

## 객체이동 및 밝기변화를 고려한 샷 전환

## 탐지 알고리즘

이준구\*, 한기선\*\*, 유병문\*\*\*, 황두성\*

\*단국대학교 컴퓨터학과

\*\*강동대학교 방송영상미디어과

\*\*\*(주)엘앤와이비전

e-mail: leejg01679@hanmail.net

Shot Detection robust to object movement and  
brightness changes

Joon-Goo Lee\*, Ki-Sun Han\*\*, Byoung-Moon You\*\*\*, Doo-Sung Hwang\*

\*Dept. of Computer Science, Dan-Kook University

\*\*Dept. of Visual Broadcasting Media, Gang-Dong College

\*\*\*L&amp;Y Vision Technologies, Inc.

## 요 약

기존의 히스토그램을 이용한 샷 전환 탐지 방법은 연속적인 두 프레임의 전체 또는 대응되는 동일한 크기의 소 영역의 히스토그램을 사용하며, 객체의 이동이나, 프레임의 밝기 변화에 취약한 문제점이 있다. 본 논문에서는 이 문제점들을 해결하기 위하여 연속적인 두 프레임(현재와 참조 프레임)을 소 영역으로 분할한 후, 현재 프레임의 소 영역과 두 프레임사이에서 발생할 수 있는 객체의 이동을 고려한 참조 프레임에서의 소 영역의 비교, 그리고 참조 프레임의 소 영역에서 얻은 화소 밝기 히스토그램에 밝기 변화를 보정한 후, 현재 프레임의 소 영역에서 얻은 화소 밝기 히스토그램과 비교하는 방법을 제안한다. 제안한 방법은 영화와 뉴스 같은 비디오 데이터에 좋은 결과를 보였다.

## 1. 서론

디지털 기술 발전은 비디오 데이터의 빠른 생성과 더불어 컴퓨팅 자원의 활용을 증가시켰으나, 비디오 데이터의 관리를 위한 저장, 검색, 분류 등 연관 서비스가 효과적으로 제공되어 지고 있지 않은 실정이다. 비디오는 한 개 이상의 카메라를 사용하여 찍은 장면(frame)을 연속적으로 편집하여 구성이 되고, 비디오에서 하나의 카메라에 의해 찍힌 연속적인 장면을 샷(shot), 연속적으로 연결되어 있는 두 샷 사이의 전환되는 부분을 컷(cut)이라고 한다. 하나의 샷은 장면을 찍은 카메라의 파라미터(동작, 확대, 축소, 초점)와 샷의 공간 및 시간적 연속성에 의해 비디오의 내용(content)면에서 일관성을 가지는 최소의 단위가 된다. 샷 전환 탐지(shot boundary detection)는 대용량 비디오 데이터의 색인(indexing), 검색(browsing), 그리고 사건 탐지 등의 서비스를 효율적으로 제공하기 위한 기반이 된다.

샷 전환 탐지에는 프레임간의 화소 비교 방법[1,2], 히스토그램을 이용한 방법[3-7], 에지(edge)를 사용하는

방법[8], 등이 있으며, 이들 방법은 두 연속 프레임(현재와 참조 프레임)의 전체 또는 일부분에서 추출한 특성 값의 차이가 미리 설정한 임계치(threshold) 이상이면 두 연속 프레임 사이에 샷 전환이 발생된 것으로 탐지한다. Zhang 등[1]과 Xiaoquan Yi[2]은 샷 전환 탐지를 위해 두 연속 프레임사이에서 서로 대응되는 화소간의 값의 차이를 이용하는 프레임간의 평균 차이를 사용했다(1).

$$MAFD_n = \frac{1}{MN} \sum_{i=0}^M \sum_{j=0}^N |f_n(i, j) - f_{n+1}(i, j)| \quad (1)$$

여기서  $MAFD_n$ (mean absolute frame differences)은  $n$ 번째 프레임과  $n+1$ 번째 프레임사이의 화소 값의 절대 차이의 평균 값을,  $M$ ,  $N$ 은 프레임의 수직 방향과 수평 방향의 총 화소 수를,  $f_n(i, j)$ 는  $n$ 번째 프레임에서의  $(i, j)$  번째 위치한 화소의 밝기 값을 나타낸다. Zhang 등[1]은 두 연속 프레임에서 계산한 그레이-레벨 히스토그램(gray-level histogram)의 차이를, Kerhar 등[5]은 프레임의 화소값에서 구한 색도(Chromaticity) 히스토그램을, Cernekova 등[6]은 두 연속 프레임에서 계산한 칼라 히스토그램을 샷 탐지를 위해 사용했다(2).

$$CHD_i = \frac{1}{W} \sum_{r=0}^{B-1} \sum_{g=0}^{B-1} \sum_{b=0}^{B-1} |P_i(r, g, b) - P_{i+1}(r, g, b)| \quad (2)$$

이 논문은 행정안전부 국가기록원 재원으로 2012년 기록보존기술 연구개발사업의 지원을 받아 수행된 연구임.

