

# MPEG-4 압축 영역에서 형상을 이용한 키 VOP 선정

정희원 한 상 진\*, 정희원 김 용 철\*\*

## Key VOP by Shape in MPEG-4 Compressed Domain

Sang-Jin Han\*, Yong Cheol Kim\*\* *Regular Members*

### 요 약

본 논문에서는 MPEG-4 동영상을 완전히 디코딩하지 않은 상태에서 근사화된 형상 정보를 추출하여 형상의 유사도에 의해 VOP를 클러스터링 하여 VO의 동작을 요약, 검색하는 새로운 방법을 제안한다. 제안하는 형상 추출 방법은 Erol의 방법을 단순화한 것으로서, 키 VOP 형상의 정확도는 개선되면서도 실행시간은 감소한다. 제안하는 방법에서는 근사화된 형상 사이의 유사도를 Normalized Mean Hausdorff Distance로 구하며, 이를 2-means 클러스터링하여 VOP를 선정한다. VOP 검색은 질의 형상과 키 VOP의 형상 사이의 Mean Hausdorff Distance를 계산하여 가장 일치하는 키 VOP를 선택하는 방식으로 이루어진다. 제안하는 방법을 표준 MPEG-4 테스트 영상 열에 대해 시험한 결과, 기존의 방법에 비해 근사화된 형상 정보를 더 짧은 시간에 얻을 수 있고, 재귀적 클러스터링 방법을 사용하여 키 VOP들의 정확도도 개선되었다.

key Words : MPEG-4, Key VOP, Hausdorff distance, clustering.

### ABSTRACT

We propose a novel method of selecting key VOPs from MPEG-4 compressed domain without fully decoding the compressed data. Approximated shapes of VOPs are obtained from the shape coding mode and then VOPs are clustered by shape similarity to generate key VOPs. The proposed method reduces the computation time of shape approximation, compared with Erol's method. Nevertheless, the resulting VOPs have a good summarizing capability of a video sequence. NMHD (normalized mean Hausdorff distance) values are 2-means clustered to generate key VOPs. In the video search, the MHD of a query VOP from key VOPs are computed and the VOP with the lowest distance is returned. Tests on standard MPEG-4 test sequences show that the computational complexity is very low. Recursive clustering proved to be very effective for generating suitable key VOPs.

### I. 서 론

동영상 데이터는 용량이 크면서도 비정형적이어서 기존의 문자기반 데이터베이스 관리 방법으로는 분류와 검색이 어려우므로, 영상의 내용을 분석하여 키 프레임 추출하여 사용하는 것이 효과적이다. 키 프레임은 동영상의 여러 프레임 중에서 특징적인 장면들을 대표하는 프레임들을 의미한다. 키 프레임을 이용하면 동영상의 효율적 요약이 가능하며,

영상 검색에 있어서도 색상이나 텍스처 등과 같은 내용적 특징을 질의 입력으로 이용할 수 있다.

압축 동영상과 비압축 동영상 모두로부터 키 프레임의 추출이 가능하며, 비압축 동영상은 매 프레임이 완전한 영상 정보를 갖고 있으므로, 각 프레임의 색상, 밝기 혹은 모션 히스토그램 등을 구하여 프레임 사이의 변화가 큰 부분들을 선택하는 방법들이 흔히 사용된다. MPEG 등과 같은 압축 동영상으로부터 압축을 풀지 않고 직접 키 프레임을 추

\* (주)에어코드 기술 연구소(sjhan@aircode.com) \*\*서울시립대학교 컴퓨터비전연구실(yckim@uos.ac.kr)

논문번호 : 020410-0927, 접수일자 : 2202년 9월 27일

※ 본 연구는 산업자원부 중기거점기술개발 대화형미디어 솔루션 개발사업 (20020415-2-1-001) 지원으로 수행되었음.

출하는 방법으로는, 각 프레임의 매크로블록의 DC 계수의 급격한 변화를 검출하거나 텍스처 코딩 모드 (인트라코딩/ 인터코딩) 혹은 움직임벡터의 변화를 고려하여 장면변화를 감지하는 방법들이 있다 [5,7-13].

영상의 하나의 프레임 내에는 여러 개의 객체가 포함되어 있다. 어떤 특정한 객체를 검색하기 위해서는 동영상의 내용이 객체 단위로 서술되어 있어야 한다. 이런 점에서, 키 프레임만으로는 영상의 내용을 비디오 객체 별로 나타내기에 적합하지 않다. MPEG-2에서는 하나의 프레임이 영상의 내용과는 상관없이 매크로블록 단위로 부호화되지만, MPEG-4에서는 하나의 프레임 내의 VO (Video Object)들을 개별적으로 부호화 할 수 있다. 하나의 VO는 등장하여 소멸될 때까지 여러 프레임에 걸쳐 존재하며, VOP (Video Object Plane)은 어느 특정한 하나의 프레임에서의 VO를 의미한다. 따라서 동영상의 요약에 위해서는 키 VOP가 효과적이다.

객체 기반 비디오 요약에 대한 기존의 연구에 대해서는 다음 장에서 상술하며, 이 장에서는 본 논문에서 제안하는 방법과 직접 비교가 되는 Erol<sup>[3,4]</sup>의 방법만을 소개한다. Erol은 MPEG-4 영역에서 객체에 대한 VOP의 선정 방법을 제안하였는데, 이 방법에서는 압축된 동영상의 텍스처와 움직임 코딩 모드 정보를 이용하여 VO의 형상을 근사적으로 구하여 프레임의 진행에 따른 형상의 변화의 정도가 어느 임계값 보다 크면, 그 프레임의 VOP를 키 VOP로 선택한다.

