

# MPEG으로 압축된 뉴스 데이터에서의 DC성분 추출과 컬러 클러스터링을 이용한 앵커 프레임 검색 기법

정정훈, 이근섭, 오화중, 최병욱  
한양대학교 전자통신전파공학부

## An Anchor-frame Detection Algorithm in MPEG News Data using DC component extraction and Color Clustering

Jeong-hoon Jeong, Geun-seop Lee, Hwa-jong Oh, Byung-uk Choi  
Dept. of Electrical and Computer Engineering, Hanyang Univ.  
E-mail : laskill@hymail.hanyang.ac.kr

### 요약

대용량 비디오 데이터의 이용에 있어 효과적인 비디오 검색을 위해서는 비디오 데이터의 색인 과정이 필요하다. 효과적인 비디오 데이터의 색인을 위해서는 의미적 단위인 씬(scene)으로 이루어진 비디오 데이터를 물리적인 경계면인 컷(장면전환점)으로 검출하는 기법이 필수적이며 각 샷에서의 키 프레임 추출 또한 필수적이다. 본 논문에서는 뉴스 비디오 데이터의 키 프레임인 앵커 프레임의 효과적인 검색을 위해 DC 성분 추출과 이진 검색기법, 그리고 컬러 클러스터링을 이용하고 있다.

본 논문에서 제안하고 있는 방법을 검증하기 위해서 47분 10초 분량의 MPEG-2 로 압축된 뉴스 비디오 데이터에 적용한 결과

91.3%의 정확도와 84.0%의 재현율을 보여 제안한 방법의 우수성을 증명하고 있다.

### I. 서론

최근 멀티미디어 데이터에 대한 압축기술과 전송기술의 발달로 엄청난 양의 멀티미디어 데이터들이 매일 생산되고 있다. 이에 따라 멀티미디어 데이터의 저장과 색인의 필요성이 대두되고 있다. 한편 지속적으로 생산되며 많은 정보를 담고 있는 멀티미디어 데이터로는 뉴스 데이터를 꼽을수 있으며 뉴스 데이터에 대한 색인은 주로 기사 단위로 이루어진다. 뉴스 프로그램의 특성상 각각의 기사는 앵커 화면에 의해 나누어진다.

본 논문에서는 MPEG으로 압축된 뉴스데이

터의 색인을 위한 장면 전환점 및 앵커프레임 검출 방법으로 DC 성분 추출과 컬러 클러스터링을 이용한다.

본논문의 구성은 다음과 같다. II장에서는 I 프레임 추출과 DC 성분 추출에 대한 이유에 대해서 설명하고, III장에서는 장면 전환점 검출을 위한 이진 검색기법에 대해서 소개하며, IV장에서는 앵커 프레임들의 그룹화를 위한 컬러 클러스터링을 설명한다. 그리고 V장에서는 제안한 방법을 이용한 실험 결과를 소개한다.

## II. I 프레임과 DC 성분 추출

먼저 뉴스 프로그램에 대한 통계학적인 특징을 살펴본 후 I-프레임과 DC 성분 추출에 대한 타당성과 검출기법에 대해서 소개한다.

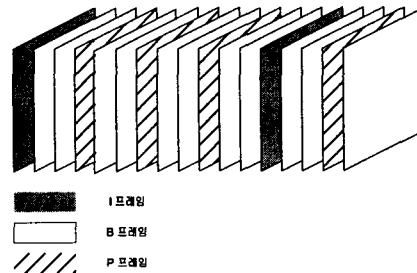
### 1. 뉴스 프로그램의 특징

뉴스 프로그램은 기사 장면 이전에 반드시 앵커 장면이 나오며 기사 장면은 여러개의 샷으로 이루어져 있다. 그리고 평균적인 샷의 길이는 약 2초이다. 즉 NTSC기준으로 약 60프레임마다 한번씩 장면 전환이 일어난다고 볼 수 있다. 그리고 뉴스 프로그램에서 사용되고 있는 디졸브(점진적인 장면 전환)의 경우 하나의 기사를 이루는 여러개의 샷들 사이에 다소 나타나며 약 17프레임에 걸쳐서 일어난다. 그러나 기사 장면과 앵커 장면 사이에는 디졸브가 나타나지 않으므로 디졸브에 의한 장면전환은 앵커 프레임의 검출에 있어서 무시할 수 있다.

### 2. I 프레임과 DC 성분 추출

현재 MPEG-2로 압축된 동영상의 경우 대부분이 12개의 GOP로 이루어져 있다. 이 때, I 프레임간의 시간 간격은 약 0.4초이다. 또한 뉴스 비디오에서의 평균 샷의 길이는 약 2초이므로 하나의 샷에는 약 5개의 I 프레임이 존재한다.

