

MPEG-7 기술자를 이용한 디지털 비디오 카메라 기반 실시간 샷 경계 검출

심상훈, 양승지, 윤정현, 노용만
한국정보통신대학원대학교 멀티미디어정보통신그룹

Real-time Shot Boundary Detection Based On Digital Video Camera Using The MPEG-7 Descriptor

Sang-Heun Shim, Seung-Ji Yang, Jeong-Hyun Yoon, Yong-Man Ro

Multimedia Group, Information and Communication University (ICU)

E-mail: ssheun@icu.ac.kr

요약

최근 대용량 디지털 비디오의 효율적인 이용 및 관리를 위한 데이터 베이스 구축이 절실히 요구되고 있다. 이를 위해서, 디지털 비디오의 내용 기반 요약이 필요하며, 선행 기술로서 비디오의 샷 경계 검출이 이루어져야 한다. 기존의 샷 경계 검출 방법들은 압축 파일로 저장된 비디오 데이터에 대해 샷 경계 검출을 수행한다. 이러한 기존 방법과 달리, 본 논문에서는 디지털 비디오 카메라로부터 비디오 데이터를 얻어 저장하면서, 실시간으로 샷 경계 검출을 수행하고자 한다. 그리고, 실시간 샷 경계 검출에 있어서 본 논문은 MPEG-7의 Homogeneous Texture와 Edge Histogram 기술자를 적용하여 추출된 각 비디오 프레임의 질감과 에지 정보를 이용한다. 이 방법은 비디오의 샷 경계 검출에 있어서 시간적으로 기존 방법들보다 효율적이고, 내용 기반 검색에 유용하다.

I. 서론

최근 디지털 비디오의 효율적인 이용 및 관리를 위해서, 디지털 비디오의 내용 정보에 대한 데이터 베이스 구축이 절실히 요구되고 있다. 이를 위해서, 비디오의 내용 기반 요약이 선행되어야 한다. 내용 기반 요약은 사용자가 원하는 비디오 정보에 비순차적인 접근이 가능하게 하고, 내용 기반 검색이 가능케 한다. 비디오를 내용별로 분할하기 위한 최소 단위는 비디오의 각 샷(shot)이다. 따라서, 비디오 내용 정보의 데이터 베이스화에 있어서 비디오의 샷 경계 검출(Shot Boundary Detection)은 필수적인 기술이다[1,2].

비디오의 샷 경계 검출이란 비디오에서 연속적으로 주사되는 각 프레임에 담겨있는 정보를 이용하여 샷이 전환되는 지점을 찾아내는 것이다. 이에 대한, 기존 연구는 크게 두 가지로 분류된다. 하나는 비디오를 이루고 있는 영상 자체를 이용하여 샷 경계를 검출하는 방법이고[2,3], 다른 하나는 압축된 비디오 정보로부터 복원과정을 거치지 않거나[2,4], 최소한의 복원 처리만을 통하여 샷 경계를 검출하는 방법이다[1,2]. 이러한 기존 방법들은 편집되어 압축 저장된 비디오에 대해서 샷 경계 검출을 수행함으로써, 내용 기반 요약에 있어서 부가적인 처리 시간을 필요로 한다.

본 논문에서는 압축 파일로 저장된 비디오에 대해서 샷 경계 검출을 하지 않고, 실시간으로 디지털 카메라로부터 얻어지는 비디오 데이터에 대해서 국제 표준 기술인 MPEG-7의 Homogeneous Texture와 Edge Histogram 기술자를 이용하여 각 비디오 프레임 데이터의 질감과 에지 정보를 추출하고, 추출된 질감과 에지 정보를 기반으로 샷 경계 검출을 수행한다. 따라서, 비디오의 샷 경계 검출 처리 시간에 있어서 기존 방법들 보다 매우 효율적이며, 큰 이득을 얻을 수 있다. 그리고, 국제 표준 기술인 MPEG-7을 적용함으로써, 비디오의 내용 기반 검색에도 유용하다.