본 논문은 Erol의 방법에서 다음과 같은 점들을 개선하였다. 첫째, Erol의 방법을 개선하여 형상의 근사화가 더욱 빠르게 수행되게 하였다. Erol의 방법에서도 MPEG-4 압축된 상태에서 근사화된 형상을 얻을 수 있으나, 이를 위해 P-VOP나 B-VOP의 기준이 되는 I-VOP의 블록 형상과 움직임의 코딩 모드를 확인하는 과정이 필요하다. 제안하는 방법에서는 이러한 과정 없이 MPEG-4 압축 데이터로부터 근사화된 형상 정보를 얻을 수 있다. 둘째, Erol의 방법에서는 VOP 형상 사이의 유사도의 임계값에 따라 불안정해질 수 있는데, 제안하는 방법에서는 VOP의 형상을 BAB (Binary Alpha Block) 단위로 NMHD (Normalized Mean Hausdorff Distance) 계산한 값을 재귀적으로 2-means 클러스터링하여 키 VOP의 선정을 안정화하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는

MPEG-4 형상 코딩에 관하여 서술하고 기존 의 Erol의 방법을 비교 설명한 후, 제안하는 키 VOP 선정 알고리즘과 검색 알고리즘을 제 3장에서 상술한다. 제 4장에 실험 결과를 제시하였다.

## II. MPEG-4 형상코딩과 기존의 연구

기존의 객체 기반 비디오 인덱싱은 주로 프레임 전체의 색상, 움직임 벡터, DCT 계수 등의 통계적 특성이나 VO의 형상, 색상, 움직임 등을 이용한다. Gunsell은<sup>[5,10]</sup> 비압축 동영상에 대해서 적응적 임계값을 적용하는 방법으로 VO를 클러스터링하여 키 VOP를 선정하는 방법을 제안하였다. 특성 벡터로는 DC 이미지의 색상 히스토그램과 평균 히스토그램을 사용하였으며, 키 프레임을 검출하기 위해 k-means 클러스터링 기법과 적응 임계 값 기법의 두 방법을 사용하였다.

Ferman은<sup>[7-9]</sup> MPEG-4 압축 동영상으로부터 압축을 풀지 않고 키 VOP를 얻는 방법을 제안하였는데, 색상 히스토그램과 공간 정보를 K-means 클러스터링 하였다. 이 방법은 전체 매크로블록의 중에서 텍스처 코딩에서 인트라 모드를 사용한 비율이 임계값을 넘으면 장면에 변화가 발생한 것으로 간주하여 키 VOP를 결정하는데, 실험 결과 VO의 변화 여부를 인트라 코딩의 비율에만 의존하는 것은 정확도가 낮은 것으로 나타났다.

Zhuang은<sup>[11]</sup> K-means 클러스터링을 적용하여 장면의 경계를 적응적 추출하는 방법을 제안하였다. Lee<sup>[12]</sup>는 밝기 정보와 색상 히스토그램을 특성 벡터로 하는 2-means 클러스터링 기법을 제안하였다. 초기 클러스터링 과정 후, 전체 프레임을 동적 분할 하여 대표 키를 검출하였다.

### 1. MPEG-4 형상코딩

MPEG-4에서 VO의 형상 정보와 텍스처 정보는 분리되어 부호화된다. 텍스처 코딩은 MPEG-2 방식에서와 같은 방법으로 이루어져 있으며, 형상 코딩은 텍스처 코딩과 유사한 방법으로 이루어진다.

16 x 16 크기의 BAB 단위로 움직임 보상을 수행하고, 형상 정보는 CAE(Content-based Arithmetic Encoding)를 사용하여 부호화된다<sup>[2, 6]</sup>. MPEG-4의 형상 정보는 객체 내부와 외부 중 어느 쪽에 포함되는지를 나타내는데, Rectangle, Binary Alpha Planes, Gray Scale Alpha Planes 세 가지 방식이 있다. Rectangle 방식은

한 프레임의 영상 전체를 의미하며 VO의 영역 분할 정보가 없는 경우에 사용한다. Binary Alpha Planes 방식에서는 각 픽셀의 값이 0일 때는 객체의 내부에, 1일 때는 객체의 외부에 있음을 나타낸다. 제안하는 방법은 Binary Alpha Planes를 사용하는 경우에 적용된다.

그림 1은 Binary Alpha Plane을 BAB단위로 구분한 그림이다. Transparent BAB는 객체가 존재하지 않는 영역을 나타내며, Opaque BAB는 객체 내부에 존재하는 블록을 의미한다. Transparent와 Opaque의 경우는 모든 픽셀이 동일하게 0 혹은 1의 값을 가지므로 정보량이 적다. VOP의 경계 부분에 위치한 Boundary BAB의 경우는 정보량이 많으므로, I-VOP의 경우에는 CAE 코딩을 행하며 P-VOP와 B-VOP의 경우에는 움직임 보상과 CAE 코딩을 병행한다.