다음의 [그림 1]은 MPEG에서의 프레임 배열을 보여주고 있다.



[그림 1] MPEG에서의 프레임 배열

그림과 같이 I 프레임 사이에는 11개의 B 프레임과 P 프레임이 있으나 I 프레임만을 추출 하더라도 I 프레임 사이의 시간 간격이 0.4초이므로 I 프레임만으로 앵커 프레임을 검출하였을 때의 프레임 오차는 비디오 내용 이해에 큰 영향을 미치지 않는다.

또한 동일한 샷에서의 I 프레임을 추출할 경우 프레임간의 미세한 색상 차이로 인하여 다른 샷의 프레임으로 판별할 수 있다. 따라서 이러한 오류를 없애고 연산량을 감소 시키기 위해서 I 프레임에서의 DC 성분만을 추출하여 I 프레임을 재구성 한다. DC 성분은 8X8 픽셀 당 하나씩 존재하며 DC 성분으로 프레임을 재구성하면 원래 프레임의 1/64의 크기로 줄어든다.

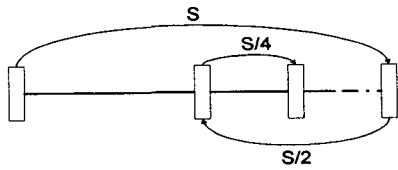


[그림 2] 원 영상과 DC 영상

[그림 2]는 원래의 I 프레임과 DC 성분만으로 재구성된 I 프레임을 보여주고 있다. 원래의 프레임보다 훨씬 작아진 프레임 크기를 확인할 수 있다.

### III. 이진검색 기법

DC 성분으로 재구성된 I 프레임으로부터 각각의 샷에 대한 키 프레임들을 추출하는 방법으로 영상검색에 필요한 계산량의 최적성을 위해서 이진 검색기법을 적용한다. 이진 검색 알고리즘을 적용하면 동일 장면내의 물체의 움직임의 영향으로 인한 후보구간의 검색에서의 오검출을 방지할 수 있다. 또한 순차검색에 의한 불필요한 검색을 줄일 수가 있어 검색속도를 향상시킬 수 있다.



[그림 3] 후보구간내 2진 검색 알고리즘

[그림 3]은 본 논문에서 사용한 이진 검색 알고리즘을 도식화해 보여주고 있다. S는 검색간격을 나타낸다. 뉴스 비디오에 평균 장면 전환은 2초 간격으로 일어난다. 즉 6개의 I 프레임 간격으로 장면전환이 일어나므로 검색간격은 6으로 고정한다. 이 경우 검색속도는 다음의 (식 1)에서 보이는 바와 같이 하나의

$$\text{속도향상} = \frac{n}{\log_2 n} \quad (\text{식 1})$$

검색구간에 대해서 순차 검색에 비해 약 1.89배 정도 빨라진다.

한편 프레임간의 유사성 판별을 위해서 히스토그램 비교법을 사용하며 그 식은 다음과 같다.

$$d(I_i, I_{i+1}) = \sum_{j=0}^{N-1} |H_i(j) - H_{i+1}(j)| \quad (\text{식 2})$$

$d(I_i, I_{i+1})$ 는 이웃 프레임과의 유사성을 측정하는 함수로,  $H_i(j)$ 는 i번째 프레임의 j번째 히스토그램 요소 값이라는 의미이다. 즉, 앵커 장면에서는 배경이 일정하므로 에지 히스토그램의 모양이 거의 비슷하다. 그러므로 차이값이 크게 나타나는 곳은 앵커 장면에서 기사

장면으로 넘어가는 곳이라고 할 수 있다. 그리고 기사 장면에서는 일반적으로 장면의 변화가 많아 이웃한 프레임과의 차이가 크게 나타나는 경향이 있는데, 이럴 경우 첫 프레임을 키 프레임으로 선택한다.

