

# 형상 근사화와 클러스터링 기법을 이용한 MPEG-4 영역에서의 키 VOP 선정

한 상 진, 김 용 철  
서울시립대학교 전자전기컴퓨터 공학부

## Selection of Key VOP by Clustering of Approximated Shape in MPEG-4 Compressed Domain

Sang Jin Han, Yong Cheol Kim  
Dept. of Electrical and Computer Engineering, University of Seoul  
E-mail: yckim@uos.ac.kr

### Abstraction

본 논문에서는 MPEG-4 비디오 스트림의 형상 정보를 클러스터링 하여 VO(Visual Object)의 동작을 요약하는 새로운 방법을 제안한다. 제안하는 방법은 MPEG-4 비트 스트림의 디코딩 없이 형상 정보를 근사화 한다. 그리고 사용자가 입력한 질의 VOP(Video Object Plane)와 VO의 각 VOP와의 NMHD(Normalized Mean Hausdorff Distance)를 구한 후 클러스터링을 수행 하여 키 영역을 분리해낸다. 클러스터링에 의해 시간적으로 분리된 영역의 지속성을 고려하여 Rank를 매김으로써 사용자가 원하는 수의 키 VOP를 선택할 수 있게 한다. 제안하는 방법은 클러스터링을 사용함으로써 키 VOP를 선정하였으며, Rank와 질의 VOP를 사용하여 사용자와의 상호작용이 가능하다.

### I. 소개

최근 멀티미디어의 양이 방대해짐에 따라 가시적 특성을 기반으로 비디오 데이터의 분석 또는 요약의 자동화로 해결 필요가 있다. 기존의 프레임 기반 비디오 인덱싱 방법은 비디오의 씬을 자동으로 나누어주어 요약을 해준다. 그러나 개별적인 VO의 움직임, 색, 모양 등을 구체적으로 묘사해주진 못한다. VO를 기반으로 인덱싱을 하면 개별적인 VO의 정확한 묘사가 가능하다. 객체기반 비디오 인덱싱은 프레임내의 개별적인 VO의 형상, 색상, 움직임 등의 다양한 특성을 이용하여 VO의 일생을 묘사해준다.

최근 비디오와 이미지의 인덱싱을 위해 형상 정보를 이용하여 많은 연구[1, 2]가 이루어지고 있으며, 클러스터링을 사용한 방법도 많이 사용되어지고 있다[3, 4]. 형상정보는 VO의 작은 변화도 쉽게 검출해낼 수 있기 때문에 많이 사용되어지며, 클러스터링 기법은 데이터를 임계치 없이 자동으로 분리해줄 수 있다.

본 논문에서는 새로운 형상 근사화 방법을 제안하며, 윤곽정보의 NMHD를 구하여 2-means 클러스터링을 수행하고 영역의 지속성을 고려하여 Rank를 매기어 키 VOP를 선정한다. 새로운 형상 근사화 방법을 사용함으

로써 MPEG-4 비트 스트림의 디코딩 없이 간단히 근사화 할 수 있다. 클러스터링을 사용함으로써 키 VOP를 선택할 수 있으며, Rank를 사용하여 사용자가 원하는 수의 키를 선정할 수 있다.

본 논문에서는 우선 MPEG-4의 형상 코딩에 대하여 알아보고 형상 정보를 근사화 하는 새로운 방법을 제시한다. 근사화된 형상의 k-means 클러스터링과 Rank를 사용하여 키 VOP를 선정하는 방법에 대하여 서술한 후 실험 결과를 살펴보고 결론을 맺는다.

### II. MPEG-4의 형상 정보

MPEG-4는 VO의 형상 정보와 텍스처 정보를 사용한다. 형상 코딩은 텍스처 코딩과 유사한 방법으로 이루어진다. 16 \* 16 BAB(Binary Alpha Block)단위로 움직임 보상을 수행하고 CAE(Content-based Arithmetic Encoding)를 수행하여 BAB 내부 갱신을 한다. 이러한 형상 정보는 형상 코딩을 하여 전송되어지는데 각 블록의 코딩 모드는 표 1과 같다. 형상 정보는 그림 1과 같이 Transparent, Opaque, Intra 세가지 종류가 있다. Transparent 블록은 VO가 없는 배경 부분이며 Opaque 블록은 VO 내부 영역이다. Boundary 블록은 Transparent와 Opaque를 경계 지어주는 블록이다. 자세한 내용은 ISO의 MPEG-4 VM[5]을 참조하기 바란다.