MPEG-4의 BAB의 형상 코딩은 표 1에 보인 바와 같이 일곱 종류의 BAB 모드가 정의되어 있다. I-VOP는 0(all\_0, transparent) 및 1(all\_255, opaque), 2(인트라코딩, boundary)의 코딩 모드만을 가진다. 모드가 0(Transparent)일 때와 모드가 1(Opaque)일 때에는 코딩 모드 정보의 표현이 간단하다. 모드가 2일 때는 VO의 Boundary에 해당하는 경우이므로 하나의 BAB 내의 16 x 16개의 픽셀 정보를 인트라 코딩한다.

P-VOP와 B-VOP는 인접하는 프레임에 있는 어떤 기준 VOP와의 차이를 구하여 압축하는 방법을 사용하므로 다양한 코딩 모드가 있다. 0~2 코딩 모드는 I-VOP의 경우와 동일하며, 3~6 코딩 모드에서는 현재 프레임에서 VOP의 형상이 기준 프레임의 VOP로부터 효율적으로 예측하는 것이 어려운 경우에만 인트라 코딩을 수행한다. 즉, 기준 VOP로부터 예측한 형상(MVPS)과 현재 프레임의 형상(MVDS)의 차이(MVDS=MVS-MVPS)가 허용

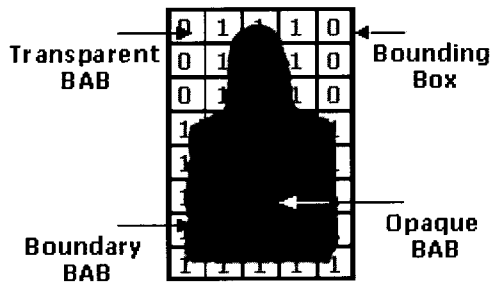


그림 1 형상코딩의 타입

표 1 형상 코딩 모드

형상코딩모드	코딩 타입	Used in
0	Transparent(all_0)	I,P,B VOP
1	Opaque(all_255)	I,P,B VOP
2	Intra(CAE)	P,B VOP
3	MVDs=0 & no inter update	P,B VOP
4	MVDs!=0 & no inter update	P,B VOP
5	MVDs=0 & inter coded	P,B VOP
6	MVDs!=0 & inter coded	P,B VOP

오차 범위 이내에 있으면 움직임 보상만 행하며(No Inter Update), 그 차이가 허용 오차 범위를 초과하면 그 차이의 값에 대해 인트라 코딩한 결과와 현재 프레임 VOP 형상을 인트라 코딩한 결과를 비교하여 비트 수가 적은 쪽을 택하여 전송한다.

2. Erol의 형상 근사화 기법

Erol은 각각의 BAB 블록이 VOP 내부,외부 및 경계선 중에서 어느 쪽에 포함되는지를 MPEG-4 형상 코딩에서 발생하는 여러 경우를 확률적으로 유추하여 근사적으로 얻는 방법을 제안하였다.

MPEG-4의 BAB는 Transparent, Opaque, Boundary 중 하나의 타입에 속한다. I-VOP의 BAB로부터는 직접 Transparent, Opaque, Boundary 정보를 얻을 수 있다. 그러나 P-VOP와 B-VOP의 경우에는 모두 일곱 가지의 코딩 모드가 존재하며, 기준 VOP의 형상으로부터 움직임 예측과 보상을 수행하므로 기준 VOP의 BAB 형상과 현재 VOP의 예측 형상을 디코딩한 후에야 BAB의 타입을 구할 수 있다. Erol은 이같은 처리 과정 없이 BAB 타입을 근사적으로 얻는 방법을 제안하였다.

표 2에 Erol의 방법이 요약되어 있다. 우선, BAB 타입이 Transparent, Opaque, Boundary 일 경우에는 그 정보를 직접 이용한다(1,2,3). 형상의 움직임 벡터나 예측 형상을 갱신하는 경우에는, 예측 BAB는 기준 블록의 형상 코딩 모드에 상관 없이 VO의 Boundary 블록일 가능성이 높다고 본다(4,5,6).

움직임 벡터 정보와 갱신 정보를 모두 코딩하지 않는 MVDs == 0 & No Inter Update 모드의 경우에는(7,8,9), MPEG-4의 형상 코딩에 대한 가변장 부호표<sup>[6]</sup>의 구성을 참조한다. 그 기준에 의하면,

표 2 P-VOP와 B-VOP에 대한 Erol의 형상 근사화 방법

	Block coding mode in the reference VOP	Block coding mode in the predicted VOP	Approximated block coding mode
1	Transparent, opaque, intra	Transparent	Transparent(outside)
2	Transparent, opaque, intra	Opaque	Opaque(inside)
3	Transparent, opaque, intra	Intra	Intra(boundary)
4	Transparent, opaque, intra	MVDs!=0 & no inter update	Intra(boundary)
5	Transparent, opaque, intra	MVDs=0 & inter coded	Intra(boundary)
6	Transparent, opaque, intra	MVDs!=0 & inter coded	Intra(boundary)
7	Transparent	MVDs=0 & no inter update	Intra(boundary)
8	opaque	MVDs=0 & no inter update	Opaque(inside)
9	intra	MVDs=0 & no inter update	Intra(boundary)

기준 블록과 예측 블록이 모두 Opaque한 경우, 예측 블록은 MVDs=0 & No Inter Update 모드로 코딩하는 것이 효율적이다. 그러므로 이 경우에 기준 블록이 Opaque이면 예측되는 블록도 Opaque일 가능성이 크다. 한편, 기준 블록과 예측 블록이 모두 Transparent이면 예측 블록은 Transparent 코딩이 효율적이다 따라서, 기준 블록이 Transparent 이고 MVDs=0 & No Inter Update 모드일 경우, 예측 블록의 타입은 Transparent이거나 Opaque일 가능성이 낮으므로 Boundary일 확률이 제일 크다.