여기서  $CHD_i$ (color histogram difference)은  $i$ 번째 프레임과  $i+1$ 번째 프레임사이의 칼라 히스토그램의 절대 차이이며,  $P_i(r,g,b)$ 는  $W$ 개의 화소를 가진 프레임의  $color(r,g,b)$ 의 개수이고,  $B$ 는 화소가 표현할 수 있는 칼라를 수이다. Zabih 등[8]은 연속적인 두 프레임에서 에지들의 생성 또는 소멸 정도를 측정하는 에지 변화율(edge change ratio)을 사용하여 두 연속 프레임사이의 샷 전환을 탐지했다.

두 연속 프레임사이의 동일한 화소 또는 구역을 비교하는 화소 비교 방법과 히스토그램을 이용한 방법은 계산상의 간편함으로 인해 보편적으로 사용되나 객체의 이동이나 화소의 밝기 등의 변화에 취약하고, 에지와 같은 특징을 이용한 방법은 밝기 변화나 객체 이동에는 강인하나 탐지를 위한 계산 량이 많고 잡음에 민감한 문제점이 있다.

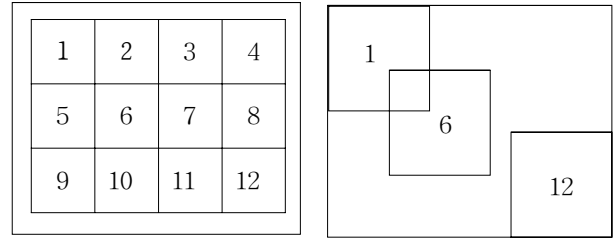
본 논문은 객체의 이동이나 화소의 밝기 변화에 강인한 히스토그램 기반의 급진적 샷 전환(abrupt shot change) 탐지 알고리즘을 제안한다. 제2장에서는 샷 전환 탐지를 위해 제안한 알고리즘을 설명하고, 제3장에서는 제안한 알고리즘을 국가기록원 소장 디지털화 영상에 적용한 결과를, 그리고 마지막 제4장에서는 결론을 맺는다.

## 2. 샷 전환 탐지 알고리즘

본 논문은 프레임 내에 존재하는 지역 정보를 활용하기 위해 각 프레임을 여러 개의 소 영역으로 분할한 후, 두 연속 프레임의 각 소 영역에서 획득한 화소의 밝기 값 히스토그램을 비교하여 샷 전환 탐지를 한다. 본 논문은 객체의 이동에 취약한 기존의 히스토그램 비교 방법의 문제점을 해결하기 위하여 참조 프레임의 소영역의 크기를 현재 프레임의 소 영역의 크기에 두 연속 프레임사이에서 발생하는 객체의 최대 이동 가능한 거리를 더한 크기로 결정한다. 또 참조 프레임의 소 영역에서 구한 밝기 값 히스토그램에 모폴러지의 그레이-레벨 팽창(morphological gray-level dilation) 연산을 적용하여 프레임간 밝기 변화에 강인한 방법을 사용한다.

### 2.1 객체의 이동을 고려한 영역 분할

한 대의 카메라에 의해 촬영된 샷은 시간적으로 일정한 간격 차를 두고 획득된 여러 개의 비디오 프레임으로 구성이 되며, 이러한 시간적 차로 인해 현재 프레임과 참조 프레임사이 존재하는 객체의 움직임은 샷 전환 탐지에 오 탐지의 한 원인이 된다. 현재 프레임과 참조 프레임은 그림 1과 같이 각각 12개의 소구역으로 분할되고, 이때 두 연속 프레임사이에서 발생하는 객체의 이동을 보상하기 위해 참조 프레임의 소구역의 크기는 현재 프레임의 소구역의 크기에 객체가 두 연속 프레임사이에서 이동 가능한 거리를 더한 크기로 결정이 된다.