표 1. Shape Coding Mode

Coding Mode	Coding Type	Used in
0	MVDs=0 & no inter update	P,B VOP
1	MVDs!=0 & no inter update	P,B VOP
2	Transparent	I,P,B VOP
3	Opaque	I,P,B VOP
4	Intra coded	I,P,B VOP
5	MVDs=0 & inter coded	P,B VOP
6	MVDs!=0 & inter coded	P,B VOP



(Normalized Mean Hausdorff Distance)를 사용한다. Hausdorff Distance는 두 VO의 윤곽 사이의 거리를 나타낸다. 식 (1)은 NMHD를 계산하는 수식이다. A, B는 비교하고자 하는 두 VO의 윤곽 블록의 집합이며, a, b는 각각 A, B 집합의 원소이다.  $N_C$ 는 Boundary BAB의 개수이며,  $D_B(a)$ 는 B 집합의 원소를 기준으로 A 집합의 원소들과의 Euclidean Distance를 구한 값 중 최소값을 나타낸다. MHD는  $D_B(a)$  값을 A의 모든 원소에 대하여 구한 후 평균을 구하여 얻는다. 그리고 MHD를 정규화 하여 NMHD를 얻는다. 이 때 두 VO의 형상간의 거리를 구할 때 VO의 약간의 이동은 큰 거리 차이를 발생시킬 수 있다. 그러므로 무게중심을 일치시킨 후 거리를 계산하였다[1].

$$NMHD = \frac{1}{N_C D_{B,max}} \sum_{a \in A} D_B(a) \quad \dots (1)$$

$$D_B(a) = \min_{b \in B} \{d(a, b)\}$$

기존에는 연속된 프레임 또는 선택된 키 프레임과 현재 프레임 사이의 거리를 측정하였다. 바로 이전 프레임과 현재 프레임 사이의 거리 측정[3,4]은 연속된 동작을 갖는 VO의 변화를 측정하기에 적합하지 않다. 또한 선택된 키 VOP와 현재 VOP 사이의 거리를 측정하는 방법[1,2]은 클러스터링 기법을 사용하는 본 논문의 방법에 적합하지 않다. 그러므로 본 논문에서는 임의의 질의 VOP를 기준으로 거리를 측정하는 방법을 사용한다.

질의 VOP와 VO 일생의 모든 VOP의 비교는 그 VO의 형태를 요약할 수 있게 해준다. 이 때 질의 VOP는 사용자가 정하거나 VO의 시작 VOP를 사용한다. 이 방법을 사용하면 VO의 형태에 따른 요약을 얻을 수 있을 뿐만 아니라 역으로 전체 VOP에서 유사한 형태를 갖는 VOP의 위치를 찾아 낼 수도 있게 된다.

### 3.3 k-means 클러스터링

본 논문에서는 각 VOP의 NMHD를 구한 값을 2-means 클러스터링을 이용하여 SVOP와 NSVOP로 분류한다. k-means 클러스터링은 특성 벡터에 따라 분포된 샘플들을 k개의 영역으로 나누어주는 기법이다. 먼저 초기 Centroid를 k개 설정하고 각 샘플과 Centroid 사이의 거리를 구하여 가장 가까운 클러스터에 포함시킨다. 그리고 포함시킨 클러스터의 Centroid를 매번 다시 갱신함으로써 최종 Centroid를 얻게 되며 그 값을 기준으로 샘플들을 분리한다.

k-means 클러스터링을 사용하면 임계치를 적용하지 않고 분류할 수 있는 장점이 있지만 초기 Centroid에 의해 결과가 Local Minima에 빠질 수 있다는 단점이 있다. 본 논문에서는 초기 Centroid를 모든 NMHD 값들의 평균(m)과 표준편차(SD)를 구하여 설정하였다. SVOP의 초기 Centroid는  $m - SD$ , NSVOP는  $m + SD$ 로 설정한다.

