^j| < Th_H \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$W(I_1, I_2) = \alpha W_H + \beta W_L \quad (4)$$

위 식에서  $I_1, I_2$ 는 입력 영상을 나타내고,  $l_k^i$ 는 i프레임의 k번째 고주파 부분 에너지 성분,  $d_k^i$ 는 i프레임의 저주파 부분 k번째 명도 히스토그램 빈 값을 각각 나타낸다. 또한 식 4의  $\alpha, \beta$ 는 가중치를 나타낸다. ( $\alpha + \beta = 1$ )

#### 4. 실험 및 성능 분석

##### 4.1 실험 환경

본 논문에서 제안한 주파수 정보를 이용한 대표 프레임 군집화 방법의 성능을 평가하기 위해 Windows NT 운영체제 환경에서 Pentium 166MHz PC를 사용하여 실험하였다. 실험에 사용된 비디오 데이터는 드라마 데이터로써, 실험 데이터의 종류와 크기는 표 1과 같다. 표 1에서  $N_{total}$ 은 총 프레임의 수,  $N_{cut}$ 은 컷의 개수를 나타내고 있다. 드라마 비디오 데이터의 경우에는 등장 인물이 서로 대화하는 장면에서 비슷한 대표 프레임을 갖는 것이 많이 나타나게 되어 이러한 비슷한 컷의 군집화 여부를 실험을 통해 알 수 있었다.

표 1. 실험 비디오 데이터

종류	분량(분)	$N_{total}$	$N_{cut}$
드라마 1	14.5	26190	601
드라마 2	16.9	30496	675

##### 4.2 실험 결과 및 성능 분석

표 2는 표 1의 드라마 1을 대상으로 대표 프레임의 저주파 영역과 고주파 영역의 가중치를 조정해 가면서 실험해 본 결과를 보여주고 있는데, 저주파 영역에 가중치를 많이 주면 유사도가 높아져서 비교적 많은 수의 대표 프레임이 군집화됨을 알 수 있고, 고주파 영역에 가중치를 많이 주면 유사도가 낮아져서 적은 수의 대표 프레임이 군집화 되는 것을 볼 수 있다.  $N_{correct}$ 는 정확히 군집화된 대표 프레임의 수를 나타내고,  $N_{false}$ 는 잘못 군집화된 대표 프레임의 수를 나타낸다.

표 2. 가중치 변화에 따른 군집화 성능 비교 성능

$\alpha$	$\beta$	$N_{correct}$	$N_{false}$
0.3	0.7	484	117
0.4	0.6	511	90
0.5	0.5	433	168
0.6	0.4	401	200
0.7	0.3	325	276

표 3은 표 1의 드라마 1을 대상으로 성능이 우수하다고 알려진 HSV 색상 히스토그램을 이용한 대표 프레임 군집화 방법과 본 논문에서 제안한 방법과의 정확도를 비교한 결과이다. 가중치를 조절하면서 비교해 본 결과 HSV 색상 히스토그램을 사용한 방법의 성능과 제안한 방법의 성능이 비슷함을 볼 수 있다.

표 3. HSV 색상 히스토그램을 이용한 방법과 제안한 방법의 비교

종류	$N_{correct}$	$N_{false}$
HSV 색상 히스토그램	514	87
제안한 방법 ( $\alpha = 0.3 \beta = 0.7$ )	484	117
제안한 방법 ( $\alpha = 0.4 \beta = 0.6$ )	511	90
제안한 방법 ( $\alpha = 0.5 \beta = 0.5$ )	433	168

대표 프레임이 군집화되는 정도에 대한 요구는 경우에 따라 유동적일 수 있다. 다시말해, 많은 수의 대표 프레임 서로 군집되기 원하는 경우와 어느정도 정확히 일치하지 않으면 군집되지 않기를 원하는 경우가 있을 수 있다. 이는

원하는 작업의 성격에 따라 달라질 수 있으므로 군집화되는 정도를 조절할 수 있어야 한다. 본 논문에서는 웨이블릿 변환된 대표 프레임의 저주파 영역과 고주파 영역에 가중치를 두어 군집화 함으로써 군집화 정도가 조절될 수 있음을 보였다

#### 5. 결론 및 향후 연구 방향

본 논문에서는 효율적인 비디오 검색과 브라우징을 위한 대표 프레임 군집화 방법에 대하여 제안하고 그 성능을 평가해 보았다. 본 논문에서는 주파수 정보를 이용한 군집화 방법을 제안하였는데, 웨이블릿 변환을 통해 각 주파수 대역별 에너지 정보를 사용함으로써 정확도를 높였으며, 성능이 우수하다고 알려진 HSV 색상 히스토그램을 사용한 방법과 성능을 비교하여 비슷한 성능을 보였다. 또한, 제안한 방법은 저주파 영역과 고주파 영역의 가중치를 부여하여 군집화 정도를 조절할 수 있게 하였으며, 이를 통해 비디오 검색과 브라우징의 효율성을 상호조절이 가능하도록 하였다.

그러나 제안된 대표 프레임 군집화 방법은 저주파 영역에서 명도 히스토그램을 사용하였으므로 명도 변화에 민감하여 잘못된 군집화할 수 있다는 문제점을 나타내었으며, 순차적으로 대표 프레임을 비교하며 유사도를 측정하는 방법을 사용하여 비디오 데이터의 크기가 커짐에 따라 속도가 떨어짐을 볼 수 있었다. 따라서 대표 프레임 정합 방법 뿐 아니라 군집화 알고리즘에 대한 연구가 필요하다. 마지막으로 본 연구에서는 압축 비디오 데이터를 복호화 과정을 거쳐 비압축 상태에서 실험을 수행하였으나, 대표 프레임 군집화의 효율성을 높이기 위해 DCT 계수나 움직임 벡터 등을 사용하여 압축 환경에서 직접 군집화 하는 연구도 수행되어야 할 것이다.

#### 감사의 글

본 연구는 한국전자통신연구원 위탁 과제 연구비 지원에 받았음

#### 참고 문헌

- [1] 이 미숙, 황 본우, 이 성환, "내용 기반 비디오 검색을 위한 장면 전환 검출 방법의 성능 분석," 한국정보과학회 인공지능연구회 춘계 학술대회 발표논문집, 서울, pp. 103-112, 1997년 3월.
- [2] M. M. Yeung, B.-L. Yeo, W. Wolf, and B. Liu, "Video Browsing using Clustering and Scene Transitions on Compressed Sequences," *Proc. of SPIE - Multimedia Computing and Networking 1995*, Vol. 2417, San Jose, pp. 399-413, 1995.
- [3] D. Zhong, H. J. Zhang, and S.-F. Chang, "Clustering Methods for Video Browsing and Annotation," *Proc. of SPIE - Storage and Retrieval for Image and Video Database IV*, Vol. 2670, San Jose, pp. 239-246, 1996.
- [4] W. Xiong, R. Ma, and J. C.-M. Lee, "A Novel Technique for Automatic Key Frame Computing," *Proc. of SPIE - Storage and Retrieval for Image and Video Databases V*, Vol. 3022, San Jose, pp. 166-174, 1997.