(a) 현재 프레임 (b) 참조 프레임

(그림 1) 현재 프레임과 참조 프레임의 분할된 소 영역

현재 프레임의 소 영역의 크기는 식 (3)과 같다.

$$B_w = \frac{(f_w - 2d)}{4}$$

$$B_h = \frac{(f_h - 2d)}{3} \quad \text{----- (3)}$$

여기서  $B_w$ ,  $B_h$ 는 소 영역의 가로 및 세로 크기,  $f_w$ ,  $f_h$ 는 프레임의 가로 및 세로 크기,  $d$ 는 두 연속 프레임 사이에서 객체가 이동할 수 있는 거리이다.

참조 프레임의 소구역의 크기는  $(B_w+2d)*(B_h+2d)$ 이며, 참조 프레임의 소구역의 중심점은 현재 프레임의 소구역의 중심점은 일치한다.

### 2.2 히스토그램 추출 및 밝기 보상

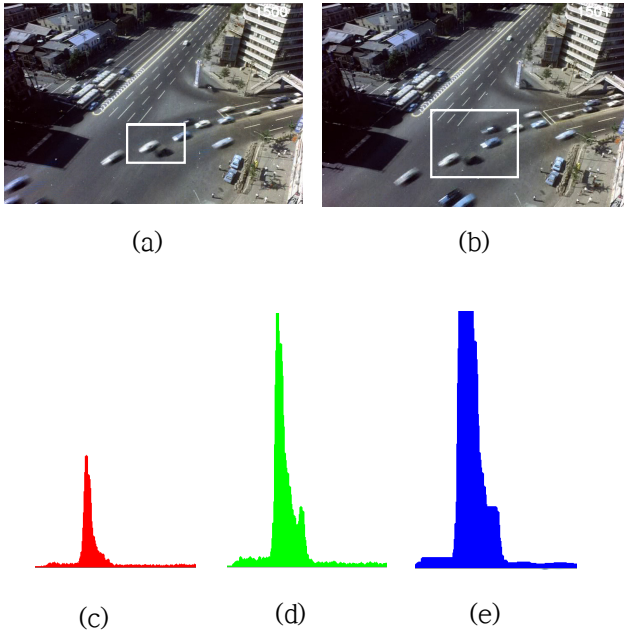
현재 프레임의 소구역(a)로 부터 화소 밝기 히스토그램  $H_c(c)$ 와, 참조 프레임의 소구역(b)로 부터 화소 밝기 히스토그램  $H_r(d)$ 가 생성된다(그림 2). 히스토그램  $H_r$ 에 모폴러지의 그레이-레벨 팽창 연산을 적용하여 샷 탐지를 위한 프레임간 히스토그램 비교 시에 밝기 변화에 취약한 점을 해결한다.

$$H_r(j) = \text{Max}(H_r(j+x) \oplus S(x)) \quad \text{----- (4)}$$

여기서  $H_r(j)$ 는  $H_r$  히스토그램의  $j$  번째 빈(bin)을,  $S$ 는 모폴러지를 연산을 위한 1차원의 모폴러지 마스크를, 그리고  $x$ 는 마스크의 크기를 나타낸다. 만약 구성 요소  $S$ 가 7 이고, 히스토그램의  $j$  번째 빈에 있으면,  $H_r(j)$ 는  $H_r(j-3)$ ,  $H_r(j-2)$ ,  $H_r(j-1)$ ,  $H_r(j)$ ,  $H_r(j+1)$ ,  $H_r(j+2)$ ,  $H_r(j+3)$  중에서 최대치 값을 취한다.

### 2.3 구역 히스토그램 매칭

그림 2에서 참조 프레임의 소 영역으로부터 얻은 히스토그램(d)이 현재 프레임의 소 영역으로부터 얻은 히스토그램(c)보다 더 많은 정보를 포함하고 있다. 참조 프레임의 소 영역으로부터 얻은 히스토그램은 현재 프레임의 소 영역으로부터 얻은 히스토그램이 포함하고 있는 정보와 함께 연속적인 두 프레임사이의 객체의 움직임과 밝기 변화를 보상한 정보를 가지고 있다. 그러므로 만약 연속적인 두 프레임이 같은 샷에 속한다면, 현재 프레임의 소



(그림 2) 현재 프레임과 참조 프레임의 소 영역과 밝기 히스토그램의 예 (a) 소 영역의 예를 보이는 현재 프레임, (b) (a)의 소 영역과 대응되는 소 영역을 보이는 참조 프레임, (c) 현재 프레임의 소 영(a)의 밝기 히스토그램, (d) 참조 프레임의 소 영역(b)의 밝기 히스토그램, (e) (d)에 그레이-레벨 팽창 연산이 적용된 후의 밝기 히스토그램

영역으로부터 얻은 히스토그램은 참조 프레임의 소 영역으로부터 얻은 히스토그램에 포함되어야 한다(식 5).

$$B_{p,j} \subseteq B_{p+1,j} \quad \text{----- (5)}$$

$B_{p,j}$ 는  $p$ 번째 프레임의  $j$ 번째 소 영역이고,  $B_{p+1,j}$ 는  $(p+1)$ 번째 프레임의  $j$ 번째 소 영역이다.