영상의 질감과 에지 정보는 색, 형태, 움직임 등과 함께 영상의 내용 정보를 기술하는 중요한 시각적인 특징이다. 이 질감과 에지 정보는 영상의 구조, 방향성, 거친 정도 등의 특성을 나타내고, 비디오 데이터를 내용 기반 요약하기 위한 샷 경계 검출에 있어서 중요한 특징으로 이용될 수 있다[5,6,7,8].

본 논문의 구성은 다음과 같다. II 장에서 비디오 프레임의 질감과 에지 정보 추출 알고리즘에 대해

기술하고, III 장에서는 비디오 프레임간의 유사도 거리값 측정, IV장에서는 측정된 유사도 거리값을 이용한 샷 경계 검출 알고리즘에 대하여 설명한다. 그리고, V 장에서 비디오에 대한 실시간 샷 경계 검출 시스템에 대해서 기술하고, VI장에서 실험 및 결과를 설명한다. 마지막으로, VII장에서 결론을 맺는다.

II. 비디오 프레임의 질감과 에지 정보 추출

2.1 Homogeneous texture 기술자

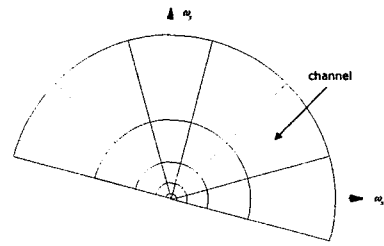
영상의 질감 정보는 옷감의 무늬, 벽, 타일 등의 균질 또는 비균질의 패턴을 말하는 것이고, 색상, 에지, 형태, 움직임 등과 함께 영상의 내용 정보를 기술하는 중요한 시각적 특징의 하나이다. 이러한, 영상 질감 표현의 국제 표준 기술은 MPEG-7 Homogeneous Texture 기술자이다.

MPEG-7의 Homogeneous Texture 기술자는 2차원 영상을 라돈 변환을 통해 얻어진 1차원의 시노그램(sinogram)을 1차원 DFT하여 직교 좌표계에 배치한 것이다. 직교 좌표계에서 라돈(Radon) 변환된 데이터는 저주파와 중간 주파수 영역에서 보다 조밀한 샘플 데이터를 가지게 되며, 고주파 영역에서는 적은 샘플 데이터를 가지게 된다. 이것은 인간 시각 특성(HVS: Human Visual System)과 관련된 저주파 정보의 강조 및 고주파 정보의 중요도 감소로 설명된다. [그림 1]에서 영상의 주파수 영역을 180도를 기준으로 방사 방향으로 각각 5개, 각 방향으로 6개씩 총 30개 채널로 나눠 인간 시각 특성을 이용하는 것을 보여준다. 이렇게 나눈 각 채널 에너지의 평균값과 분산값을 추출하여 Homogeneous texture 기술자의 특징 벡터로 이용한다[5,6].

Homogeneous Texture 기술자의 특징 벡터는 영상 전체의 평균값과 분산값, 각 채널 에너지의 평균값과 분산값으로 구성되는 총 62개의 특징값으로 구성된다. 이 특징 벡터는 식 (1)과 같다.

$$TD = \{dc, std, e_1, e_2, \dots, e_{30}, d_1, d_2, \dots, d_{30}\} \quad (1)$$

식(1)에서 TD 는 질감 기술자(Texture descriptor)의 특징값 벡터를 나타낸다. 그리고, dc 는 전체 영상의 평균값, std 는 전체 영상의 분산값, e 는 각 채널 에너지의 평균값, d 는 각 채널 에너지의 분산값을 의미한다.

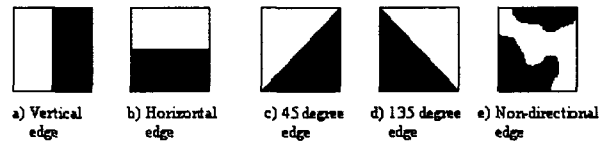


[그림 1] HVS에 의한 주파수 공간 분할

2.2 Edge Histogram 기술자

MPEG-7의 edge histogram은 영상이 지니고 있는 비 에지 정보를 추출하는 기술자이다. 이러한 edge histogram을 이용하여 실시간으로 얻어지는 비디오 프레임의 에지 정보를 추출한다. 그리고, 추출된 에지 정보를 이용하여 샷 경계 검출을 수행한다.