### 3. 키 VOP 선정 방법

Erol의 방법에서는 기준 프레임에 있는 VOP (VOP-1)와 다른 프레임의 VOP(VOP-2)의 유사도의 크기를 계산하여 그 유사도가 어느 임계 값보다 작으면 그 둘은 다르다고 간주하여 VOP-2를 키 VOP로 선정한다. 유사도 측정 방법으로는 Hamming Distance와 Hausdorff Distance를 사용하였다. Distance의 계산에 있어서 Transparent, Boundary, Opaque BAB는 각각 0, 1, 2 값으로 대체된다. Distance가 적응적 임계값보다 커지면 그 프레임의 VOP가 키 VOP로 선정된다.

### III. 제안하는 키 VOP 선정 방법

제안하는 방법에서는, Erol의 방법을 개선하여 근사화된 형상을 얻은 후, VOP 사이의 유사도를 각 프레임 간 측정한다. BAB 단위로 이루어진 형상 간의 NMHD 값을 유사도로 이용하여 재귀적 2-means 클러스터링을 행한다. 클러스터들의 경계 프레임과 중간 프레임의 VOP 들이 후보 키 VOP로 선택되며, 이들의 적합성 순위 (Rank)를 정한다.

#### 1. 제안하는 형상 근사화 방법

P-VOP, B-VOP의 형상 근사화 과정에서 기준 VOP의 형상 정보를 참조하는 Erol의 방법과는 달리, 제안하는 방법에서는 형상 코딩 모드의 Transparency와 8-Neighbor 블록의 연결성만을 고려하여 BAB 단위로 형상을 근사화하여, I-VOP, P-VOP 및 B-VOP 모두에 대해서 우수한 근사화 형상을 얻는다.

제안하는 방법을 그림 2에 보였다. 형상은 BAB 단위로 나타내며, O 로 표시된 것은 Transparent BAB를, N 으로 표시된 것은 Non-Transparent BAB를, N/O 로 표시된 부분은 둘 중의 임

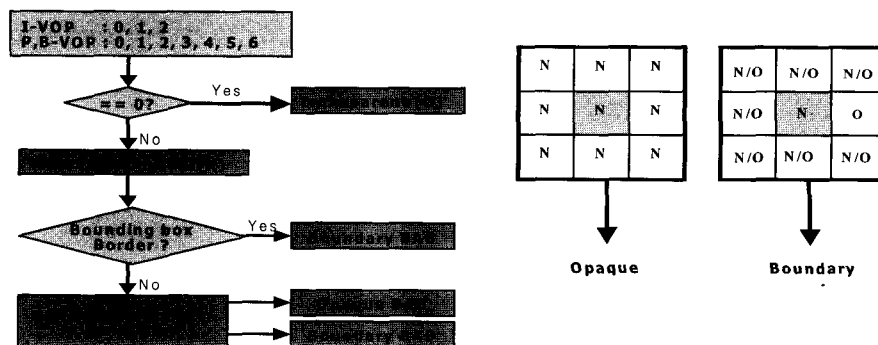


그림 2 형상 근사화 알고리즘

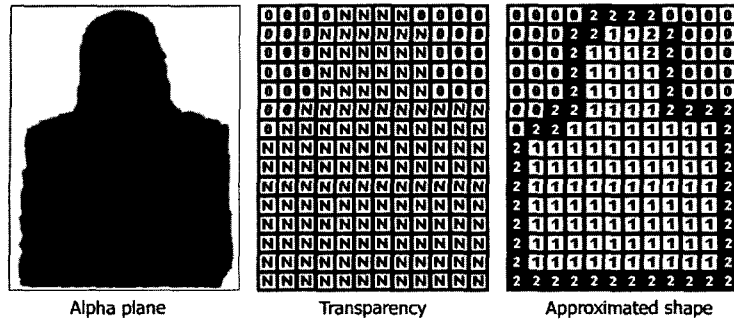


그림 3 형상 근사화의 예시

의의 것이라는 것을 의미한다.

표 2의 여러 경우를 간략화하면 Transparent BAB와 Non-Transparent BAB (Boundary, Opaque)의 두 종류로 나눌 수 있으며, Transparent의 경우는 Erol의 방법에서와 마찬가지로 간단하다. Non-Transparent의 경우에는 다음과 같이 결정한다. 첫째, Non-Transparent로 근사화되는 BAB가 바운딩 박스의 테두리와 인접한 경우 그 BAB는 Boundary로 설정한다. 왜냐하면 Transparent 블록과 Opaque 블록의 경계는 바로 Boundary 블록이기 때문이다. 둘째, Non-Transparent로 근사화되는 BAB들의 8-Neighbor 주위에 Transparent한 블록이 존재하는지 조사한다. Neighbor 블록이 모두 Non-Transparent일 경우 그 BAB는 VOP의 내부가 되므로 Opaque로 설정한다. 만약 8-Neighbor 중 하나 이상이 Transparent한 블록이라면 그 블록은 Transparent 블록과의 경계 부분일 것이므로 Boundary BAB로 설정한다.

