

# MEPG-7 서술자를 이용한 향상된 샷 경계 검출

강호경, 노용만  
한국정보통신대학원, 멀티미디어 그룹

## Enhanced shot boundary detection using MPEG-7 descriptors

Ho Kyung Kang, Yong Man Ro  
Multimedia Group  
Information and Communication University  
E-mail : kyoung@icu.ac.kr

### Abstract

본 논문에서는 MPEG-7 서술자들을 이용하여 효과적인 샷 경계 검출을 수행하는 방법에 대하여 제안한다. 기존의 샷 경계를 분석하는 방법은 기본적으로 각 프레임의 특징을 추출하여 거리 비교를 통하여 샷을 검출한다. 그러나 이러한 샷 검출은 각 프레임에서의 한가지 특징만을 이용하여 샷 경계를 찾지 때문에 샷의 변화와 프레임에 대한 특징 값의 변화가 일치하지 않는 경우 문제점이 발생된다. 이러한 문제점을 해결하기 위하여 MPEG-7 표준을 통하여 그 성능이 증명된 다양한 특징들을 동시에 이용하여 정확한 샷 경계를 추출하는 방법을 제안한다. 실험결과 제안한 방법은 급격한 샷 변화와 점진적인 샷 변화를 동시에 검출하였고 플래쉬와 같은 비디오의 순간적인 변화에 강인하였다.

### I. 서론

현재 다양한 멀티미디어 매체의 증가, 디지털 기술의 발전으로 방대한 량의 비디오 데이터가 생성되고 있다. 이러한 비디오 데이터의 효율적인 관리를 위하여 비디오에 대한 요약 및 분석이 필요하다. 하나의 샷은 동일한 비디오 정보를 담고 있으며 비디오의 내용기반 요약을 수행하는데 있어서 최상단위이다. 따라서 비디오 데이터의 관리를 위하여 샷 경계 검출은 필수적인 기술이다[1-3].

샷은 카메라의 시선이 바뀌는 것을 의미한다. 하나

의 카메라로부터 입력에서 다른 카메라로 변경되거나 편집을 통하여 샷이 생성된다. 샷은 편집 방법에 의하여 급격한 변화와 점진적인 변화로 나뉘어 진다. 급격한 변화는 컷으로 불리우고 편집시 다른 두개의 비디오를 바로 연결하는 경우 생성된다. 그리고 점진적인 변화의 경우 두 개의 비디오를 연결하는 부분에 특수한 효과를 집어넣어 샷에 대한 경계가 연속적으로 변화하는 것을 의미한다. 대표적으로 디졸브(Dissolve), 와이프(Wipe), 페이드 인/아웃(Fade in/out) 등이 존재한다.

이러한 샷 경계를 검출하기 위하여는 프레임에 대한 정보가 필요하다. [1-3]에서는 프레임에서의 칼라 히스토그램을 이용하거나 프레임 간의 특징으로 모션벡터를 사용하는 방법이 제안되었다.

본 논문에서는 MPEG-7에서 표준으로 완성된 서술자를 이용하여 비디오 프레임에 대한 정보를 추출한다. MPEG-7 서술자는 이미지에 대한 다양한 기술을 포함하고 있기 때문에 정확하고 효과적인 프레임 특징의 추출이 가능하다[4,5]. 또한 샷 검출시 샷의 변화와 각 프레임의 특징값 변화가 일치하지 않는 문제점에 대한 해결 방법을 제안한다.

본 논문은 우선 2장에서 샷 경계를 검출하는 시스템에 대하여 소개를 하고 3장에서 실험을 수행하고 그 결과를 고찰하였다. 그리고 4장에 결론을 내린다.

### II. Theory

#### 2.1 샷 경계 검출 알고리즘

그림 1 은 본 논문에서 제안하는 샷 경계 추출하는 시스템의 알고리즘을 나타내고 있다.

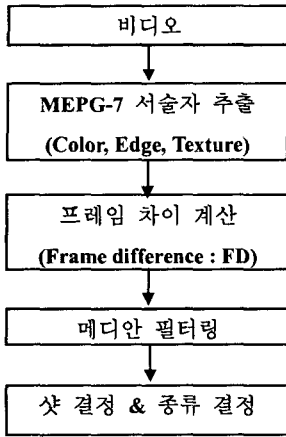


그림 1 샷 경계 추출 알고리즘

제안하는 시스템은 비디오 파일의 각 프레임에 대하여 MPEG-7 서술자들을 추출하고 추출된 서술자들을 이용하여 프레임간의 차이를 구하고 메디안 필터링을 통하여 노이즈 성분의 제거한다. 그리고 필터링된 프레임 차이를 이용하여 샷 경계를 구분한다.