Edge histogram은 원 영상을 16개의 local block으로 나눠서 각 블록에 대해 [그림 2]에서 보이는 것과 같이 수직, 수평, 45도, 135도, 무 방향성의 5개 에지 정보를 추출하여 local edge histogram을 구한다.



[그림 2] 5가지 형태의 에지

그리고, 샷 경계 검출을 위한 비디오 프레임의 유사도 거리값 추출 과정에서, 5개의 local edge histogram을 이용하여, global과 semi-global edge histogram을 추출한다. 결과적으로 샷 경계 검출을 위하여 16 x 5 개의 local edge, 5개의 global edge 13 x 5 개의 semi-global edge 성분을 사용한다 [7,8].

III. 프레임간 유사도 측정

3.1 Homogeneous texture 기술자

비디오 카메라로부터 비디오 데이터를 얻어 II장에서 설명했듯이, 각 프레임의 총 62개의 질감 특징값들을 추출한다. 샷 경계 검출을 위해서, 이 값들을 이용하여 프레임 간의 유사도 측정을 한다.

유사도 측정으로 인접한 두 프레임 i 번째 프레임 F_i 와 $i-1$ 번째 프레임 F_{i-1} 의 유사도 거리값 $D_i = d(F_i, F_{i-1})$ 를 구한다. 이 유사도 거리값은 식

(2)와 (3)으로 정의된다.

$$D_i = d(F_i, F_{i-1}) \quad (2)$$

$$= Wb + \sum_{m=0, n=0}^{S-1, K-1} \left[W_{e, mn} \left| \frac{e_{mn}^{(F_i)} - e_{mn}^{(F_{i-1})}}{\alpha(e_{mn})} \right| \right] + \sum_{m=0, n=0}^{S-1, K-1} \left[W_{d, mn} \left| \frac{d_{mn}^{(F_i)} - d_{mn}^{(F_{i-1})}}{\alpha(d_{mn})} \right| \right]$$

여기서,

$$b = \left| \frac{dc^{F_i} - dc^{F_{i-1}}}{\alpha(dc)} \right| + \left| \frac{std^{F_i} - std^{F_{i-1}}}{\alpha(std)} \right| \quad (3)$$

식 (2), (3)의 dc , std , e_{mn} , d_{mn} 는 식 (1)의 질감 벡터의 특징값들이고, S 는 5, K 는 6으로 정의된다. 그리고, W , $W_{e, mn}$, $W_{d, mn}$ 는 특징값 간의 가중치이고, $\alpha(\cdot)$ 는 비디오의 각 프레임에 대한 각 특징값들의 평균으로 특징값들의 정규화 과정에서 사용된다 [5,6].

3.2 Edge Histogram 기술자

비디오 프레임 간의 edge histogram 유사도 거리 값 측정법은 식 (4)와 같이 가중치가 부여된 유클리디안 거리 측정법이 사용된다. edge histogram 기술자는 II장에서 설명했듯이, 단지 local edge histogram 성분만 포함한다. global edge histogram과, semi-global edge histogram은 단지 유사도 거리 값 측정을 위해서만 사용되며, global edge histogram과 semi-global edge histogram은 local edge histogram으로부터 얻어진다[7,8].

$$D_i = d(F_i, F_{i-1}) \quad (4)$$

$$= \sum_{n=0}^{79} |Local_Edge_F_i[n] - Local_Edge_F_{i-1}[n]| +$$

$$5 \times \sum_{n=0}^{64} |Global_Edge_F_i[n] - Global_Edge_F_{i-1}[n]| +$$

$$\sum_{n=0}^{64} |Semi_Global_Edge_F_i[n] - Semi_Global_Edge_F_{i-1}[n]|$$

IV. 샷 경계 검출 알고리즘

3.1 급진적 샷 경계 검출: 컷

비디오의 샷 경계 중에서 컷(Cut)은 인접한 두 프레임 사이에서 급진적(abrupt) 변화를 보이는 샷 경계이다. 따라서, MPEG-7 기술자 특징값을 이용해 얻어진 인접한 두 프레임 간의 유사도 거리값이 주위의 유사도 거리값보다 갑자기 커지면, 이 부분에서 컷이 발생한 것으로 본다.