그림 3에 형상 근사화 과정을 나타내었다. 왼쪽 그림은 알파 평면이며, 중간 그림은 투명성 판단의

결과이며, 오른쪽 그림은 8-Neighborhood의 연결성을 고려하여 윤곽 정보를 추출한 결과이다.

### 2. 유사도의 측정

본 논문에서는 클러스터링의 특성 벡터로서 NMHD를 사용한다. 평균값을 사용한 이유는 작은 오류에 덜 민감하도록 하기 위해서이다. 식 (1)에서 A와 B는 비교하고자 하는 두 VO의 윤곽의 BAB 블록 단위의 집합이며,  $N_c$ 는 Boundary BAB의 개수를 나타낸다.  $D_b(a)$ 는 Hausdorff Distance를 나타내며, A의 모든 BAB에 대한 평균 값을 정규화 하여 NMHD를 얻는다. 이 때 두 VO의 형상간의 거리를 구할 때 무게 중심을 일치시킨 후 거리를 계산하였다<sup>[3]</sup>.

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

$$D_B(a) = \min_{b \in B} d(a, b) \tag{1}$$

VOP들의 유사도를 계산할 때, 기준 VOP의 선정에 따라 결과가 다를 수 있다. 본 논문에서는 기준 VOP로서 초기에는 첫 프레임의 VOP를 사용하나, 이 후 재귀적 클러스터링을 수행함으로써 이와 같은 문제를 피할 수 있다.

### 3. 재귀적 2-means 클러스터링

재귀적 키 VOP 선정 방법을 그림 4에 보였다. BAB 단위의 형상들은 프레임 단위로 기준 VOP와의 NMHD 값에 의해 2-means 클러스터링된다. 프레임들은 SVOP(Similar VOP) 클러스터와 NSVOP (Non-Similar VOP) 클러스터로 분리되며, NSVOP 클러스터의 크기에 따라 재귀적 실행 여부를 판단한다. SVOP와 NSVOP는 각 클러스터의 시간적 지속성을 고려하여 후보 키 VOP로서의 적합성 순위(Rank)를 부여한다.

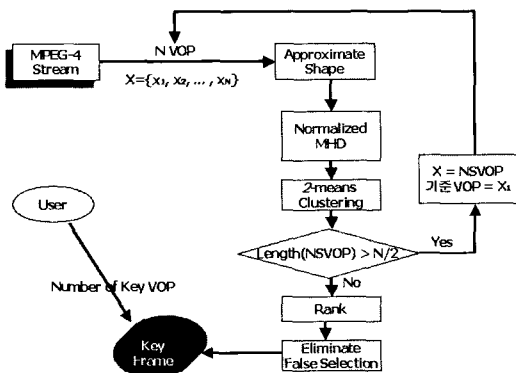


그림 4 제안하는 재귀적 키 VOP 선정 방법

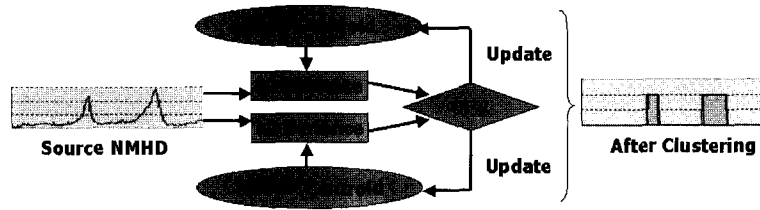


그림 5 SVOP 클러스터와 NSVOP 클러스터의 분리

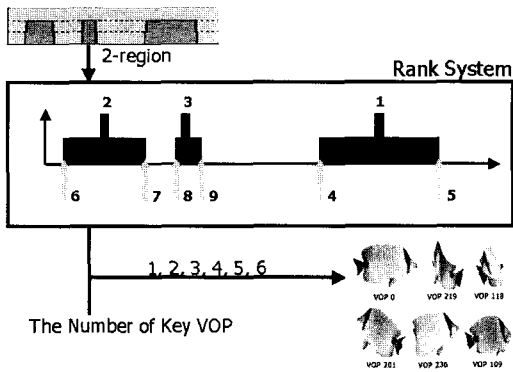


그림 6 후보 키 VOP의 Rank 결정

$X = \{x_1, x_2, \dots, x_N\}$  : VOP 특성벡터 집합  
 기준 VOP : 첫째 VOP

Step 1. 기준 VOP와 X와의 NMHD 계산  
 Step 2. NMHD를 2-means 클러스터링하여 SVOP, NSVOP영역 분리  
 Step 3. If  $\text{length}(\text{NSVOP}) > N/2$ ,  
 $X = \{x_i, x_{i+1}, \dots, x_j\}$ , 기준 VOP =  $x_j$ ,  
 go to Step 1.  
 Else go to Step 4.  
 Step 4. Rank 계산 및 Key VOP 선정.

2-means 클러스터링에서 초기 중심값은 모든 NMHD 값들의 평균( $m$ )과 표준편차( $\sigma$ )로부터 구한다. SVOP의 초기 중심값은 ( $m - \sigma$ ), NSVOP는 ( $m + \sigma$ )로 한다. 그림 5에 2-means 클러스터링에 의한 SVOP 클러스터와 NSVOP 클러스터의 분리 과정을 나타낸다.