### 2.2 MPEG-7 descriptors

본 논문에서는 비디오 프레임에 대한 정보를 추출하기 위한 MPEG-7 서술자로 프레임의 색 정보를 나타내는 Scalable Color Descriptor(CD), 프레임에서 경계 정보를 나타내는 Edge Histogram Descriptor(HD), 그리고 프레임의 거친 정도를 나타내는 Homogeneous Texture Descriptor(TD)를 사용한다.

CD 는 이미지 식별을 위한 중요한 특징 값으로써 사용되는데 프레임 내의 칼라의 분포를 나타내는 히스토그램으로 표현된다. 이미지의 ED 를 추출하기 위하여는 이미지를 16 개의 하위블록으로 나누고, 각 블록에 대해서 모두 5 개의 에지 성분, 수직, 수평, 45°, 135°, 그리고 무방향성의 에지 성분을 구한다. TD 는 이미지의 균질 또는 비균질의 패턴을 말하는 것으로 질감 특징 정보 표현을 위한 서술자는 주파수 영역을 6 개의 방향 성분과 5 개의 크기 성분으로 나누어, 모두 30 개의 채널을 구성하고 각 채널에 대한 질감 특징을 추출한다.

각 프레임에 대하여 추출된 MPEG-7 서술자는 이 이미지에 대한 특징을 표현하는 효과적인 방법이다.

### 2.3 Frame difference (FD)

본 논문에서는 샷에 대하여 급격한 변화와 점진적인 변화에 대하여 하나의 방법으로 우선 샷을 수행하고 샷의 종류를 분류한다.

급격한 샷의 경우 그 경계가 뚜렷하기 때문에 프레임 차이(FD)를 계산할 때 바로 이전 프레임과의 차이를 이용하면 된다. 이와 달리 점진적인 샷의 경우는 특수한 효과를 사용하여 다음 샷으로 변화하는 것을 의미한다. 이 경우 바로 이전 프레임과의 FD 를 측정하면 점진적인 샷의 변화를 알아내기 어렵다. 따라서 점진적인 샷이 끝난 후의 일정한 간격이(k) 떨어진 프레임간의 FD 를 계산한다. 점진적인 샷의 경우 15~25 프레임에 대하여 변화가 일어나는데 k 의 값이 큰 경우 같은 샷에서 많은 변화가 있는 샷에 대하여 오류가 발생되고 너무 작은 경우 점진적인 샷을 구별할 수가 없다.

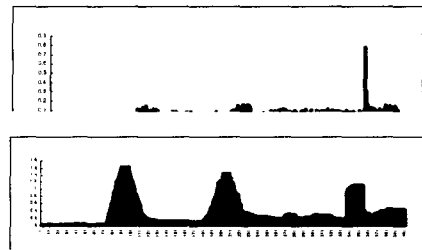


그림 2. k 값에 따른 FD, k=1 (상) k=20 (하)

그림 2 는 k 값에 따른 FD 를 보여준다. 그림 2 의 FD 는 프레임 순서별로 점진적인 샷 2 개와 급진적인 샷 1 개를 가지고 있는 구간이다. k 값이 1 인 경우 점진적인 샷에 대한 FD 가 매우 작음으로 검출이 불가능하다.

그리고 FD 를 구할 때 MPEG-7 서술자를 동시에 사용하기 위하여 식 1 과 같이 비디오에 대한 CD, ED, TD 의 평균을 구하여 평균화(Normalization)를 수행하여 거리를 측정한다.

식 1 에서 FD 는  $i, i-k$  사이의 프레임 차이를 나타내고  $\frac{1}{N} 는 i 프레임의 CD 를 의미한다.  $\frac{1}{N} 는 비디오에서 해당 서술자에 대한 평균값을 의미한다.$$

$$FD(i, i-k) = \left| \overline{CD}(i) - \overline{CD}(i-k) \right| / \alpha(\overline{CD}) + \left| \overline{ED}(i) - \overline{ED}(i-k) \right| / \alpha(\overline{ED}) + \left| \overline{TD}(i) - \overline{TD}(i-k) \right| / \alpha(\overline{TD}) \quad (1)$$

### 2.4 Median filtering

메디안 필터링은 FD 에 대하여 잡음 성분을 제거하는 역할을 수행한다. 메디안 필터링을 수행함으로써 같은 하나의 샷에서 플래쉬 등의 순간적인 변화에 의한 오류 샷의 발생을 막을수 있다. 또한 FD 를 구하는 간격인 k 값보다 작은 포인트로 필터링을 수행함으로써 급격한 샷과 점진적인 샷의 검출이 가능하다.