비디오의 급진적 샷 경계인 컷을 검출하기 위해서, 두 프레임 F_i 와 F_{i-1} 사이의 유사도 거리값 D_i 를

구한다. 여기서, $D_i = d(F_i, F_{i-1})$, $i=1, 2, \dots, N$ 이고, i 는 프레임 순서, N 은 전체 프레임 수를 나타낸다. 얻어진 유사도 거리값 D_i 전체에 m 간격의 sliding window를 한 프레임 씩 이동하면서 샷 경계 검출을 수행한다. 이때, 급진적 샷 경계인 컷에 대한 샷 경계 검출 조건은 다음과 같다. D_i 값이 아래 조건 1), 2)를 만족하면 i 와 $i-1$ 번째 프레임 사이에서 컷이 발생한 것이다[1].

- 1) D_i 가 m 간격 구간의 영역 내에서 가장 크다
- 2) D_i 가 sliding window 영역 내의 두 번째로 큰 유사도 값보다 n 배 이상이다.

여기서, m 은 20 ~ 40 사이의 일정한 값이고, n 은 1.5 ~ 3.5 사이의 일정한 값이다[1].

3.2 점진적 샷 경계 검출: 페이드, 와이프, 디졸브

비디오의 샷 경계 중에서 페이드(fade), 와이프(wipe), 디졸브(dissolve) 같은 점진적(gradual) 샷 경계는 인접한 두 샷 사이의 여러 개의 프레임들에서 발생한다. 그래서, 이러한 점진적 샷 경계는 인접한 두 프레임 간의 변화가 크지 않고, 급진적 샷 경계 검출 방법과 같이 인접한 두 프레임간의 유사도 거리값을 이용하는 방법은 점진적 샷 경계 검출에 효과적이지 않다.

비디오의 점진적 샷 경계를 검출하기 위해서, 각 프레임 데이터의 추출된 Homogeneous Texture와 Edge Histogram 기술자 특징값을 이용해 k 간격의 두 프레임 F_i 와 F_{i-k} 간의 유사도 거리값 D_i^k 을 구한다. 여기서, $D_i^k = d(F_i, F_{i-k})$, $i=k+1, 2, \dots, N$ 이고, i 는 프레임 순서, N 은 전체 프레임 수, k 는 16 ~ 24 사이의 일정한 값이다. 점진적 샷 경계 검출 조건은 아래 1), 2)와 같다[1].

- 1) 유사도 거리값 D_i^k 가 일정 간격 s 구간 동안 유사한 값을 갖는다.
- 2) 유사도 거리값 D_i^k 가 일정 간격 s 구간 동안 $D_{i-(k/2)-1}^k$ 와 $D_{i+(k/2)+1}^k$ 보다 l 배 이상 크다.

위의 두 조건을 만족시키면 i 프레임에서 s 간격의 구간 동안의 점진적 샷 경계 변화가 일어난 것이다. 여기서, 값 s 는 3 ~ 7 사이의 일정한 값이고, l 은 1.5 ~ 3.5 사이의 일정한 값이다[1].

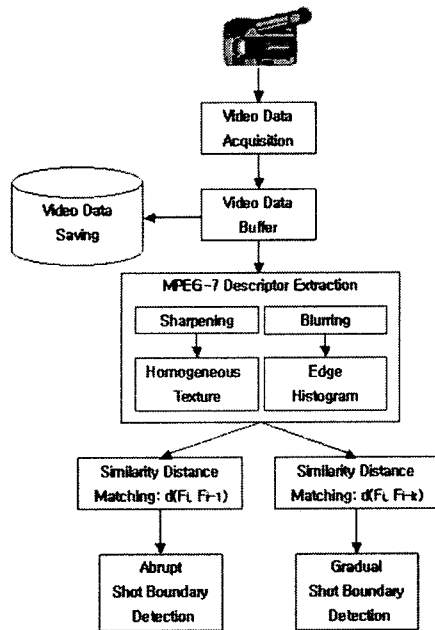
V. 실시간 비디오 샷 경계 검출 시스템

실시간으로 샷 경계를 검출할 디지털 비디오 데이터를 얻기 위해서, 디지털 비디오 카메라가 비디오 데이터를 처리할 수 있는 컴퓨터에 연결된다. 이 디지털 비디오 카메라를 이용하여 비디오 데이터가 실

시간으로 얻어지고, 이 입력되는 비디오의 각 프레임 데이터에 대해서 샤프닝과 블러링 효과를 준 후 Homogeneous Texture와 Edge Histogram 특징값이 추출된다. 추출된 각 프레임의 질감과 에지 특징값 간의 유사도 거리값을 구해서, 샷 경계를 검출한다.