### IV. 컬러 클러스터링

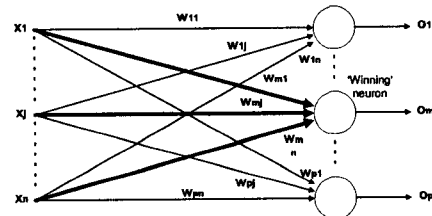
검출된 키프레임들 중에서 앵커프레임만을 그룹화하기 위해서 컬러클러스터링을 적용한다. 컬러 클러스터링을 적용하기에 앞서 추출된 키 프레임들의 Color space를 사람의 눈과 가장 유사한 Color space인 HSV로 변환한다. RGB와 HSV와의 관계식은 다음과 같다.

$$H = \text{CON}^{-1} \frac{\frac{1}{2} [(R-G) + (R-B)]}{\sqrt{[(R-G)^2 + (R-B)(G-B)]}}$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)]$$

$$I = \frac{1}{3} (R+G+B) \quad (\text{식 3})$$

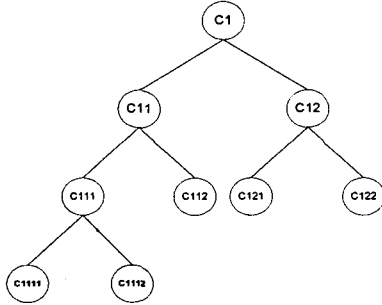
Color space가 HSV로 변환된 키 프레임들에 대해서 Kohonen net을 이용하여 유사한 이미지들을 클러스터링한다. 다음의 [그림 4]는 본 논문에서 사용한 Kohonen network 학습방법을 보여주고 있다.



[그림 4] Kohonen network 학습방법

[그림 4]에서 보이는 바와 같이 Single-layerd forward 뉴럴 네트워크의 형태를 가지며 무감독 학습 방법이다. 이 방식은 초기 weight 벡터 갯수에 의해서 출력의 갯수가 결정되므로 적절한 수의 클러스터를 위해서 이진 트리 클러스터기법을 동시에 적용한다. 다음의 [그

림 5]은 이진 트리 클러스터의 도식화된 모습을 보여주고 있다.



[그림 5] 이진 트리 클러스터

그림에서 보이는 바와 같이 크러스터링이 되는 특정 벡터가 어떠한 경계값을 만족하지 못할 때는 새로운 클러스터를 생성하여 그룹화된다.

### V. 실험결과 및 결론

실험에 사용된 뉴스 데이터는 MPEG-2로 압축이 되었으며 영상의 크기는 352 X 240 이고 총 47분 10초의 길이이며 84,836개의 프레임으로 구성되어 있다.

실험을 위한 프로그램으로는 MPEG 소프트웨어 그룹에서 제공한 MPEG-2 디코더 소스 프로그램을 이용하여 I 프레임과 DC 성분을 추출하였다. 이진 검색기법의 적용에 있어서 사용된 임계값은 실험에 의하여 2425로 하였으며 292차원의 HSV 벡터 내적값의 정규화된 수치 0.7을 클러스터링 경계값으로 하였다.

전체 프레임의 수	84,836
전체 I 프레임의 수	7,069
전체 기사 수	23
전체 샷의 수	637

[표 1] 실험에 사용된 뉴스 데이터 수치  
각 샷에 대한 키프레임 검출에 있어서 어두운 프레임일수록 밝기에 민감한 반응을 보였으며, 또한 기사 장면에서의 자막에서도 다른 샷으

추출될 키 프레임	637
추출된 키 프레임	608
오검출된 키 프레임	36

[표 2] 키 프레임 추출 결과

로 분류해 내었다. 또한 앵커장면과 인접한 기사 장면이 앵커장면과 유사할 경우 키프레임을 정확히 분류해 내지 못하였다.

그룹화될 프레임	23
그룹화된 프레임	25
그룹화 되지않은 프레임	2

[표 3] 클러스터링 결과

클러스터링에서는 앵커 프레임과 유사한 기사 프레임이 다소 앵커 프레임 그룹에 포함되었으며 앵커 프레임 중에서도 앵커의 오른쪽에 나오는 슈퍼포지션 캡션이 없거나 컬러가 지나치게 두드러지는 경우 앵커 프레임 그룹에서 제외되는 결과를 보였다.

위의 [표 3]에 따른 전체 정확도와 재현율은 다음의 표와 같다.

전체 정확도	91.3%
전체 재현율	84.0%

[표 4] 전체 정확도와 재현율

본 논문에서 제안한 방법으로 실제 뉴스 데이터에서 앵커프레임을 검색한 결과 전체 23개의 앵커 장면에서 21개의 앵커 장면에 해당되는 앵커프레임을 검출하였다.

### VI. 참고문헌

- [1] B.-L. Yeo and B. Liu, "Rapid scene analysis on compressed video", IEEE Tr. on CSVT, vol.5, no6, pp. 533-544, Dec.1995
- [2] B. Shaharay, "Scene Change Detection and Contents-based Sampling of Video Sequences", Proc. of SPIE, Vol. 2419, pp. 2-13, Feb. 1995