2-means 클러스터링 후 어느 클러스터의 사이즈가 전체의 절반보다 클 경우 그 클러스터에 대하여 재귀적 클러스터링을 행한다. 재귀적 클러스터링을 사용하면 유사도의 측정에서 기준 VOP의 임의 선정에 따른 문제점을 피할 수 있다. 유사 코드로 나타내면 다음과 같다. Step 3의  $x_i, x_j$ 는 각각 하나의 NSVOP 클러스터의 시작 VOP와 끝 VOP의 프레임 번호이다. 재귀적 알고리즘은 NSVOP 클러스터가 전체의 절반보다 크지 않은 보통의 경우에는 비재귀적 알고리즘과 동일한 결과를 산출한다.

#### 4. Rank

각 NSVOP 클러스터들을 시간적인 분리와 지속성을 고려하여 VOP들의 Rank를 부여한다. 하나의 클러스터의 중간과 시작, 끝 위치의 프레임의

VOP들이 후보 키 VOP로 선정되며, 긴 지속 시간을 갖는 클러스터의 VOP들이 높은 값을 받는다. 하나의 클러스터 내에서는 중간의 VOP가 제일 높은 Rank를, 시작과 끝 프레임의 VOP들이 그 다음의 Rank를 갖는다. 선정된 후보 키 VOP 집합에서 다시 NMHD를 계산하고 클러스터링을 하여 서로 유사한 형상의 VOP들은 하나만 남기고 제거한다. 그림 6에 세 개의 클러스터가 존재하는 경우를 나타냈다.

#### 5. VOP 검색

요약된 키 VOP들을 대상으로 사용자의 질의 VOP와의 MHD를 유사도로 사용하여 가장 유사한 VOP를 검색하였다. 검색은 데이터베이스 사이즈에 따라 복잡도가 증가하므로, 요약된 키 VOP들만을 대상으로 검색함으로써, 검색 시간을 줄일 수 있다.

Step 1. VOP 요약하는 키 VOP 추출  
 Step 2. 질의 VOP와 키 VOP간의 MHD 측정  
 Step 3. 가장 작은 Distance를 갖는 키 VOP를 Retrieval

#### IV. 실험 결과

제안한 방법을 이용하여 MPEG-4 동영상에 대한 비디오 인덱싱과 검색을 행하였다. 실험은 인텔 펜티엄 III 800MHz CPU의 PC에서 Visual C++ 6.0으로 코딩하여 이루어졌다. ISO의 MPEG-4 공개 소스 코드<sup>[15]</sup>에서 Shape 코딩 부분만 변경하여 사용하였다. MPEG-4 동영상으로부터 근사화된 형상 정보를 얻은 후 계산 시간과 결과의 정확도를 기존의 방법과 비교한다. 이 후 재귀적 클러스터링에 의한 키 VOP의 선정 결과를 기존의 방법에 의한 결과와 비교한다. 마지막으로 VOP 검색 결과를 보이고 이를 분석하였다.

실험에 사용한 테스트 영상 열은 IPBBPBBPBB 포맷으로 코딩된 352x288크기의 Bream, Weather, Stefan, Cyclamen 이다. 각각 300 프레임으로 구성되어 있으며 하나의 VO 만을 가지고 있다.

##### 1. 형상 근사화 결과

제안하는 형상 근사화 방법과 Erol의 방법에 의한 실험 결과를 비교하였다. Bream, Weather, Stefan 클립의 매 프레임을 두 가지 방법으로 근사화 한 후 윤곽선의 형상을 BAB 단위로 다음의 식을 이용하여 구한 일치도의 값을 표 3에 나타내었다. 세 경우 모두, 두 방법에 의한 결과가 97% 이상으로서 거의 일치함을 알 수 있다.

$$\text{일치도} = \frac{\sum |S_{\text{Erol}}(b) - S_{\text{제안}}(b)|}{\#(\text{BAB in Bounding\_box})} \quad (2)$$

$$\text{, where } S(b) = \begin{cases} 0 & b: \text{transparent} \\ 1 & b: \text{Boundary} \\ 2 & b: \text{Opaque} \end{cases}$$

##### 2. 비디오 인덱싱 결과

먼저, Erol의 방법을 이용하여 키 VOP 들을 구하였다. 그림 7과 그림 8은 Bream 데이터에 대하여 각각 {T=0.25, T=0.4}를 사용하여 구한 결과이다. 그림 7의 결과는 키 VOP를 정확히 선정하였으나, 그림 8의 결과에서는 미검출된 키 VOP가 존재한다. Erol의 방법을 이용하여 Bream, Weather, Stefan 테스트 데이터 모두에 대해 키 VOP를 빠르게 선정할 수는 있었으나, T의 크기에 따라 미검출과 오검출되는 키 VOP가 발생한다. 임계값 T는 데이터에 따라 다르게 선정하였다.

표 3 Erol의 방법에 의해 근사화된 형상과의 비교

시험 데이터	일치도
Bream Clip	99.3 %
Weather Clip	97.3 %
Stefan Clip	99.2 %

제안하는 방법으로 구한 결과를 비교하였다. 첫째 시험에서는 일률적으로 각 객체의 등장 프레임에서의 VOP를 기준 VOP로 유사도를 계산하여 키 VOP를 선정하였다.