이러한 실시간 디지털 비디오 획득, 저장, 샷 경계 검출 과정에서, 디지털 비디오 카메라로 USB PC 카메라를 사용한다. 그리고, 디지털 비디오 데이터는 VFW(Video For Windows) 인터페이스를 이용하여, 30fps 속도로 압축되지않은 QCIF(176x144), RGB24 형식으로 얻어지고, AVI 비디오 파일 형식으로 저장된다. 그리고, 샷 경계 검출을 하기 위한 비디오 콘텐츠는 다른 컴퓨터에서 실행되는 비디오 화면을 디지털 카메라로부터 직접 얻는다. 이러한 비디오 획득 방식을 사용한 이유는 비디오 데이터를 획득하여 저장하는 동시에 페이드, 와이프, 디졸브 같은 편집 효과를 줄 수 있는 디지털 비디오 카메라가 현재 존재하지 않기 때문이다.

[그림 3]은 실시간 디지털 비디오 획득, 저장 및 샷 경계 검출 알고리즘의 전체 블록도를 보여준다.



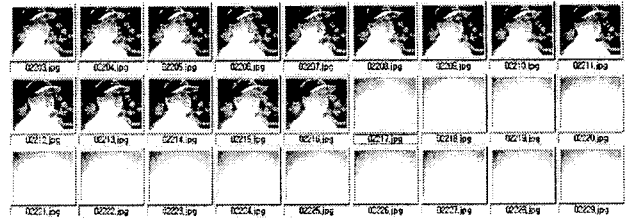
[그림 3] 실시간 비디오 획득 및 샷 경계 검출 시스템

VI. 실험 및 결과

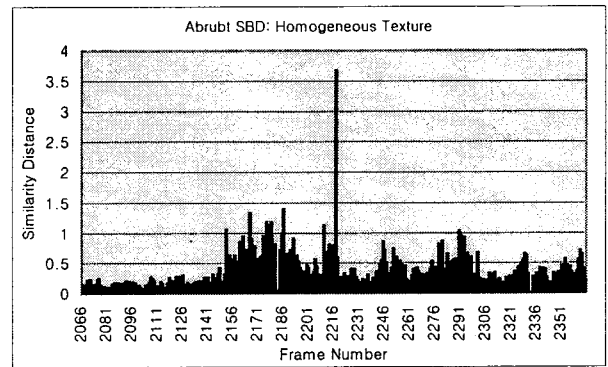
5.1 급진적 샷 경계 검출: 컷

[그림 4]는 비디오 카메라로부터 획득한 실험 비디오의 샷 경계 중에서 급진적 샷 경계인 컷을 보여준다. 그리고, [그림 5]와 [그림 6]은 [그림 4]의 컷에 대한 Homogeneous Texture와 Edge Histogram

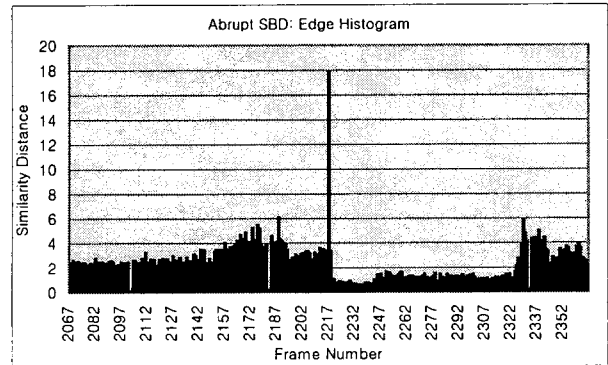
특징값을 이용해 구해진 인접한 두 프레임 간의 유사도 거리값 D_i 를 도시한 그래프이다.