### 2.5 샷 검출

샷 검출은 식 2 와 같이 FD 값의 변화가 급격히 일어난 부분에 대하여 샷이 검출을 수행한다.

$$shot : \text{if } FD(i, i-k) > L \times FD(j, j-k) \quad (2)$$

여기서 i 프레임에서 구한  $FD(i, i-k)$  의 값이 j 프레임에서 구한  $FD(j, j-k)$  의 값의 L 배 이상 차이나는 경우, 즉 한계치를 넘어가는 경우 샷 경계가 검출되었음을 의미한다. j 값의 경우 점진적인 샷의 발생한 부분에 대하여 포함하기 위하여  $j = i + 2 \times k + 3$  으로 정한다.

샷의 종류를 구별하기 위하여 FD 값의 미분을 수행한다. 급격한 샷의 경우 FD 값 역시 빠르게 변화하여 점진적인 샷과의 구별이 가능하다.

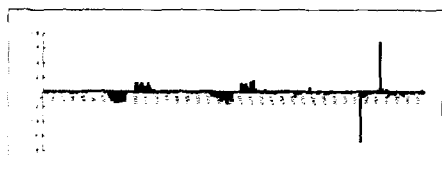


그림 3. FD 값의 미분결과

그림 3 은 그림 2 에 대한 FD 값의 미분결과를 보인다. 즉 81, 200 프레임에서 점진적인 변화가 발생하고 330 프레임에서 급격한 샷이 발생한다.

$$CUT : \text{if } FD'(i, j-k) < th \text{ and } FD'(i+20, i+20-k) > th \quad (3)$$

식 3 에서 보이듯이 샷의 급격한 변화인 컷의 경우 그림 3 의 330 프레임과 같이 급격한 변화를 가지고 있다. 따라서 컷에 대한 한계치(th = 11.3)에 의하여 검출이 가능하다.

## III. 실험결과

샷 경계 검출의 성능은 다큐먼트 비디오를 통하여 수행하였다. 선택한 다큐먼트 비디오는 8867 프레임의 총 45 개의 샷을 가지고 있고 25 개의 급격한 샷과 점진적인 샷으로는 12 개의 디졸브와 함께 8 개의 와이프가 포함되어 있다. 다큐먼트 비디오는 샷 검출에 효과적인 구조로 되어 있다.

$$Recall = \frac{N_{Correct}}{N_{Correct} + N_{Miss}} \quad (4)$$

$$Precision = \frac{N_{Correct}}{N_{Correct} + N_{False}} \quad (5)$$

식 4 의 Recall 은 샷을 검출에서 놓치는 경우를 나타내고 있고 식 5 는 샷이 아닌 경우를 샷으로 인식하는 것을 나타낸다.

제안된 알고리즘의 성능 비교를 위하여 CD 만을 사용한 경우와 CD, ED, TD 를 동시에 사용한 경우의 결과를 보인다.

표 1 은 CD 만을 사용했을때의 샷 경계 검출 결과이다. 한계치 L 값의 변화에 따라서 Recall 과 Precision 은 Trade-off 관계에 있다. 25 개의 급격한 샷은 모두 검출한 반면에 점진적인 샷에 대하여 1 개 검출하지 못하였고 또한 2 부분에서 샷이 아닌곳에 대하여 샷으로 인정한 부분이 존재하였다.

표 1. CD 를 이용한 샷 검출 결과

	L	샷 검출				정확도	
		Ori.	Cor.	Miss	False	Recall	Precision
CD	3.4	45	44	1	3	0.978	0.936
	3.5	45	44	1	2	0.978	0.960
	3.7	45	44	1	2	0.978	0.960
	3.9	45	44	1	2	0.978	0.960
	4.0	45	44	1	2	0.978	0.960
	4.1	45	43	2	2	0.956	0.956

표 2 는 CD, HD, TD 를 동시에 사용하여 샷을 검출

한 결과이다. CD 만을 사용하여 검출한 부분에서 샷이 아닌 부분에 대하여 에러를 줄일수 있음을 보여준다.