그림 9부터 그림 14는 Bream, Stefan, Weather 데이터에 대한 결과를 나타내고 있다. NMHD는 첫째 프레임의 VOP와의 거리를 나타내며, 각 클러스터의 0 의 값은 SVOP를, 1 은 NSVOP를 의미한다. 그림 9와 그림 10에서 NMHD 클러스터링에 의해 얻어지는 키 VOP는 Bream 클립에서 물고기가 방향전환을 하는 시간 축에서의 위치를 잘 검출한다. 그림 11과 그림 12 에 나타난 Stefan 클립에서 얻어진 키 VOP는 다양한 형상을 잘 나타내고 있으며, 그림 13과 그림 14의 Weather 클립의 결과는 앵커의 형상 변화를 잘 나타내었다. 결론적으로 세 종류의 데이터 모두에 대해 키 VOP 선정 결과는 Erol의 방법에서 적절한 T의 값을 사용하여 얻은 결과와 유사하며 프레임의 변화에 따른 형상의 변화를 빠르게 검출한다.

그림 15의 Cyclamen 클립은 화초를 촬영한 동영상으로서 카메라 움직임을 가지고 있으며 객체의 형상이 화면 전체를 차지한다. 재귀적 클러스터링 방법으로 얻는 키 VOP들은 이러한 경우에도 형상의 변화를 제대로 검출한다. 그림 16은 제안하는 방법으로 얻은 키 VOP 들이며, 화초의 모습을 제대로 요약했음을 알 수 있다.

##### 3. VOP 검색 결과 분석

VOP 검색의 성능을 판단하기 위해 요약된 키 VOP들 모두를 질의 VOP로 입력하여 검색 결과를 비교하였다. VOP 검색의 정합도로는 Swain 의 AMP(Average Match Percentile)를 사용한다<sup>[9]</sup>. MP는  $(n - R)/(n - 1) \times 100$  의 식으로 주어지는데, 여기서 n은 VOP의 수이며, R은 각 VOP를 질의 VOP로 사용하였을 때의 검색 결과에서의 유사도의 순위이다. AMP는 MP들의 평균을 취하여 구한다.

표 4는 Stefan, Cyclamen, Bream, Weather에

표 4 VOP 검색의 정합도

Test Stream	Correct Match Placement				AMP
	1st	2nd	3rd	>3rd	
Bream	15	0	0	0	100%
Stefan	48	0	0	0	100%
Weather	12	0	0	0	100%
Cyclamen	24	0	0	0	100%

대하여 AMP를 측정된 결과이다. 검색 대상은 모든 프레임의 VOP 전부이며, 각 키 VOP를 질의 입력으로 하여 실험하였다. 질의 입력이 검색 대상에 포함되어 있으므로 AMP는 100%가 된다. 모든 VOP를 질의 입력으로 시험한 경우에도 같다.

**4. Computation Time의 비교**

Erol의 방법과 제안하는 방법의 계산 시간은 두 부분에서 차이가 있다. 첫번째는 근사화된 형상을 얻는 과정이고 두번째는 클러스터링 과정이다. 제안하는 방법은 MPEG-4 압축 스트림에서 형상 코딩 모드의 값을 추출한 후 Thining 과정을 수행하면 근사화된 형상을 얻을 수 있다. 그러나 Erol의 근사화 방법은 기준 VOP와 움직임 벡터를 디코딩한 후, BAB단위에서의 움직임 보상을 해주어야만 한다. 네 종류의 데이터에 대한 시험 결과, 전체 300 프레임에 대해 근사화된 VOP 형상을 구하는 시간이 60~100 msec 이었으며, Erol의 방법을 사용할 경우에는 1.5배에서 4배까지의 시간이 소요되었다.

키 VOP를 얻기 위해 사용하는 재귀적 클러스터링 방법은 Erol의 임계값을 적용하는 Erol의 방법보다 더 많은 계산을 필요로 한다. 그러나, 전술한 바와 같이 재귀적 클러스터링에 의하여 키 VOP의 선정이 안정적으로 이루어질 수 있었다.

**V. 결론**

본 논문에서는 MPEG-4 압축 동영상으로부터 압축을 풀지 않고 근사화된 형상 정보를 추출하여 형상의 유사도에 의해 VOP를 클러스터링하여 키 VOP를 구하여 VO의 동작을 요약, 검색하는 새로운 방법을 제안하였다. 제안하는 방법을 대표적인 MPEG-4 테스트 영상 열에 대해 시험한 결과, 기존의 방법에 비해 근사화된 형상 정보를 더 짧은 시간에 얻을 수 있었고, 재귀적 클러스터링 방법을 사용하여 키 VOP들의 정확도를 높일 수 있었다.

**참고 문헌**

- [1] G. Pass, *et al.*, "Comparing Images Using Color Coherence Vectors", ACM Conference on Multimedia. Boston, November 1996.
- [2] ISO/IEC, "Information technology- Generic coding of moving pictures and associated audio information: Video", 13, 1995
- [3] B. Erol and F. Kossentini, "Automatic Key Video Plane Selection Using the Shape Information in the MPEG-4 Compressed Domain", Multimedia, IEEE Trans. On, 2000, pp 129-138.
- [4] B. Erol and F. Kossentini, "Retrieval of Video Objects by Compressed Domain Shape Features", IEEE 7th ICECS, 2000, pp 667-670
- [5] B.Gunsel, A. Murat Tekalp, van Beek and P.J.L., "Moving Visual Representations of Video Objects for Content-Based Search and Browsing", IEEE ICIP, 1997, pp 502-505.
- [6] ISO / IEC, "MPEG-4 Video Verification Model version 18"
- [7] A.Ferman, "Effective Content Representation for video", ICIP 1998, pp 521-525
- [8] A. Ferman, *et al.*, "Group-of-Frames/Pictures Color Histogram Descriptors for Multimedia Applications", ICIP, 2000. pp.65-68
- [9] M.Naphade, *et al.*, "A High-Performance Shot Boundary Detection Algorithm Using Multiple Cues", IEEE, ICIP 1998, pp 884-887
- [10] B.Gunsel and A. Tekalp, "Content-based Video Abstraction", ICIP, 1998, pp 128-132
- [11] Y. Zhuang, *et al.*, "Adaptive Key Frame Extraction Using Unsupervised Clustering", IEEE, ICIP 98, 1998, pp 866-870
- [12] H.Lee, *et al.* "Abrupt shot change detection using an unsupervised clustering of multiple features", ICASSP, 2000, pp 2015-2018.
- [13] K. Oh *et al.*, "New Hausdorff Distances Based on Robust Statistics for Comparing Images", IPIC, 1996, pp 21-24
- [14] ISO/IEC, "MPEG-4 video encoder/decoder"
- [15] <http://www.iso.ch/iso/>