[그림 4] 급진적 샷 경계: 컷



[그림 5] Homogeneous Texture를 이용한 Abrupt SBD: 컷



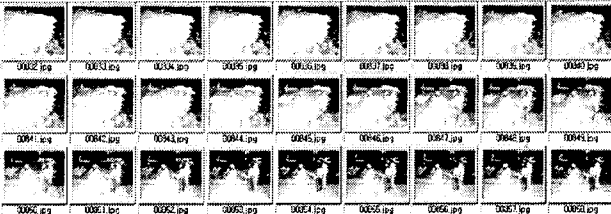
[그림 6] Edge Histogram을 이용한 Abrupt SBD: 컷

5.2 점진적 샷 경계 검출: 디졸브

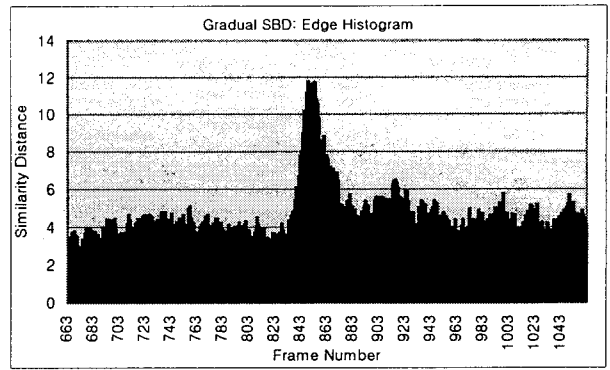
[그림 7]은 비디오 카메라로부터 획득한 실험 비디오의 샷 경계 중에서 점진적 샷 경계인 디졸브를 보여준다. 그리고, [그림 8]과 [그림 10]은 [그림 7]의 디졸브에 대한 Homogeneous Texture와 Edge Histogram 특징값을 이용해 구해진 인접한 두 프레임 간의 유사도 거리값 D_i 를 도시한 그래프이다. 이 두 그래프를 보면, 점진적 샷 경계인 디졸브를 검출하는데 어려움이 있다.

[그림 9]과 [그림 11]은 [그림 7]의 디졸브에 대한 Homogeneous Texture와 Edge Histogram 특징값을

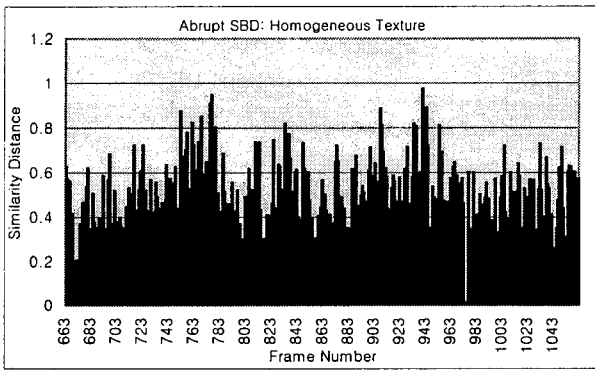
이용해 구해진 20 프레임 간격을 갖는 두 프레임 간의 유사도 거리값 D_i^{20} 를 도시한 그래프이다. 이 두 그래프를 보면, 점진적 샷 경계인 디졸브를 검출할 수 있다.