그러나 이러한 방법으로도 클러스터링 초기에 매우 크거나 작은 NMHD(1 or 0)가 발생하면 Centroid가 갑자기 커지거나 작아져 Local Minima에 빠지게 된다. 그러므로 제안하는 방법에서는 Centroid를 갱신시킬 때, Centroid 값과 너무 큰 차이가 나는 값은 갱신에 포함시키지 않았다.

### 3.4 Rank

2-means 클러스터링에 의해 SVOP와 NSVOP 클러스터로 분리한 후 각 NSVOP 영역을 시간적인 분리와 지속성을 고려하여 우선권을 주어 Rank를 매기었다. 먼저 분리된 영역의 시간적인 지속성이 긴 순서로 우선권을 준다. 분리된 영역에서 후보 Key VOP로 중간 값과 시작, 끝 VOP를 선정하며 중간 값을 갖는 VOP가 시작과 끝 VOP보다 높은 우선권을 갖는다. 이러한 우선권은 VO가 동작을 취할 때 가장 오랫동안 변화하는 구간이 가장 중요한 구간이 되기 때문이다. 구간의 중간 값을 모두 후보 키 VOP로 취한 후 구간의 시작과 끝 VOP를 후보 키 VOP로 선정한다. 선정된 후보 키 VOP 집합에서 다시 NMHD를 측정하고 클러스터링을 하여 유사한 VOP를 제거함으로써 오 검출을 제거한다.

2-means 클러스터링을 하여 얻은 NSVOP에 우선권을 부여함으로써 효과적으로 키 VOP를 선정할 수 있으며, 사용자가 원하는 수만큼의 VOP를 얻을 수 있게 된다.

## V. 실험결과

본 논문에서는 제안한 형상 근사화 방법과 클러스터링 기법을 이용하여 MPEG-4로 압축된 영상을 인덱싱하였다. MS사의 MPEG-4 디코더[7]를 사용하여 MPEG-4 비트 스트림에서 형상 코딩 정보를 얻었다. MPEG-4 스트림의 I, P, B-VOP 모두를 고려하여 비디오 인덱싱을 수행하였다. 실험에 사용한 테스트 스트림은 352 x 288 크기의 Bream, Weather, Stefan이다. 각각의 스트림은 300 프레임으로 구성되어 있으며 하나의 VO만을 가지고 있다. Stefan은 테니스를 치는 모습을 담고 매우 활동적이다. Weather는 앵커가 일기예보를 하는 장면이다. Bream은 물고기가 수영하는 장면이며 두 번의 방향 전환을 한다.

그림 4, 5, 6는 각각 Bream, Weather, Stefan의 MHD 히스토그램과 2-means 클러스터링 후의 결과를 나타낸다. Bream은 물고기가 두 번 방향전환을 하는 동안 큰 거리 값을 갖는다. Weather는 일기예보 화면을 손으로 가리키며 설명하는 부분에서 큰 변화가 있음을 볼 수 있다. 마지막으로 Stefan은 테니스를 치는 장면이기 때문에 자주 큰 변화가 일어나며 특히 스트림의 끝 부분에서 큰 변화를 볼 수 있다. 그림과 같이 2-means 클러스터링은 MHD의 변화가 큰 영역을 잘 분리한다.

그림 7, 8, 9은 각각 Bream, Weather, Stefan의 키 VOP 선정 결과이다. 2-means 클러스터링 후의 SVOP와 NSVOP 영역을 시간적인 분리와 지속성을 이용하여 Rank를 매긴 후 Rank가 가장 높은 6개를 선정한 결과이다. Weather의 키 VOP는 4개뿐이므로 4개의 키 VOP를 모두 선택하였다.