현재 프레임의  $j$ 번째 소 영역 히스토그램과 참조 프레임의  $j$ 번째 소 영역 히스토그램의 비교는 식 (6)와 같다

$$B_j = \sum_{k=0}^{n-1} B_{j,k},$$

$$B_{j,k} = 1 \text{ if } H_{j,r}(k) - H_{j,c}(k) \geq 0 \quad \text{---- (6)}$$

$$B_{j,k} = 0 \text{ if } H_{j,r}(k) - H_{j,c}(k) < 0$$

여기서  $n$ 는 히스토그램의 빈의 수이고,  $H_{j,r}(k)$ 와  $H_{j,c}(k)$ 는 참조 프레임과 현재 프레임의  $j$ 번째 소 영역에서 밝기의 값이  $k$ 인 화소의 개수를 나타낸다.

$B_j$ 의 값이 설정한 임계치( $t_1$ )보다 크면 비교하는 참조 프레임의 소 영역이 현재 프레임의 소 영역은 모든 정보를 포함하고 있다고 간주한다 (식 7).

$$d_j = \begin{cases} 1, & \text{if } B_j > t_1 \\ 0, & \text{otherwise} \end{cases} \quad \text{----- (7)}$$

마지막으로 현재 프레임의 소 영역에 모든 정보를 포함하고 있는 참조 프레임의 소 영역이 개수가 설정한 임계치 ( $t_2$ )보다 크면 연속한 두 프레임은 동일한 샷에, 아니면 샷 전환이 발생한 것으로 한다.

$$C_{p,p+1} = \begin{cases} 1, & \text{if } \sum_{i=0}^{n-1} (d_i) \geq t_2 \\ 0, & \text{if } \sum_{i=0}^{n-1} (d_i) < t_2 \end{cases} \quad \text{---- (8)}$$

$n$ 는 각 프레임의 분할된 소 영역의 개수이고,  $d_i$ 는 현재 프레임의역 소 영역의 모든 정보를 포함하는 참조 프레임의 수이다.

### 3. 실험 결과

본 논문에서 제안한 알고리즘의 성능을 평가하기 위하여 국가기록원이 소장하고 있는 디지털화 영화 필름에서 다음의 6개 디지털화 영화 필름(칼라 영화 필름 4개, 흑백 영화 필름 2개)을 사용하였다. 테스트에 사용된 영화는 1960년대부터 1970년대 사이에 촬영된 것으로 ‘텔레시네’ 장치를 사용하여 아날로그 필름의 영화를 디지털화 (720 x 480 크기의 MPEG2 포맷 영상) 한 것으로 오직 급진적인 샷 전환만을 포함하고 있다. 사용된 디지털화 영화 필름의 프레임 수, 샷 전환 수는 표 1과 같다.

<표 1> 테스트 비디오

	총 프레임 수	샷 전환 수
영화 1	21,127	96
뉴스 1	2,879	24
뉴스 2	2,497	18
뉴스 3	34,029	235
기록영화 1	9,607	31
기록영화 2	29,946	167

디지털화된 칼라 비디오는 먼저 흑백 비디오로 변환이 되었고, 인접한 두 프레임의 비교에는 그레이-레벨 히스토그램이 사용되었다. 연속적인 두 프레임의 사이에서 발생할 수 있는 객체의 이동거리는 20화소를, 그리고 허용할 수 있는 밝기의 변화를 보상하기 위해 길이가 10인 1차원의 모폴러지 마스크를 사용했다.

제안한 샷 탐지 알고리즘의 성능을 평가하는 기준으로 는 재현율(Recall)과 정확도(Precision)을 사용한다.

$$\text{재현율(Recall)} = \frac{\text{Correct}}{\text{Correct} + \text{Missed}} \quad \text{-- (9)}$$

$$\text{정확도(Precision)} = \frac{\text{Correct}}{\text{Correct} + \text{False}} \quad \text{--- (10)}$$

여기서 Correct는 탐지한 샷 전환 수, Missed는 탐지하지 못한 샷 전환 수, 그리고 False는 실제 샷 전환이 발생하지 않았으나 샷 전환으로 인식한 오 탐지한 샷 전환 수를 나타낸다. 식 (9)에서 재현율의 값이 “1”이면 모든 샷 경계를 탐지한 것을 말하며, 식 (10)에서 정확도의 값이 “1”이면 오 탐지된 샷 경계가 없다는 것을 의미한다.