[그림 7] 점진적 샷 경계: 디졸브



[그림 11] Edge Histogram을 이용한 Gradual SBD: 디졸브

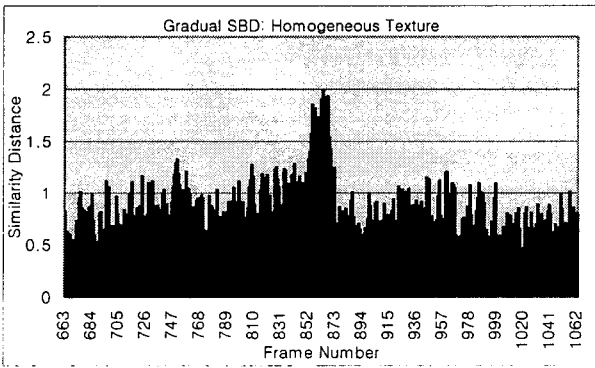


[그림 8] Homogeneous Texture를 이용한 Abrupt SBD: 디졸브

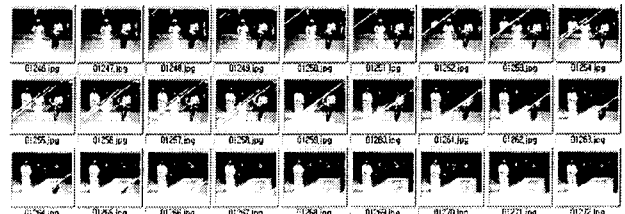
5.3 점진적 샷 경계 검출: 와이프

[그림 12]는 실시간으로 얻은 실험 비디오의 샷 경계 중에서 점진적 샷 경계인 와이프를 보여준다. 그리고, [그림 13]과 [그림 15]은 [그림 12]의 와이프에 대한 Homogeneous Texture와 Edge Histogram 특징값을 이용해 구해진 인접한 두 프레임 간의 유사도 거리값 D_i 를 도시한 그래프이다. 이 두 그래프를 보면, 급진적 샷 경계 검출 방법으로 점진적 샷 경계인 와이프를 검출하는데 어려움이 있다.

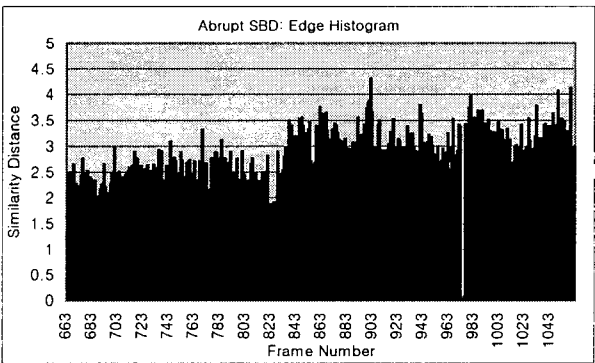
[그림 14]과 [그림 16]은 [그림 12]의 디졸브에 대한 Homogeneous Texture와 Edge Histogram 특징값을 이용해 구해진 20 프레임 간격의 두 프레임 간 유사도 거리값 D_i^{20} 를 도시한 그래프이다. 이 두 그래프를 보면, 점진적 샷 경계인 와이프를 검출할 수 있다.



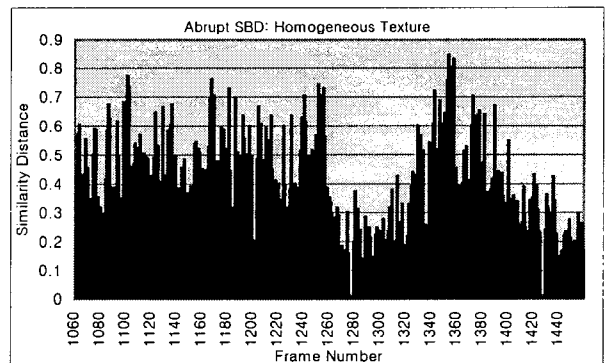
[그림 9] Homogeneous Texture를 이용한 Gradual SBD: 디졸브