표 2. CD, ED, HD 를 이용한 샷 검출 결과

	L	샷 검출				정확도	
		Ori.	Cor.	Miss	False	Recall	Precision
CD ED HD	3.2	45	44	1	1	0.978	0.978
	3.3	45	44	1	1	0.978	0.978
	3.4	45	44	1	0	0.978	1.000
	3.5	45	43	1	0	0.978	1.000
	3.6	45	42	2	0	0.933	1.000
	3.7	45	44	4	2	0.911	0.953

샷의 종류별로 구분하면 최적의 한계치 L 값에서 컷의 경우 25 개 모두 검출하였고 점진적인 변화 중 디졸브에서 샷의 경계 검출에 대한 에러가 발생되었다.

다음의 그림 4 는 CD 만 이용한 경우 오류가 에러가 발생하고 CD, ED, HD 를 동시에 사용하여 극복되는 부분의 프레임을 보여준다. 물체가 갑자기 변화하는 경우 CD 만을 사용하는 경우 샷이 검출되었다. 그러나 CD, ED, HD 를 동시에 이용하는 경우 CD 만을 사용한 에러를 줄일 수 있다. 따라서 다양한 특징을 조합하여 보다 효과적인 샷 검출이 가능하다.

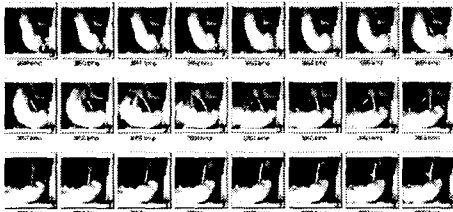


그림 4. CD, ED, HD 를 이용하여 극복된 부분

다음의 그림 5 는 하나의 샷 중간에 플래쉬 효과를 통하여 나타나는 FD 를 보인다.

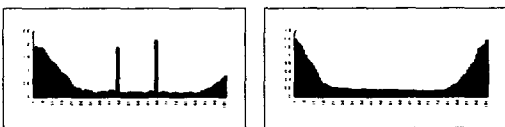


그림 5. 플래쉬에 대한 FD(좌), 메디안 필터후(우)

플래쉬가 생기는 경우 그림 5 와 같이 20 프레임

간격으로 피크가 발생된다. 이 경우 FD 에 대하여 간단히 메디안 필터링을 수행을 통하여 이러한 피크가 사라짐을 볼 수 있다.

#### IV. 결론

본 논문에서는 MPEG-7 서술자들을 이용하여 Shot 을 분석하는 시스템에 대하여 제안하였다.

기본적으로 칼라 정보를 이용한 샷 경계 검출 방법은 좋은 성능을 보인다. 그러나 실험을 통하여 보다 강인하고 효과적인 샷 검출을 위하여는 칼라정보 이외에 다른 정보들을 부가적으로 이용하는 것이 필요함을 보였다. 제안된 알고리즘은 각 프레임에 대하여 효과적인 정보를 추출하고 이를 복합적으로 이용하여 샷을 검출하였다. 이러한 정확한 샷 검출은 비디오 데이터의 요약 및 관리에 효율적으로 이용될수 있다. 또한 샷검출에 사용한 MPEG-7 서술자들은 비디오의 내용기반 검색의 특징으로 응용이 가능하다.

#### Acknowledgement

본 논문은 한국 과학 재단의 "이동 인터넷 기반의 내용 기반 검색"에 관한 특정 목적 기초 연구과제 수행의 일환으로 얻어진 연구 결과입니다.

#### 참고문헌

- [1] John S. Boreczky and Lawrence A. Rowe, " Comparison of video shot boundary detection techniques," *Proc. SPIE Conference on Storage and Retrieval Still Image and Video Databases IV*, San Jose, CA, February, 1996, pp.170-179.
- [2] R. Zabih, J. Miller, and K. Mai, " A feature-based algorithm for detecting and classifying scene breaks," in *Proc. ACM Multimedia 95*, 1995, pp.189-200.
- [3] H. Zhang, C. Y. Low, and S. W. Smoliar, " Video parsing and browsing using compressed data," *Multimedia Tools Applicat.*, Vol. 1, pp. 89-111, 1995.
- [4] Sang Heun Shim, Seung Ji Yang, Jeong Hyun Yoon, Ki Hyun Kim and Yong Man Ro, "Real-time Shot boundary detection for digital video camera using the MPEG-7 descriptor," *Real-Time Imaging VI*, Proceeding on SPIE Electronic Imaging, Vol. 4666, pp.161-171, 2002-01
- [5] Ho Kyung Kang, Yong Ju Jung, Hee Kyung Lee and Yong Man Ro, "Multimedia database system with embedding MPEG-7 meta data," *SPIE Electronic Imaging*, Jan 20.2001