## VI. 결론

본 논문에서는 근사화된 영상 정보를 2-means 클러스터링을 수행하여 VO의 동작을 요약하는 방법을 제안하였다. MPEG-4 스트림에서 영상을 근사화 하는 기존의 방법을 간단하게 개선하여 복잡도를 줄였다. 그리고 질의 VOP와의 거리를 이용하여 2-means 클러스터링함으로써 서로 다른 영상을 추출하고 분리된 영역의 지속성을 고려하여 Rank를 매기었다. 본 논문에서는 클러스터링을 사용함으로써 임계치를 적용하지 않고 키 VOP를 선정할 수 있으며, 질의 VOP와 Rank를 사용하여 사용자와의 상호작용을 가능하게 한다. 앞으로 영상 뿐만이 아닌 색상과 움직임을 함께 사용하여 클러스터링 하는 방법을 연구할 계획이다.

## Reference

- [1]Berna Erol, *et al*, "Automatic Key Video Plane Selection Using the Shape Information in the MPEG-4 Compressed Domain", Multimedia, IEEE Trans. On, 2000, pp 129-138.
- [2]Gunsel.B, *et al*, "Moving Visual Representations of Video Objects for Content-Based Search and Browsing", IEEE ICIP, 1997, pp 502-505.
- [3]Hun Cheol Lee, *et al*, "Abrupt shot change detection using an unsupervised clustering of multiple features", IEEE ICASSP, 2000, pp 2015-2018.
- [4]Ferman A.B., *et al*, "Effective Content Representation for video", IEEE ICIP, 1998, pp 521-525.
- [5]ISO/IEC, "MPEG-4 Video Verification Model version 18".
- [6]Earl Gose, *et al*, "Pattern recognition and image analysis", Prentice Hall, 1996, pp 199-215.
- [7]MS, "MPEG-4 video encoder/decoder", 2000.

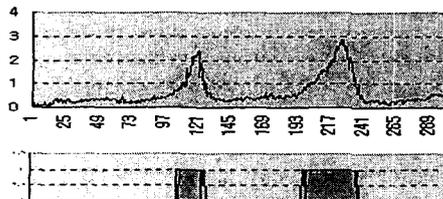


그림 4. Bream의 MHD, 클러스터링 결과

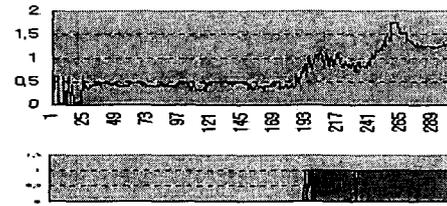


그림 5. Weather의 MHD, 클러스터링 결과

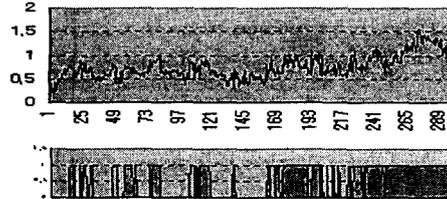


그림 6. Stefan의 MHD, 클러스터링 결과

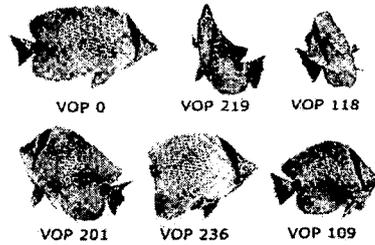


그림 7. Bream에서 6개의 키 VOP 선택

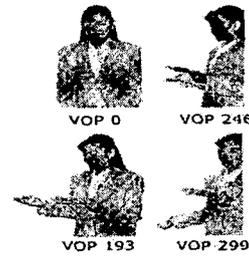


그림 8. Weather에서 4개의 키 VOP 선택

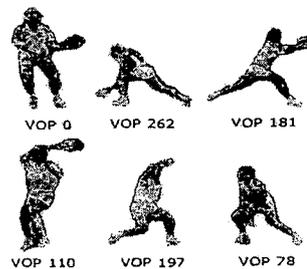


그림 9. Stefan에서 6개의 키 VOP 선택