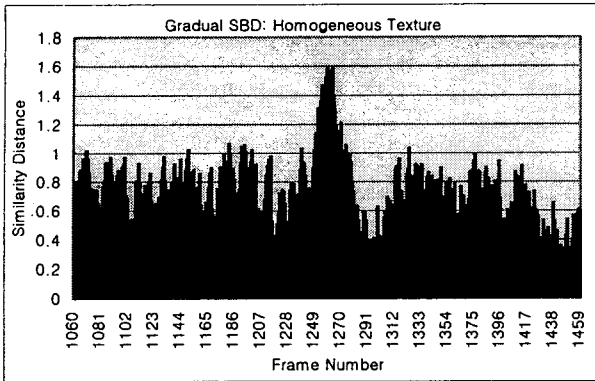
[그림 12] 점진적 샷 경계: 와이프



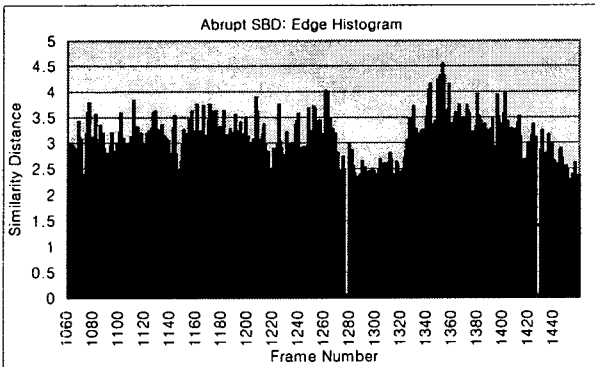
[그림 10] Edge Histogram을 이용한 Abrupt SBD: 디졸브



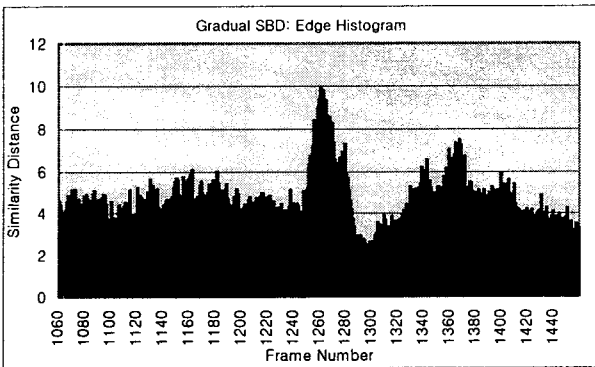
[그림 13] Homogeneous Texture를 이용한 Abrupt SBD: 와이프



[그림 14] Homogeneous Texture를 이용한 Gradual SBD: 와이프



[그림 15] Edge Histogram을 이용한 Abrupt SBD: 와이프



[그림 16] Edge Histogram을 이용한 Gradual SBD: 와이프

VII. 결론

최근 대용량 디지털 비디오의 효율적인 이용 및 관리를 위해서 내용 기반 요약 및 검색이 중요하게 인식되고 있다. 이를 위해서, 본 논문은 MPEG-7의 Homogeneous Texture와 Edge Histogram 기술자를 이용하여 디지털 비디오 카메라로부터 얻어지는 비디오 데이터에 대한 실시간 샷 경계 검출 방법을 제안하였다.

제안된 디지털 비디오 카메라 기반 실시간 샷 경계 검출 방법은 효과적이고, 기존 방법보다 시간적

으로도 매우 효율적이다. 그리고, 제안된 비디오의 샷 경계 검출 방법은 디지털 비디오에 대한 MPEG-7의 내용 기반 요약 및 검색에도 효과적으로 이용될 수 있다. 또한, 디지털 감시 카메라 분야에도 효과적으로 응용될 수 있다[9].

참고문헌

- [1] B.Yeo and B.Liu, "Rapid scene analysis on compressed video", *IEEE Trans. Circuits Systems Video Technology*, vol.5, no.6, pp.533—544, 1995.
- [2] 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.
- [3] 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.
- [4] H. Zhang, C. Y. Low, and S. W. Smoliar, "Video parsing and browsing using compressed data," *Multimedia Tools Applcat.*, vol. 1, pp. 89—111, 1995.
- [5] Ho-Kyung Kang, Yong-Man Ro, "Texture feature extraction based on HVS for MPEG-7 homogeneous texture descriptor", *SPIE 2000*, Vol. 4210, p72-81.
- [6] Y. M. Ro, K.W. Yoo, M.C. Kim, and J.W. Kim, Texture Description using Radon transform, *ISO/MPEG*, m4703, Vancouver, 1999
- [7] MPEG-7 Visual part of eXperimentation Model Version 8.0, *ISO/MPEG*, w3673, La Baule, Oct 2000.
- [8] Visual Working Draft 4.0, *ISO/MPEG*, w3522, Beijing, July 2000.
- [9] Elena Stringa and Carlo S. Regazzoni, Real-Time Video-Shot Detection for Scene Surveillance Applications, *IEEE Transactions on Image Processing*, Vol. 9, No. 1, January 2000