<표 2> 테스트 결과

	탐지 샷 수	미탐지 샷 수	오탐지 샷 수	재현율	정확도
영화 1	91	5	5	0.95	0.95
뉴스 1	24	0	0	1.0	1.0
뉴스 2	17	1	1	0.94	0.94
뉴스 3	222	13	30	0.94	0.88
기록영화 1	30	2	1	0.92	0.97
기록영화 2	161	6	73	0.96	0.69

#### 4. 결론 및 향후 계획

기존의 히스토그램을 이용한 샷 전환 탐지 방법은 연속적인 두 프레임의 전체 또는 대응되는 동일한 크기의 소 영역의 히스토그램을 사용하며, 객체의 이동이나, 프레임의 밝기 변화에 취약한 문제점이 있다.

본 논문에서는 이 문제점들을 해결하기 위하여 연속적인 두 프레임을 여러 개의 소 영역으로 분할한 후, 각 소 영역에서 획득한 화소의 밝기 값 히스토그램을 비교하여 샷 전환을 탐지하는 방법을 제안하였다. 참조 프레임의 크기는 현재 프레임의 소 영역의 크기에 두 프레임사이에서 발생할 수 있는 객체의 이동과 밝기의 변화를 보상한 크기로 하여 객체의 이동에 의한 문제를 해결하였다. 또 참조 프레임의 소 영역에서 계산된 화소 밝기 값 히스토그램에 모폴러지의 그레이-레벨 팽창 연산을 적용한 후, 두 히스토그램을 비교함으로써 프레임사이의 밝기 변화에 의한 문제점을 해결했다.

영화나 뉴스와 같은 비디오 데이터의 경우, 연속적인 프레임사이에서의 객체의 이동 거리가 실험에 의해 정해진 임계치 범위 내에 있었으며, 샷 전환 탐지 결과가 우수했다. 그러나 기록 비디오의 경우, 손동작 기술과 같은 부분을 상세히 기록하기 위하여 줌인(zoom-in)하여 촬영한 연속적인 프레임에서의 객체 이동 거리는 설정한 임계치 범위 밖에 있었으며 오 탐지의 원인이 되었다. 이 문제를 해결하기 위하여 본 논문에서 사용한 지역적인 정보와 프레임 전체에서 추출한 프레임 정보를 함께 사용하는 것이 필요하다.

#### 참고문헌

- [1] H. J. Jhang, A. Kankanhalli, S. W. Smoliar, "Automatic Partitioning of Full-Motion Video," in *Multimedia*, Vol. 1, No. 1, pp. 10-28, 1993.
- [2] Xiaoquan Yi and Nam Ling, "Fast Pixel-Based Video Scene Change Detection," in *IEEE Int. Symposium on Circuits and Systems*, pp. 3443-3446, 2005.
- [3] A. Nagasaka, and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," in *IFIP Proceeding of Visual Database Systems*, pp. 113-127, 1992.
- [4] 한영욱, 정성일, 김성재, 이시영, 김승호, "움직임 벡터를 이용한 히스토그램 장면 전환 검출 기법의 개선," *한국정보과학회 1999년도 가을 학술발표논문집*, Vol. 26, No. 2, pp. 410-412, 1999.
- [5] B. H. Shekar, K. Holla K. Raghurama, Kumari M. Sharmila, "Video Cut Detection Using Chromaticity Histogram," in *International Journal of Machine Intelligence*, Vol. 3, No. 4, pp. 371-375, 2011
- [6] Z. Cernekova, C. Kotropoulos, and I. Pitas, "Video shot segmentation using singular value decomposition," in *Proc. 2003 IEEE Int. Conf. Multimedia and Expo, Baltimore, Maryland*, vol. 2, pp. 301 - 302, 2003.
- [7] Priyadarshinee Adhikari, Neeta Gargote, Jyothi Digge, and B.G. Hogade, "Abrupt Scene Shange Detection," in *World Academy of Science, Engineering and Technology* 42 2008.
- [8] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classification production effects," *ACM Multimedia Syst*, vol. 7, no. 1, pp. 119 - 128, 1999.