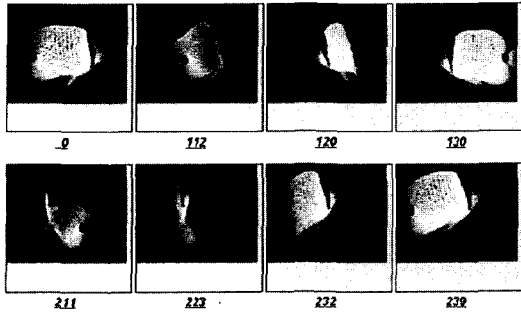


그림 7 Erol의 방법(T=0.25, 총 8개중 8개)

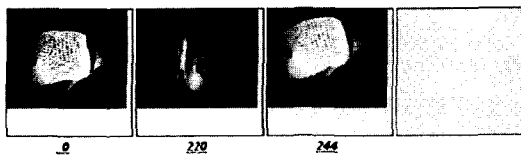


그림 8 Erol의 방법(T=0.4, 총 3개중 3개)

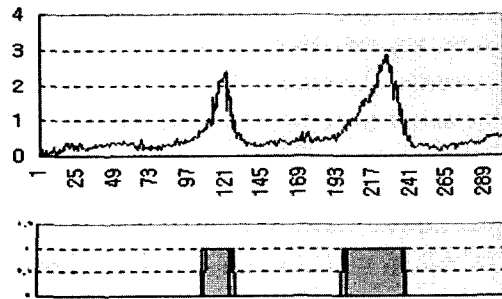


그림 9 Bream(위:NMHD, 아래: 클러스터링 결과)

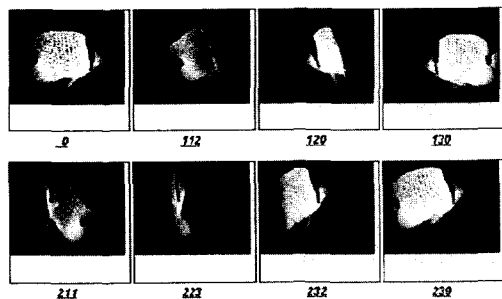


그림 10 제안하는 방법에 의한 키 VOP(Bream)

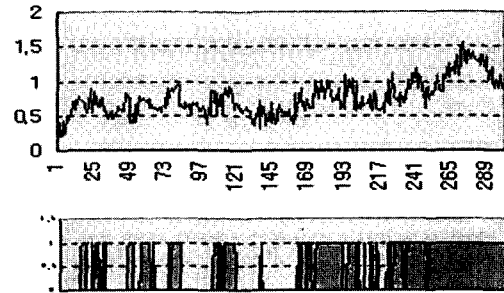


그림 11 Stefan(위: NMHD, 아래: 클러스터링 결과)

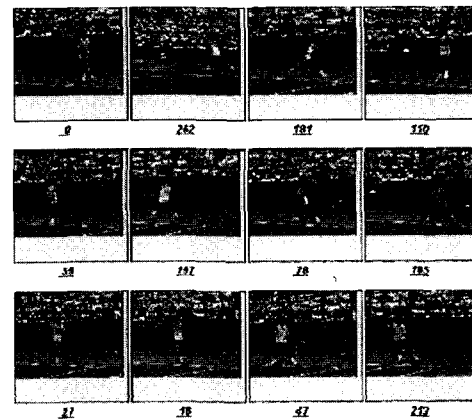


그림 12 제안하는 방법에 의한 키 VOP결과(Stefan)

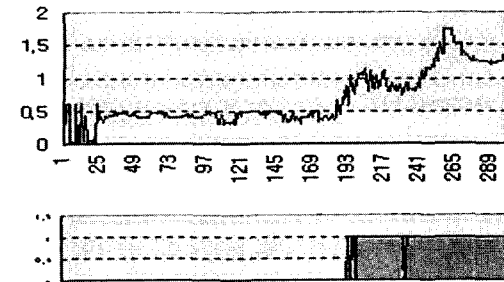


그림 13 Weather(위:NMHD, 아래: 클러스터링 결과)

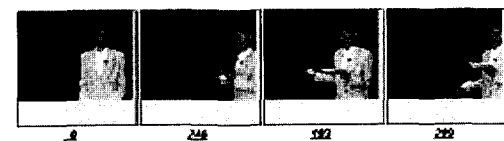


그림 14 제안하는 방법에 의한 키 VOP 결과(Weather)

